Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Biologia Cellulare e Molecolare

Ciclo XXXIV

Settore Concorsuale: 03/B1

Settore Scientifico Disciplinare: CHIM/03

Nickel dependent carcinogenesis and infections: structural and biophysical characterization of NDRG1 and *Sg*SrnR

Presentata da: Ylenia Beniamino

Coordinatore Dottorato

Prof. Vincenzo Scarlato

Supervisore

Prof.ssa Barbara Zambelli

Esame finale anno 2022

Abstract

In prokaryotic organisms, lower eukaryotes and plants, some important biological reactions are catalyzed by nickel-dependent enzymes, making this metal ion essential microelement for their life. On the other hand, excessive concentration of nickel within the cell, or prolonged exposure to nickel compounds, has toxic effects in living organisms. In addition, nickel has been classified by IARC as Group I human carcinogen, because of the correlation between its inhalation and increased incidence of nasal and lung cancers.

The aim of this work was to investigate the nickel impact on human health, considering both its direct role on human cells and its indirect effect as essential element for human important bacteria. In humans, nickel induces N-myc downstream regulated gene 1 (NDRG1) expression, recently proposed as new target in cancer therapy. CD, light scattering and ITC were applied on the recombinant full-length protein and its C-terminal intrinsically disordered domain, for studying the NDRG1 structural and functional properties. In particular, the fold and dynamics of the C-terminal region were examined by NMR spectroscopy and site-directed spin labeling coupled to EPR, showing the features of an intrinsically disordered region.

In nickel-dependent bacteria, nickel metabolism is strictly regulated, through the activity of different transcription factors. In *Streptomyces griseus* the expression of two superoxide dismutases (SODs) is antagonistically regulated by nickel thanks to the transcriptional complex *Sg*SrnR/*Sg*SrnQ. The *Sg*SrnR protein was heterologously expressed and its activity as possible nickel sensor studied. DNase I footprinting and β-galactosidase gene reporter assays revealed that *Sg*SrnR functions as transcriptional activator, prompting the hypothesis of a new model to describe the activity of this complex. In addition, ITC, NMR and X-ray crystallography demonstrated that *Sg*SrnR presents the fold typical of ArsR/SmtB transcription factors and low metal binding affinity, non-compatible with a role as a nickel-sensor, function probably played by its partner *Sg*SrnQ.

1. Introduction

1.1 Nickel as essential element in living organism

Metal ions such as zinc, magnesium, iron and manganese are essential nutrients for all living organisms. Although metals are necessary for performing important biological functions, such as DNA replication, enzymatic catalysis, electron transfer and signal transduction, they can have cytotoxic effects and lead to the activation of an immune response, especially in humans (Wang and Dai 2013). For this reason, all organisms have systems for the regulation of metal trafficking, in order to ensure their availability and their correct localization in the subcellular compartments, always keeping their concentration within physiological limits (Finney and O'Halloran 2003).
Nickel, in particular, is one of the most important trace metals in biology, naturally present in air, water and soil as nickel compounds and complexes, released from natural source or human activity (A.R et al. 2012). This metal plays fundamental biological roles for plants, bacteria, archaea ad unicellular eukaryotes, where it is part of the catalytic centers of some important metalloenzymes, involved in different microbial pathways. Nine nickel enzymes are known: urease, NiFe-hydrogenase, Co-dehydrogenase, Acetyl-CoA synthase, Methyl-CoM reductase, Ni-superoxide dismutase, acireductone dioxygenase, glyoxalase I and lactate racemase. Some of these enzymes work as virulence factors for some pathogenic organisms and are essential for their growth and pathogenesis (Zambelli and Ciurli 2013). One of this is the urease enzyme (Maroney and Ciurli 2021), which catalyzes the hydrolysis of urea into ammonia and bicarbonate, promoting the colonization of the host's stomach, gut and urinary tract, habitats naturally hostile and inhospitable, because of their low pH (Ragsdale 2009). Reactive oxygen species (ROS) are the most widely distributed free radicals. The most important ROS are the superoxide anion O^{2-}, the hydrogen peroxide H_2O_2 and the hydroxyl radical •OH. These molecules are produced during aerobic metabolism and can cause oxidative damage to DNA, proteins and membrane lipids. Aerobic organisms have different defense systems to protect themselves from ROS, maintaining their level low within the cells (Storz et al. 1990). These include the conversion of the superoxide radical (O^{2-}) to hydrogen peroxide H_2O_2 and molecular oxygen (Fridovich 1983), catalyzed by the superoxide dismutase (SOD) enzyme. Depending on the metal ion present in the active site, it is possible to distinguish four groups of SODs: Cu/Zn-SOD, Fe- SOD, Mn-SOD (prokaryotes and eukaryotes) and Ni-SOD (prokaryotes). A nickel-dependent SOD (Ni-SOD) and an iron-zinc-dependent SOD (FeZn-SOD), encoded by the *sodN* and *sodF* genes, respectively, were identified in *Streptomyces griseus* and *Streptomyces coelicolor* (Kim, Kang, and Lee 2003). The expression of these enzymes is reported to be antagonistically regulated by nickel, increasing Ni-SOD (E. J. Kim et al. 1998) expression and downregulating that of FeZn-SOD (E.-J. Kim et al. 1998).

1.2 Nickel and human health

Inhalation and ingestion of water and food or contact with nickel containing products, together with tobacco cigarette, are the main sources of human exposition to nickel compounds. The nickel level into the cells must be strictly regulated. This metal is an essential micronutrient for numerous important biological functions like reproductive and sensory functions, bone composition, strength and energy metabolism (Kumar S. and Trivedi A.V 2016). Nevertheless, excessive exposure to nickel compounds leads to adverse effects on humans, including immune reactions, respiratory, cardiovascular and kidney diseases and the development of some tumor forms (Zambelli and Ciurli 2013).

1.2.1 Immune reactions and nickel-induced dermatitis

Allergic contact dermatitis (ACD) is the most common form of dermatitis worldwide. Nickel is one of the main contact allergens, with more than 30% of the population showing sensitivity to this metal (Girolomoni et al. 2004). The primary cause is the facility of being exposed to this metal, widely used in modern industry in jewelry products and cell phones. The most common sign of nickel allergy is skin inflammation, in the contact area, with erythema, itching and blisters. The ingestion or the inhalation of nickel by a person sensitized to epi-cutaneous contact with nickel may lead to systemic contact dermatitis (SCD), characterized by generalized eczematous reaction, including headache, fever and diarrhea. The first event in the immunological response to nickel is a silent sensitization phase, initiated at the first contact with the antigen, which leads to the generation of allergen-specific T cells and the activation of lymphocytes. Upon re-exposure to the allergen, these triggered T cells will be activated, exerting cytotoxic functions and secreting inflammatory mediators, such as chemokines and cytokines, to amplify the inflammatory response and resulting in the elicitation phase and clinical presentation of the Ni-ACD (Peana et al. 2017; Rothenberg 2010). The only way to avoid the development of these allergic forms is to avoid nickel exposure.

1.2.2 Carcinogenic potential of nickel compounds

The chronic exposition to nickel compounds is related to the progression of numerous cancers. The carcinogenic potential of nickel can be attributed both to soluble and insoluble compounds, which lead, in individuals who are daily exposed, to an increase in cases of nose and lung cancer. Genetic predispositions, including metabolic variations, may be important in nickel-induced carcinogenesis, whose compounds, when co-administered with other carcinogens, produce a synergistic effect, favoring cellular transformation both *in vivo* and *in vitro*. Nickel can promotes the

carcinogenic transformation of cells acting at the level of DNA and gene expression (Costa et al. 2003) or alternatively at the level of proteins and transcriptional regulators involved in important cellular processes, leading to a general change in metabolism and cellular homeostasis.

Nickel compounds are considered weakly clastogenic and have previously been shown to generate specific chromosomal damage, particularly in heterochromatic regions of the genome. From an epigenetic perspectives nickel is also able to induce DNA methylation, commonly found in the heterochromatic regions of chromosomes and associated with transcriptional repression (Lande-Diner et al. 2004). Ni(II) can replace Mg(II), normally bound to DNA, causing the formation of new heterochromatic regions and favoring the activity of methylases (Salnikow and Kasprzak 2007), which recognize the newly generated heterochromatic and unmethylated DNA (Lee et al. 1995). This property is due to the same charge and similar ionic rays of the two ions. Nickel also influences the global levels of histone modifications, which are important for controlling the access of regulatory proteins to DNA and the chromatin transcriptional status (Salnikow and Zhitkovich 2008). In this case, nickel acts by decreasing the acetylation levels of some histones and increasing the methylation (Chen et al. 2006) and ubiquitination (Karaczyn, Golebiowski, and Kasprzak 2006) of others, through the JHMD histone demethylases (Chen et al. 2010), dependent on iron and α-ketoglutarate. These changes in DNA methylation can alter the expression levels of genes that play a role in the progression of cancer cell transformation (Klein et al. 1991).

Exposure to nickel produces a pattern of gene expression identical to that induced by hypoxia and the transcription factor HIF-1 (Maxwell and Salnikow 2004), belonging to the PAS family and consisting of the two subunits: HIF-1α and HIF-1β. The α subunit, with 826 aa and a MW of 120 kDa, is the regulatory component of the HIF-1 complex. Under normoxic conditions the protein is rapidly degraded by the proteasome, contrary to the β subunit which is constitutively expressed. Hypoxia is the physiological factor that activates HIF-1 and only the α subunit responds to changes in oxygen tension, through enzymatic hydroxylation of specific proline and asparagine residues. The capture by VHL requires the hydroxylation of HIF-1α at Pro402 or Pro564 residues, favoring the hydrogen bond with Ser111 and His115 of VHL and the ubiquitination of HIF-1α. At the same time, the interaction with p300 is prevented by hydroxylation of the β carbon of Asn803, inhibiting transcriptional activation (Huang and Bunn 2003). The hydroxylation of these residues is mediated by the prolyl (PHD 1-3) and asparaginyl hydroxylase (FIH-1) enzymes, which act by sensing the oxygen levels inside the cell and going to hydroxylate specific residues of HIF-1α in the presence of oxygen, Fe(II) and 2-oxoglutarate (2-OG). PHDs requires ascorbate to function (Salnikow et al. 2004). The depletion of intra-cellular ascorbate blocks the PHD activity, because Fe(II) cannot be maintained in the reduced state. In addition, the Fe(II) ion, normally present in these enzymes, can

be replaced by Ni(II) or Co(II) (Salnikow, Su, et al. 2000) by competition with the DMT1 divalent metal carrier. Under these conditions, PHD and FIH enzymes are no longer able to catalyze the splitting of the oxygen molecule and to proceed with the hydroxylation of HIF-1α, mimicking the hypoxic condition. In this way, exposure to Ni(II) and Co(II) leads to accumulation of HIF-1α which can thus dimerize with HIF-1β. The heterodimer recalls p300/CBP and binds to the 5'-RCGTG-3' sequences, called *hypoxia response elements* (HRE), present in the promoter or enhancer regions of the hypoxia responsive genes, activating their expression. This phenomenon is known as *metal-induced hypoxia* and the gene coding for NDRG1 is one of the genes induced by this pathway (Cangul 2004) (Figure 1).

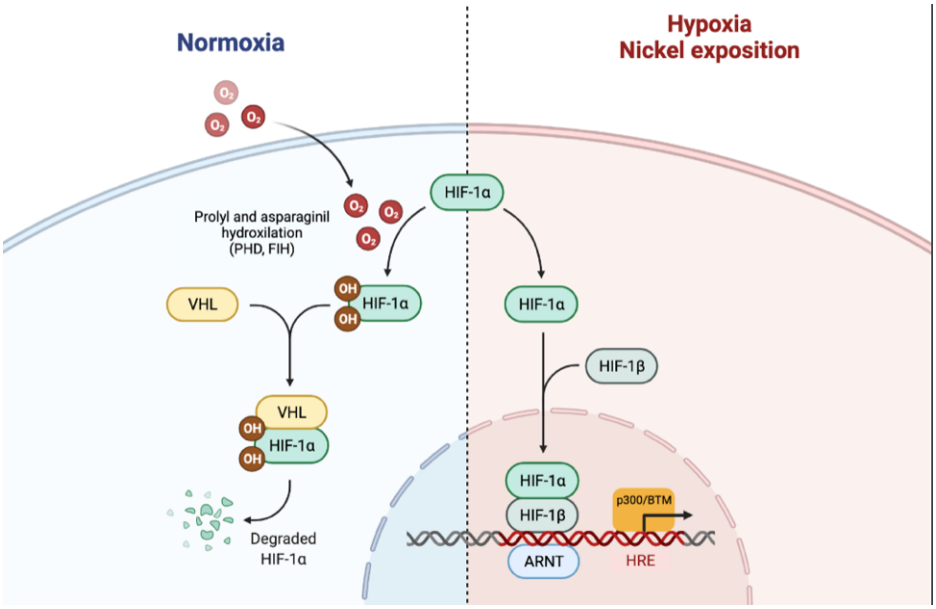


Figure 1: *Nickel induced hypoxia pathway. Adapted from "HIF signaling" by BioRender.com (2021). Retrieved from https://app.biorender.com/biorender-templates.*

Activation of the hypoxic signaling pathway and the transition of cellular metabolism to a state that mimics permanent hypoxia can cause nickel-induced carcinogenesis. Hypoxia is a common state in tumors because transformed cells grow faster than the blood vessels that supply them with oxygen. This state can activate genes that allow cells to overcome the adverse conditions, stimulating angiogenesis and inhibiting cell proliferation, favoring on the contrary apoptosis. Consequently, hypoxic conditions, like those induced by nickel, promote the survival of neoplastic cells that exhibit altered metabolism and cell cycle control and increased resistance to apoptosis (Salnikow and Kasprzak 2007) (Figure 2).

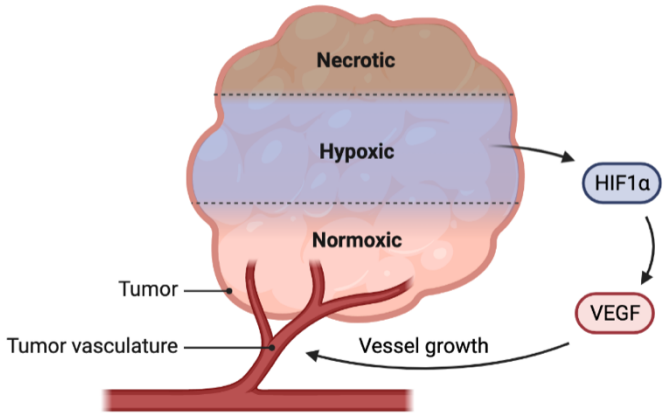


Figure 2: *Hypoxic signaling contributes to tumor vessel growth. Reprinted from "Hypoxic signaling in tumors" by BioRender.com (2021). Retrieved from https://app.biorender.com/biorender-templates.*

1.3 N-myc downstream regulated gene 1

1.3.1 Structure of NDRG1 and protein expression regulation

The human protein N-myc downstream regulated gene 1 (NDRG1) belongs to the NDRG family, including NDRG2, NDRG3 and NDRG4 proteins, sharing 53-65% sequence identity with each other. The NDRG1 gene is localized on the long arm of chromosome 8, specifically at the 8q24.22 locus, transcribed into a 3-kB mRNA corresponding to a protein of 43 kDa and 394 aa. The NDRG1 mRNA is ubiquitous, although its levels are highly heterogeneous. The protein expression is strictly regulated at transcriptional, post-transcriptional and translational levels (Lachat et al. 2002). It is primarily a cytoplasmic protein (47.8%), but it was also identified in the nucleus (26.1%) and mitochondria (8.7%). In NDRG1 it is possible to distinguish a N-terminal domain (aa 1-311), including an α/β hydrolase motif, common to all NDRG proteins, and a unique intrinsically disordered region. This region is 83 amino acidic residues long (312-394) and it is located at the C-terminal end of the protein. It is characterized by a three repeated ten amino acids sequence (GTRSRSHTSE), which shows binding activity toward nickel ions (Fang et al. 2014) and some serine and threonine residues, undergoing functional phosphorylation by the PKA, SGK1, PKC, CK-II and CAMK kinases. The phosphorylation status of this residues is reported to be linked to the localization and biological functions of the protein (Bae et al. 2013) (Figure 3).

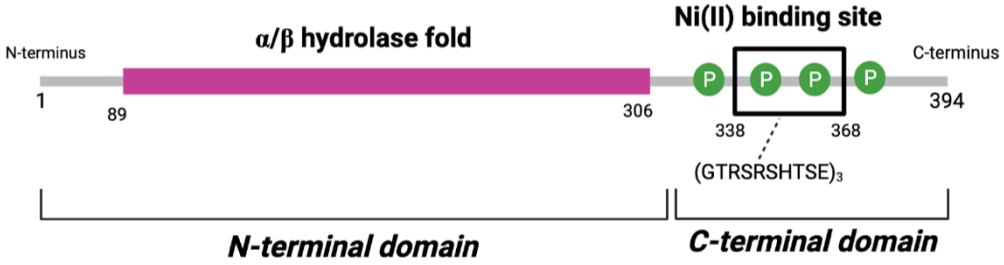


Figure 3: *Structure of the human NDRG1 protein. The N-terminal domain (aa 1-310, 34.2 kDa) including the α/β hydrolase fold and the C-terminal domain (aa 311-394, 8.8 kDa) with the phosphorylation sites (green circles) and the nickel binding sequence are shown. Created with BioRender.com.*

The crystallographic structure of the NDRG1 core (aa 31-319) was recently resolved at 2.96 Å resolution, resulting in two chains in the asymmetric unit. The NDRG1 core includes the canonical α/β hydrolase motif, common to all NDRG proteins, with eight central β-sheets surrounded by α-helices. This motif is covered by a cap domain with three helices and some disordered regions. In particular, the two regions spanning from 170 to 184 and from 197-201 residues were not detectable, because of their high flexibility (Mustonen et al. 2020) (Figure 4).

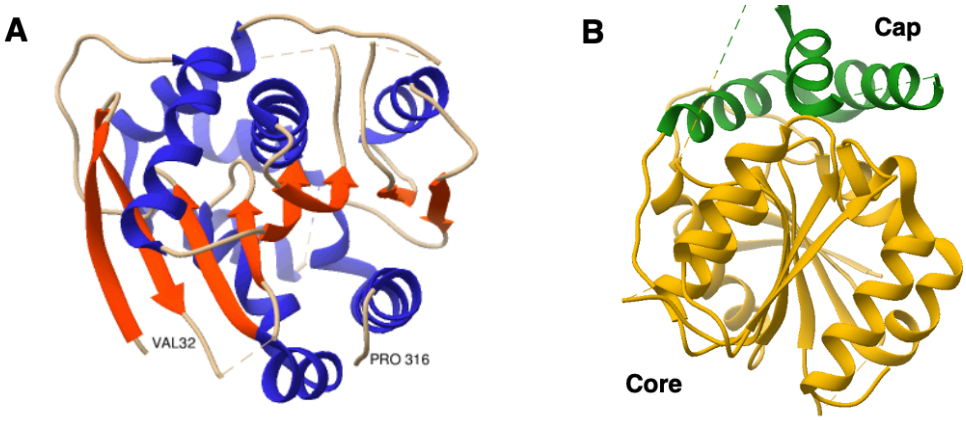


Figure 4: *Structure of one molecule of the NDRG1 core, in orange β-sheets and in blue α-helices (**A**). The canonical α/β hydrolase fold (gold) shows eight-central β-sheets surrounded by α-helices and protected by a cap domain (green) (**B**). The dashed lines indicate the residues which were not detectable. Images of 6ZMM (Mustonen et al. 2020) created with Chimera software (Pettersen et al. 2004).*

NDRG1 belongs to the α/β hydrolase superfamily. The sequence consisting of a nucleophile, an acidic amino acid and a histidine, necessary for proper hydrolasic activity, is absent in NDRG1 (Mustonen et al. 2020). The absence of this sequence, despite the presence of the hydrolytic domain in NDRG1, suggests its evolution from a hydrolytic enzyme, maintaining the structure but not the functionality. On the other hand, there is the possibility of a convergent evolution, where the NDRG genes have developed this molecular architecture for a different purpose (Hwang et al. 2011). However, a nearly equivalent site in NDRG1, including Asp64 and His194 was recently identified (Figure 5) (Mustonen et al. 2020).

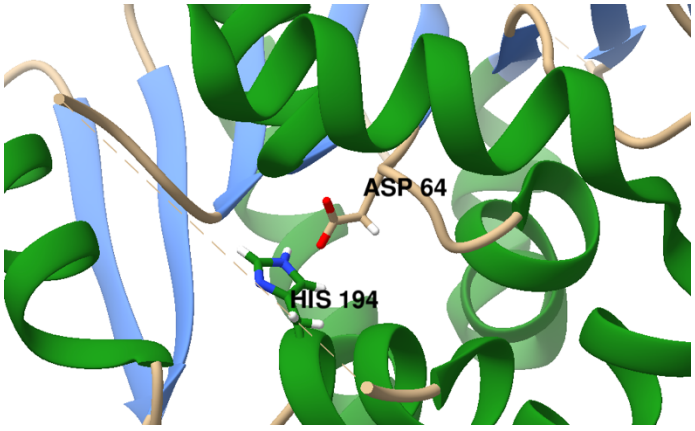


Figure 5: *Possible catalytic residues of the hydrolytic domain in NDRG1: D64 and H194. Images of 6ZMM (Mustonen et al. 2020) created with Chimera software (Pettersen et al. 2004).*

NDRG1 expression is regulated by different factors. Calcium, iron chelators (Kovacevic et al. 2016), the tumor suppressors p53 (Stein et al. 2004) and PTEN, the oncogene MYC (Zhang et al. 2008), hypoxia and nickel are the most important. NDRG1 expression can be induced in many tumor forms by novel anti-cancer agents, such as Dp44mT and DpC, through mechanisms involving iron depletion. Iron is an essential element, important for DNA proliferation and synthesis and its depletion by these new chelating agents inhibits cell cycle and induces apoptosis, affecting several targets. These include: iron-dependent enzymes such as ribonuclease reductase (RR), which is critical for DNA synthesis, NDRG1, p53, cyclin D1, p21 and many others. These compounds, unlike typical chelators, do not induce total iron depletion and their mechanism of action requires iron binding, whose reduction induce the formation of active redox complexes. The latter react with oxygen to form reactive oxygen species that mediate oxidative insults in the cell (Bae et al. 2013).

1.3.2 NDRG1 post-translational modifications

NDRG1 has numerous phosphorylation sites for several kinases, both in the core of the protein and in the C-terminal portion. SGK1 (Serum/Glucocorticoid Regulated Kinase 1) phosphorylates NDRG1 at Thr346, Thr356 and Thr366 in the three tandem repeats and at Thr328 and Ser330, the residues conserved in all NDRG proteins. Phosphorylation of NDRG1 by SGK1 makes this protein an excellent substrate for GSK3 (glycogen synthase kinase 3), which could then phosphorylate serine/threonine residues located in its consensus sequence: S/T-X-X-X-S/T(P). This consensus sequence suggests that GSK3 can phosphorylate serine/threonine residues located four residues before a pre-phosphorylated serine/threonine (Ser342, Ser352 and Ser362) (Murray et al. 2004).
NDRG1 is phosphorylated also by the kinases PKA, PKC, and CaMK-II. Phosphorylation of NDRG1 *in vitro* by PKA or PKC is reported to increase when the three C-terminal tandem repeats are deleted, suggesting that this C-terminal region masks the PKA or PKC phosphorylation sites, located in the NDRG1 core (Sugiki et al. 2004). The role of the phosphorylation of NDRG1 remains unclear, but some studies have tried to highlight the importance of this process for the activity of this protein, showing that this post-translational modification can affect NDRG1 subcellular localization (McCaig et al. 2011; Park et al. 2018)

1.3.3 Biological functions of NDRG1

NDRG1 is intricate in different biological functions (Figure 6). Stimulating the differentiated phenotype, inhibiting cell cycle progression (Bae et al. 2013) and promoting myelin sheath formation and maintenance are some of them. NDRG1 is also a stress response gene, aiding cells to defend

themselves against adverse conditions such as iron depletion, hypoxia, exposure to nickel compounds and DNA damage agents, commonly used in chemotherapies. A recent study has shown that NDRG1 can inhibit autoimmune T-cell-mediated inflammation (Oh et al. 2015), by maintaining clonal T-cell anergy, a hypo-responsive condition in which T cells are unable to respond to certain stimuli or antigens. In addition to the functions described above, NDRG1 appears to be involved in other physiological mechanisms that occur in cells, such as lipoprotein regulation and formation, vesicle transport, genome stabilization, placental formation, fetal development and urinary and reproductive system organ formation. In addition to various physiological roles, NDRG1 is also involved in the development of certain diseases. Recently, the NDRG1 gene has been identified as an antiviral factor for hepatitis C virus (HCV), while mutations of the NDRG1 gene cause the neuropathy Charcot Marie Tooth type 4D (CMT4D) (Kalaydjieva et al. 2000), a disorder of the peripheral nervous system, characterized by progressive weakness and atrophy. The identification of the NDRG1 gene as an antiviral factor is particularly interesting because HCV infection is one of the conditions leading to hepatocellular carcinoma (HCC), in which NDRG1 has been reported to be oncogenic.

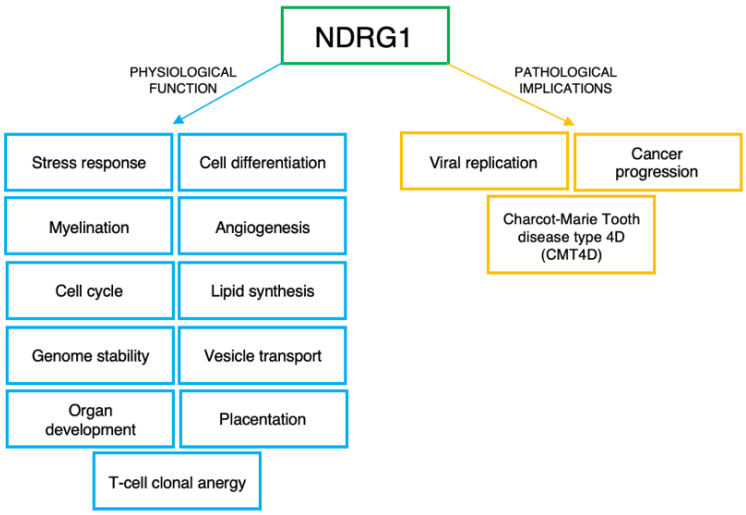


Figure 6: *Biological roles of NDRG1 associated with physiological or pathological conditions.*

The different pleiotropic roles of NDRG1 in various types of cancer will be discussed in detail below (Park et al. 2020).

1.3.4 NDRG1 and its role in cancer: a new possible target in cancer therapy

NDRG1 works as tumor suppressor or oncogene in a tissue dependent manner. For example, it has an anti-oncogenic role in brain, breast, colon, rectum, esophageal, prostate, pancreatic and ovarian tumor forms (Wang et al. 2014). Conversely, in other cancers, such as those of the liver, skin and

cervix, its expression is increased, suggesting that its role in these tissues is to promote tumorigenesis (Song and Cao 2013). Studies examining the prognosis of patients with prostate, pancreatic, breast, or colon cancers, where NDRG1 expression is negatively correlated with tumor progression, have shown that those with higher NDRG1 levels have a better prognosis when compared with those in whom NDRG1 levels are lower. As a result, NDRG1 can be used as a prognostic indicator and molecular target for the treatment of these tumors (Kovacevic et al. 2016). Cell proliferation and invasion are particularly activated by NDRG1 silencing, promoting tumor cell progression, which is instead decreased when NDRG1 is overexpressed, favoring in this case differentiation (Chang et al. 2016). NDRG1 inhibits the oncogenes RAS, c-SRC, PI3K, WNT, ROCK7pMLC2 and NF-k and promotes the expression of key molecules with tumor suppressor function, such as phosphatase and E-cadherin. NDRG1 has also a suppressor role in tumor angiogenesis, decreasing the expression of vascular endothelial growth factor (VEGF) and interleukin-8 (IL-8) in some forms of cancer, such as prostate cancer (Park et al. 2020) (Figure 7).

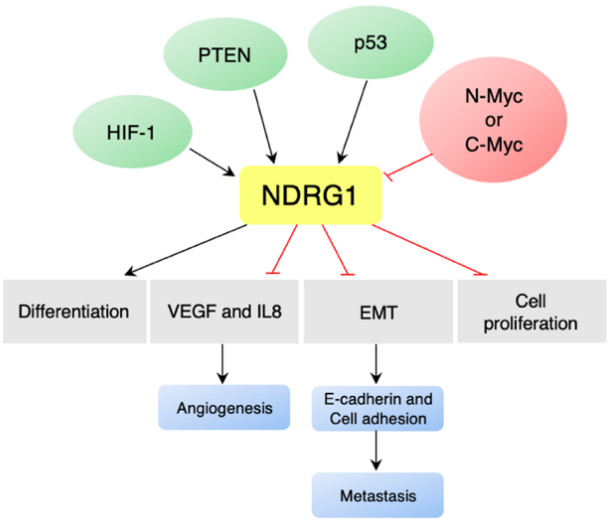


Figure 7: *NDRG1 regulation and its effects in tumor progression. Adapted by permission from BMJ Publishing Group Ltd. Redrawn from Bae et al. 2013 with License number 5234871131470 Copyright 2013, BMJ Publishing Group Ltd & Association of Clinical Pathologists.*

One of the main roles attributed to NDRG1 as tumor suppressor is certainly to inhibit the development of metastases. Cancer metastasis is a complex biological process, involving pronounced changes in gene expression and cell morphology. Tumor cells must migrate, survive in the circulatory system, invade a secondary site and start to proliferate. Profound changes in gene expression and cell morphology are required, representing in their totality the epithelium-mesenchyme transition (EMT), the initial stage of metastasis development (Kovacevic et al. 2016), involved also in embryonic growth and tissue remodeling and repair (Sahni, Krishan, and Richardson 2014). Cells undergoing EMT tend to lose their epithelial markers (E-cadherin) and express those typical of the mesenchymal phenotype, acquiring a great migratory and invasive capacity. NDRG1 mediates its metastasis-suppressing effect by modulating TGF-B, which through SMAD proteins, can induce

the expression of transcription factors that facilitate EMT in certain tumors (Sahni et al. 2014). E-cadherin, one of the proteins most involved in EMT regulation, is a transmembrane protein whose ectodomains in adjacent epithelial cells interact to form homodimeric bridges, known as adherents junctions, essential for providing structural integrity to epithelial sheets. The cytoplasmic domains of E-cadherin bind to the actin fibers of the cytoskeleton through a complex of α and β-catenins, giving resistance to mechanical forces to epithelial cells. Loss of E-cadherin results in the presence of free β-catenin in the cytosol, which then translocates into the nucleus, where through interaction with other transcription factors leads to the expression of genes involved in EMT, such as cyclin D1 and c-myc. The over-expression of NDRG1 maintains the membrane localization of E-cadherin and β-catenin, inhibiting consequently the β-catenin translocation and its interaction with the TCF/LEF1 complex to up-regulate various Wnt responsive genes (Liu et al. 2012)(Figure 8).

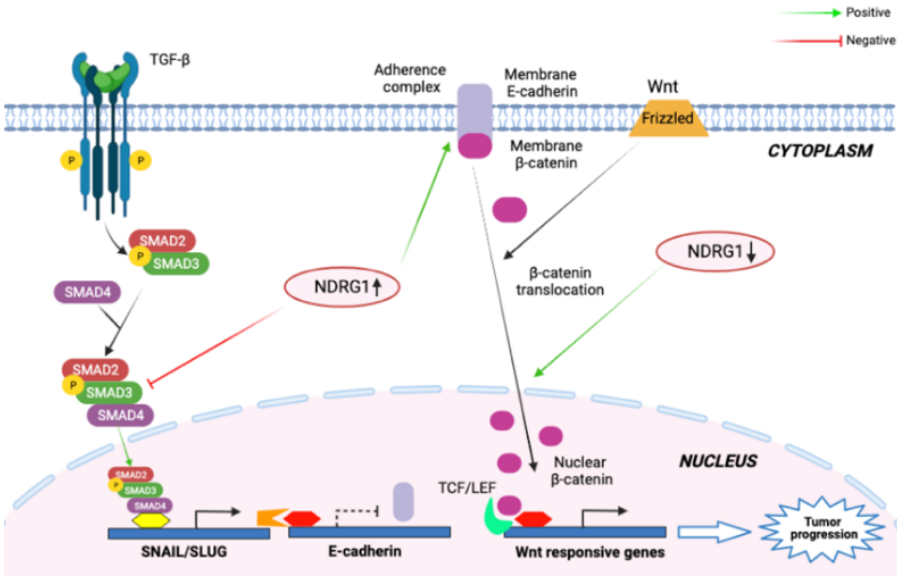


Figure 8: *Regulation of EMT by NDRG1. Created with BioRender.com.*

Recent studies have shown that NDRG1 has an active role in lung cancer and it is a possible target for its treatment (Nagai et al. 2011). In A549 line lung cancer cells, overexpression of NDRG1 was found to lead to increased proliferation and decreased apoptosis of cells (Wang et al. 2013). In addition, NDRG1 can contribute to cisplatin resistance, one of the most used chemotherapeutic agents in the treatment of non-small cell lung cancer, probably reducing the expression of the stress-inducible gene ATF3, associated to apoptosis in some human cancers and induced by cisplatin (Du, Jiang, and Fan 2018). Therefore, targeting NDRG1 may be critical to attack chemoresistance in lung cancer treatment (He et al. 2018).

1.4 Metal ion-dependent transcriptional regulators

Uptake, excretion and utilization of metal ions are strictly regulated. This is carried out by transcriptional regulators called metal sensors, that can specifically recognize metal ions and selectively bind to them (Zambelli, Musiani, and Ciurli 2012). The specific interaction between metal ions and metal sensors affects their structure, causing conformational changes that alter their affinity and specificity for DNA, leading to the repression or activation of genes encoding ion channels, transporters for specific metal ions, metal-dependent enzymes or other metal-dependent regulators (Silver and Phung 1996). Metal sensors translate the concentration of a specific metal ion into a change in gene expression, causing an adaptive response. They are often orthologues of other transcriptional regulators, whose ligands are small organic molecules, such as intermediates of metabolic reactions, amino acids, and lipophilic molecules. The difference between metal sensors and these metal-independent transcriptional regulators lies mainly in the evolution of the cavity in which the ligand is arranged, which has been adapted to recognize a specific metal ion (Pennella and Giedroc 2005). Seven main families of metal sensors have been identified in bacteria to date. Proteins belonging to different families may be able to recognize the same metal ions. The functional units of metal sensors are generally homodimers or tetramers of homodimers (Zambelli et al. 2012). Typically, each monomer has a structure consisting of an N-terminal working as DNA binding domain (DBD), often presenting the winged helix-turn-helix (wHTH) motif, and a dimerization domain at the C-terminal domain (Gajiwala and Burley 2000). The loop between α2 and α3 and the N-terminal end of the beta-wing binds the major groove of the DNA double helix at the operator level. In other cases, DBD has a ribbon-helix (RHH) pattern consisting of a short beta-strand followed by two α-helices (Schreiter and Drennan 2007). Metal-binding sites are often located at the interface between two monomers or between the N-terminal and C-terminal domains. In most cases, metal sensors exert a negative control on transcription by acting on the promoter of the regulated gene and thus preventing the action of RNA polymerase. Binding of the metal ion to the transcriptional regulator, which acts as a repressor, can lead to repression of the genes under its control, SmtB/ArsR and CsoR/RcnR families, or can cause allosteric activation of transcription, MerR family. The remaining families are involved in switching off metal ion uptake systems in response to high metal ion levels. In this case, the metal ions act as co-repressors, binding the transcriptional regulator and causing repression of gene expression. The SmtB/ArsR and MerR family members are capable of recognizing the widest range of metal ions, but also metallic organic compounds. Members of the MerR family activate gene expression by distorting the structure of the operator when bound to the respective metal ion. This allows RNA polymerase to initiate transcription from an otherwise weak promoter (Brown et al. 2003). The family name is

derived from the study of the merR gene of Tn501 transposon from *Pseudomonas aeruginosa* and Tn21 from *Shigella flexneri*. In gram-negative bacteria MerR regulates the transcription of an operon involved in mercury resistance (Barkay, Miller, and Summers 2003). Several MerR orthologues are known to bind many inducers including metal ions, lipophilic compounds, and nitric oxide. This is possible by careful changes in the position and type of residues that determine the binding site. The structure of some MerR transcriptional regulators has been resolved. It is possible to analyze the structure of some *Escherichia coli* orthologues. *Ec*ZntR predominantly binds Zn(II) and regulates the expression of the zntA gene, which encodes for a Zn(II)/Cd(II)/Pb(II) ATPase. The operator is characterized by a 20 bp spacer between the -35 and -10 regions of the promoter. *Ec*ZntR is present in homodimeric form and, like Hg(II)-MerR, exerts a promoter distortion to activate expression of the gene under its control (Browning and Busby 2004). The structure of the protein shows two Zn(II) ions bound to two cysteine/histidine residues (Cys114 and Cys124 for the first ion Zn(II), Cys115 and His119 for the second one), forming part of the metal binding domain (MBD) and another cysteine residue (Cys79) belonging to the N-terminus of the dimerization helix of the opposite subunit. A phosphate ion binds the two zinc ions in a bidentate manner. NMR spectroscopy studies conducted on Hg(II)-MerR suggest that the Hg(II) ion possesses a planar coordination geometry involving three cysteine residues corresponding to the residues coordinating the metal ions of *Ec*ZntR (Utschig, Bryson, and O'Halloran 1995). This suggests that binding to metal ions occurs at similar positions in different proteins binding different metal ions. Instead, the *Ec*CueR metal sensor binds monovalent Cu(I), Au(I) or Ag(I) metal ions. In this protein, the residue analogous to Cys79 is a Ser77 (Changela et al. 2003). This substitution blocks a possible coordination position and makes the binding pocket hydrophobic. In this case Cu(I) has a linear S-Cu-S coordination geometry involving the cysteine residues Cys112 and Cys120. It can be concluded that the selectivity of the MerR family members derives from the difference in the type of coordination.

The SmtB/ArsR family contains more than 500 sequenced members in over 200 bacterial genomes, many of which are not functionally characterized (Campbell et al. 2007). The name of this family is derived from the first two identified members: ArsR from *E. coli* (Shi, Wu, and Rosen 1994) and SmtB from *Synechococcus* PCC7942 (Morby et al. 1993). ArsR is an arsenic and antimony-sensitive transcriptional regulator that represses the expression of the operon encoding an arsenate reductase. SmtB is a zinc-sensitive transcriptional regulator that represses the transcription of an operon containing both the smtA gene, which encodes a metallothionein, and the smtB gene. In addition to controlling proteins that regulate the flow of metal ions, this family has several functions including sulfur oxidation, bioluminescence, and the response to oxidative stress

(Busenlehner, Pennella, and Giedroc 2003). Another member of this family is ZiaR from *Synechococcus* PCC 6803 which binds Zn(II) and shares ~50% sequence identity with SmtB. It regulates the expression of the zia operon, encoding ZiaR and ZiaA, an ATPase used for metal excretion (Thelwell, Robinson, and Turner-Cavet 1998). In contrast, the transcriptional regulator CzrA from *Staphylococcus aureus* is sensitive to Zn(II) and Co(II) and regulates the expression of the czr operon encoding both the repressor and the zinc membrane transporter CzrB (Kuroda, Hayashi, and Ohta 1999). The proteins in this family, form stable dimers and present at least five α-helices and two or three stranded β-sheets. Two of the five helices, (α3 and α4) are directly involved in DNA binding, while the remaining three establish hydrophobic interactions, important for stabilizing and properly orienting the DNA-protein complex or are located in the dimer interface, inducing the dimerization process (α1 and α5) (Zambelli et al. 2012). In contrast to MerR family regulators, the position and coordination of the metal ion binding site (MBS) are very different in SmtB/ArsR family proteins. Two different main types of MBS can be distinguished depending on their position on the secondary structural elements. One of these is near the N-terminal portion and includes the α3 helix. This site is generally rich in cysteine residues and favors ions such as Cd(II), Pb(II) and As(III). This category is related to *Ec*ArsR. The second type of MBS is located at the C-terminus of the protein sequence and comprises residues belonging to the α5 helix. These are generally histidine, aspartate and glutamate residues and form a tetrahedral or distorted tetrahedral coordination site with Zn(II), Ni(II) or Co(II) ions.

1.4.1 The SrnR-SrnQ transcriptional regulatory complex

Reactive oxygen species (ROS) are produced during cellular respiration and can cause oxidative damage to biological macromolecules. Aerobic organisms have developed enzymatic defense systems to protect themselves from ROS including superoxide dismutase (SOD), catalase and peroxidase (Storz et al. 1990). In *Streptomyces griseus* and *Streptomyces coelicolor* a nickel-dependent SOD (Ni-SOD) (13 kDa) and an iron-zinc-dependent SOD (FeZn-SOD) (22 kDa) have been identified (Kim et al. 2003). Expression of Ni-SOD, encoded by the *sodN* gene (E. J. Kim et al. 1998), increases in these organisms in the presence of nickel, while FeZn-SOD expression is negatively regulated by the presence of the metal ion (E.-J. Kim et al. 1998). A first study identified that the region potentially affected by binding to the sodF repressor was a sequence overlapping the transcription start site of the gene and consisting of inverted repeated sequences, identified by performing a gel mobility shift assay (Kim et al. 2003).

*Sg*SrnR and *Sg*SrnQ are two proteins involved in the transcriptional regulation of the *sodF* gene in *S. griseus.* This gene codes for the FeZn-SOD, used as defense from ROS.

*Sg*SrnR consists of 114 residues, has a molecular mass of 12.33 kDa and a pI of 6.6. A comparison of *Sg*SrnR amino acid sequence with protein databases revealed its homology with SmtB/ArsR transcriptional family regulators.

```
>tr|Q8L1Y3|Q8L1Y3_STRGR Transcriptional regulator SrnR OS=Streptomyces griseus
OX=1911 GN=srnR PE=4 SV=1
MESRALPADEEASAFRAVADPTRRQILEDLRGGELAAGEIAGRFPISAPSISRHLGVLKG
AGLVTERRDANRILYSLAEERLALCVGRFLSAVCPEQIVLRTTKWRSAPEGDAS
```

*Sg*SrnQ consists of 110 residues, has a molecular mass of 12.49 kDa and a pI of 12.3. No protein has been identified that shows sequence homology with *Sg*SrnQ. It has a high arginine content (26%) and an intrinsic disorder (ID) rate of ~80%. *Sg*SrnR, on the other hand, is visibly more ordered, with an ID rate of 25-30% (Zambelli, Uversky, and Ciurli 2016).

```
>tr|Q8L1Y2|Q8L1Y2_STRGR Transcriptional regulator SrnQ OS=Streptomyces griseus
OX=1911 GN=srnQ PE=4 SV=1
MIRPRLASVIERRLLVNYRVDPHVAAALLPAPLRPQLVARPGGGRDLSAPYRRRPSRLGP
RRGRADQRERGAPDRGGVGRARRRRAWRLHPASRHRLPAQRLRGRPDLPG
```

The two open reading frames (ORFs) coding for these proteins are located downstream of *sodF* and the stop codon of *Sg*SrnR and the start codon of *Sg*SrnQ overlap. A possible mechanism of action of this complex involves an interaction between *Sg*SrnR, *Sg*SrnQ and Ni(II) ions, resulting in the binding to the *sodF* operator. This leads to the inhibition of the *sodF* transcription by RNA polymerase detachment (Figure 9). According to this model, *Sg*SrnR acts as repressor, binding directly to DNA, whereas *Sg*SrnQ, containing a Ni(II) binding motif, as co-repressor (Kim et al. 2003).

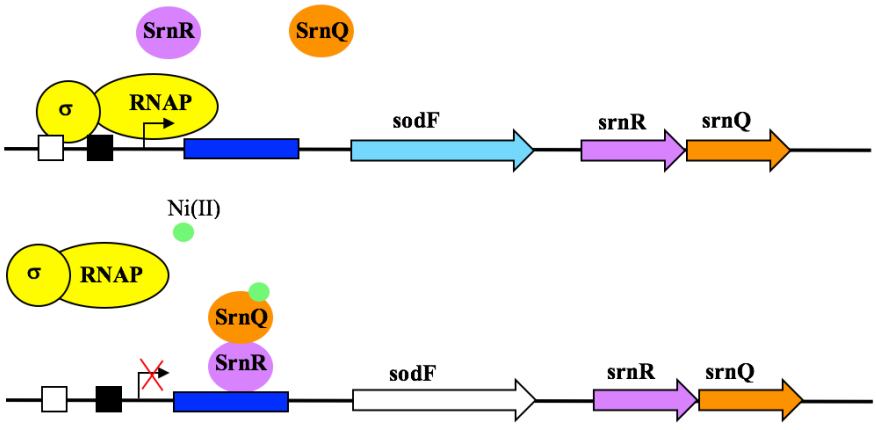


Figure 9: *Mechanism of function of SgSrnR/SgSrnQ protein complex, which act as repressor according to the model proposed in J. S. Kim, Kang, and Lee 2003.*

2. Aims of the study

Nickel is an essential element for bacteria and lower eukaryotes to catalyze important reactions. Nickel trafficking is tightly regulated in order to keep its intracellular levels stable and avoid cytotoxic effects. Furthermore, in humans, an excessive exposure to nickel, can lead to allergic reactions and favor the tumorigenesis process. The present work is divided into two parts. The first part of this work is focused on the structural and biophysical characterization of the human NDRG1 protein, involved in numerous cellular processes, some of which implicated in the control of the tumorigenesis process. The interactions and the mechanism of action of NDRG1 are still unclear and the understanding of these elements at the molecular and cellular level is of great importance, in order to design drugs specifically targeting this protein. Among the biological activities of NDRG1 its activity as oncogene or tumor suppressor in a tissue-dependent manner is of particular interest. Especially in lung cancer, one of the leading causes of death in the world, due to excessive exposure to nickel compounds, the overexpression of NDRG1 is associated to increased aggressiveness of the tumor and worse prognosis. For this reason, NDRG1 has been recently proposed as tumor marker and new target in cancer therapy. Particular attention was given to the C-terminal intrinsically disordered region of NDRG1, reported to be involved in the regulation of the biological functions of the protein. To reach this goal, protocols to purify both the full-length protein and its disordered C-terminal domain via heterologous expression in *E. coli* were developed. Then, the purified proteins in solution were structurally and biophysically characterized using circular dichroism, light scattering and isothermal titration calorimetry. For the C-terminal region, NMR spectroscopy was applied. The dynamics of the latter protein was further investigated by site directed spin labeling coupled to EPR technique (SDSL-EPR) in collaboration with Elisabetta Mileo of the laboratory of Bioénergétique et Ingénierie des Protéines of the CNRS, in Marseille.

The second aim was the determination of the biophysical and functional properties of *Sg*SrnR. This protein derives from the Gram-positive bacterium *Streptomyces griseus,* the main producer of natural antibiotics*,* and with *Sg*SrnQ it is responsible of the regulation of the nickel dependent expression of the FeZn-superoxide dismutase (*sodF*). The structural and metal-binding properties were investigated on the purified protein. DNase I footprinting and β-galactosidase gene reporter assays were carried out to study its binding to the *sodF* promoter region and its functional role for protein expression.

3. Results part 1: Biophysical characterization of NDRG1

3.1 Protein expression and purification

The *h*NDRG1 canonical sequence (UniProt code Q92597-1) was reverse translated to obtain the nucleotide sequence and cloned into the *pET15b::StrepTag* expression vector (Miraula, Ciurli, and Zambelli 2015). *E. coli* BL21-CodonPlus (DE3) RIL competent cells were used to overexpress the Strep-Tagged *h*NDRG1 protein (45.28 kDa) in an auto-induction medium containing both glucose and lactose as carbon sources. The expressed polypeptide was clarified from the soluble fraction of the cell lysate through two chromatographic steps: an affinity chromatography with a prepacked StrepTactin Sepharose High Performance column (Figure 10) followed by a size exclusion chromatography.

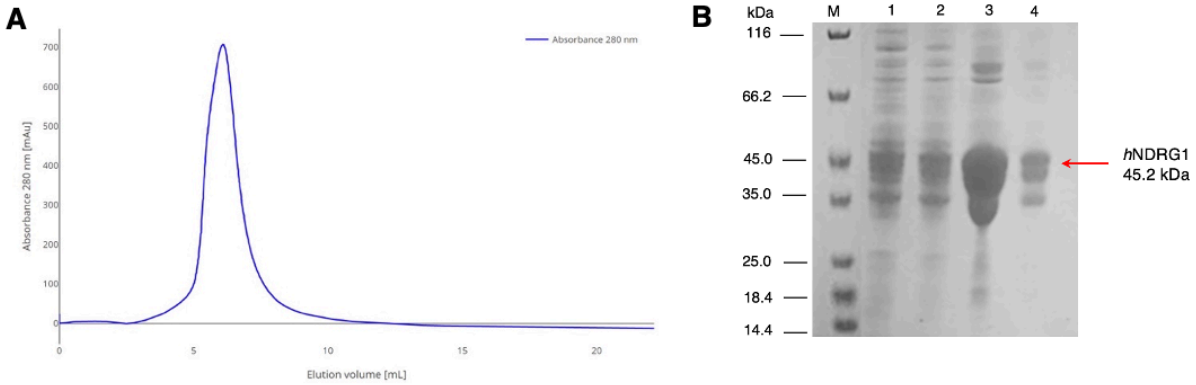


Figure 10: *StrepTrap™ HP elution of hNDRG1 (**A**). SDS-Page, from the left to the right: lane M marker, lanes 1-4 fractions of the StrepTrap™ HP column eluted with 2.5 mM desthiobiotin (**B**) .*

The elution profile of the protein in the size exclusion chromatography showed two elution peaks, both containing, according to the SDS-page, the protein of the expected molecular weight (Figure 11). The initial hypothesis, later verified by light scattering (Section 3.2.4), was that different oligomeric forms of this protein coexist in solution. Considering the difficulties in separating the different eluted species, fractions containing *h*NDRG1 were combined and concentrated, resulting in a final yield of 6-8 mg L^{-1} of initial culture.

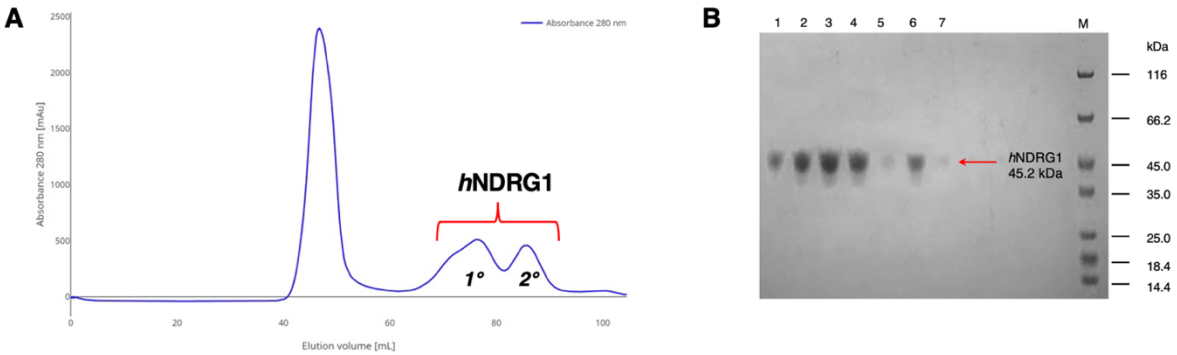


Figure 11*: Size exclusion chromatography of hNDRG1 performed on a Superdex 200 16/60 (**A**). SDS-Page, from the left to the right: lanes 1-4 fractions of the first peak, lanes 5-6-7 fractions of the second peak and lane M marker (**B**).*

The C-terminal domain of NDGR1 *(h*NDRG1*C) was previously expressed as N-terminal Strep-Tagged protein. Unfortunately, expression levels were low and a different expression approach was needed. For improving the yields, the solubility and the conformational stability of the protein, the best approach resulted to be the *h*NDRG1*C expression in *E. coli* BL21-CodonPlus (DE3) RIL competent cells as His-Tag/ZZ-tag/Strep-Tag/*h*NDRG1*C fusion polypeptide (27.98 kDa) and induction with IPTG. The fusion polypeptide and its single and double cysteine variants, used in EPR experiments (Section 3.2.6), were purified from the soluble fraction of the cellular extract with a Ni(II)-based affinity chromatography (Figure 12).

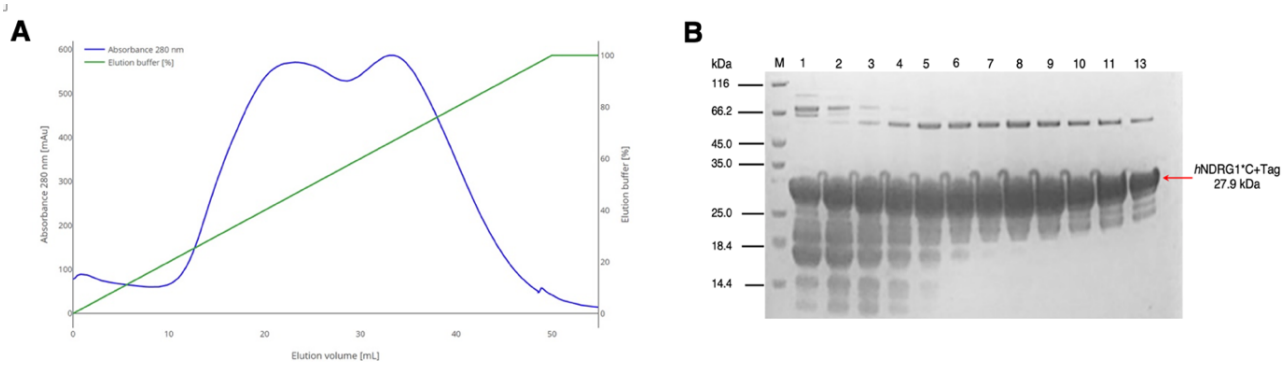


Figure 12: *Elution of hNDRG*C from the HiTrapChelating HP column pre-packed with 100 mM $NiSO_4$. The elution was performed with a gradient from 20 to 500 mM of imidazole (**A**). SDS-Page, from the left to the right: lane M marker, lanes 1-13 fractions of the elution of the HiTrapChelating HP column (**B**).*

According to the SDS-Page, the fractions containing the polypeptide were pooled and incubated with TEV protease for removing the tag, leaving the C-terminal sequence with the addition of a N-terminal Gly-His dipeptide. Subsequently, the His-Tag/ZZ-tag/Strep-Tag polypeptide was separated from the protein of interest by cation exchange chromatography and a gradient from 0 to 1 M NaCl. Two major peaks were present: one around 14% of buffer B (140 mM NaCl) with tag and uncut fusion polypeptide and another around 20% of buffer B (200 mM NaCl) containing *h*NDRG1*C (Figure 13).

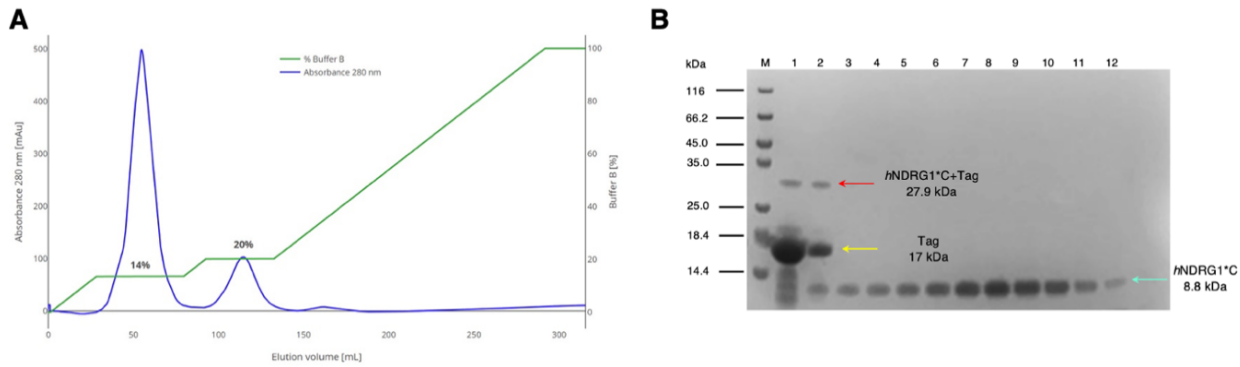


Figure 13*: Cation exchange chromatography of hNDRG1*C with the peaks corresponding to the tag (14%) and to the protein of interest highlighted (20%)(**A**). SDS-Page of the fractions of the cation exchange chromatography, from*

the left to the right: line M marker, lanes 1 and 2 fractions of the peak at 14% of buffer B, lanes 3-12 fractions of the peak at 20% of buffer B.

Final *h*NDRG1*C (8.8 kDa) polishing was obtained with size exclusion chromatography (SEC). The protein eluted in a single peak between 75 and 85 mL, reaching a final yield of 8-10 mg L^{-1} of initial culture (Figure 14).

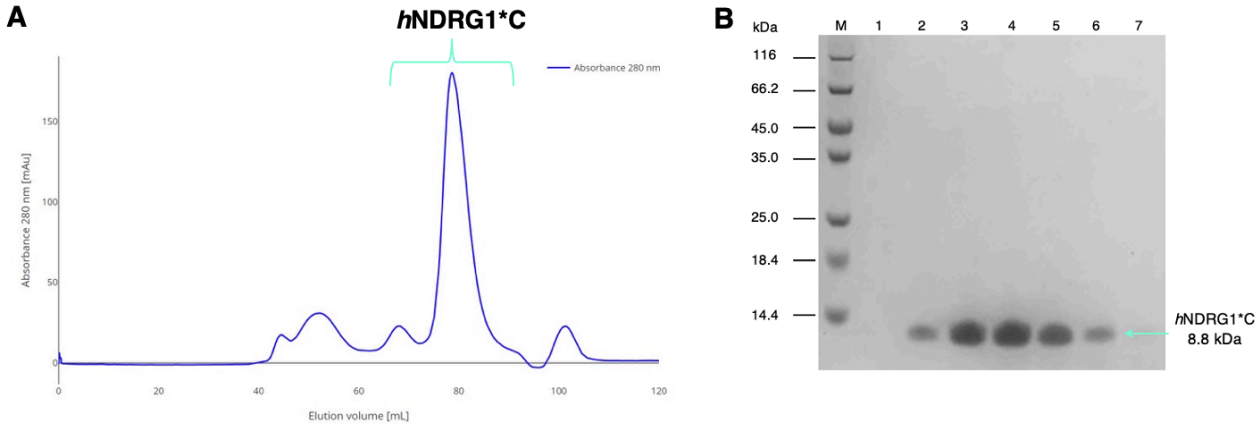


***Figure 14**: Size exclusion chromatography of hNDRG1*C performed on a Superdex 75 16/60 (**A**). SDS-Page, from the left to the right: lane M marker, lanes 1-7 fractions of the peak eluted between 75 and 85 mL corresponding to hNDRG1*C (**B**).*

3.2 Structural and biophysical characterization of NDRG1

3.2.1 NDRG1 structure and disorder prediction

The disorder prediction performed on the canonical sequence of *h*NDRG1 (Uniprot accession number Q92597-1), using the Database of Disordered Protein Predictions (D^2P^2) (https://d2p2.pro/) (Oates et al. 2013), confirmed the presence of two disordered domains: a small region at the N-terminal end of 30 residues (residues 1-30) and a longer region, 83 residues, in the C-terminal region (*h*NDRG1*C, residues 312-394), separated by a well-structured α/β hydrolase domain (Figure 15A) (Mustonen et al. 2020). The disordered C-terminal region is unique among NDRG proteins, as confirmed by the NDRG sequences alignment (Figure 15B). In addition, this region presents a nickel binding site (Zoroddu et al. 2004; Zoroddu et al. 2009) (Figure 15B) and some residues undergoing functional phosphorylation (Fang et al. 2014). The charge-hydropathy plot (CH- plot), where the absolute mean net charge is plotted against the mean scaled hydropathy, calculated from the hydropathy scale developed by Kyte-Doolittle (Huang et al. 2014; Uversky, Gillespie, and Fink 2000), confirmed that the full-length protein is a well-structured protein, conversely to the C-terminal domain, falling in the region of polypeptides with an intrinsically disordered behavior (Figure 15C).

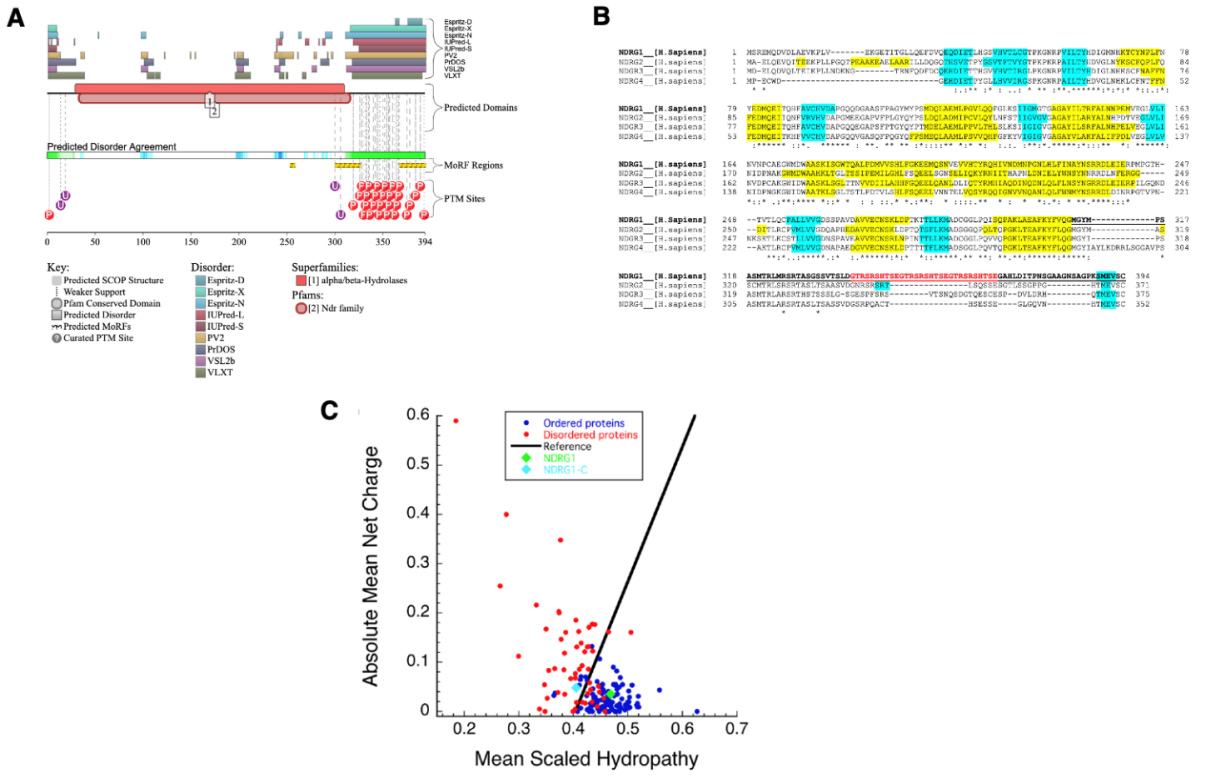


Figure 15: *Disorder prediction analysis performed with D^2P^2 database (https://d2p2.pro/) (**A**). hNDRG protein sequences alignment carried out with the PROMALS3D multiple sequence and structure alignment server (http://prodata.swmed.edu/promals3d/promals3d.php) (Pei, Kim, and Grishin 2008), showing in bulk the C-terminal domain of NDRG1 and in red the three-fold repeated sequence. Secondary structure prediction is reported, with in yellow α-helices and in cyan β-strands (**B**). CH-plot where data for ordered and disordered proteins are shown as blue and red circles respectively, whereas the positions of hNDRG1 and hNDRG1*C are indicated in green and cyan squares. The black line represents a boundary between compact and extended proteins. The Predictor of Natural Disordered Regions (PONDR®)(http://www.pondr.com/) was used for the analysis (**C**).*

3.2.2 Nickel binding activity of NDRG1

The isothermal titration calorimetry (ITC) technique was applied to investigate the nickel binding activity of *h*NDRG1 and *h*NDRG1*C, reported in literature and previously assumed (Fang et al. 2014; Zoroddu et al. 2004; Zoroddu et al.2009).

Both the full-length protein and its C-terminal domain bind nickel with and exothermic reaction, as shown by the negative peaks, following metal addition. Two inflection points resulted in the binding isotherm of the Ni(II) titration on *h*NDRG1, suggesting the presence of two sets of binding sites, one of which with higher affinity for the metal. Consequently, two binding events occurred. The calorimetric data were: $N_1 = 0.4$ and $N_2 = 2$, $K_{a1} = 6 \pm 4 \times 10^5$ ($K_{d1} = 1.7 \pm 1.13$ μM), $\Delta H_1 = -5 \pm 2$ kcal mol^{-1}, $\Delta S_1 = +13.1$ cal mol^{-1} K^{-1}, and $K_{a2} = 4 \pm 2 \times 10^4$ ($K_{d2} = 25 \pm 12.5$ μM), $\Delta H_2 = -2.4 \pm 0.80$ kcal mol^{-1}, $\Delta S_2 = +10.7$ cal mol^{-1} K^{-1}(Figure 16A and 16B).

The titration of Ni(II) on *h*NDRG1*C (Figure 16C) produced a binding isotherm with a single inflection point. It was fitted with a single set of site model, giving results very similar to the higher affinity site observed for *h*NDRG1, suggesting that this protein region is responsible for this binding event: N = 0.8, $K_a = 3.3 \pm 0.1 \times 10^5$ ($K_d = 3.0 \pm 0.1$ µM), $\Delta H = -5.5 \pm 0.4$ kcal mol^{-1}, $\Delta S_1 =$ +6.91 cal mol^{-1} K^{-1} (Figure 16D). The pH dependence of the affinity of the C-terminal for nickel was investigated in the pH range 6.5-8.5, demonstrating that the affinity increases at higher pH values (Figure 16E).

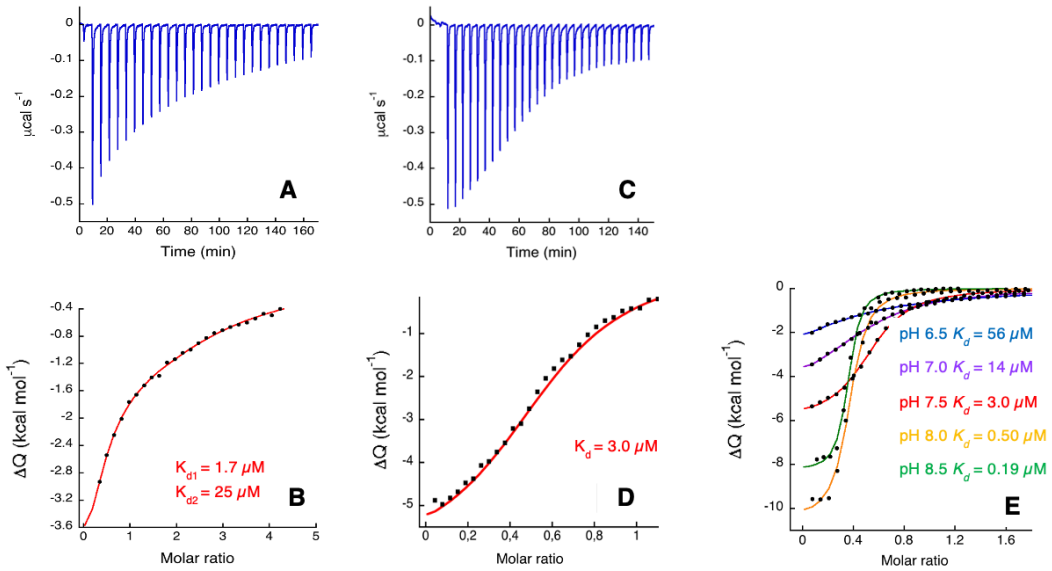


Figure 16: *ITC titration data for the binding of Ni(II)) to hNDRG1 (**A**,**B**) and to hNDRG1*C (**C**, **D, E**) in 20 mM Hepes, 150 mM NaCl, 1 mM TCEP. Raw titration data (**A**, **C**) represent the thermal effect of 10 µL injections of metal ion solution (200-600 µM) into protein solution (23-70 µM) at pH 7.5. Normalized heat reaction data for the binding of Ni(II) to hNDRG1 at pH 7.5 (**B**) or hNDRG1*C at pH 7.5 (**D**) and in the pH range between 6.5 and 8.5 (**E**) were obtained integrating raw data and subtracting the ligand heat of dilution into the buffer. The best fits of the integrated data, continuous lines, were obtained applying the two set of independent sites model for hNDRG1 (**B**) and the single site model for hNDRG1*C (**D, E**).*

3.2.3 NDRG1 secondary structure

The *h*NDRG1 CD spectra, recorded in this work in the presence and in the absence of Ni(II), are very similar to those previously reported (Mustonen et al. 2020). The quantitative analysis, performed using BestSel, hints to the prevalence of α-helices (25%), while β-strands relative amount are minor (19%) (Figure 17A). These data are in accordance with the secondary structure quantity extracted from the crystal structure (25% α-helices and 12.5% β-strands) (Mustonen et al. 20202 doi: 10.2210/pdb6zmm/pdb). Contrarily, the CD spectrum of *h*NDRG1*C is typical of an intrinsically disordered protein, with a pronounced negative peak around 198 nm (Figure 17B). The quantitative analysis reported 10.7% α-helices, 13.8% β-strands, 20.5% turns and 55% random

coil. The same analysis was performed in the presence of Ni(II), whose addition slightly influences the proteins secondary structure (Figure 17A and 17B). The absence of structure was confirmed also by the analysis performed using the online tool CAPITO (https://capito.uni-jena.de/) (Wiedemann, Bellstedt, and Görlach 2013). This analysis, showing the ratio between ellipticity at 200 nm and at 222 nm, placed *h*NDRG*C in the pre-molten globule region (Figure 17C), typical of intrinsically disordered proteins containing few elements of secondary structure. Differently, *h*NDRG1 fall in the region of the plot typical of well-folded proteins (Figure 17C).

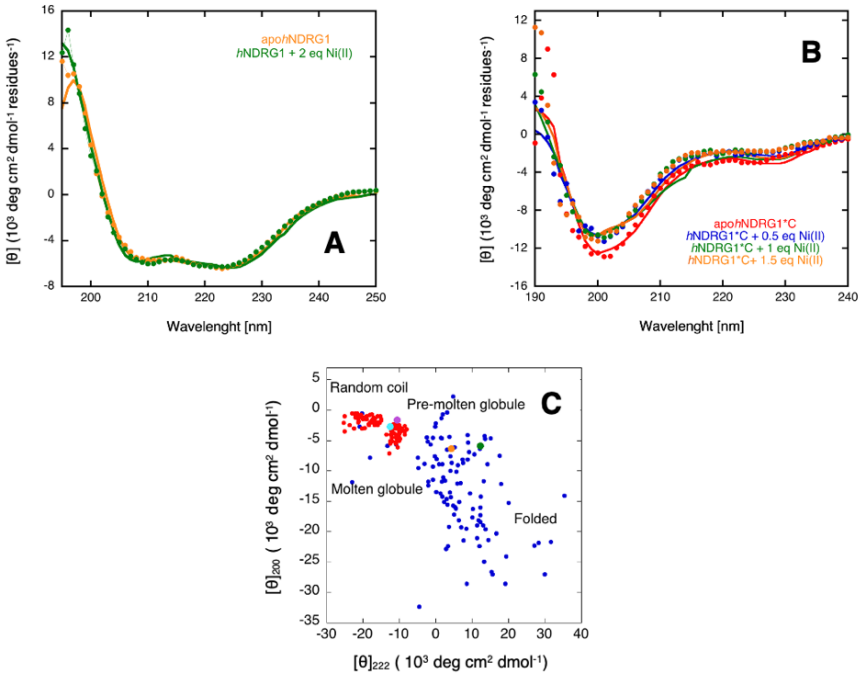


Figure 17: *Far-UV CD spectra of hNDRG1 (30 µM) in the absence (orange) and in the presence of two equivalents (green) of Ni(II) (**A**). Far-UV CD spectra of hNDRG1*C (200 µM) in the absence (red) and in the presence of half (blue), one (green) and one and a half equivalents (orange) of Ni(II) (**B**). Analysis of structural content of apo (orange) or holo (green)-hNDRG1 and of apo (cyan) or holo (violet)-hNDRG1*C according to its CD spectrum performed with the software CAPITO (https://capito.uni-jena.de/) (**C**).*

The CD signals in the near-UV region are sensitive to the overall tertiary structure of the protein in proximity of the aromatic residues. *h*NDRG1wt shows signals in the 250-270 nm range, probably referred to the ten phenylalanine moieties. The addition of two Ni(II) equivalents to *h*NDRG1wt increased the band intensity, suggesting a gain of rigidity (Figure 18A). The absence of bands in the 270-290 nm and 280-300 nm ranges, that could be attributed to the nine tyrosine and to the three tryptophan residues, respectively, suggests that these residues are located in asymmetric or flexible regions of the protein. *h*NDRG1*C has a single tyrosine as aromatic residues. The near-UV signal nearly zero confirms that a defined three-dimensional structure lacks in the C-terminal domain (Figure 18B).

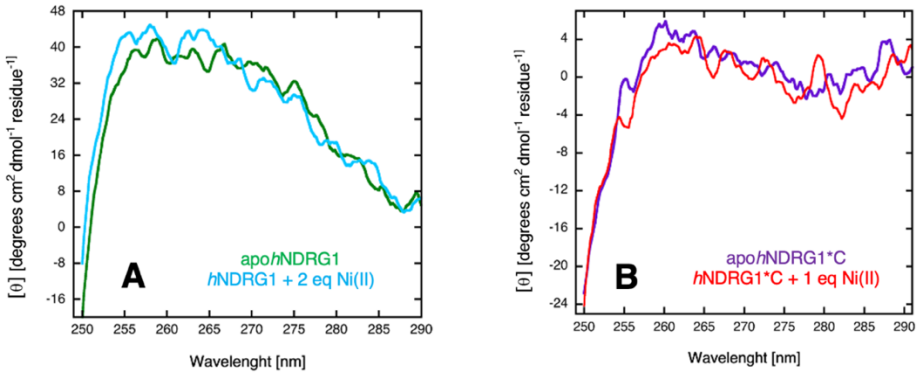


Figure 18: *Near-UV CD spectra of hNDRG1 (31 µM) in the absence (green) and in the presence of two equivalents (cyan) of Ni(II) (**A**). Near-UV CD spectra of hNDRG1*C (900 µM) in the absence (purple) and in the presence of one equivalent (orange) of Ni(II) (**B**).*

Mutations made on *h*NRG1*C, to obtain single cysteine variants to be used in EPR spectroscopy experiments (Section 3.2.6), did not affect the secondary structure of the protein. A small and not particularly pronounced change was noted only for the S378C_C394A variant, which showed a higher content in β-strands, as shown by the more positive transitions at approximately 192 nm and 218 nm (0.8% α-helices, 29.9 % β-strands, 14.4% turns, 54.8 % random coil) (Figure 19).

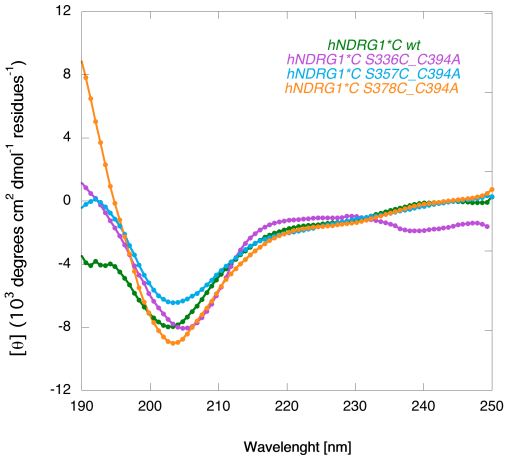


Figure 19: *Far-UV CD spectra of hNDRG1*C wild type (green) and single cysteine variants: S336C_C394A (purple), S357C_C394A (cyan) and S378C_C394A (orange). The protein concentration is 200 µM.*

TFE is an additive known to increase the content of α-helices. In the presence of 30% TFE, a more pronounced negative transition at 222 nm was observed, both on the native protein and on the single cysteine variants. This indicates an increase in the component in α-helices and a rearrangement of the composition in secondary structure elements (Uversky 2009). Under these conditions CD analysis calculated: 18.2% α-helices, 6.2% β-strands, 18.1% turns, 57.6% random coil (native *h*NDRG1*C), 13.1% α-helices, 18.1% β-strands, 17.7% turns, 51.1% random coil (S336C_C394A variant), 15.9% α-helices, 18.3% β-strands, 17.8% turns, 47.9% random coil (S357C_C394A

variant), 16.3% α-helices, 18.2% β-strands, 18.8% turns, 46.7% random coil (S378C_C394A variant) (Figure 20).

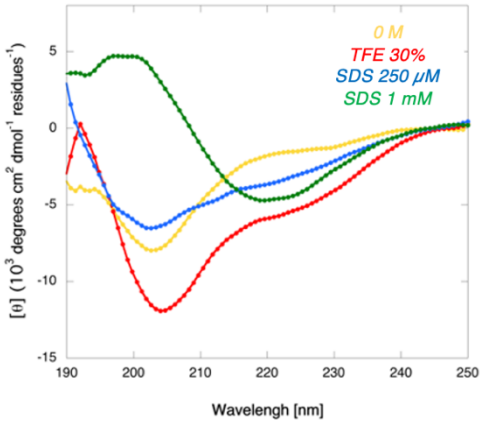


Figure 20: *Effects of TFE and SDS followed by CD on the native hNDRG1*C. Concentration of additives tested: 0M (yellow), 30% TFE (red), 250 µM SDS (blue), 1 mM SDS (green). Similar curves were obtained for S336C_C394A, S357C_C394A and S378C_C394A variants of hNDGR1*C and therefore not shown.*

Sodium dodecyl sulfate (SDS) is often used to stabilize protein structures in IDPs. Titration of SDS to protein solutions resulted in a gain in secondary structure elements. The maximum effect was measured in the presence of 1 mM SDS, a value higher than the calculated critical micellar concentration (CMC) for the detergent under these buffer conditions (CMC = 0.43 mM in 20 mM Hepes pH 7.5, 150 mM NaCl, 1 mM TCEP) (Figure 21).

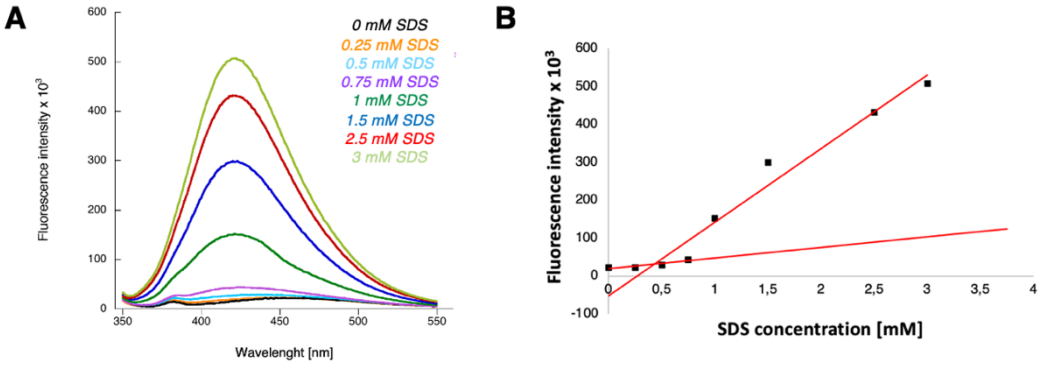


Figure 21: *Fluorescence emission spectra of 1 µM NPN in 20 mM HEPES pH 7.5 containing 150 mM NaCl and 1 mM TCEP in the presence of 0 mM, 0.25 mM, 5 mM, 0.5 mM, 0.75 mM, 1 mM, 1.5 mM, 2.5 mM and 3 mM SDS at room temperature (**A**). Fluorescence emission intensity of NPN as a function of SDS concentration. The point of intersection of the two straight lines corresponds to the critical micellar concentration of SDS in this buffer condition (CMC = 0.43 mM)(**B**).*

The change in secondary structure affects both the α-helical content, which decreases, and the β-strands content, which increases. This is true for the native protein and for the S336C_C394A and S357C_C394A variants. For the S378C_C394A variant both components increase. The secondary structure calculated in the presence of 1 mM SDS by circular dichroism was found to be: 0.6% α-

helices, 43.4% β-strands, 17.6% turns, 38.5% random coil (native *h*NDRG1*C), 0% α-helices, 54.1% β-strands, 17% turns, 29% random coil (S336C_C394A variant), 0% α-helices, 47.3% β-strands, 15.8% turns, 36.9% random coil (S357C_C394A variant), 5.4% α-helices, 39.8% β-strands, 17% turns, 37.9% random coil (S378C_C394A variant) (Figure 20).

3.2.4 NDRG1 oligomeric state

The multiple-angle light scattering (MALS) and quasi-elastic light scattering (QELS) were combined with size-exclusion chromatography column (SEC) for investigating the hydrodynamic and oligomeric properties of *h*NDGR1 and *h*NDRG*C. The results indicate that *h*NDRG1wt elutes in two peaks, corresponding to different oligomeric forms of the protein in solution and with the following molecular masses and hydrodynamic radii: MW = 150 kDa and R_h = 5.3 nm for the first eluted species, MW = 90.5 kDa and R_h = 4 nm for the second one and MW = 47.4 kDa and R_h = 2.4 nm for the last one. (Figure 22A). The first eluted species has a MW value intermediate between the molar mass of the tetramer (180 kDa) and that of the trimer (135 kDa). Considering the profile of the chromatogram, indicating a superimposition of this species with one at lower MW, an underestimation of the calculated molecular weight can be supposed, leading to the assertion that it corresponds to the tetrameric form. Instead, the others eluted species match with the dimer and the monomer respectively. Conversely, a single elution peak, equivalent to the monomeric form was observed for *h*NDRG1*C (MW = 10.5 kDa, R_h = 1.7 nm) (Figure 22B). The results were similar when one equivalent of Ni(II) for protein monomer was added to *h*NDRG1wt (MW = 145 kDa, R_h = 4.8 nm; MW = 92 kDa, Rh = 3.7; MW = 47 kDa, Rh = 2.3 nm) or *h*NDRG1*C (MW = 10.5 kDa, R_h = 1.2 nm), demonstrating that the addition of Ni(II) ions does not significantly modify the oligomeric conformation.

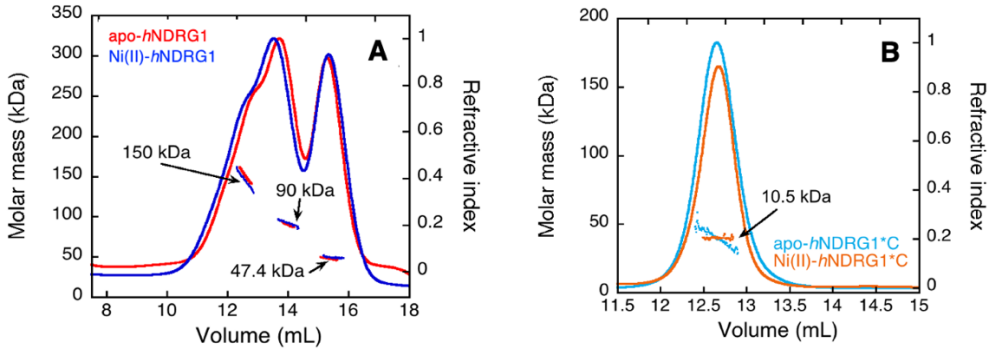


Figure 22: *Size exclusion profiles monitored by the refractive index detector (lines) and weight-averaged molar mass distribution (dots) of hNDRG1 (300 µL, 140 µM) (**A**) and hNDRG1*C (300 µL, 1.2 mM) (**B**) obtained with a combination of SEC–MALS–QELS in the absence (red and cyan) and in the presence of one equivalent of Ni(II) (blue and orange).*

3.2.5 NMR spectroscopy

2D and 3D high resolution NMR spectra were acquired to perform the assignment of the NMR signals of *h*NDRG1*C. The spectra were recorded in 20 mM HEPES, 150 mM NaCl, 1 mM TCEP both at pH 7.5 and 6.5. The quality of the spectrum at pH 7.5 was very low, with few visible signals, due to a greater exchange of protons between protein and solvent. The visible signals were just those corresponding to portions of the protein less exposed, having a greater secondary structure propensity (Figure 23A). On the contrary, 82 signals were visible at pH 6.5 and therefore subsequent NMR spectroscopy analyses were performed at this pH value. Most of the signals fell in the random coil region, with a ^{1}H chemical shift dispersion between 8.0 and 8.6 ppm, as expected from an intrinsically disordered protein (Kosol et al. 2013). 76 signals of the 82 visible ones were assigned (92.7%) and numbered from 1 to 85, with Met3 and Cys85 corresponding to Met312 and Cys394 in *h*NDGR1. The peptide also contains Gly1 and His2, resulting from the protease digestion because of the cloning protocol (Figure 23B).

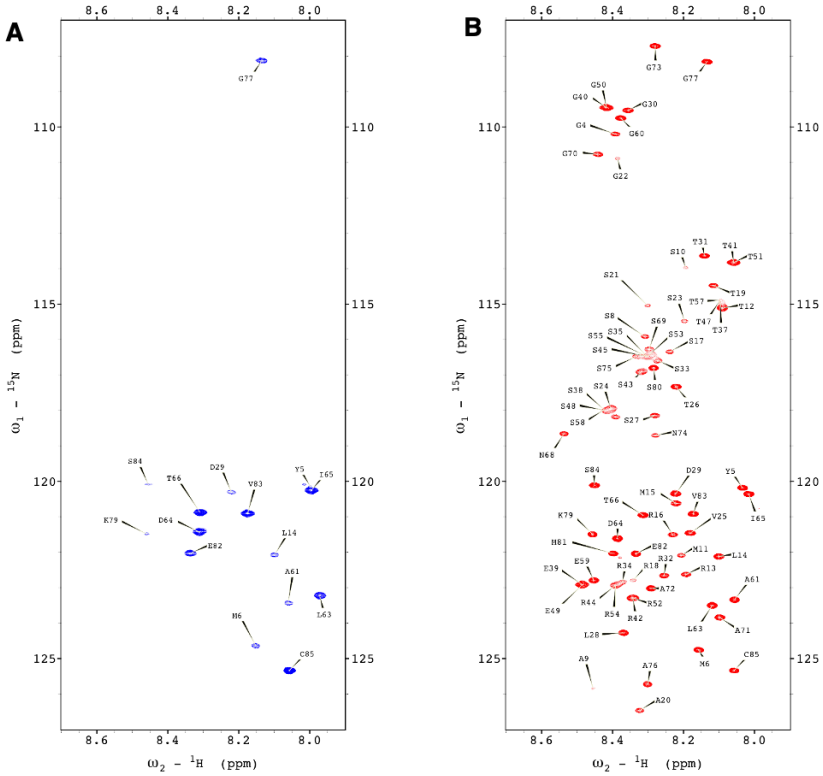


Figure 23: *^{1}H-^{15}N HSQC spectra of hNDGR1*C acquired on a 1.2 GHz NMR spectrometer at 298 K at a concentration of 0.8-1 mM in 20 mM Hepes 150 mM NaCl, 1 mM TCEP at pH 7.5 (**A**) and pH 6.5 (**B**). Each assigned peak is indicated by a letter, corresponding to the single-letter amino acid code and the corresponding residue number in the protein sequence. The non-native GH sequence is included. At pH 7.5 there is a disappearance of signals due to increased proton exchange between protein and solvent. The visible signals correspond to the portions of the protein that are less exposed to the solvent, as they are located in areas of higher secondary structure propensity.*

The unassigned NH signals include Gly1, His2, His36, His46, His56 and His62, not observable because of fast proton exchange with water, process demonstrated to significantly increase in the presence of a protonated histidine imidazole ring (Molday, Englander, and Kallen 1972). Considering the possible importance of histidine residues in Ni(II) binding, a combination of different NMR spectra (^{1}H-^{1}H TOCSY and ^{1}H-^{13}C HSQC spectra in the aromatics region, CBHD heteronuclear spectra, ^{13}C-detected CON, CACO and CBCACO experiments) was used to assign these residues. The measured chemical shift values (CE1 ~138 ppm and CD2 ~120 ppm) indicate a doubly protonated state of all imidazole rings at pH 6.5 (Kjaergaard, Brander, and Poulsen 2011). Met3 signal was not observed, probably because its NMR signal is broadened beyond detection, due to conformational exchange phenomena occurring with rates comparable to the frequency differences among the different conformers. Pro7, Pro67, and Pro78 signals are characteristically missing in the ^{1}H-^{15}N HSQC spectrum, but their signals were observed in triple resonance experiments. The deviation of the chemical shifts of CA, CB, and HB from those generated by random coil, was used to establish secondary structure propensity (SSP) (Wishart and Sykes 1994) using the TALOS-N algorithm (Shen and Bax 2013), which confirmed that *h*NDRG1*C in solution is largely disordered, with the exception of the region between residues 8 and 18 near the N-terminus, which exhibits a small but significant propensity to assume an α-helix conformation. The conformational flexibility of the main chain, measured as the amplitude of the movements of the amide groups of the "backbone" in the ps-ns to μs time scale, was predicted empirically using TALOS-N and following the Random Coil Index (RCI) (Berjanskii and Wishart 2005) approach and reported as S^2 order parameter, where S^2= 1 corresponds to fully restricted movements, while S^2 = 0 to unrestricted movements. The calculated average value of S^2 was 0.37 ± 0.10 (Figure 24), indicating the presence of a conformationally flexible polypeptide. This data is in line with the predictions made using the D^{2}P^2 web server (Figure 15) and the CD spectrum (Figure 17).

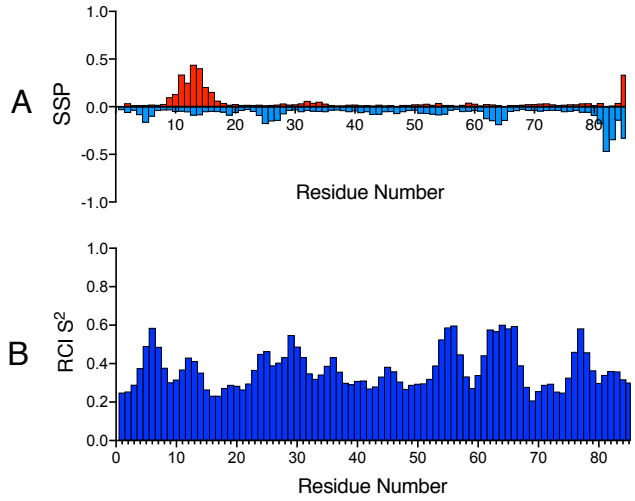


Figure 24: *Secondary structure propensity (SSP) of hNDRG1*C derived from NMR backbone chemical shifts using TALOS-N (+1 equals to 100% helical propensity, red; −1 equals to 100% beta-sheet propensity - cyan) plotted vs. residue number (**A**). Random Coil Index (RCI) order parameters S^2 predicted by TALOS-N plotted vs. residue number (**B**).*

The quality of a NMR spectrum is strictly dependent on the conditions of the protein sample and factors such as temperature, salt concentration and pH can influence the results. For example, the pH dependence of random coil chemical shifts of histidine residues is known (Kjaergaard et al. 2011). For this reason, the pH-induced signals modifications were followed by increasing the pH in steps of 0.25 pH units in the 6.5-7.5 range, allowing the assignment of the signals at pH 7.5 and monitoring Ni(II) binding effects. Progressive additions of $NiSO_4$, from 0.25 to 1.5 equivalents, changed only the signals of Leu63 and Asp64 in the ^{1}H-^{15}N HSQC spectrum recorded at pH 6.5. At this pH value, the binding of Ni(II) to histidines, monitored by ^{1}H-^{13}C HSQC spectra in the aromatic region, resulted in progressive small shifts of the signals of all histidine residues, indicating a continuous transition from the bound to the unbound state. This is consistent with the pH dependence of the binding constant as measured by calorimetry (Figure 16D), indicating a relatively weak binding at pH 6.5 (K_d = 65 μM). The binding of Ni(II) to *h*NDGR1*C was monitored also at pH 7.5. At this pH value, progressive disappearance of CE and HE signals of His2 and His62 was observed, as the Ni(II) equivalents increase. A similar decrease, but with a lower extent, was observed for His36, His46 and His56 (Figure 25). These preliminary data indicate that the five histidines are involved in metal ion coordination and further analysis are underway.

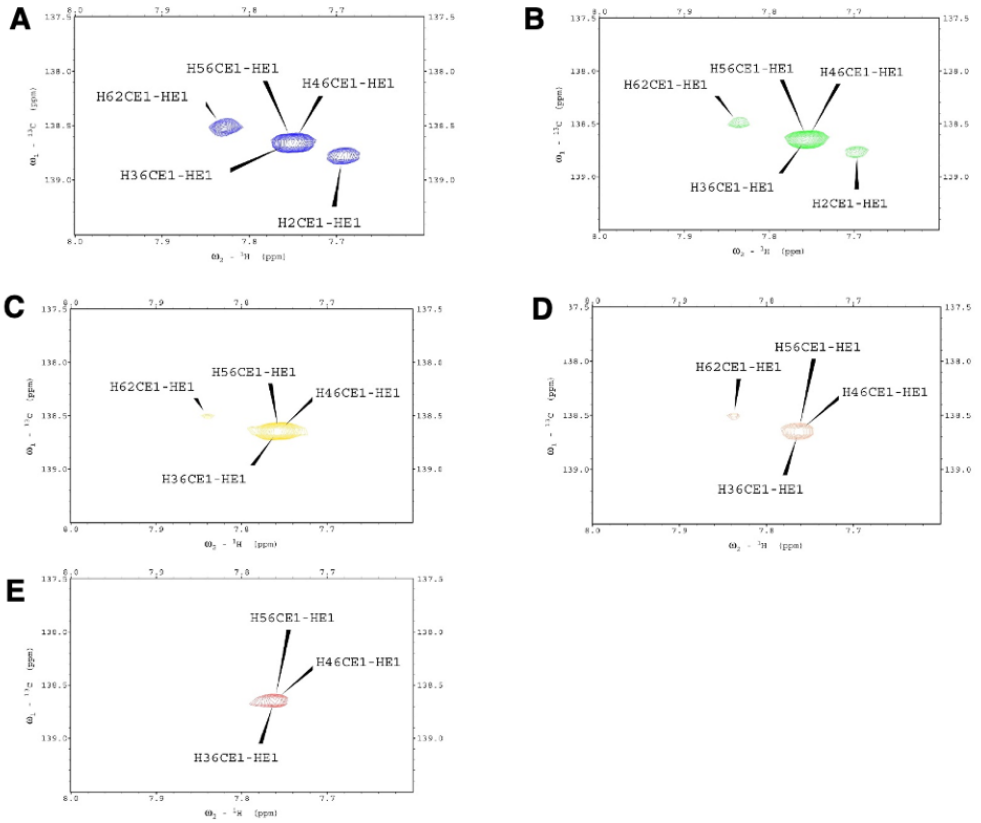


Figure 25: *^{1}H-^{13}C HSQC spectra of hNDRG1*C at a concentration of 800 µM in 20 mM Hepes, 150 mM NaCl, 1 mM TCEP at pH 7.5 in the absence of Ni(II)(**A**), + 0.25 Ni(II)(**B**), + 0.5 Ni(II)(**C**), + 0.75 Ni(II)(**D**), + 1 Ni(II)(**E**).*

3.2.6 EPR spectroscopy

Electron Paramagnetic Resonance (EPR) spectroscopy can be employed to characterize different paramagnetic species present in the biological macromolecules. This technique can be coupled to the site directed spin labeling approach (SDSL), based on the covalent attachment of a paramagnetic tag, called spin label, onto a specific position of a diamagnetic biomolecule. The resulting paramagnetic protein can be characterized through EPR, obtaining crucial information about protein structural dynamics, conformational changes, and interactions between biomolecules. SDSL coupled to EPR is a non-destructive and non-invasive technique that can be carried out at room temperature and in solution. Nitroxide-based spin labels, generally used as spin label, present high sensitivity to the local environment. The EPR spectrum of a nitroxide in water solution is composed by three lines arising from the Zeeman interaction (with the magnetic field) and the hyperfine interaction between the unpaired electron (S= ½) and the nitrogen nucleus (I=1). The shape of the EPR spectrum is strictly dependent on the mobility of the nitroxide and can be a sensitive reporter of the dynamics of the biomolecule region to which the nitroxide is attached. If the nitroxide mobility decreases, the EPR line shape becomes broader (*broad signal*), while if the nitroxide exhibits higher mobility the signal becomes sharper (*sharp signal*) (Figure 26) (Breton et al. 2015). Changes in the nitroxide mobility can be observed in the ns-µs time scale and therefore the SDSL-EPR technique is an excellent method to study the disorder-order transitions that occur in intrinsically disordered proteins (IDPs).

Several types of EPR spectroscopy can be applied to proteins. The main distinction can be made between continuous wave EPR spectroscopy (CW-EPR) or pulsed EPR. In CW-EPR, the sample is continuously irradiated by electromagnetic radiation at a fixed frequency. The X-band (9.9 GHz) is the most commonly used microwave frequency in CW-EPR. The Double electron-electron resonance (DEER) is the most used pulsed technique, allowing to obtain important information on the organization and structural transformations that occur in protein complexes. Requiring the presence of a protein with two nitroxide labels or two interacting mono-labeled proteins and working at cryogenic temperatures (60 K) in the Q-band (34 GHz), thanks to this technique it is possible to obtain a distance distribution between spin pairs and detect a distance range up to 8 nm (Drescher 2012).

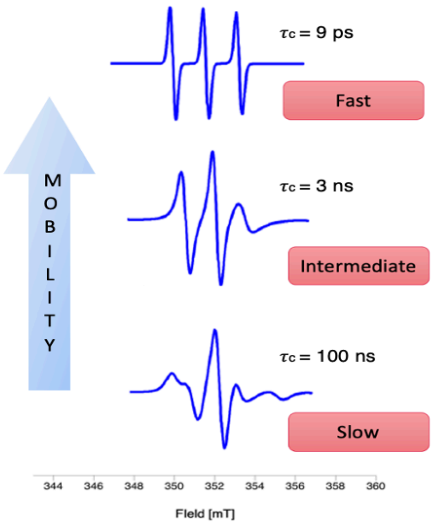


Figure 26: *EPR spectra of a nitroxide in different regimes of mobility. Spectra were recorded in CW-EPR and simulated with EasySpin.*

Both continuous wave EPR and DEER can be applied to perform structural studies in a cellular context (*in cell* EPR), where different factors can significantly modify the dynamics of the protein under study, compared to the *in vitro* conditions (Bonucci et al. 2020).

In this work, SDSL-EPR was applied to complete the structural characterization of the C-terminal domain. The Maleimido Proxyl (M-Prox) nitroxide was used as spin label and specifically attached to cysteine residues introduced by mutagenesis, replacing three serine residues not involved in the phosphorylation processes. The natural terminal cysteine was kept, obtaining three variants with two sites available for spin labeling to be used in DEER experiments (Ser336Cys/Cys394 ⇒ S336C^{Prox}_C394^{Prox}, S357Cys/Cys394 ⇒ S357C^{Prox}_C394^{Prox} and Ser378Cys/Cys394 ⇒ S378C^{Prox}_C394^{Prox}) or mutagenized in order to get single cysteine variants (Ser336Cys/Cys394Ala ⇒ S336C^{Prox}, Ser357Cys/Cys394Ala ⇒ S357C^{Prox} and Ser378Cys/Cys394Ala ⇒ S378C^{Prox}) for continuous wave (CW) and *in cell* EPR experiments. The labeling yields were 20-40% for both single-labeled and double-labeled *h*NDGR1*C variants. These values were calculated by comparing the spin concentration with the protein concentration, estimated by measuring the Abs at 280 nm. A possible explanation for these yields is that part of the protein aggregates during the labeling reaction steps and is no longer available to be labeled with the nitroxide probe. The efficiency of the labeling reaction was confirmed by mass analysis (MALDI-ToF).

3.2.6.1 *In vitro* CW-EPR and crowding effect

The EPR spectra of the single (Figure 27) and double (Figure 28) labeled variants in buffer solution are very sharp and similar to those already obtained for intrinsically disordered regions or proteins (Breton et al. 2015; Weickert, Cattani, and Drescher 2019).

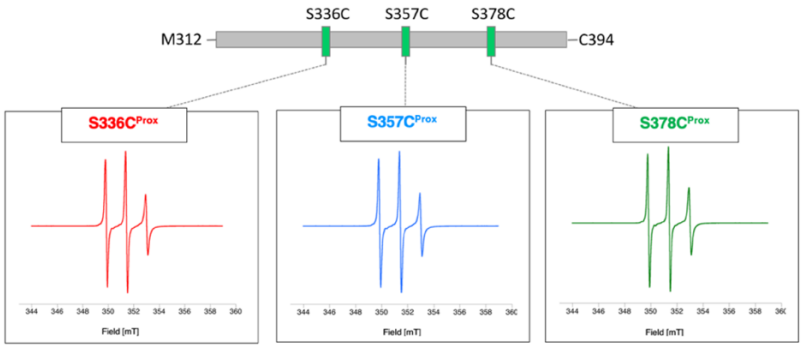


Figure 27: *Normalized CW-EPR spectra of hNDRG1*C single labeled variants with M-Prox nitroxide label. The spectra were recorded at the X-band (9.9 GHz) and at room temperature using the following parameters: microwaves power = 10 mW; magnetic field modulation amplitude = 0.1mT; field sweep =15 mT; receiver gain = 60 dB. The proteins are in 10 mM Tris pH 7.4.*

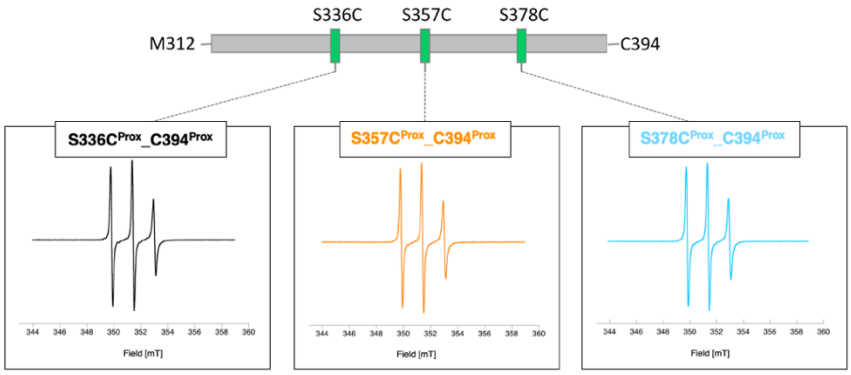
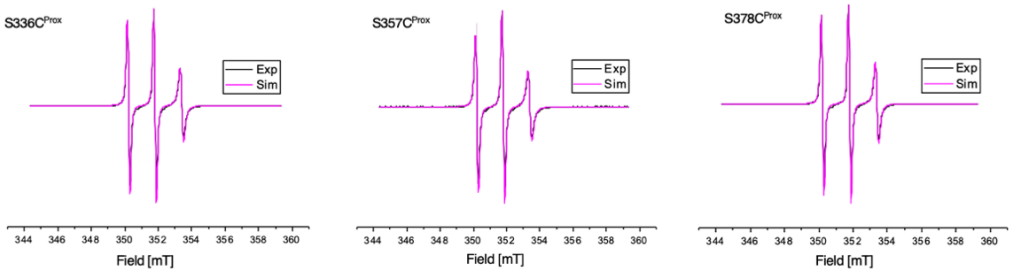


Figure 28: *Normalized CW-EPR spectra of hNDRG1*C double labeled variants with M-Prox nitroxide label. The spectra were recorded at the X-band (9.9 GHz) and at room temperature using the following parameters: microwaves power = 10 mW; magnetic field modulation amplitude = 0.1mT; field sweep =15 mT; receiver gain = 60 dB. The proteins are in 20 mM Tris pH 7.5 containing 200 mM NaCl, 1 mM TCEP.*

The nitroxide dynamics was quantitatively characterized by simulating the EPR spectra with Sim-Label program (Etienne et al. 2017), a MatLab graphical user interface of EasySpin software (Stoll and Schweiger 2006). This program allows to simulate and fit the EPR spectra, determining the presence of different components and for each of them calculating the magnetic parameters (g-factor and hyperfine A-tensors) and the dynamic parameter τ_c corresponding to the rotational correlation time (Torricella et al. 2021).

The simulation of the EPR spectra of the single cysteine variants revealed the presence of a single motility component for the single labeled variants: τ_c = 0.32 ns, τ_c = 0.31 ns and τ_c = 0.27 for S336C^{Prox}, S357C^{Prox} and S378C^{Prox} respectively. These results confirms the high flexibility and lack of secondary and/or tertiary structure in the region surrounding the label (Figure 29).



	S336C^{Prox}	S357C^{Prox}	S378C^{Prox}
τ_c	0,32 ns	0,31 ns	0,27 ns
g	2,00820 2,00671 2,00224	2,00781 2,00745 2,00218	2,00914 2,00512 2,00226
a_N (mT)	0,41 0,41 3,97	0,32 0,32 4,13	0,37 0,37 4,03
a_N iso (mT)	1,59	1,59	1,59
Lw (mT)	G:0,12 L:0,04	G:0,12 L:0,05	G:0,13 L:0,03

Figure 39: *Upper panel overlay of the CW-EPR spectra of hNDRG1*C single cysteine variants in 10 mM Tris pH 7.4 (black) and the corresponding simulated spectra. In the bottom panel the parameters resulting from the simulation are reported.*

The intracellular environment is crowded with macromolecules, reducing the volume (Bonucci et al. 2021; Gnutt et al. 2015) available to any macromolecule, and affecting their dynamics.

Different chemical additives, here identified as crowders, can mimic crowding effects in cell and work as fold-promoting agents (Dhar et al. 2010; Homouz et al. 2008), while osmolytes interact with protein backbone and surfaces, forcing highly unstructured protein in an aqueous environment to fold (Kumar 2009). In order to evaluate if the lack of secondary and tertiary structure could be related to the more diluted *in vitro* condition than the intracellular environment, EPR spectra were recorded in the presence of different crowders and osmolytes (Figure 30A). The results are reported below as the ratio of the peak amplitude h+1 and h0 of an EPR spectra (Figure 30B). This ratio can be used as a direct indicator of the spin label mobility, considering that lower ratio corresponds to lower mobility of the spin label, and therefore to a greater crowding effect (Belle et al. 2008).

A

Crowders and osmolytes	Concentration
Ficoll-PM70	300 mg/mL
BSA	200 mg/mL
Sucrose	30% w/v
Glycerol	30%
TFE	30%
TMA0	2M
Potassium glutamate	200 mM
SDS	2 mM, 4 mM, 8 mM

B

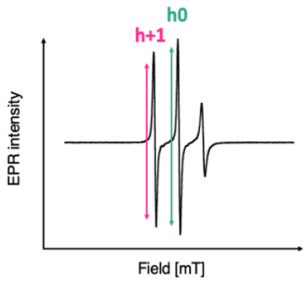


Figure 30: *Crowders and osmolytes used to mimic the cellular environment with indicated the concentrations tested (**A**). EPR spectrum of a macromolecule labeled with a nitroxide probe. The two peaks whose ratio of the amplitudes will be used to analyze the effect of different crowders and osmolytes on the structure of hNDRG1*C are denoted as h+1 and h0, respectively (**B**).*

The significance of this analysis is that all crowders have a little effect on the three positions studied, as demonstrated by the small changes observed in the ratios (Figure 31).

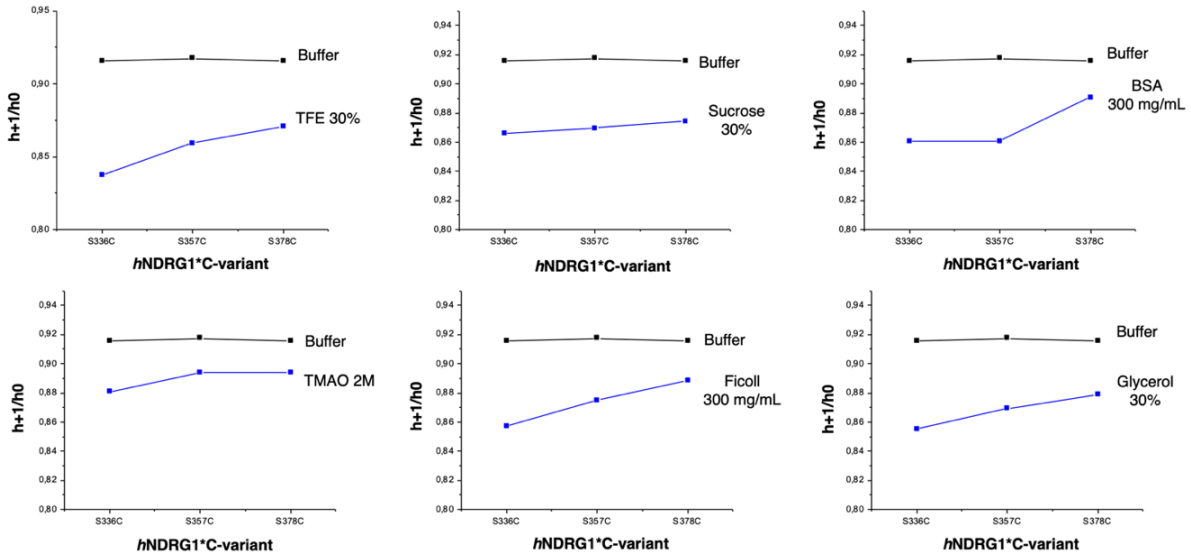


Figure 31: *Ratio of the amplitude of the peaks h+1/h0 of the three single cysteine variants of hNDRG1*C in buffer (10 mM Tris pH 7.4) (black lines) and in the presence of different crowding agents (blue lines).*

The most affected position is S336C^{Prox} and the crowder with the greatest effect is TFE, an α-helix promoting agent (Figure 32).

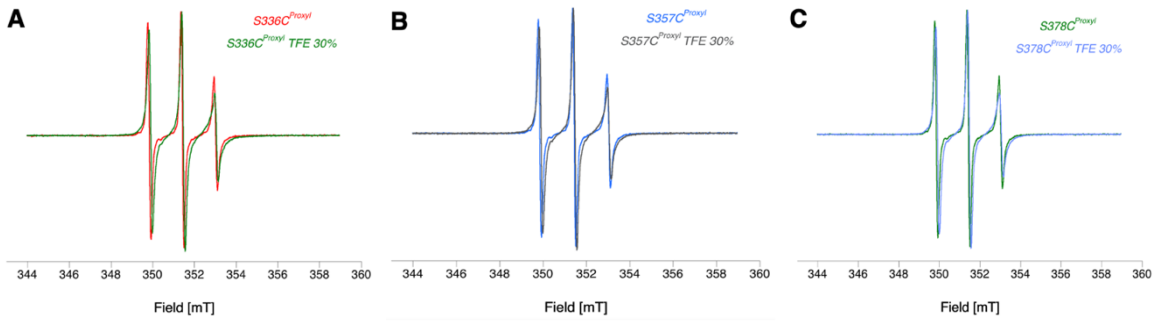


Figure 32: *Overlay of the normalized EPR spectra of S336C^{Prox} (**A**), S357C^{Prox} (**B**) and S378C^{Prox} (**C**) in buffer (10 mM Tris pH 7.4) and in the presence of 30% of trifluoroethanol (TFE). Spin concentration in the samples is 40 µM. The spectra were recorded at the X-band (9.9 GHz) using the following parameters: microwaves power = 10 mW; magnetic field modulation amplitude = 0.1mT; field sweep =15 mT ; receiver gain = 60 dB, 12 scans.*

Although NDRG1 is a predominantly cytoplasmic protein, its localization at the level of cell membranes is reported in literature, as well as its translocation from the cytoplasm to the nucleus in response to certain stimuli, working as a stress responsive gene. For this reason, the dynamics of the C-terminal domain was tested in the presence of increasing amounts of sodium dodecyl sulfate (SDS), in order to mimic the native lipid bilayer environment. The critical micellar concentration (CMC) calculated for SDS in working buffer conditions (10 mM Tris pH 7.4) is 3.83 mM (Figure 33).

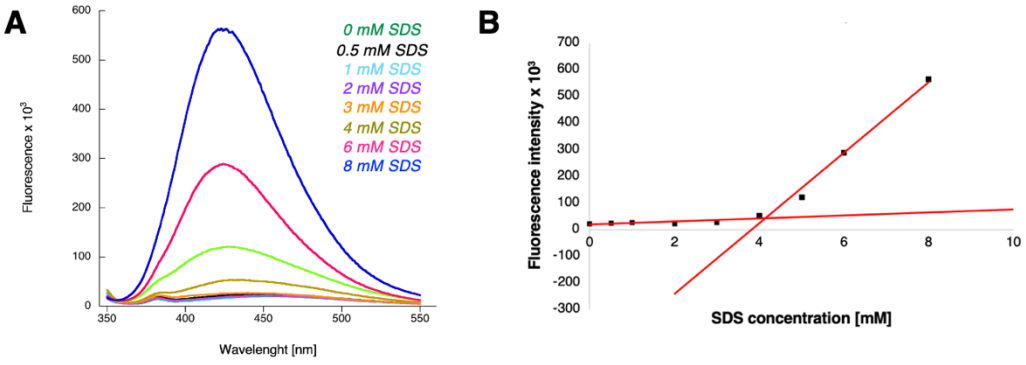


Figure 33: *Fluorescence emission spectra of 1 µM NPN in 10 mM Tris pH 7.4 in the presence of 0 mM, 0.5 mM, 1 mM, 2 mM, 3 mM , 4 mM, 6 mM and 8 mM SDS at room temperature (**A**). NPN fluorescence emission intensity as a function of SDS concentration. The point of intersection of the two straight lines corresponds to the critical micellar concentration of SDS in this buffer condition (CMC = 3.82 mM)(**B**).*

Considering the calculated CMC value, three concentrations of SDS were tested (2 mM, 4 mM and 8 mM) (Figure 34).

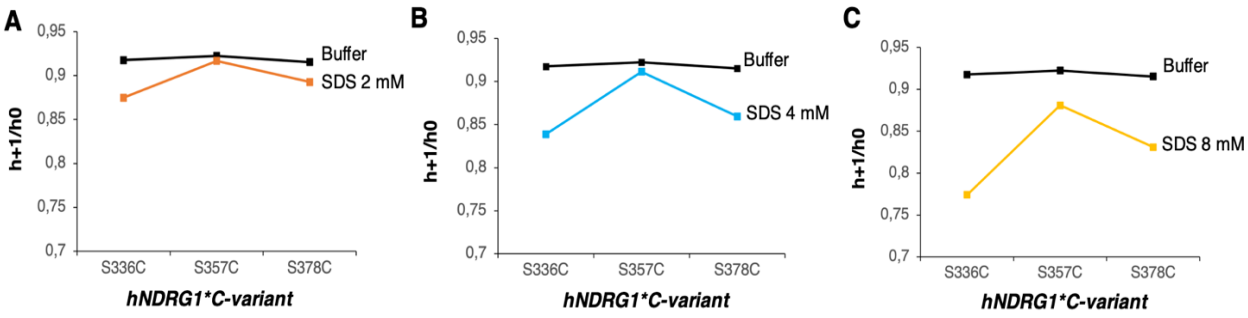


Figure 34: *Ratio of the amplitude of the peaks h+1/h0 of the three single cysteine variants of hNDRG1*C in buffer (10 mM Tris pH 7.4) (black lines) and in the presence of 2 mM SDS (orange lines)(**A**), 4 mM SDS (blue lines)(**B**) and 8 mM SDS (yellow lines)(**C**).*

Although the EPR spectra of all variants were found to be affected by the presence of this surfactant, with a maximum effect in the presence of 8 mM SDS, a concentration well above the calculated CMC (3.43 mM), the most sensitive variant was found to be again S336C^{Prox} (Figure 35).

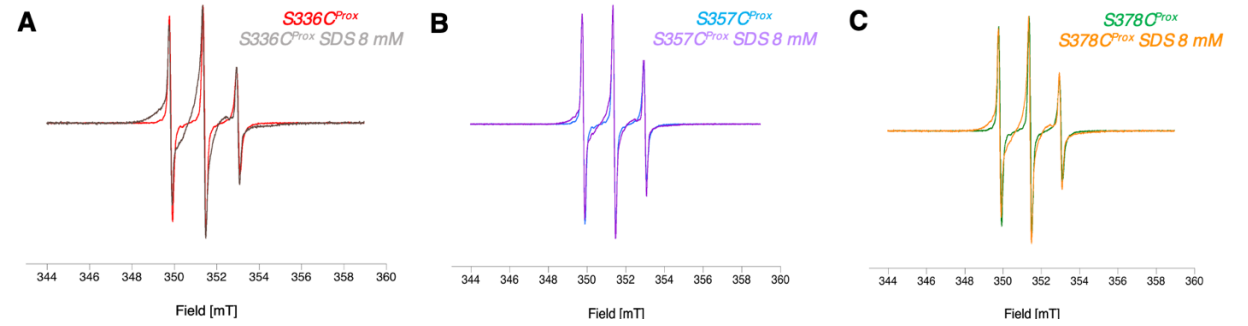


Figure 35: *Overlay of the normalized EPR spectra of S336C^{Prox} (**A**), S357C^{Prox} (**B**) and S378C^{Prox} (**C**) in buffer (10 mM Tris pH 7.4) and in the presence of 8 mM SDS, a concentration above the calculated CMC. Spin concentration in the samples is 40 µM. The spectra were recorded at the X-band (9.9 GHz) using the following parameters: microwaves power = 10 mW; magnetic field modulation amplitude = 0.1mT; field sweep =15 mT ; receiver gain = 60 dB, 12 scans.*

3.2.6.2 DEER

DEER data of the variant S336C^{Prox}_C394^{Prox} gave a distance distribution with 3 peaks centered at 3, 4.5 and 5 nm. Traces from the variants S357C^{Prox}_C394^{Prox}, including a spin-label in the residue 357 located in the supposed nickel binding sequence, and S378C^{Prox}_C394^{Prox} displayed a broad distance distribution with a single maximum at 3.5 and 3 nm respectively. These results confirmed that this region of *h*NDRG1 is highly dynamic, and a large number of different conformers are present in solution, as suggested by NMR studies (Figure 36).

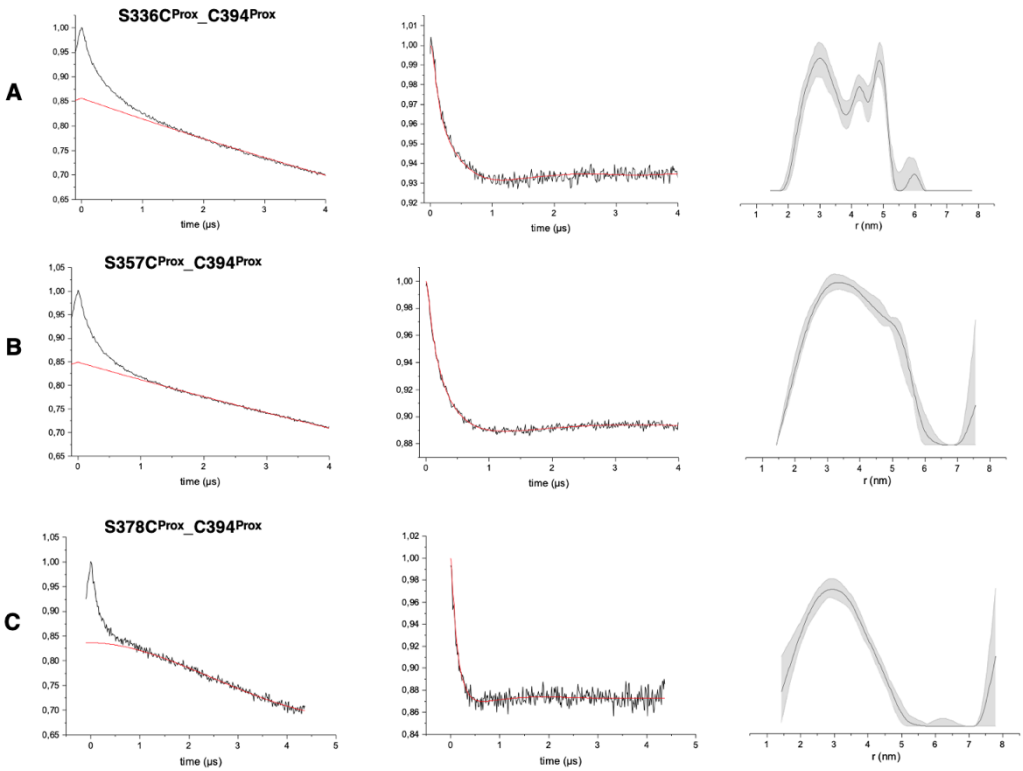


Figure 36: *Q-band DEER experiments at 60 K of the double labeled variants of hNDRG1*C: S336C^{Prox}_C394^{Prox} (**A**), S357C^{Prox}_C394^{Prox}(**B**) and S378C^{Prox}_C394^{Prox}(**C**) in 20 mM Tris pH 7.5, 200 mM NaCl, 30% Glycerol D_8. The spin concentrations of the samples were 112 µM, 92 µM and 90 µM respectively for S336C^{Prox}_C394^{Prox}, S357C^{Prox}_C394^{Prox} and S378C^{Prox}_C394^{Prox} variants. Left panel: experimental DEER traces recorded at 60 K. Red lines indicate the baseline used for background correction. Central panel: corrected DEER traces (black) with superimposed fits derived from Tikhonov regularization (red). Right panel: Tikhonov derived distance distributions obtained using DeerAnalysis2019 (black) superimposed with gray lines that are the full variation of the probability of given distances, whereas the thickness is respectively the upper and lower error estimation corresponding to the mean value ± two times the standard deviation of the different Tikhonov fits.*

The addition of TFE, one of the crowders with higher effect, does not significantly change the DEER spectra and, although more defined peaks have been observed, the prevalence of a population at a specific distance is still absent (Figure 37).

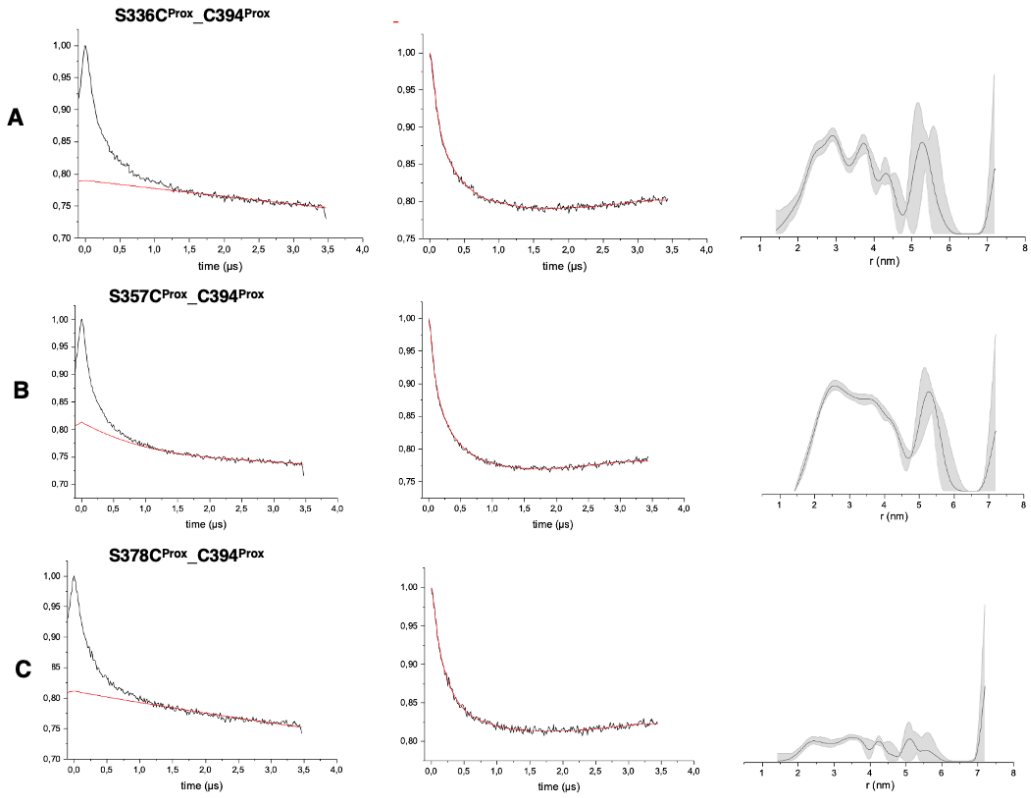


Figure 37: *Q-band DEER experiments at 60 K of the double labeled variants of hNDRG1*C in the presence of 30% TFE: S336C^{Prox}_C394^{Prox}* ***(A)****, S357C^{Prox}_C394^{Prox}* ***(B)*** *and S378C^{Prox}_C394^{Prox}* ***(C)*** *in 20 mM Tris pH 7.5, 200 mM NaCl, 30% Glycerol D_8 and 30% TFE . The spin concentrations of the samples were 78 µM, 64 µM and and 90 µM respectively for S336C^{Prox}_C394^{Prox}, S357C^{Prox}_C394^{Prox} and S378C^{Prox}_C394^{Prox} variants. Left panel: experimental DEER traces recorded at 60 K. Red lines indicate the baseline used for background correction. Central panel: corrected DEER traces (black) with superimposed fits derived from Tikhonov regularization (red). Right panel: Tikhonov derived distance distributions obtained using DeerAnalysis2019 (black) superimposed with gray lines that are the full variation of the probability of given distances, whereas the thickness is respectively the upper and lower error estimation corresponding to the mean value ± two times the standard deviation of the different Tikhonov fits.*

3.2.6.3 *In cell* EPR

The *in cell* EPR approach allows to study the dynamics and structure of biological macromolecules in their physiological environment. Factors such as viscosity, cytosol pH and possible specific and non-specific interactions with other macromolecules could affect the conformation and dynamics of proteins in cells, which may be different from those they would have *in vitro*.

The *h*NDGR1*C single cysteine variants, were labelled with M-prox and delivered in *E. coli* DH5α cells by electroporation and maintained in suspension. The structural dynamic was then followed by EPR at room temperature. The signal to noise ratio was good and the flexible behavior seen *in vitro*, was maintained in the bacterial cytoplasm, although a broadening of the EPR spectrum was observed. The change in the spectra shape can be observed overlaying the *in vitro* and the *in cell* spectra of each variant, resulting less obvious for the Cys 378 position (Figure 38).

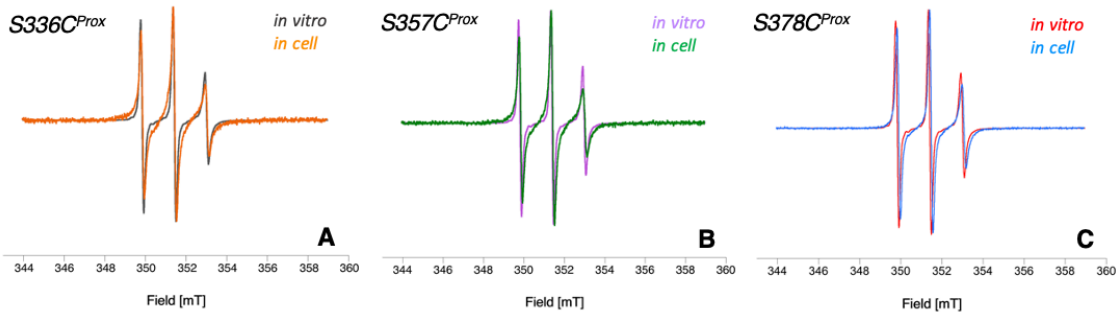


Figure 38: *Superimposition of the normalized CW-EPR spectra of hNDRG1*C S336C^{Prox} (**A**), hNDRG1*C S357C^{Prox} (**B**) and hNDRG1*C S378C^{Prox} (**C**) in vitro (10 mM Tris pH 7.4) and in cell (PBS + 1% agarose), after delivering in E. coli cell by electroporation. The spectra in cell were recorded in X-band at RT with these parameters: field sweep (10 mT), microwave power (10 mW), amplitude modulation (0.1 mT) and receiver gain (60 dB). The bulk spin concentration reached in cell was ~ 30-40 µM. The in cell spectra were recorded as a function of time and in order to improve the signal to noise the in cell data shown are the sum of the first 10 spectra, one spectra every 90 seconds.*

One of the main limitations of the *in cell* EPR with nitroxide labels is related to their reduction in hydroxylamine, a diamagnetic species, silent in EPR, by different physiological reducing agents. However, this event can be used to follow the site-specific accessibility of the nitroxide to the cellular reductants. To follow the progressive reduction of the spin-label over time, the double integral of the EPR signal for each *h*NDRG1*C^{Prox} variant in *E. coli* cells , which represents the spin concentration of the sample, was plotted as a function of time.

Different trends in the reduction properties of *h*NDRG1*C variants were observed. S357C^{Prox} showed the fastest reduction profile, with a half-time of ~ 150 min, followed by S336C^{Prox} with ~ 350 min. S378C^{Prox} reduction is the slowest one and, in this case, the complete reduction of nitroxides was not achieved, even after an overnight spectra acquisition (Figure 39).

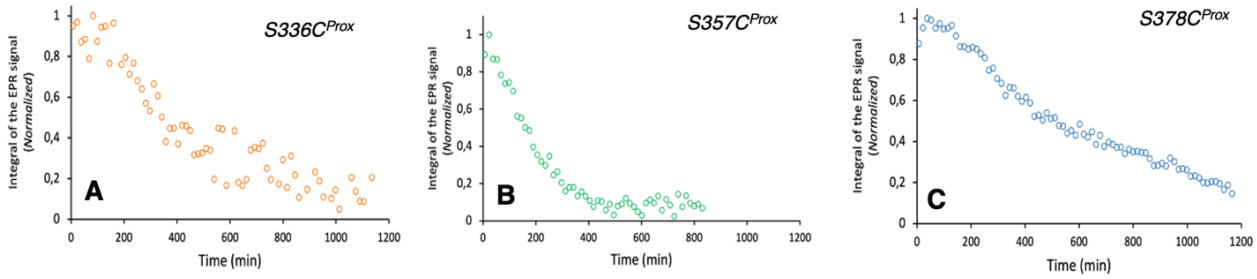


Figure 39: *Reduction profiles of hNDRG1*C single cysteine variants in E. coli in PBS + 1% agarose. Each point represents the sum of the integral of 10 consecutives EPR spectra: S336C^{Prox}(**A**), S357C^{Prox} (**B**) and S378C^{Prox} (**C**). The initial bulk spin concentration was ~ 30- 40 µM.*

The resistance to reduction of S378C^{Prox}, could be explained by increased protection of the nitroxide from reducing agents inside cells. This may result from a specific or unspecific interaction with a potential partner present in the *E. coli* cytoplasm or to a different conformation of the protein in

cell, resulting in less nitroxide exposure and a slower reduction profile compared to the S336C^{Prox} and S357C^{Prox} variants.

By measuring the height of the h0 central line of an EPR spectrum, from now called peak-to-peak intensity, it is possible to detect variations in the EPR spectral shape over time. Specifically, the increase of this value, corresponds to a rise of the mobility of the label. The profiles obtained by following the peak-to-peak intensity of all *h*NDRG1*C single cysteine variants are shown in Figure 40. As well as the double-integral of the signal, the final decrease of the peak-to-peak value is due to the progressive reduction of the nitroxide and the disappearance of the EPR signal.

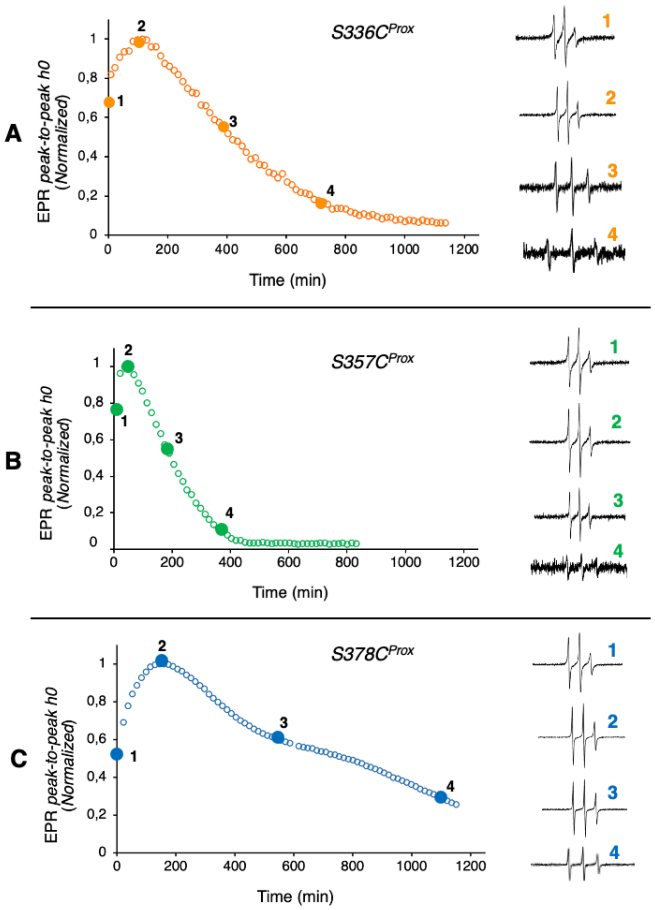


Figure 40: *Evolution of hNDRG1*C local structural-dynamics over time in E. coli cells. Peak-to-peak (h0) profiles of hNDRG1*C single labeled variants S336C^{Prox} (**A**), S357C^{Prox} (**B**) and S378C^{Prox} (**C**) and the evolution of the EPR signal (1=time 0, 2= maximum of the amplitude, 3=50%, 4=20% of the initial amplitude). The point from which the EPR spectra are extracted is highlighted with filled circles.*

For the three positions studied, the peak-to-peak-intensity increases before starting to decrease. The length of this event is position dependent and it takes, ~120 min for S336C^{Prox} ~ 60 min for S357C^{Prox} and ~150 min for S378C^{Prox}. Comparing the EPR spectra corresponding to *time 0* (1), *peak* of the curve (2), 50% (3) and 20% (4) of the initial amplitude it is possible to note a sharpening of the EPR signal related to an increase of the mobility of the labeled region. This event can be due to three possible causes: possible action of proteases on *h*NDGR1*C, further structure loss of the protein *in cell* or opening of the succinimide ring.

3.2.6.4 Control experiments of *in cell* EPR and confocal microscopy

Control experiments were performed to check if the protein is absorbed by the membrane, repeating the same protocol followed for the in cell EPR in the same conditions and skipping the electric shock. Normally, this kind of experiments should have as result the absence of signal. Surprisingly, for *h*NDRG1*C an EPR signal was recorded for all variants, reaching a bulk spin concentration of ~ 10 µM (Figure 41).

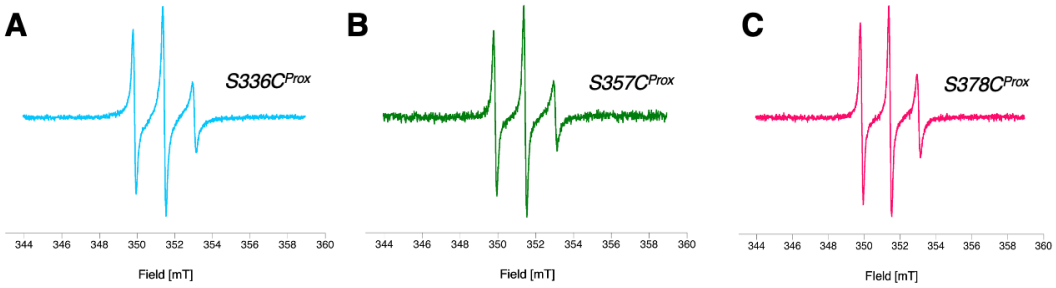


Figure 41: *Normalized sum of the first 10 spectra in CW-EPR, at RT, of the labelled hNDRG1*C S336C^{Prox} (**A**), hNDRG1*C S357C^{Prox} (**B**) and hNDRG1*C S378C^{Prox} (**C**) in (PBS + 1% agarose), after 5 min contact with E. coli cells, no electric shock and 5 washes. The parameters were: field sweep (10 mT), microwave power (10 mW), amplitude modulation (0.1 mT) and receiver gain (60 dB). The bulk spin concentration reached in cell was ~ 10 µM.*

In order to confirm these peculiar results, several parameters were modified, with no particular changes in the experimental results: washing steps, contact time, pH of the protein solution, protein concentration. The peak-to-peak intensity profiles recorded in the control experiments (Figure 42), when compared with those shown in Figure 40, exhibit the same trend, with an initial increase in intensity before starting to decrease. However, a faster complete nitroxide reduction, compared to that shown in Figure 40, was observed for the three variants, with a half-life of ~ 160 min for S336C^{Prox} and ~ 100 min for S357C^{Prox} and S378C^{Prox}. Moreover, in this case the S336C^{Prox} variant showed the slowest reduction profile.

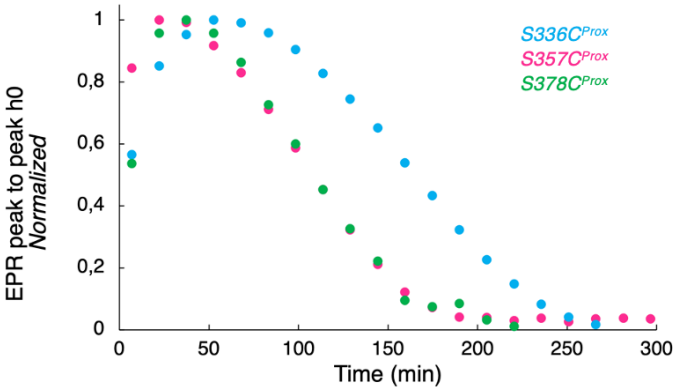


Figure 42: *Peak-to-peak (h0) profiles of hNDRG1*C variants recorded during the control experiments carried out with 5 minutes contact with E. coli cells without electric shock and 5 washes.*

Two possible events were hypothesized as possible causes: the protein interacts with the membranes, causing the lysis of the cells or it enters the cells as cell penetrating peptide.

The answer was given by confocal microscopy, after protein labeling with Alexa Fluor® 488-C5 maleimido fluorophore and incubation with non-competent *E. coli* cells, subsequently treated in the same way as the EPR control experiments.

The protein internalization and localization, followed by confocal microscopy, resulted in fluorescence inside the cells, regardless of the time of protein-cell contact (Figure 43). The fluorescence was homogeneous, suggesting the absence of a detectable sub-localization or aggregation of the protein.

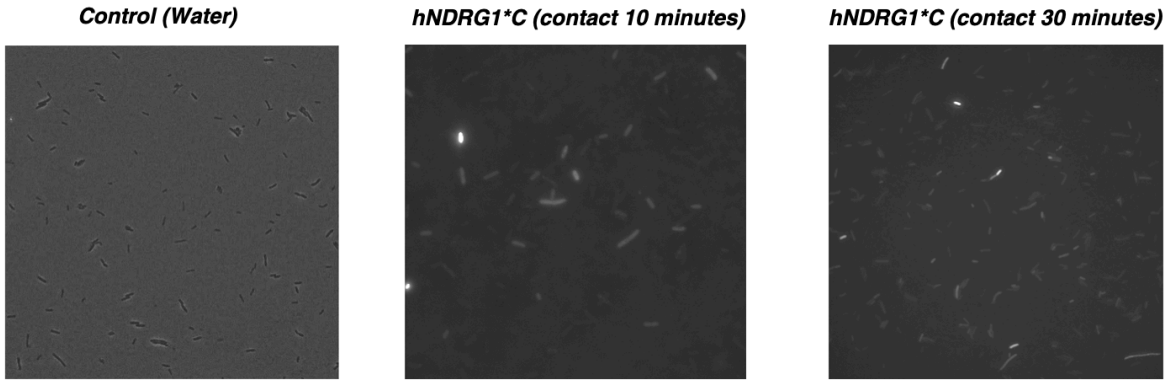


Figure 43: *Confocal microscopy of hNDRG*C labeled with Alexa Fluor® 488-C5 maleimido fluorophore. Non competent E.coli DH5α cells were incubated with water as negative control (left panel) or with hNDRG1*C for 10 (central panel) and 30 minutes (right panel). Then, the cells were treated following the in cell EPR protocol, skipping the electric shock.*

As confirmation that this was cytoplasmic fluorescence due to internalization of the protein, photobleaching experiments were performed, turning off the fluorescence signal at specific points inside the cell. A rapid redistribution of the signal was observed, demonstrating that the protein enters the cell even without electric shock and once inside it moves rapidly.

3.3 Nickel effect on *h*NDRG1 expression *in vivo*

To verify the effect that nickel exposure has on *h*NDRG1 expression and its subcellular localization, HeLa and A549 cells were treated with 1 mM $NiSO_4$, respectively for 24 hours and 2 days, a condition previously reported to increase the expression level of *h*NDRG1 (Salnikow, Su, et al. 2000). The cells, treated or untreated, were lysed under non-denaturing conditions to preserve the stability of the protein and equal amounts of cellular extracts were resolved on SDS-PAGE under non-denaturing conditions. Samples were transferred to nitrocellulose membrane overnight at 4 °C and incubated with the anti-*h*NDRG1 N-terminal region antibody.

In the untreated HeLa cells two more intense bands, probably corresponding to the monomer (ca. 50 kDa) and to the dimer (ca. 100 kDa) were visible (Figure 44A). Other minor bands were

distinguishable, possibly due to different phosphorylated forms of *h*NDRG1 or its truncated forms at the N-terminal end (Park et al. 2018). In addition, a very faint band at ca. 200 kDa, coherent with the tetramer of *h*NDRG1, appeared at longer exposition of the film. The treatment with Ni(II) for 24 hours shifts the equilibrium toward the monomeric form of *h*NDGR1, as evidenced by the increase in the intensity of the corresponding band and the decrease of the band at 100 kDa.

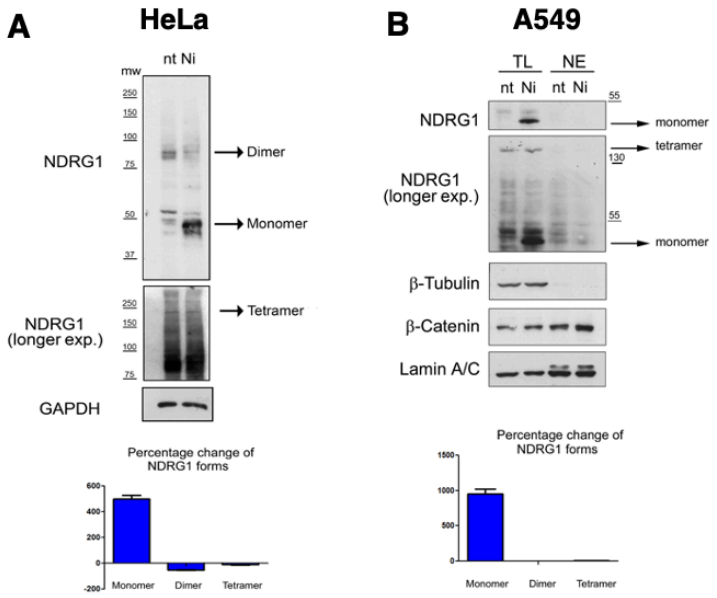


Figure 44: *Effect of nickel exposure on hNDRG1 expression in HeLa and A549 cells. HeLa cells were treated with 1 mM $NiSO_4$ for 24 hours (Ni) or left untreated (nt) and lysed under non denaturing conditions. Lysates were resolved by an SDS-gel, under non-denaturing conditions and filters were incubated with an antibody against the N-terminal region of NDRG1. GAPDH was used as equal loading marker (**A**). Lung carcinoma-derived A549 cells were treated with 1 mM $NiSO_4$ for two days (Ni) or left untreated (nt). Total lysates (TL) and nuclear extracts (NE) were analyzed in SDS-PAGE and the amount of NDRG1 was evaluated by western blot analysis. β-tubulin and lamin A/C indicate the purity of nuclear extraction and equal loading (**B**). Histograms indicate the fold variation of the indicated forms of hNDRG1 after nickel exposure.*

Contrarily, in A549 cells, obtained from a non-small cell lung cancer, the dimeric form of *h*NDGR1 was not uncovered and only the monomeric and tetrameric forms, were detectable in the presence and in the absence of Ni(II) (Figure 44B). Also in this cell line, the addition of Ni(II) caused a marked increase of the monomeric *h*NDRG1 form. No *h*NDRG1 oligomeric forms were detected in the nucleus, also in the presence of Ni(II), implying that in this cellular line and in this conditions the localization of *h*NDRG1 is cytoplasmatic.

4. Conclusions and perspectives part 1

The intrinsically disordered proteins (IDPs) are proteins completely unstructured or containing well-structured regions interspersed with long intrinsically disordered regions (IDPRs). These proteins are typically rich in disorder-promoting amino acids (alanine, glycine, serine, proline, glutamine, glutamic acid, lysine, arginine) and have high percentages of positively charged amino acids, resulting in high isoelectric points. Intrinsically disordered proteins have been recognized in all living organisms and in eukaryotes they often undergo post-translational modifications, such as phosphorylation. Upon interaction with proteins, nucleic acids, and other ligands, IDPs can undergo disorder/order transitions, and thanks to this ability they are able to bind specifically to a large variety of partners. As a result, IDPs are involved in several signaling pathways, working as regulators of important biological functions (Dyson and Wright 2005; Wright and Dyson 2015) and changes in their expression levels or mutations of these proteins are responsible for the progression of specific diseases.

The goal of this work was the structural and biophysical characterization of NDRG1, a human protein of 43 kDa, member of the NDRG family. Primary sequence alignment with secondary structure prediction analysis (Figure 15B) and the truncated crystallographic structures of NDRG1, NDRG2 and NDGR3 showed that all NDRG proteins share a well-structured N-terminal α/β-hydrolase fold with six canonical β-strands, surrounded by eight α-helices (Figure 45).

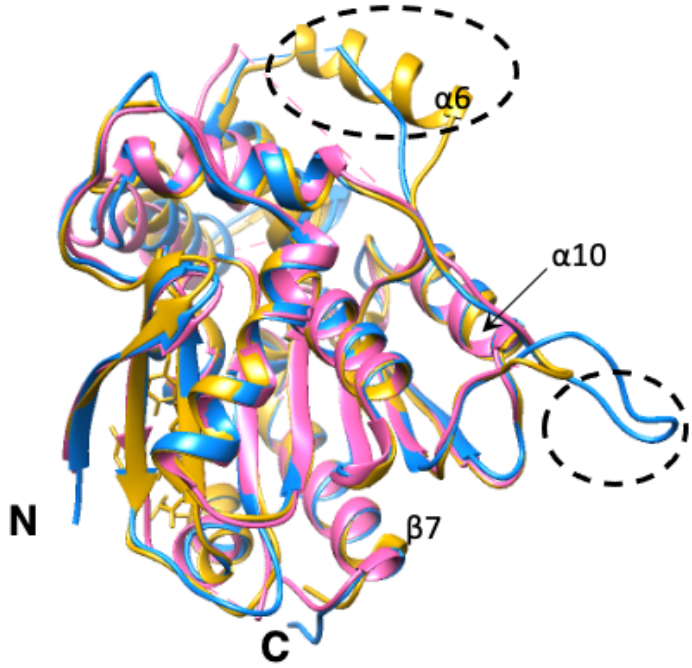


Figure 45: *Overlay of NDRG1 (6ZMM) (Mustonen et al. 2020) (pink), NDRG2 (2XMQ) (Hwang et al. 2011) (yellow) and NDRG3 (6L4B) (Kim et al. 2020) (blue) α/β hydrolase domain. The high structural similarity is confirmed, with two main differences, respectively in α6 of NDRG2 and in the loop between β7 and α10 in NDRG3. The three structures are truncated forms of the protein, lacking of some residues in the N and C-terminus, proving the predicted high flexibility of these regions. Image created with Chimera software (Pettersen et al. 2004).*

Conversely, the NDRG1 C-terminal region, unique among the NDRG proteins, shows the main properties of an IDP: high content in charged and hydrophilic amino acids, resulting in extreme pI values (pI = 10.65), the absence of a stable secondary and tertiary structures and sites for post-translational modifications. The Database of Protein Genetic and Chemical Interactions

(BioGRID, https://thebiogrid.org/) (Stark et al. 2006) shows that NDRG1 has more than 200 interactors in cells and is part of a large interaction network, as demonstrated by the STRING computational platform (https://string-db.org/) (Szklarczyk et al. 2019) (Figure 46).

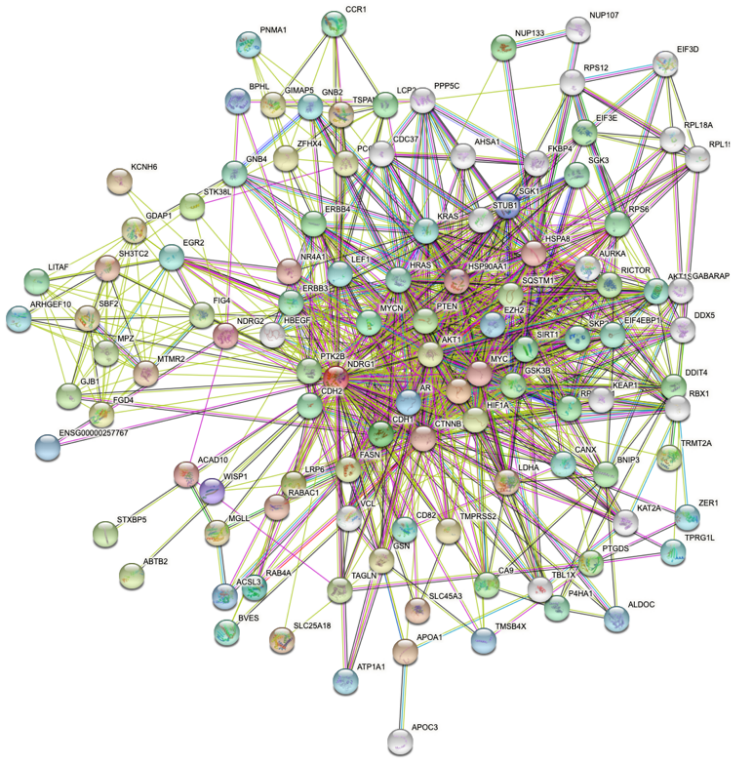


Figure 46: *Network of predicted interactions of hNDRG1 created by the STRING computational platform (https://string-db.org/). Nodes represent proteins, with the query protein indicated in a red node. The others colored nodes represent the first shell of interactors, while the white ones the second shell of interactors. The empty nodes are relative to protein of unknown 3D structure, while those whose structure is known or predicted are in filled nodes.*

It has been hypothesized that the intrinsically disordered C-terminal region is responsible for most of these interactions, resulting in various physiological and pathological implications of the protein. For this reason, NDRG1 intrinsically disordered region was given particular attention in this work.

To reach my goal, protocols for the heterologous expression and purification from *E. coli* of the full-length protein and its disordered C-terminal domain were developed and the purified proteins in solution were structurally and biophysically characterized applying the circular dichroism, light scattering, isothermal titration calorimetry, NMR spectroscopy and SDSL-EPR techniques.

Light scattering data showed that the native protein exists in solution in three abundant oligomeric forms: tetramer, dimer and monomer. This result is different from previous reports in literature (Hwang et al. 2011; Kim et al. 2020; Mustonen et al. 2020), where NDRG1, NDRG2 and NDRG3 have been identified as monomers and only for NDRG1 and NDRG3, small amounts of dimer was detected. To support my data, the tetramer and the dimer of NDRG1 were identified *in vivo* in HeLa cells treated with nickel sulfate for two days. It is important to point out that the most stable forms of these proteins in solution, as well as those crystallized, were found to be those lacking of some amino acids at the N- and C-terminus. On the contrary the C-terminal domain of NDRG1,

as well as the N-terminal domain (Figure 47), resulted to be monomer in solution from light scattering analysis. Thus, it is possible to assume that both domains are involved in the oligomerization process.

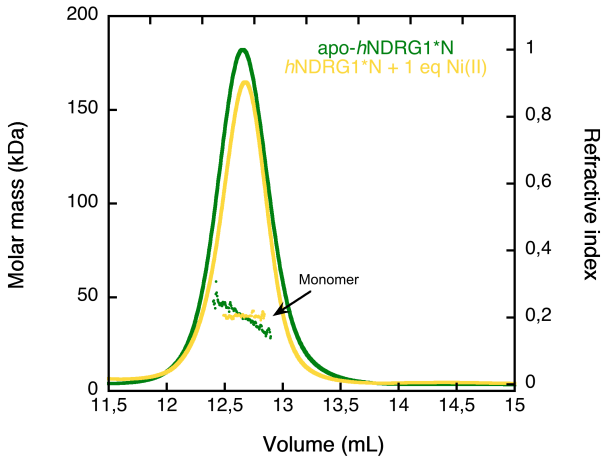


Figure 47: *Size exclusion profile monitored by the refractive index detector (lines) and weight-averaged molar mass distribution (dots) of hNDRG1*N (50 µL, 143 µM) obtained with a combination of SEC–MALS–QELS in the absence (green) and in the presence (yellow) of one equivalent of Ni(II).*

The quantitative analysis of far-UV CD spectral curves, analyzed using the new implemented and freely accessible algorithm, BeStSel (Micsonai et al. 2018), indicate that *h*NDRG1wt has a prevalence of α-helical content. At the same time, there is a relevant percentage of flexible elements, most of witch concentrated in the C-terminal region. The CD profile of *h*NDRG1*C is typical of an high flexible polypeptide. The most common methods used for protein quantification, such as the Bradford assay and the ultraviolet absorbance at 280 nm, cannot be accurate in the determination of the concentration of the IDPs. Due to their amino acidic sequence, IDPs tend to have low molar extinction coefficient at 280 nm and weak binding of the Coomassie Brilliant Blue, used as dye in the Bradford assay, and mainly based on the hydrophobic interactions with aromatic residues and electrostatic interactions between its anionic forms and the basic residues of the protein (Contreras-Martos et al. 2018). The online tool BeStSel, allowed to calculate a "Best" correction factor, useful to obtain the lowest NRMSD of the fitted curve and a more accurate prediction in secondary structure content, covering a broad of concentration variations and avoiding relevant deviations in spectra deconvolutions (Micsonai et al. 2015). The addition of TFE and SDS drastically changed the secondary structure of *h*NDRG1*C and its variants (Figure 20).

The flexibility of this region was confirmed by NMR spectroscopy. The ^{1}H-^{15}N HSQC spectra of the recombinant C-terminal domain are typical of an intrinsically disordered protein, with low signal dispersion in the proton dimensions and narrow spectral ranges (Figure 23). This confirms that this region is dominated by random coil conformations, lacking a well-defined structure, as predicted by the *in silico* secondary structure and disorder predictions analysis.

The nickel binding activity was confirmed by ITC experiments. Two binding sites have been identified in *h*NDRG1wt, one of which appears to involve the C-terminal region of the protein. The

metal binding affinity of *h*NDRG1*C resulted to be pH dependent, with low binding affinity at pH 6.5 (K_d = 56 μM) (Figure 16D), probably due to the unfavorable double protonation of the imidazole ring of all histidine residues at this pH, as determined by NMR spectroscopy.

The SDSL-EPR technique, based on the selective labeling of a protein with a local environment sensible spin label, followed by its study by EPR spectroscopy, was employed to further investigate the structure of *h*NDRG1*C, obtaining important information about protein structural dynamics and conformational changes. This technique was applied both *in vitro* (DEER and CW-EPR*)* and *in cell* (CW-EPR). All data confirmed the flexibility and the fast dynamics of *h*NDRG1*C and the presence of multiple conformations of the protein.

The study of biomolecules in their native environment is the main goal of the structural biology. *h*NDRG1 is a primarily cytosolic protein, cellular environment highly populated by other macromolecules. For example, in *E. coli* cells the concentration of macromolecules in the cytoplasm can exceed 300 g/l and occupy 30% of volume, leading to a volume exclusion effect and changing in the local viscosity and differences in polarity (Gnutt et al. 2015; Stepanenko et al. 2016; Theillet et al. 2014). These elements can influence the structure of biomolecules, favoring a more compact state. In this work the influence of the intracellular environment on *h*NDRG1*C structure and dynamics in a position dependent manner was demonstrated, by protein delivery in *E. coli* cells. For this purpose, a new protocol, developed by Dr. Elisabetta Mileo and Dr. Annalisa Pierro, was followed during my staying as a visiting PhD student in the laboratory of Bioénergétique et Ingénierie des Protéines (CNRS, Marseille). The results showed that the S378C^{Prox} variant was the lowest affected, lacking of the spectral shape broadening. This ruled out that the changes recorded were due to the impact of viscosity, which otherwise should have been observed in any protein variant. The results obtained in the cellular context were compared to those obtained *in vitro* and in the presence of crowding agents and osmolytes. Once again, the labeled regions of the protein resulted to be differently affected by crowding agents, with minimal effect on S378C^{Prox} variant and more marked on S336C^{Prox} variant. Only in the presence of different concentrations of SDS, the S357C^{Prox} variant appeared to be the variant less disturbed, behavior still uncleared.

A possible explanation derives from the main difference between EPR and CD. In fact, CD allows to detect a general change on the whole protein structure, no localized around the mutated residue. Moreover, CD spectra, recorded in the presence of different crowders and osmolytes in the far-UV region, allowed to determine the composition in secondary structure elements. Therefore, the molecules tested could have induced a general change in the secondary structure of the protein, perceived by CD, without affecting the tertiary structure and the dynamics of a specific region of the protein, whose changes are detected by EPR (Torricella et al. 2021).

The results obtained are only preliminary and further experiments are planned to complete this study, whose aim is the structural characterization and the understanding of the mechanism of action of NDRG1, in order to design drugs specifically targeting this protein for the treatment of several diseases. First of all, a protocol for *in vivo* phosphorylation of NDRG1 by PKA is under development. This is a very important part, considering the importance that this post-translational modification has in the structure of many proteins and in the regulation of the biological functions of NDRG1. For example, in glioma cells NDRG1 can mediate chemoresistance to alkylating treatments by interaction of its phosphorylated form with the O^6-methylguanine-DNA methyltransferase (MGMT) (Sun et al. 2018; Weiler et al. 2014).

Next, the dynamics of *h*NDRG1*C in the cellular environment needs to be further investigated in eukaryotic cells, for example in HeLa cells. The ability of *h*NDRG1*C to cross the bacterial cell membrane, even without electric shock, opens the possibility to use this protein as a cargo protein for *in cell* delivery of other molecules, such as drugs.

5. Results part 2: Biophysical and functional properties of SgSrnR

5.1 Protein heterologous expression and purification

The *Sg*SrnQ amino acid sequence (UniProt code: Q8L1Y3) was retro-translated, in order to obtain the corresponding nucleotide sequence to be cloned into appropriate expression vectors, following its optimization for expression in *E. coli*. The best approach resulted in the expression of *Sg*SrnR as an N-terminal His-ZZ-His tagged protein and purification with a non-native GSH sequence at the N-terminus. This construct allowed us to increase the yield and solubility of the protein thanks to the modified immunoglobulin binding domain of protein A of *Staphylococcus aureus* (ZZ-tag) (Bogomolovas et al. 2009) and to purify it by affinity chromatography (His-Tag). The His-ZZ-His-*Sg*SrnR fusion polypeptide (31.1 kDa) was overexpressed in *E. coli* BL21 (DE3) cells and induced with IPTG, following the protocols reported in literature for the unlabeled protein (Beniamino et al. 2020) and for the single-labeled (^{15}N), double-labeled (^{13}C,^{15}N), and triple-labeled (^{2}H ,^{13}C,^{15}N) variants (Mazzei et al. 2021). The first purification step of *Sg*SrnR from the soluble fraction involved a Ni-based affinity chromatography followed by column incubation with thrombin protease to remove the tag. The protein (12.6 kDa) is eluted directly into a desalting column to lower the ionic strength for the following anion exchange chromatography, resulting in a single peak between 100 and 150 mM NaCl (Figure 48) (Beniamino et al. 2020).

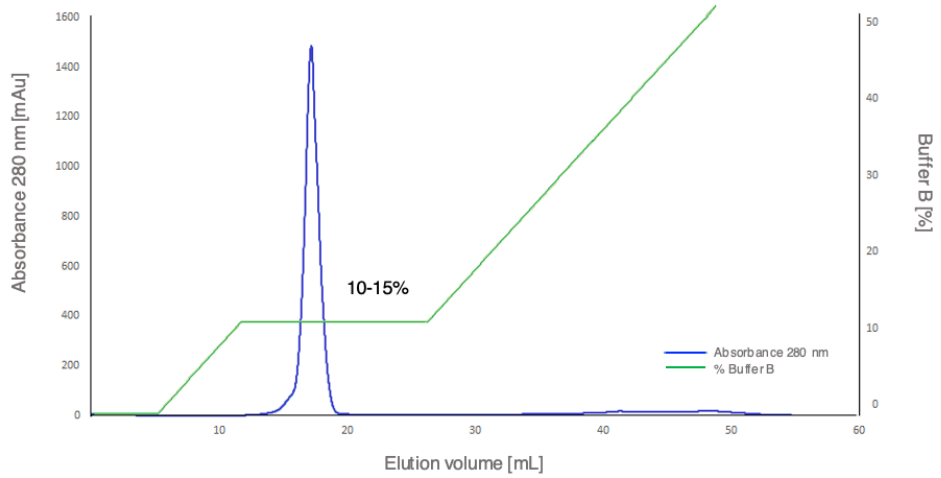


Figure 48: *Elution of the anion exchange chromatography of SgSrnR, resulting in a single peak at 10-15% of Buffer B, corresponding to 100-150 mM NaCl. The absorbance at 280 nm is shown in blue, the percentage of buffer B with 1 M NaCl in green.*

The final purification step was performed with a size exclusion chromatography (SEC) using a Superdex 75 10/300 column, resulting in a single peak around 13 mL (Figure 49A). Protein purity at the end of the purification protocol was assessed by SDS-Page (Figure 49B) and the fractions containing the protein were combined and concentrated, with a final yield of 10-15 mg L^{-1} of initial culture (Beniamino et al. 2020).

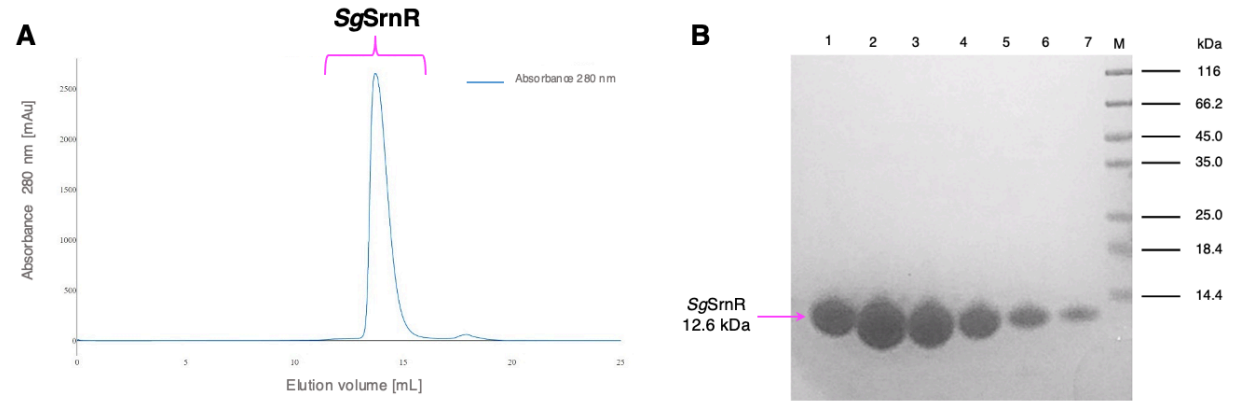


***Figure 49**: Superdex 75 10/300 of SgSrnR. (**A**). SDS-PAGE electrophoresis of the fractions of the last chromatographic step to verify the purity of the isolated SgSrnR. From the left to the right: lanes 1-7 fractions of the corresponding peak in size exclusion chromatography, lane M marker. Adapted by permission from Springer Nature (Beniamino et al. 2020) with License number 5232420214476. Copyright © 2019, Society for Biological Inorganic Chemistry (SBIC) (**B**).*

The protein was stored at -80 ° C in 20 mM Tris pH 7.5, 150 mM NaCl, 1 mM TCEP for subsequent analysis. These conditions were selected evaluating the stability of the protein under different buffer conditions carrying out a Thermal Shift Assay and selecting the condition with the highest T_m (T_m = 52.0 °C for 20 mM Tris HCl, 150 mM NaCl) (Figure 50) (Beniamino et al. 2020).

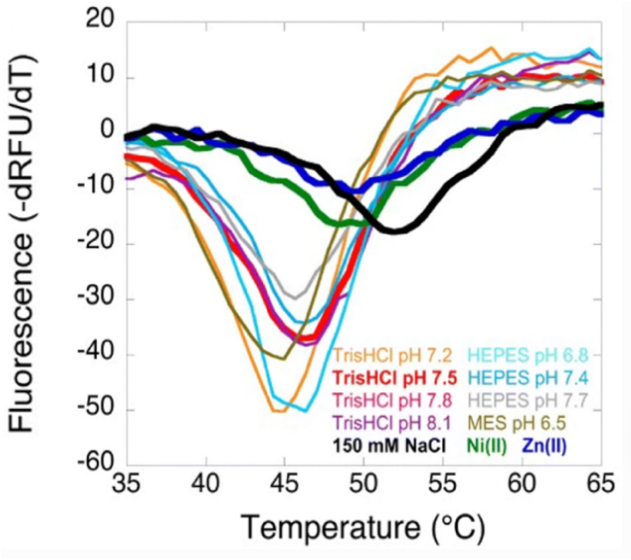


***Figure 50:** First derivative curves of the fluorescence obtained in a thermal shift assay carried out in diverse buffers, with pH values in the range of 6.5 and 8.1. The minimum of the curves corresponds to the protein melting temperatures. The effect of 1:1 Ni(II)/Zn(II):protein ratio on the protein unfolding in 20 mM TrisHCl pH 7.5, 150 mM NaCl, 1 mM TCEP is shown. The addition of metal ions decreases the stability of the protein, as suggested by the shift of the T_m to 49.3°C. Reprinted by permission from Springer Nature (Beniamino et al. 2020) with License number 5232420214476. Copyright © 2019, Society for Biological Inorganic Chemistry (SBIC).*

5.2 *Sg*SrnR metal binding properties

The Ni(II) and Zn(II) binding properties of *Sg*SrnR were investigated by isothermal titration calorimetry (ITC). Analyses were performed at room temperature by titrating 600 µM of the metal solution into 32-40 µM of protein through 55 injections, each with a volume of 5 µL. The negative

peaks observed in the calorimetric curve indicate that *Sg*SrnR binds both Ni(II) and Zn(II) with an exothermic reaction (Figures 51A and 51C). A single binding site was identified in the Ni(II) binding isotherm, with a stoichiometry of one nickel ion per monomer and a moderate affinity $K_a = 6.2 \pm 0.8 \times 10^4$ ($K_d = 16 \pm 2$ µM) (Figures 51B). Two inflection points, corresponding to two different binding sites, were identified in the Zn(II) binding isotherm, respectively, with stoichiometry $N_1 = 0.25$ and $N_2 = 1$ and affinity $K_{a1} = 7.0 \pm 0.1 \times 10^6$ ($K_{d1} = 0.14 \pm 0.02$ µM) and $K_{a2} = 2.56 \pm 0.02 \times 10^5$ ($K_{d2} = 3.91 \pm 0.03$ µM) (Figure 51D). The stoichiometry of 0.25 for the first Zn(II) binding event indicates that four monomers of the protein are required to bind a Zn(II) ion, suggesting an oligomerization process of *Sg*SrnR in the presence of the metal. The enthalpy and entropy contributions in the two reactions are different: Ni(II) binding is entropically driven ($\Delta H = - 3.3 \pm 0.2$ kcal mol^{-1}, $\Delta S = + 10.9$ cal mol^{-1} K^{-1}), while for Zn(II) binding a big contribution is given by enthalpy ($\Delta H_1 = - 11.56 \pm 0.02$ kcal mol^{-1}, $\Delta S_1 = - 7.47$ cal mol^{-1} K^{-1}, $\Delta H_2 = - 8.51 \pm 0.03$ kcal mol^{-1}, $\Delta S_2 = - 3.81$ cal mol^{-1} K^{-1}) (Beniamino et al. 2020).

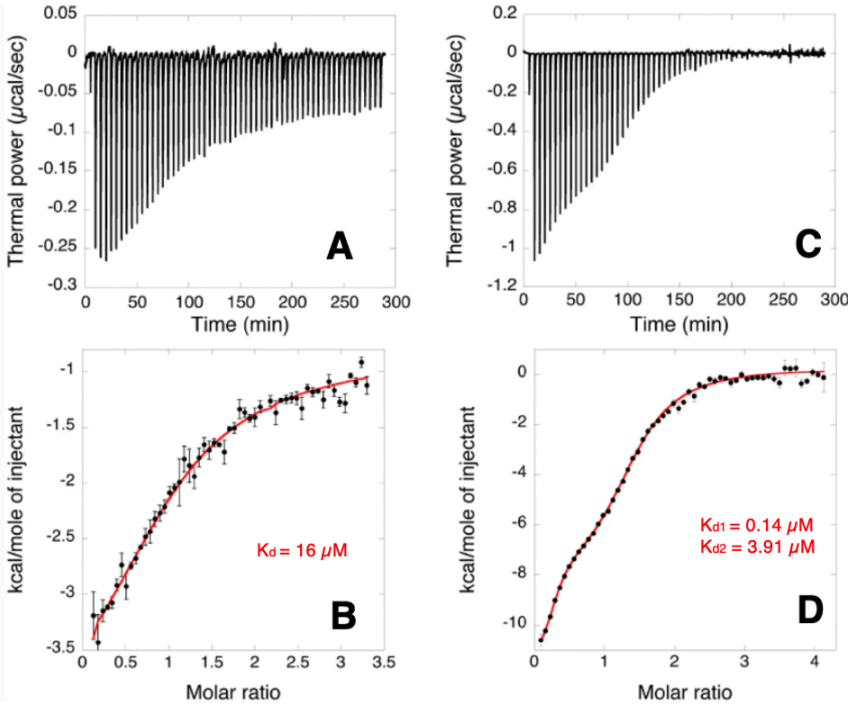


Figure 51: *ITC titration data for the binding of Ni(II) (**A**, **B**) and Zn(II) (**C**,**D**) to SgSrnR. Raw titration data (**A**, **C**) represent the thermal effect of 5 µL injections of 600 µM of metal ion solution into protein solution (30-40 mM). Normalized integrated heat reaction data (filled dots) were obtained integrating raw data and subtracting the ligand heat of dilution into the buffer. The single site and the two sets of independent sites models were applied for fitting (continuous red lines) Ni(II)(**B**) and Zn(II)(**D**) binding respectively. Adapted by permission from Springer Nature (Beniamino et al. 2020) with License number 5232420214476. Copyright © 2019, Society for Biological Inorganic Chemistry (SBIC).*

5.3 Circular dichroism and fluorescence spectroscopy

The CD spectrum of *Sg*SrnR is typical of a well-structured protein with a prevalence of α-helices: one large positive signal at 192 nm and two negative ones at 210 and 224 nm, respectively. The quantification carried out with DichroWeb (Lobley, Whitmore, and Wallace 2002; Whitmore and Wallace 2008) indicated the following composition in secondary structure: 52% α-helices, 15% β-strands, 16% turns, 16% random coil (Figure 52A). The secondary structure was not affected by the addition of up to 5 equivalents Ni(II) ions, unlike the Zn(II) ones. The addition of only one equivalent of Zn(II) resulted in a substantial change in secondary structure (18% α-helices, 29% β-strands, 23% turns and 29% random coil), more evident in the presence of 2 equivalents of Zn (II) (8% α-helices, 44% β-strands, 14% turns and 34% random coil) (Figure 52A). A partial precipitation of the protein was also observed at this concentration of Zn(II). The mean residue ellipticity value at 222 nm was monitored between 10 °C and 90 °C, obtaining a sigmoidal profile consistent with a thermal denaturation curve, in which *Sg*SrnR loses progressively and irreversibly the secondary structure (Figure 5B) (Beniamino et al. 2020).

Aromatic residues provide information on the tertiary structure of the surrounding regions when the CD spectra of a protein are recorded in the near UV region (250-350 nm). In the CD spectrum of the apo-protein, two signals in the 250-270 nm region and one in the 270-290 nm region were observed, corresponding to two of the three phenylalanines and to the only tyrosine respectively (Figure 52C and 52D). This result suggests that these residues are in rigid regions of the protein. A further increase in rigidity was recorded in the presence of 1 equivalent of Ni(II). At this Ni(II) concentration, an increase in the signal at 260 nm and the emergence of that at 280 nm, probably corresponding to tryptophan, were observed. Subsequent additions of Ni(II) decreased the signal at 280 nm, suggesting an additional effect on the tertiary structure of the protein that leads this region to gain mobility (Figure 52C). Conversely, no significant change was detected in the near-UV CD spectrum upon the addition of up to 2 equivalents of Zn(II) (Figure 52D) (Beniamino et al. 2020).

The conformational changes that occur in the tryptophan region in the presence of Ni(II) and Zn(II) ions were monitored by recording steady-state fluorescence intensities and emission spectra. Increasing concentrations of Ni(II) ions resulted in a progressive decrease of peak intensity due to an increase in residue side-chain mobility (Figure 52E). In the presence of Zn(II) ions, an increase in scattering at lengths below 300 nm was observed, confirming the previously hypothesized theory of Zn(II)-induced aggregation (Figure 52F) (Beniamino et al. 2020).

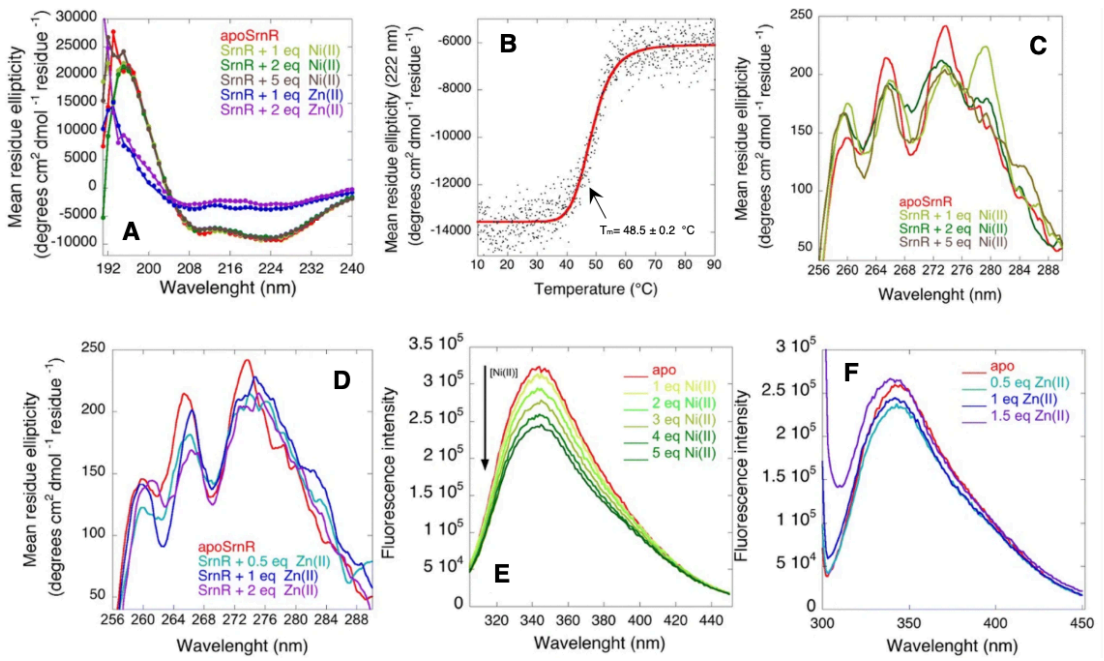


Figure 52: *CD spectroscopy of SgSrnR (90 µM) in the far-UV region in the absence (red) and in the presence of one (light green), two (green) and five equivalents (brown) of Ni(II) or one (blue) and two (purple) equivalents of Zn (II). The circles represent the experimental points in mean residue ellipticity (degrees cm^2 dmol^{-1} residue^{-1}), while the solid line is the best fit (**A**). Thermal denaturation curve obtaining monitoring the mean residue ellipticity value at 222 nm between 10 ° C and 90 ° C of the protein solution (45 µM)(**B**). Near-UV CD spectra of SgSrnR (560 µM) in the absence (red) and in the presence of one (light green), two (green) and five (brown) equivalents (**C**). Near-UV CD spectra of SgSrnR (560 µM) in the absence (red) and in the presence of half (cyan), one (blue) or two (purple) equivalents of Zn(II) (**D**). Steady-state intrinsic fluorescence intensities and emission spectra of SgSrnR (20 µM), in the absence and in the presence of increasing concentrations of Ni(II) ions (**E**) and Zn(**II**) ions (**F**). Adapted by permission from Springer Nature* (*Beniamino et al. 2020*) *with License number 5232420214476. Copyright © 2019, Society for Biological Inorganic Chemistry (SBIC).*

5.4 *Sg*SrnR oligomeric properties

Multiple-angle (MALS) and quasi-elastic (QELS) light scattering experiments were performed on *Sg*SrnR to establish its oligomeric state and hydrodynamic properties. *Sg*SrnR in solution is a dimer of 26.4 kDa and hydrodynamic radius of (Rh) 2.9 nm. The oligomeric state of the protein was not affected by the presence of Ni(II) (MW = 25.9 kDa) or Zn(II) (MW = 27.0 kDa) ions, the addition of which caused a small increase in the Rh value, 3.2 nm and 3.5 nm for Ni(II) and Zn(II), respectively (Figure 53) (Beniamino et al. 2020).

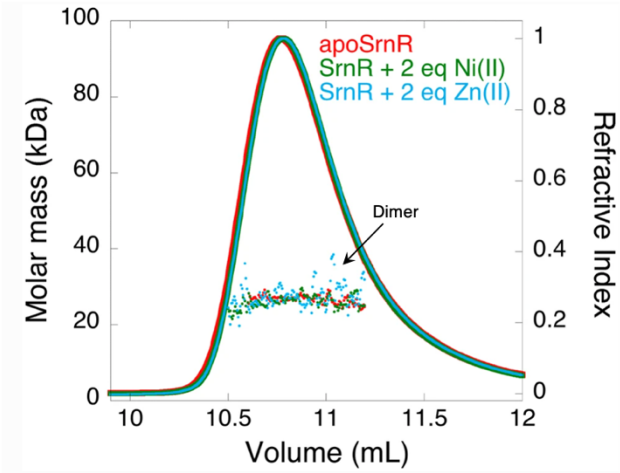


***Figure 53**: Elution profile of SgSrnR (100 µL, 560 µM) in molecular exclusion chromatography combined with MALS and QELS in the absence (red) and in the presence of one equivalent of Ni(II) (green) or Zn(II) (blue) ions. The solid lines represent the traces from the refractive index detector, while the dots are the average molecular weight. Adapted by permission from Springer Nature (Beniamino et al. 2020) with License number 5232420214476. Copyright © 2019, Society for Biological Inorganic Chemistry (SBIC).*

5.5 Structural investigation by NMR spectroscopy

3D and 4D NMR spectroscopy was applied to *Sg*SrnR, assigning 107 of the 111 cross-peaks (96%) of backbone amide groups in the $^{1}H^{15}N$ TROSY-HSQC spectrum. Unresolved signals include the N-terminal GSH sequence and Glu82. The five proline residues were nor detected (Figure 54) (Mazzei et al. 2021).

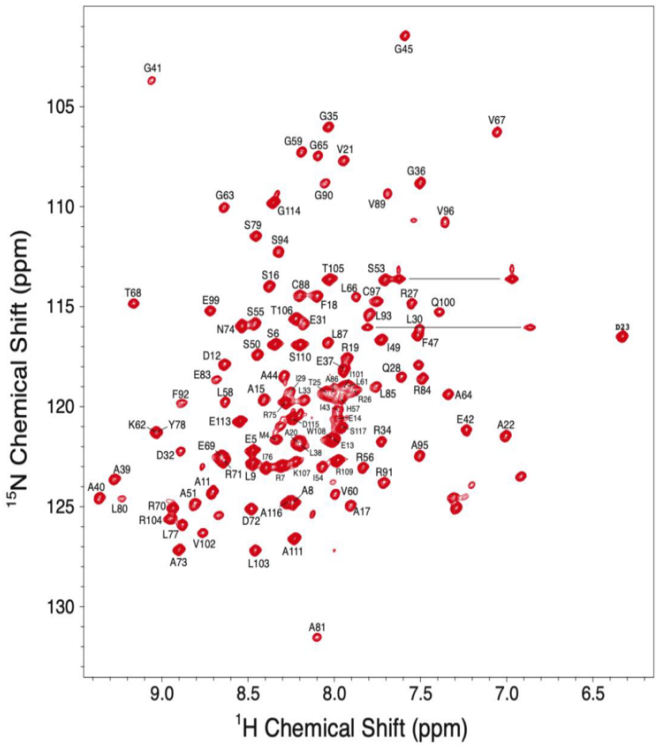


***Figure 54:** ^{1}H-^{15}N HSQC spectrum of the triple labeled variant (^{2}H ,^{13}C,^{15}N) of apo-SgSrnR (500 µM) at 800 MHz in 20 mM TrisHCl pH 7.5, 150 mM NaCl, 1 mM TCEP containing 5% D_2O. Each assigned peak is indicated by a letter, corresponding to the single-letter amino acid code and the corresponding residue number in the protein sequence. The peaks around 7/126 ppm are folded unassigned peaks from Arg sidechains; the horizontal lines link two pairs of signals from Asn and Gln sidechains; the few remaining unlabeled peaks derives from the four residues not visible in 3D and 4D spectroscopy. Reprinted by permission from Oxford University Press (Mazzei et al. 2021) with License numbers 5232421500534 and 5232431038226. Copyright © 2021, Oxford University Press.*

TalosN (Shen and Bax 2013) algorithm, used to make a prediction of the secondary structure of *Sg*SrnR (Figure 55A), confirmed the results of the circular dichroism (Figure 52A) and the disorder prediction performed with the D^2P^2 web server (http://d2p2.pro/) (Figure 55B). From these data it is possible to assume that *Sg*SrnR is a well-structured protein, with both α-helices and β-strands,

mainly localized in the central part of the protein. Only two regions present a disordered nature: residues 5-9 at the N-terminus and more significantly residues 110-117 at the C-terminus, as demonstrated by lower S^2 order parameters (Figure 55A) and higher peaks intensity in the ^{1}H-^{15}N HSQC spectrum (Figure 8C) (Mazzei et al. 2021).

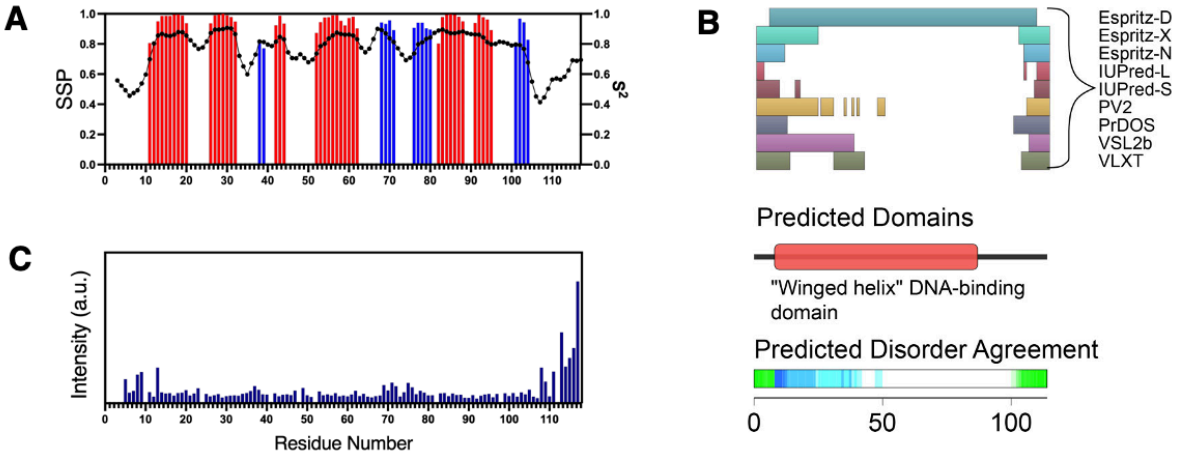


Figure 55: *Secondary structure prediction of SgSrnR derived from NMR chemical shifts and performed with TalosN. α-helices are indicated in red, β-strands in blue and the corresponding S^2 order parameters are shown as dots connected by a line: lower S^2 order parameters correspond to greater mobility in the ps−ns time scale (**A**). Disorder prediction made by disorder predictors available in the D^2P^2 web server: predicted disorder regions are indicated in the top panel, folded domains in the middle panel and disorder agreement in the bottom panel (**B**). ^{1}H-^{15}N HSQC peaks intensities of SgSrnR: the highest intensities were measured in the designated disordered regions of the protein, predominantly at the C-terminal end. The results are reported over the residue number, including the GSH non-native sequence (**C**). Reprinted by permission from Oxford University Press (Mazzei et al. 2021) with License numbers 5232421500534 and 5232431038226. Copyright © 2021, Oxford University Press*

5.6 Characterization of the binding to OP_{sodF} DNA sequence

The DNA sequence involved in nickel-mediated transcriptional regulation of the *sodF* gene in *S. griseus* was previously characterized by Kim et al. Performing a gel mobility shift assay, they identified a sequence overlapping the transcription start site of the gene and consisting of inverted repeated sequences, potentially affected by binding to the regulator of *sodF* (Figure 56).

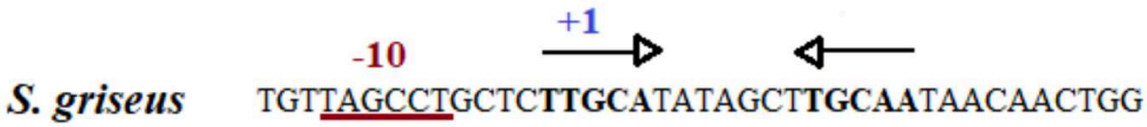


Figure 56: *sodF regulatory DNA sequence between -15 and +25 of S. griseus. The transcription start site of sodF (+1) and the inverted repeated sequences (facing arrows) are indicated (+1).*

For further studying the *Sg*SrnR-P_{sod} interaction, a Dnase I footprinting assay was performed, using the purified recombinant protein (0 nM, 150 nM, 300 nM, 600 nM, 1200 nM) and the radiolabeled P_{sod} probe, in the absence and in the presence of Ni(II)(10 µM). A protected region of DNA, from -15 to +27, relative to the transcription start site, resulted when at least 480 nM of dimeric protein was present. Addition of Ni(II) did not change the footprinting and affinity of *Sg*SrnR to DNA, suggesting that the *Sg*SrnR-P_{sod} interaction is not Ni(II) dependent (Figure 57) (Beniamino et al. 2020).

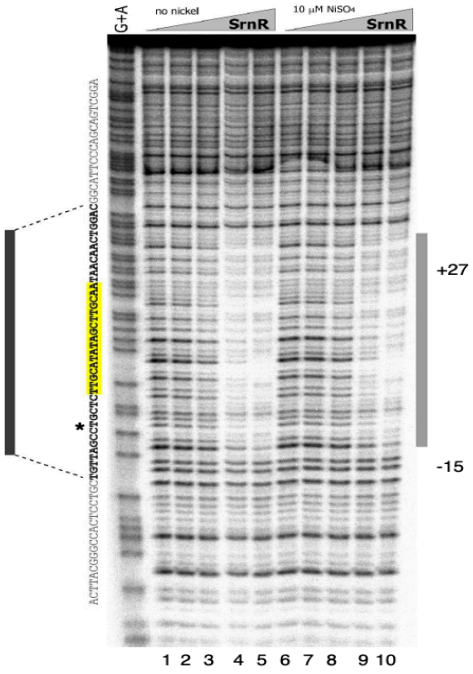


Figure 57: *Dnase I footprinting analysis to study the SgSrnR binding to P_{sodF}. The sodF promoter region was used as radiolabeled DNA probe, incubated with increasing concentration of the protein (0 nM, 120 nM, 240 nM, 480 nM, 960 nM), in the absence (lanes 1-5) or in the presence of Ni(II)(10 µM) (lanes 6-10) and subjected to partial DnaseI digestion. The grey box on the right of the autoradiograph represents the 43 bp DNA region protected from DnaseI and compatible with a molecule of 20-25 kDa. On the left the footprinting-protected region identified in this work is highlighted in bold, in yellow that individuated by J. S. Kim, Kang and Lee 2003 . The start site of sodF transcription is indicated by an asterisk. Adapted by permission from Springer Nature (Beniamino et al. 2020) with License number 5232420214476. Copyright © 2019, Society for Biological Inorganic Chemistry (SBIC).*

The OP_{sodF} sequence identified to be protected in the DNase footprinting (TGTTAGCCTGCTCTTGCATATAGCTTGCAATAACAACTGGACG) was used to further investigate the DNA binding of *Sg*SrnR by ITC and NMR spectroscopy. Calorimetric data suggested two binding events, due to the presence of endothermic peaks in the early part of the titration and exothermic peaks later on (Figure 58A). The best fit was obtained using the two sets of binding sites model (Figure 59B), with N = 0.5 each, but different affinity. One possible explanation is that the first binding event is due to the binding of a single protein monomer to the first half of the inverted DNA repeated sequence (TTGCA-N7-TGCAA9)(K_{d1} = 80 ± 10 nM, ΔH_1= + 38.17 ± 0.06 kcal mol^{-1}, ΔS_1 = +160 kcal mol^{-1} K^{-1}). This first entropically favored protein-DNA interaction is followed by the binding of the second monomer to the second half of the binding sequence, completing the formation of the protein-DNA complex. The second event has a lower affinity (K_{d2} = 1.0 ± 0.2 µM) and results in a structural variation sufficient to reduce the level of disorder of the system (ΔH_2 = -11.72 ± 0.09 kcal mol^{-1}, ΔS_2 = -11.9 kcal mol^{-1} K^{-1}) (Mazzei et al. 2021).

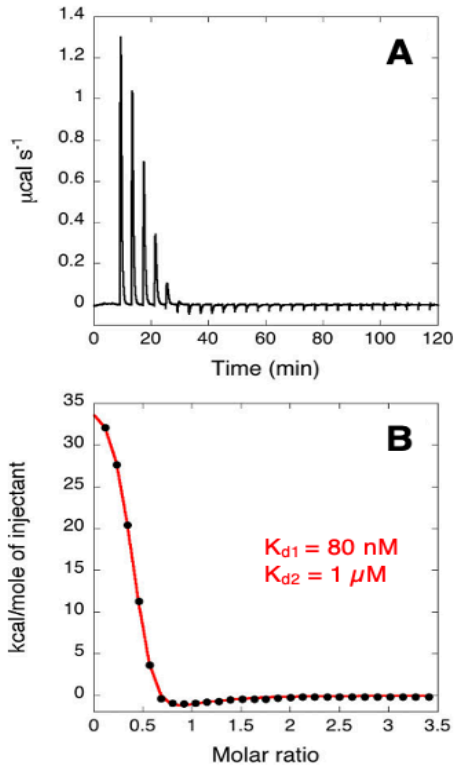


Figure 59: *Raw data of the titration of SgSrnR (13 µM) with OP_{sodF} sequence (140 µM) (**A**). Integrated heat data (filled dots) are interpolated with the two sets of binding sites model (solid red line) (**B**). Adapted by permission from Oxford University Press (Mazzei et al. 2021) with License numbers 5232421500534 and 5232431038226. Copyright © 2021, Oxford University Press.*

When an equimolar concentration of OP_{sodF} DNA sequence was added to the protein solution, the signals corresponding to the residues in the well-folded central region of the protein disappeared in the $^{1}H,^{15}N$ TROSY-HSQC (Figure 60). This is due to the formation of larger complex, with faster relaxation and consequently lower signal intensities. Persistence of NMR signals from residues located in the N-terminal and C-terminal regions suggests that these regions are minimally involved in DNA binding (Mazzei et al. 2021).

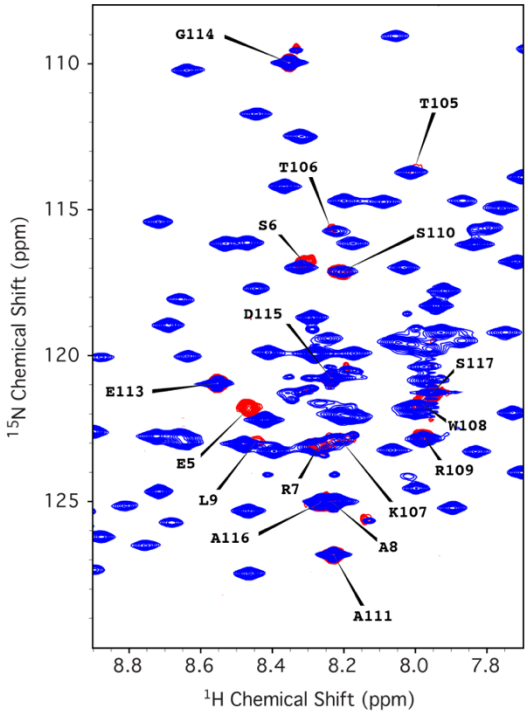


Figure 60: *Overlay of the ^{1} H-^{15}N HSQC spectrum and of the ^{1} H-^{15}N TROSY-HSQC spectrum of triply labeled SgSrnR (^{2}H ,^{13}C,^{15}N)(500 µM) at 800 MHz in 20 mM TrisHCl pH 7.5, 150 mM NaCl, 1 mM TCEP and 5% D_2O at 298 K in the absence (blue) and in the presence (red) of one equivalent of OPsodF DNA sequence. Reprinted by permission from Oxford University Press (Mazzei et al. 2021) with License numbers 5232421500534 and 5232431038226. Copyright © 2021, Oxford University Press.*

5.7 β-Galactosidase gene reporter assays

To assess the activity of *Sg*SrnR as activator or repressor in the regulation of *sodF* expression, an assay was developed using β-galactosidase (β-Gal) as a reporter gene and o-nitrophenyl-β-D-galactopyranoside (ONPG) as substrate of the enzyme, producing yellow-colored galactose and o-nitrophenol: more yellow is the solution, higher is the β-Gal concentration. The *lacZ* gene, coding for the β-gal, was placed under the control of the P_{sod} promoter. The obtained construct *pKT25::*P_{sod}*-LacZ* was co-transformed in *E. coli* BL21 (DE) cells with the *pET15b::HisTag-SgSrnR* expression vector, resulting in β-Gal activity, measured as Miller Units, only in the presence of a specific inductor. A yellow coloring of the reaction mixture was observed following induction of *Sg*SrnR with IPTG, increasing in intensity over time and indicating a corresponding increase in β-galactosidase activity. This result suggested the activity of *Sg*SrnR as activator of the transcription of *sodF*. The data were confirmed by negative control experiments and by the addition of Ni(II) up to 200 μM, supporting the hypothesis made with the DNase I footprinting, according to which Ni(II) has no effect on the capacity of *Sg*SrnR to bind DNA (Figure 61) (Beniamino et al. 2020).

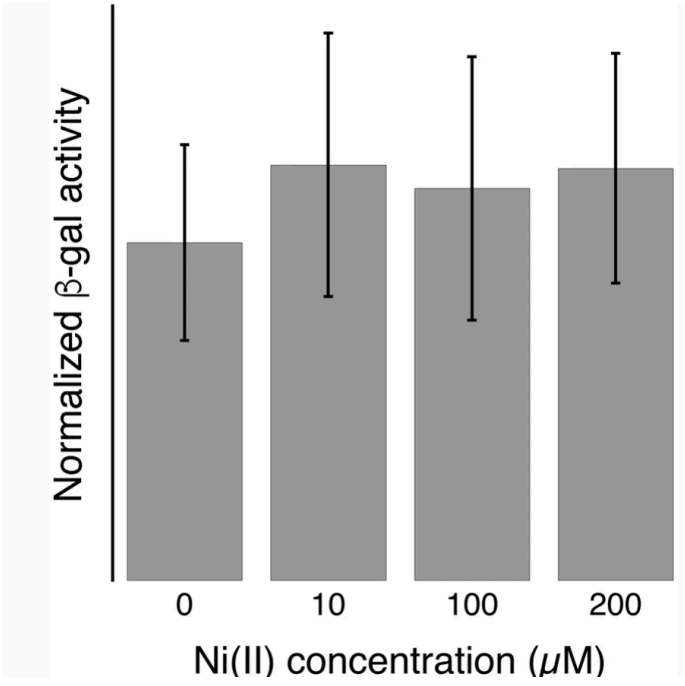


Figure 61: *Normalized β-gal activity measured as Miller units in E. coli BL21 (DE3) cells to determine the activity of SgSrnR as activator or repressor of sodF transcription in the absence and in the presence of increasing Ni(II) concentration. The latter is reported as the ratio to the activity produced by the culture without addition of Ni(II) ions. Reprinted by permission from Springer Nature (Beniamino et al. 2020) with License number 5232420214476. Copyright© 2019, Society for Biological Inorganic Chemistry (SBIC).*

6. Conclusions and perspectives part 2

The second part of this PhD thesis was focused on the characterization of the *Sg*SrnR protein, one of the two proteins of the nickel-dependent transcriptional regulatory system, involved in the regulation of the expression of FeZn-SOD in *S. griseus*. A great effort was made to develop a purification protocol that would allow to obtain enough protein to make a biophysical and structural characterization. Several techniques have been applied. The light scattering technique revealed that the protein in solution is a dimer, as well as the other proteins of the ArsR/SmtB family. The oligomeric state of *Sg*SrnR was validated by X-ray crystallography. The crystallographic structure of *Sg*SrnR, recently obtained with 1.93 Å resolution by our group (Protein data Bank accession code 7P6F), confirmed that the protein has an ArsR/SmtB-like fold in the core region with a helix-loop-helix binding motif to DNA and flexible regions at the N- and C-terminal ends, as it was confirmed by disorder prediction analysis and NMR spectroscopy (Figure 62) (Mazzei et al. 2021).

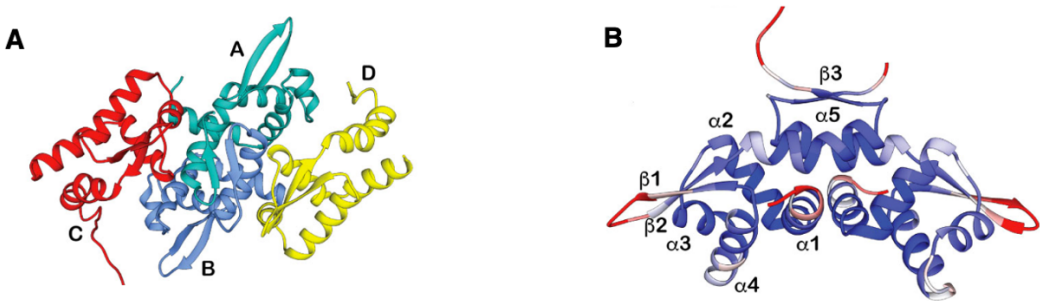


Figure 62: *Monomer of SgSrnR identified in the asymmetric unit, each colored with a different color (**A**). Dimeric biological unit of SgSrnR with indicated the elements of secondary structure for each monomer. α4 is the helix interacting with the DNA. The dimer was colored according to the B-factor (Å^2) (red, B-factor >75; white, 25 <B-factor <75; blue, B-factor <25) (**B**). Reprinted by permission from Oxford University Press (Mazzei et al. 2021) with License numbers 5232421500534 and 5232431038226. Copyright © 2021, Oxford University Press.*

The two-steps DNA binding demonstrated in this work was already found for other members of this family of transcription factors and *in silico* molecular docking analysis demonstrated that the binding involves only the central region of the protein, with α4-helix interacting with the OP_{sodF} DNA sequence 42 bp length, identified in the DNase footprinting assay (Figure 57). Conversely, the unstructured terminal tails are free. Two possible roles can be given to these regions: they might interact with *Sg*SrnQ, undergoing a disorder-order transition, or they can directly contact the RNA polymerase, bringing it closer to the promoter (Mazzei et al. 2021).

*Sg*SrnR binds Ni(II) ion with moderate affinity. Ni(II) interaction occurs at the level of the non-native GSH sequence at the N-terminus (Mazzei et al. 2021), without altering the protein structure. On the contrary, Zn(II) induces conformational changes incompatible with the biological functions of the protein, inducing its aggregation, as resulted from CD, ITC and light scattering analyses.

In the DNase footprinting assay *Sg*SrnR was found to be able to bind DNA, even in the absence of co-factors or interaction partners. This contrasts with the previously proposed model, which hypothesized the interaction between *Sg*SrnR, *Sg*SrnQ and Ni(II) ions, followed by the binding to the *sodF* operator and consequent inhibition of its transcription by detachment of the RNA polymerase (Figure 63A) (Kim et al. 2003). According to this model, *Sg*SrnR acts as a repressor, binding directly to the DNA, while the Ni(II) containing protein *Sg*SrnQ acts as co-repressor. Surprisingly, the β-galactosidase assay resulted in β-gal increased activity, following induction of *Sg*SrnR expression. Enzymatic activity was not found to be affected by the presence of nickel. Therefore, at the end of this work, it was possible to suppose a new possible mechanism, with apo-*Sg*SrnR having a role of transcription activator (Figure 63B) (Beniamino et al. 2020).

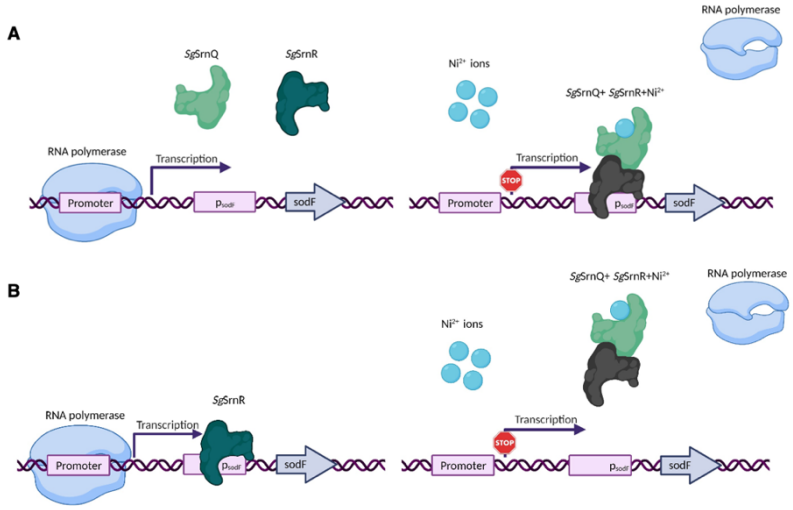


Figure 63: *Mechanism of functioning of SgSrnR as a repressor, according to the previously proposed model: SgSrnR forms a complex with SgSrnQ and binds the operator sequence, repressing the transcription (Kim et al. 2003) (**A**). Mechanism of functioning of SgSrnR resulting from this work and reported in Beniamino et al. 2020: SgSrnR binds the operator sequence and activates the transcription of sodF. In the presence of SgSrnQ, DNA release occurs and the transcription of sodF is inhibited (**B**). Created with BioRender.com.*

Future projects aim at the characterization of the *Sg*SrnQ partner. Several purification trials were conducted in the laboratory of bioinorganic chemistry of the University of Bologna, but the yields of the recombinant protein were very low. As the protein is mostly intrinsically disordered, proteolytic degradation in the recombinant host is possible. Current research is focused on the co-expression of the *Sg*SrnR-*Sg*SrnQ complex, in the attempt to characterize the complete functional unit. Indeed, obtaining the *Sg*SrnR-*Sg*SrnQ complex is of important to get a clearer picture of this nickel-dependent regulatory system and many intrinsically disordered transcription factors take on a structure when they bind their partner, becoming less susceptible to degradation.

7. Materials and Methods

7.1 Bacterial strains and growth conditions

E. coli XL10-ultracompetent cells (Agilent Technologies #200315), characterized by the high transformation phenotype (Hte) and an average transformation efficiency of 5 x 10^9 transformants/µg of supercoiled DNA Plasmid, were used for cloning and DNA ligation.

E. coli BL21 (DE3) (Agilent Technologies #200131) and BL21-CodonPlus (DE3) RIL (Agilent technologies #230245) competent cells combine efficient high-level expression of heterologous proteins with T7 promoter-driven vectors and easy induction after isopropyl-1-thio-β-D-galactopyranoside (IPTG) addition. Both strains lack the *lon* and *ampT* proteases and the BL21-Codon Plus (DE3) RIL cells are engineered with extra copies of the *argU*, *ileY* and *leuW* tRNA genes, coding tRNAs that recognize the arginine codons AGA and AGG, the isoleucine codon AUA and the leucin codon CUA respectively (Novagen 2011) (www.agilent.com).

E. coli XL10-ultracompetent cells, BL21 (DE3) and BL21-CodonPlus (DE3) RIL (Agilent technologies #230245) competent cells were grown on Luria-Bertani (LB) agar plates or LB liquid broth or ZY auto-induction medium; when required, ampicillin, chloramphenicol and kanamycin were added to the medium at the final concentration of 100 µg/mL, 34 µg/mL and 30 µg/mL respectively.

NEB® 5-alpha Electrocompetent *E. coli* cells (NEB #2989K) is a derivative of the popular DH5α and show T1 phage resistance and endA deficiency. These cells were used for in cell EPR experiments (Table 1).

***Table 1**: Escherichia coli strains used in this work*

Escherichia coli strains	Description	References
XL10-Gold ultracompetent cells	*TetrD(mcrA)183 D(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F′ proAB lacIqZDM15 Tn10 (Tetr) Amy Camr].*	(Novagen 2011)
BL21 (DE3) cells	*E. coli B F– dcm ompT hsdS(rB – mB –) gal λ(DE3)*	(Novagen 2011)
BL21 (DE3) CodonPlus RIL cells	*E. coli B F– ompT hsdS(rB – mB –) dcm+ Tetr gal λ(DE3) endA Hte [argU ileY leuW Camr]*	(Novagen 2011)
NEB® 5-alpha Electrocompetent *E. coli* cells	*fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	(Singh et al. 2014)

7.2 Gene cloning

7.2.1 Construction of the *pET15b:: StrepTag-hNDRG1* vector

The synthesis of the gene encoding for the human NDRG1 protein (Uniprot code: Q92597-1), 1197 bp, was performed by Eurofins, introducing the recognition sites for the restriction enzymes

NdeI and BamHI at 5' and 3' respectively, and cloned into the pEX-K4 sub-cloning vector. After purification from *E.coli* XL10-Gold Ultracompetent Cells (Agilent) using the NucleoSpin® plamid miniprep kit (Macherey-Nagel), this construct was double digested with FastDigest NdeI and BamHI restriction enzymes (Fermentas). The corresponding *hNDRG1* gene, using T4 DNA ligase (Promega) was ligated into a *pET15b* vector (5.7 kb) containing a N-terminal Strep-tag (WSHPQFEK) (Miraula et al. 2015) instead of the HisTag of the commercial pET15b (AddGene) and digested with the same restriction enzymes. Both the insert and the vector were previously purified from a 1% (w/v) agarose gel using the NucleoSpin® Gel and PCR clean-up kit. The obtained *pET15b::StrepTag-hNDRG1* (Figure 63) was sequenced at both strands. The performed cloning was checked by sequencing at both ends by Eurofins, employing the T7 promoter and T7 terminator primers.

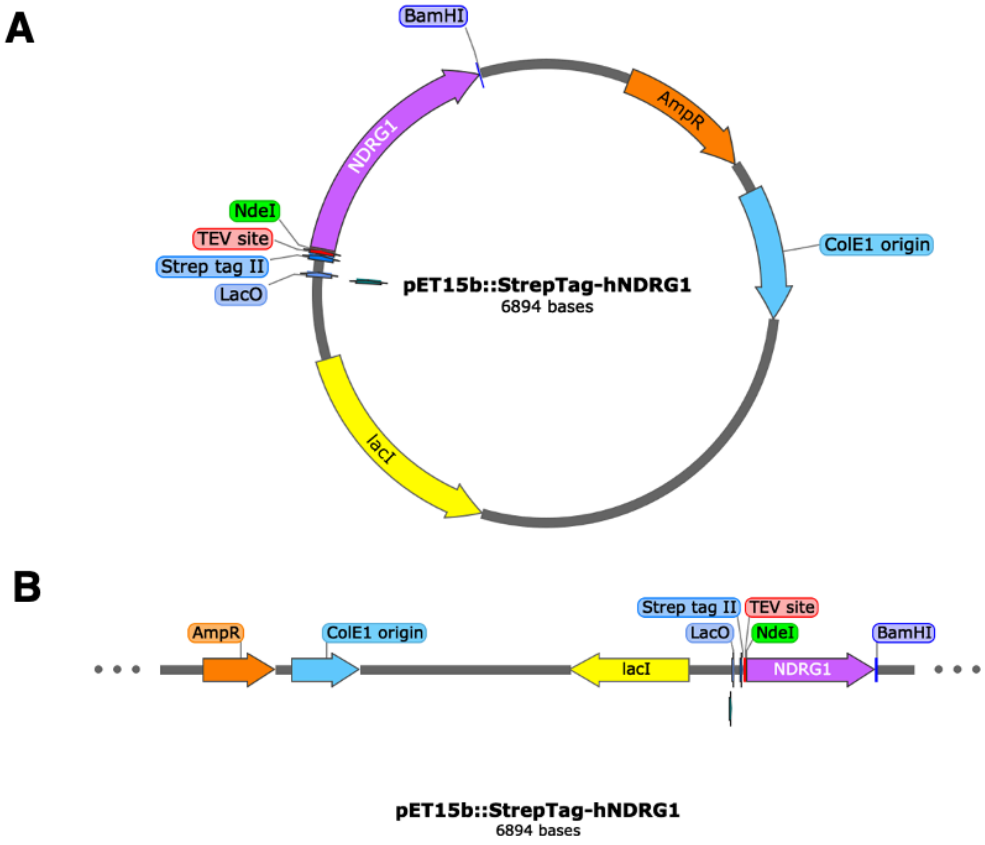


Figure 63: *pET15b::StrepTag-hNDRG1 map, circular (**A**) and linearized (**B**). Created with SnapGene® software (from Insightful Science; available at snapgene.com).*

7.2.2 Construction of the *pETZZ1a:: hNDRG1*C* vector

The *pETZZ1a::hNDRG1*C* vector was obtained by PCR amplifying the C-terminal sequence of the protein, here called *h*NDRG1*C, using the *pET15b::StrepTag-hNDRG1* construct as template, the Easy-A High Fidelity PCR Cloning Enzyme (Agilent) and the primer pairs NDRG1*C (5'-

CAACATATGGGCTATATGCCGAGCG-3')/T7-terminator_R (5'-GCTAGTTGCTCAGCGG-3'). Following double digestion with FastDigest NdeI and BamHI restriction enzymes (Fermentas), the purified PCR product was cloned into the modified *pET15b* expression vector previously described (Miraula et al. 2015). The obtained *pET15b::StrepTag-hNDRG1*C* plasmid was further cut with FastDigest NcoI and BamHI restriction enzymes (Fermentas), resulting in the 323 bp *StrepTag-hNDRG1*C* insert. This was subsequently ligated into the *pETZZ1a* expression vector (Bogomolovas et al. 2009), obtaining a recombinant protein that shows a HisTag, a fragment of the immunoglobulin binding domain of protein A of *Staphylococcus aureus* (ZZ-tag) and the StrepTag at the N-terminal region of *h*NDRG1*C (Figure 64).

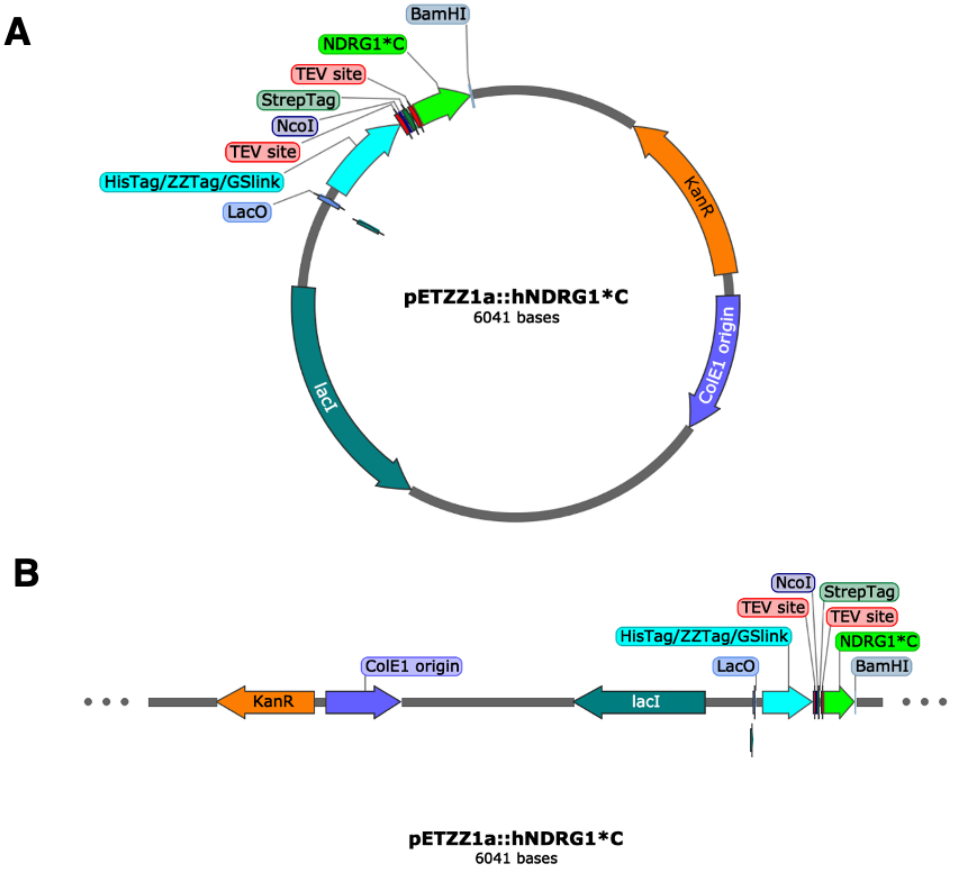


Figure 64: *pETZZ1a::hNDRG1*C map, circular (**A**) and linearized (**B**). Created with SnapGene® software (from Insightful Science; available at snapgene.com).*

7.2.3 Site directed mutagenesis

The *pETZZ1a::hNDRG1*C* plasmid was subjected to single or double site-directed mutagenesis, using the QuikChange Lightning Site-Directed Mutagenesis kit (Agilent Technologies), in order to obtain *h*NDRG1*C double or single cysteine variants, respectively, to be used in SDSL-EPR experiments. In the protein sequence of the full length protein three serine residues, not involved in the phosphorylation processes and localized in the C-terminal region, were identified as possible

targets for cysteine mutations: Ser336, Ser357 and Ser378 (Figure 65). Cys394, naturally present at the C-terminal end of the protein, was preserved, obtaining three two cysteines variants or replaced with an alanine, in order to obtain three single cysteine variants. In conclusion 6 mutants of *h*NDRG1*C were obtained (Table 2). The oligonucleotides used in the mutagenesis reactions are shown in Table 3.

Figure 65: *hNDRG1 amino acid sequence. In bulk the sequence corresponding to hNDRG1*C. The mutagenized amino acid are highlighted with orange square, while the portion with the decapeptide repeated three times is in green.*

Table 2: *Single and double cysteines variants of hNDRG1*C.*

Number of cysteines	*h*NDRG1*C variants	Number of mutations
Two	*h*NDRG1*C_S336C_C394	One
Two	*h*NDRG1*C_S357C_C394	One
Two	*h*NDRG1*C_S378C_C394	One
One	*h*NDRG1*C_S336C_C394A	Two
One	*h*NDRG1*C_S357C_C394A	Two
One	*h*NDRG1*C_S378C_C394A	Two

Table 3: *Oligonucleotides used for site directed mutagenesis of hNDRG1*C. The codons that introduce the mutations are highlighted in red.*

Mutation	Oligonucleotide forward	Oligonucleotide reverse
Ser 336 ⇒ Cys 336	5'-cagctccgtgacctgcctcgatggtacccg-3'	5'-cgggtaccatcgaggcaggtcacggagctg-3'
Ser 357⇒ Cys 357	5'-tctcgtagccatacctgcgaaggcacgcgtagtc-3'	5'-gactacgcgtgccttcgcaggtatggctacgaga-3'
Ser 378⇒ Cys 378	5'-ctggacattactccgaactgtggcgcagcc-3'	5'-ggctgcgccacagttcggagtaatgtccag-3'
Cys 394⇒ Ala 394	5'-cgatggaagtgtctgcctaataaggatccg-3'	5'-cggatccttattaggcagacacttccatcg-3'

7.3 Protein expression and purification

7.3.1 Expression and purification of *h*NDRG1

The recombinant Strep-Tagged *h*NDRG1 protein (residue 1-394), was expressed *in E. coli* BL21-CodonPlus (DE3) RIL competent cells and in 1L of ZY auto-induction medium (10 g L^{-1} triptone, 5 g L^{-1} yeast extract, 5 g L^{-1} glycerol, 3.3 g L^{-1} $(NH_4)_2SO_4$, 6.8 g L^{-1} KH_2PO_4, 7.1 g L^{-1} Na_2HPO_4, 0.120 g L^{-1} $MgSO_4$, 0.5 g L^{-1} glucose, 2 g L^{-1} lactose), enriched with 100 µg/mL of ampicillin and 34 µg/mL of chloramphenicol. Protein expression started after 4 h at 37 °C with vigorous stirring and was conducted at 26 °C for 18 h. When protein expression was completed, the cells were centrifuged at 9000 x g for 20 min and the cellular pellet was resuspended in 60 mL of lysis buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA, 5 mM DTT, 20 µg/mL DNAseI, 10 mM MgCl2, 1 mM PMSF, 1 mM benzamidine hydrochloride, 5 µg/mL pepstatin A).

Cell lysis was carried out with two passages in French Press SLM Aminco from which the crude was subjected to centrifugation at 25000 x g for 30 min at 4 °C , to remove the insoluble part. Subsequently the *h*NDRG1 protein was affinity purified by loading the soluble fraction onto a StrepTrap™ HP 5 mL column (GE Healhcare), pre-equilibrated with the binding buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA, 5 mM DTT) and eluted with the elution buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA, 5 mM DTT, 2.5 mM D-desthiobiotin).

The fractions containing the protein were pooled and concentrated using a 10-kDa cut-off membrane ultra-filtration units (Millipore) to be loaded on a Superdex 200 16/60 column equilibrated with 20 mM Hepes pH 7.5, 150 mM NaCl and 1 mM TCEP. The purity of the protein was checked by SDS-Page using the NuPAGE 4–12 % Bis–tris acrylamide gels. The final protein concentration, referred to the monomeric form of the protein, was determined using the theoretical molar extinction coefficient at 280 nm of 38,890 M^{-1} cm^{-1} and calculated using the ProtParam website (https://web.expasy.org/protparam/). The protein was stored at -80°C, waiting to be used for the next biochemical and biophysical analysis.

7.3.2 Expression and purification of *h*NDRG1*C

The fusion polypeptide His-Tag/ZZ-tag/Strep-Tag/*h*NDRG1*C was expressed *in E. coli* BL21-CodonPlus (DE3) RIL competent cells, grown in 1-2 L of lysogeny broth (LB) with the addition of 30 µg/mL kanamycin and 34 µg/mL chloramphenicol, at 37 °C and vigorous stirring. When OD_{600} was 0.6, 0.5 mM IPTG was added and temperature decreased to 26 °C for 18 h, to induce protein expression. For labeled *h*NDRG1*C production, before induction, the cells were harvested and resuspended in 2 x M9 minimal medium, (1.26 g L^{-1} $(NH_4)_2SO_4$, 12 g L^{-1} Na_2HPO_4, 6 g L^{-1}

KH_2PO_4, 1 g L^{-1} NaCl, 4 g L^{-1} glucose, 0.240 g L^{-1} $MgSO_4$), containing either ^{15}N or $^{15}N/^{13}C$ isotopes, reaching a cell concentration 4x greater than that in LB (Azatian, Kaur, and Latham 2019; Marley, Lu, and Bracken 2001). The cells were incubated for 30 min to allow growth recovery and the clearance of unlabeled metabolites, before induction with 0.5 mM IPTG.

Cells were pelleted by centrifugation at 9000 x g for 20 min at 4 °C, resuspended in 30 mL of the lysis buffer (50 mM Tris-HCl pH 7.6, 300 mM NaCl, 20 mM imidazole, 20 µg/mL DNAseI, 10 mM MgCl2, 1 mM PMSF, 1 mM benzamidine hydrochloride, 5 µg/mL pepstatin A) and disrupted by two passages through a French pressure cell (SLM Aminco) at 20,000 pounds/square inch. The lysate was centrifuged at 25000 x g for 30 min at 4 °C and the soluble part subjected to the first purification step, represented by an affinity chromatography performed using the 5 mL HisTrap HP column (GE Healthcare), pre-equilibrated with the binding buffer (50 mM Tris-HCl pH 7.6, 300 mM NaCl, 20 mM imidazole). The elution protein was performed with a linear gradient from 0% to 100% of the elution buffer (50 mM Tris-HCl pH 7.6, 300 mM NaCl, 500 mM imidazole).

Fractions containing the fusion protein were pooled and incubated with TEV protease for 3 h at room temperature, with a 1:50 protease:protein ratio. Next, the protein buffer was changed for the next purification step. The buffer change was performed using the GE Healthcare HiPrep 26/10 Desalting column, previously equilibrated with the buffer 20 mM Tris-HCl pH 7.5 containing 2 mM EDTA and 1 mM DTT. The peak fractions were pooled and the His-Tag/ZZ-tag/Strep-Tag polypeptide was separated from *h*NDRG1*C with a cation exchange chromatography, performing a linear gradient from 0 to 1M NaCl at pH 7.5.

Fractions containing the protein, those between 200 and 280 mM NaCl, were pooled and concentrated with 3 kDa MWCO Centricon ultra-filtration units (Millipore) and loaded into a Superdex 75 16/60 column (GE Healthcare) equilibrated with 20 mM Hepes containing 150 mM NaCl and 1 mM TCEP at pH 7.5 or 6.5 for the labeled protein. Protein concentration was determined using the absorption coefficient at 280 nm, experimentally calculated by inductively coupled plasma emission spectroscopy (ICP-ES), quantifying the sulfur concentration in a protein sample with known absorbance at 280 nm (Miraula et al. 2015). This technique was used also to check the absence of any metal bound in protein sample (Stola et al. 2006). The same purification protocol was followed also to purify the single and double *h*NDRG1*C cysteine variants.

7.4 Isothermal titration calorimetry

*h*NDRG1 and *h*NDRG1*C protein solutions in the sample cell (1.4093 mL) were titrated with 29 x 10 µL injections of $NiSO_4$ (300-500 µM) using a high-sensitivity VP-ITC microcalorimeter (MicroCal) equipped with a computer-controlled 310-µL microsyringe. To reach the final

concentrations, protein and nickel solutions were diluted in the same buffer, containing 20 mM HEPES, 150 mM NaCl and 1 mM TCEP at pH 7.5 (*h*NDRG1) or at the pH range 6.5-7.5 (*h*NDGR1*C) (Table 2). An interval of 6 minutes was applied between the injections to allow the system to reach thermal equilibrium. The metal ion heat of dilution in buffer alone under identical conditions was checked as control experiment and found to be non-significant. The Origin software package (MicroCal) was used to analyze and fit the data with a non-linear least-squares minimization algorithm to theoretical curves corresponding to different binding models, giving as result information about ΔH (reaction enthalpy change), K_A (binding affinity constant), n (number of binding sites), N (binding stoichiometry) and ΔS (reaction entropy). The values given for ΔH and ΔS are apparent, and include contributions from Ni(II) binding, as well as from events like deprotonation of the residues and consequent change in the buffer ionization state. The chi-square parameter $\chi 2$ was used to establish the best fit.

Table 4: *Isothermal titration calorimetry conditions*

	Protein concentration	Nickel concentration	pH
*h*NDRG1	23 µM	500 µM	7.5
*h*NDRG1*C	70 µM	600 µM	6.5
*h*NDRG1*C	70 µM	600 µM	7
*h*NDGR1*C	60 µM	300 µM	7.5
*h*NDRG1*C	40 µM	200 µM	8
*h*NDRG1*C	40 µM	200 µM	8.5

7.5 Determination of the SDS critical micellar concentration

The critical micellar concentration (CMC) of SDS was calculated in 10 mM Tris pH 7.4 and 20 mM HEPES pH 7.5 containing 150 mM NaCl and 1 mM TCEP, using the fluorescent dye N-phenyl-1-naphthylamine (NPN) (Brito and Vaz 1986). Stock solution of SDS, with a concentration of 100 mM, was prepared in the two buffers tested, while the working 500 µM NPN solution was made in 99% ethanol. For the determination of the CMC, various tubes with increasing concentrations of the surfactant were prepared in the chosen buffer, adding 1 µM NPN and kept at room temperature for 30 min, before measuring the emission intensity. The excitation wavelength was 340 nm and the emission was registered from 350 nm to 550 nm. The maximum NPN emission intensity was plotted as a function of SDS concentration. In this plot, two straight lines, defining

the NPN emission intensity, were traced and their intersection point associated to the CMC of SDS. (Brito and Vaz 1986).

7.6 Circular dichroism

The JASCO J-810 spectropolarimeter flushed with N_2 was used to investigate the *h*NDRG1 and *h*NDRG1*C secondary structure by circular dichroism. The protein concentration was 30 µM and 200 µM respectively for *h*NDRG1 and *h*NDRG1*C, both in 20 mM HEPES pH 7.5, 150 mM NaCl and 1 mM TCEP. The cuvette with 0.1 mm path length was used and the spectra were recorded at 25 °C from 260 nm to 190 nm at 0.2 nm intervals and, to reach a good signal to noise ratio, ten spectra were accumulated and averaged. The experiments were carried out in the absence and in the presence of increasing concentration of $NiSO_4$. The CD spectra of the single cysteine variants of *h*NDRG1*C were recorded in the presence of 30% trifluoro-ethanol (TFE) and 250 uM and 1 mM of sodium dodecyl sulfate (SDS). The software Best Structure Selection (BeStSel) was used to quantify the secondary structure compositions of *h*NDRG1 and *h*NDRG*C (Micsonai et al. 2018). The effect of nickel ions on *h*NDRG1 (31 µM) and *h*NDGR1*C (900 µM) protein conformation, involving the side chains of the aromatic residues, was evaluated using near-UV CD spectroscopy (250-350 nm) in a 1 mm path length cuvette.

7.7 Light scattering

The light scattering data were collected using a Superdex 200 10/300 GE column (GE Healthcare) (*h*NDRG1) or a Superdex 75 10/300 GE column (GE Healthcare) (*h*NDRG1*C), connected to a multiple angle laser-light (690 nm) scattering DAWN EOS photometer (Wyatt Technology) and to a quasi-elastic light scattering apparatus (Wyatt QELS) (Challener 2014; Tarazona and Saiz 2003). *h*NDRG1 (300 µL, 140 µM) and *h*NDRG1*C (300 µL, 1.2 mM) in 20 mM HEPES pH 7.5, 150 mM NaCl and 1 mM TCEP were injected into the system, equilibrated with the same buffer, in the absence and in the presence of equimolar concentrations of Ni(II). In this case, 28 µM and 240 µM of $NiSO_4$ were added to the buffer for *h*NDRG1 and *h*NDRG1*C, respectively. The elution was performed at room temperature with a flow rate of 0.5 mL/min. Values of 1.33 and 0.185 mL/g were used for the refractive index of the solvent and the specific refractive index increment (dn/dc) respectively. Data analysis was performed with the ASTRA software (version 5.3.4)(Wyatt Technology) and a Rayleigh-Debye-Gans (RGD) light scattering model, according to manufacturer's instructions. The uncertainties on the MW, calculated across the entire elution profile in

the intervals of 0.2 s, are a measure of the statistical consistency of the MALS data, obtained combining the standard deviations calculated for each part of the analyzed peaks.

7.8 NMR spectroscopy

200-300 µL of U-^{15}N or U-^{13}C-^{15}N *h*NDRG1*C samples with a concentration of 0.8-1 mM in NMR buffer (20 mM Hepes, 150 mM NaCl, 1 mM TCEP in the pH range 6.5-7.5) enriched with 10% D_2O and in 3-mm NMR tubes, were used for the NMR assignment. The Bruker AVANCE NEO/III spectrometer (CERM - Centre for Magnetic Resonance – University of Firenze, Italy) operating at 28.2 T (1200.73 MHz ^{1}H Larmor frequency) and equipped with a 3-mm triple resonance inverse TCI z-gradient cryo-probe was used to collect at 298K NMR spectra for the assignment of *backbone* [^{1}H-^{15}N BEST-TROSY (Transverse Relaxation Optimized SpectroscopY) ^{1}H-^{15}N and ^{1}H-^{13}C Heteronuclear Single Quantum Correlation (HSQC) spectra, HNCO, HNcaCO, HNCA, HNCACB, CBCAcoNH, HNcoCACB, HBHANH and HBHAcoNH], *aliphatic side-chains* [^{1}H-^{13}C HSQC, hCCH-TOCSY, HCcH-TOCSY and CcoNH] and aromatic side chain*s* [2D ^{1}H-^{1}H TOCSY and hbCBcgcdHD (CBHD)]. ^{13}C-detected NMR spectra (CON, hCACO and hCBCACO)(Bermel et al. 2009)(Felli and Pierattelli 2014) were acquired at 298 K using a 16.4 T Bruker AVANCE NEO spectrometer operating at 16.5 T (700.06 MHz ^{1}H Larmor frequency), equipped with a 3 mm TXO cryoprobe optimized for ^{13}C direct detection.

The NMRpipe (Delaglio et al. 1995) and the SMILE (Sparse Multidimensional Iterative Line-shape-Enhanced) reconstruction algorithm (Ying et al. 2017) plug-in module implemented in NMRpipe were used to process all NMR spectra. The PINE server (Lee et al. 2009) and LACS (Linear Analysis of Chemical Shifts) (Wang et al. 2005) were applied for the initial automated assignments and the validation of the final assignment, respectively. The assignment was deposited in the Biological Magnetic Resonance Bank (BMRB) with the accession code 50803. ^{1}H-^{13}C HSQC spectra in the presence of $NiSO_4$ from 0.25 to 1 equivalents were acquired in the aromatic region to study change in chemical shifts of the histidine residues.

7.9 EPR spectroscopy

7.9.1 Site directed spin labeling reaction (SDSL)

Before the spin-labeling reaction, in order to avoid the reduction of the nitroxide spin label, 100 nmol of *h*NDRG1*C single or double cysteines variants were loaded into a PD10 desalting columns to remove the reduced agent TCEP, present in the buffer used to store the protein after the purification process. The PD10 desalting column was equilibrated with 20 mM Tris-HCl pH 7.5

and 150 mM NaCl. The fractions containing the protein were pooled and incubated with Maleimido-Proxyl nitroxide (M-Prox) in 10-fold molar excess. The incubation was carried out for 4 hours at 4 °C under gentle stirring. The molar excess was increased up to 20-fold for *h*NDRG1*C variants containing two cysteine residues, performing a second injection after 2 hours. A second PD10 desalting column against 20 mM Tris-HCl pH 7.5 and 150 mM NaCl, for the double cysteine variants or 10 mM Tris-HCl pH 7.4 for the single cysteine variants, was performed to remove the unbound spin-label excess. The collected fractions were then checked by EPR spectroscopy and polished by centrifugation in 2 mL Vivaspin concentrators (3kDa MWCO, Sartorius), reaching a final spin concentration of ~300 µM (Figure 66).

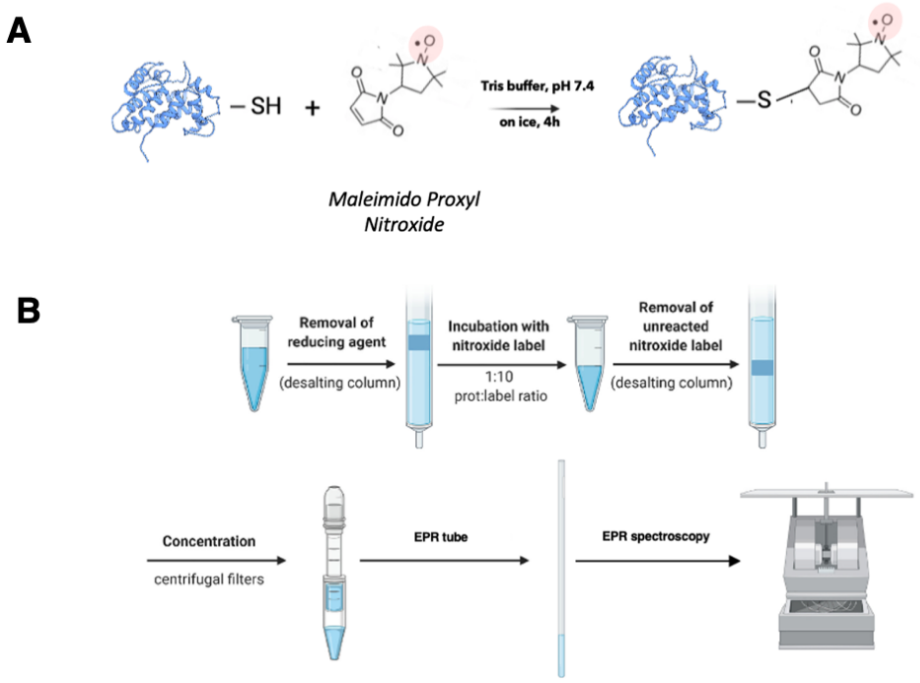


***Figure 66**: Site directed spin labeling reaction (SDSL) (**A**) and workflow of the reaction (**B**). Created with BioRender.com.*

7.9.2 Mass spectrometry

After the labeling reaction, mass analysis (MALDI-ToF) was performed to confirm the labeling with Maleimido Proxyl nitroxide (M-Prox). Expected mass increment is of 237 Da or 474 Da respectively for one or two molecules of nitroxide grafted on the protein. Samples of ~40 pmol of unlabeled *h*NDRG1*C and labeled *h*NDRG1*C variants were prepared by dilution in 10 µL of 0.1 % of trifluoro-acetic acid (TFA) in water (v/v) before being spotted onto a MALDI target plate (1 µL). A saturated solution of α-Cyano- 4-hydroxycinnamic acid matrix (1 µL) in 70 % acetonitrile/water, 0.1% TFA (v/v) was added. The MALDI-ToF mass spectrometer Microflex II from

Bruker Daltonics was used to measure in a positive linear mode the global mass in the range from 2 to 65 kDa. The signals from the Protein standard I (Bruker Daltonics) were employed to perform the external mass calibration. The error on the measurement was of +/-5 Da.

7.9.3 In vitro CW-EPR

7.9.3.1 Determination of the protein pool spin concentration

The spectrometer Elexsys 500 Bruker equipped with a Super High Q sensitivity resonator operating at X band (9.9 GHz) at room temperature was used to determine the protein pool spin concentrations. Protein samples were analyzed in a quartz capillary whose sensible volume was 53 µl. The experiments were carried out setting the following parameters: microwaves power 10 mW, magnetic field modulation amplitude 0.1mT, field sweep 15 mT and receiver gain 60 dB. The spin concentration was calculated from the double integration of the CW-EPR spectrum at room temperature and under non-saturating conditions and compared with that of a standard solution of known concentration. The simulations of the spectra were performed using SimLabel (Etienne et al. 2017), a MatLab graphical user interface for EasySpin (Stoll and Schweiger 2006).

7.9.3.2 Crowding effect

*h*NDRG1*C single cysteine variants labeled with M-Prox nitroxide label were tested in the presence of several osmolytes and crowders, working at X band (9.9 GHz). The protein concentration in the EPR capillary was 40 µM in 10 mM Tris pH 7.4 in the presence of osmolytes or crowders tested: TFE and Glycerol 30%, Ficoll 70 PM 300 mg/mL, Sucrose 30% w/v, Potassium Glutamate 200 mM, TMAO 2M, BSA 300 mg/mL and SDS 2 mM, 4 mM and 8 mM (Table 5). The EPR spectra parameters were: microwaves power 10 mW, magnetic field modulation amplitude 0.1 mT, field sweep 15 mT, receiver gain 60 dB and 12 scans.

Table 5: *Stock conditions of the tested osmolytes and crowders*

Crowders and osmolytes	Stock concentration	Buffer
TFE	100%	-
FICOLL PM70	300 mg/mL	10 mM Tris pH 7.4
Sucrose	30% w/v	10 mM Tris pH 7.4
KGlu	200 mM	10 mM Tris pH 7.4

TMAO	6 M	Milli Q
Glycerol	100%	-
SDS	50 mM	10 mM Tris pH 7.4
BSA	300 mg/mL	10 mM Tris pH 7.4 + 300 mM NaCl

7.9.4 In cell EPR

7.9.4.1 Preparation of electrocompetent cells

An overnight pre-culture of *E. coli* NEB® 5-alpha cells was diluted to a final OD_{600}= 0.05 into 50 mL of LB medium and cultured at 37 °C until reaching OD_{600}= 0.9. The cells were chilled at 4 °C on ice for at least 30 minutes to stop the growth. Bacterial cells were harvested by centrifugation at 4000 x g for 10 minutes at 4 °C and resuspended in 50 mL of sterile MilliQ water + 10% glycerol. This washing step was repeated two times, reducing the volume of the washes to 40 mL and 30 mL, respectively. Finally, bacterial cells were resuspended in sterile MilliQ water + 10% glycerol to obtain a final concentration between $4-7 * 10^{10}$ cells/mL and stored at −80 °C in 25 µL aliquots.

7.9.4.2 In cell EPR by electroporation

20 µL of homemade electrocompetent cells were incubated with 20 µL of labeled protein, 5 minutes on ice. The protein-cells mixture was, then, transferred into a pre-chilled 1 mm-gap cuvette (VRW) and electroporated in a Gene Pulser Xcell™ from Bio-Rad, using the following parameters: 1800 V/cm, 200 Ω, 25 µF, 1 msec pulse. The cells membrane integrity was recovered by addition of 500 µL of pre-warmed SOC. Cells were then washed by centrifugation at 3200 x g, 2 minutes at 5 °C four times: 500 µl of PBS (Phosphate Buffer 10 mM pH 7.5, KCl 2.7 mM, NaCl 137 mM) + 0.005% Triton solution in the first wash and 500 µL of PBS in the other washes. As final step, the pellet was resuspended in 20 µL of PBS + 30 µL of PBS with 1% -Agarose (Euromedex) and transferred in an EPR quartz capillary for CW-EPR at room temperature (Figure 3). The EPR analysis were performed over time, until the EPR signal was turned off, and the EPR spectra parameters were: microwaves power 10 mW, magnetic field modulation amplitude 0.1 mT, field sweep 10 mT and receiver gain 60 dB (Figure 67).

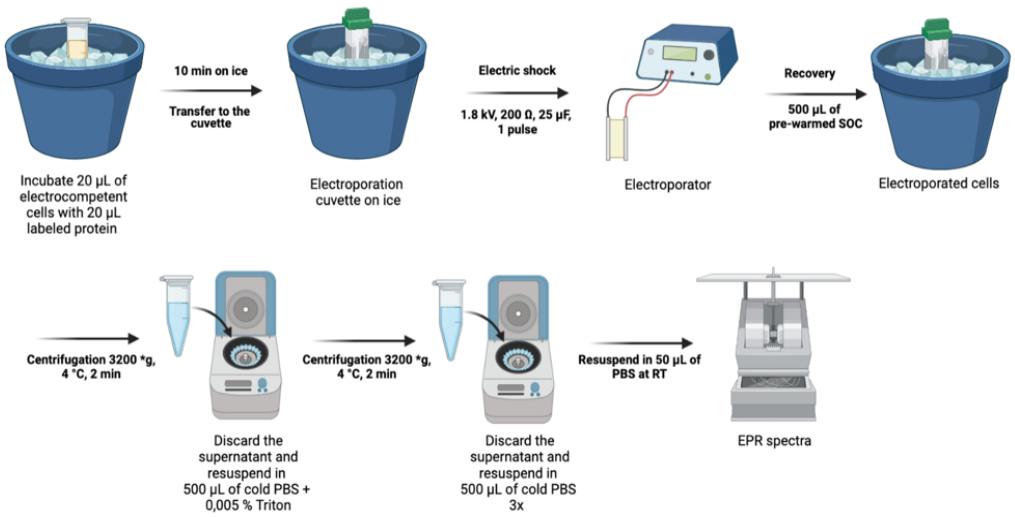


***Figure 67**: Graphic protocol of the in cell EPR experiments. Created with BioRender.com.*

Different negative control experiments were done following the same protocol in the same conditions or modifying some parameters such as number of the washing steps, contact time, protein concentration and pH of the protein solution. These samples, defined *in contact*, were analyzed by EPR to monitor if the protein was absorbed by the membrane.

7.9.5 DEER

Double labeled *h*NDRG1*C variants were used to obtain inter-label distance distributions using the four-pulse DEER sequence. Experiments were performed on a Bruker ELEXSYS E580 spectrometer equipped with an Oxford helium temperature regulation unit at Q-band (34 GHz) using the standard EN 5107D2 resonator. All measurements were carried out at 60 K on 20 µL of samples with a final protein concentration between 80 and 100 µM and in 20 mM Tris pH 7.5 containing 150 mM NaCl and 30% of glycerol-D_8. 30% TFE was added in some cases. The samples were loaded into quarts capillaries and flash-frozen in liquid nitrogen. The DEER traces were, then, analyzed using DeerAnalysis2019 software and the distance distributions were extracted through a Tikhonov regularization after baseline correction (http://www.epr.ethz.ch/software/index Jeschke G. 2011. DeerAnalysis. ETH, Zürich, Switzerland).

7.10 Confocal microscopy

7.10.1 Protein labeling with Alexa Fluor™ 488 C_5 Maleimide

*h*NDRG1*C_S336C_C394 (100 nmol) in 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP, was buffer exchanged using PD 10 desalting column (GE Healthcare) equilibrated with 10 mM

Tris pH 7.4. The fractions containing the protein were pooled and incubated with Alexa Fluor™ 488 C_5 Maleimide (stock 5 mM in DMSO) (Invitrogen) in 5-fold excess for 4h at 4 °C under gentle stirring. After reaction completion, the solution is loaded again into a PD10 desalting column equilibrated with the same buffer previously used, to remove the unreacted fluorophore. The bright yellow fractions with the labeled protein were combined and dialyzed in 10 mM Tris pH 7.4 overnight using the Slide-A-Lyzer Dialysis Cassette G2 2000 MWCO of 3 mL capacity, for eliminating the excess of the unbound fluorophore. The protein solution was concentrated and diluted 4 times using Amicon Ultra-2 mL Centrifugal Filters MWCO 3 kDa. The degree of labeling was calculated using the following formula:

$$\frac{moles\ of\ dye}{moles\ of\ protein} = \frac{A_x}{\varepsilon} x \frac{\text{MW of protein}}{mg\ protein/mL}$$

Where A*x* is the absorbance value of the dye at the absorption maximum wavelength (493 nm), ε is the molar extinction coefficient of the dye at the absorption maximum wavelength (72000 cm^{-1} M^{-1}).

7.10.2 Protein localization inside the cell

E. coli DH5α non competent cells were incubated with buffer 10 mM Tris pH 7.4 (control) or *h*NDRG1*C_S336C_C394 labeled with Alexa Fluor™ 488 C_5 Maleimide (final protein concentration 500 μM) for 30 minutes at room temperature. The cells were treated as in the *in cell* EPR control experiment and observed at the confocal microscopy over time.

7.11 Cellular culture and protein extracts

Human A549 and HeLa cell lines were cultivated in Dulbecco's Modified Eagle Medium (DMEM) GlutaMAX (Gibco, ThermoFisher Scientific, Monza, Italy) with 10% of heat-inactivated fetal bovine serum (FBS, Gibco), preserved in a heat-inactivated fetal bovine serum (FBS, Gibco) and sub-cultured twice a week. When necessary, $NiSO_4$ (Sigma Aldrich) treatment was performed. Cells were lysed in SDS lysis buffer (20 mM Tris-HCl, pH 7.5, 1% SDS, 1 mM Na3VO4, 1 mM PMSF, 5% β-Mercapto Ethanol and protease inhibitors from Thermo-Fisher Whatman USA), while for the preparation of the nuclear extracts the cells were previously trypsinized, collected and then resuspended in hypotonic buffer (10 mM Tris-HCl pH 7.8, 5 mM $MgCl_2$), adding 0.2% Triton X-100 (Merck Life Science S.r.l., Milan, Italy). After centrifugation, nuclei were lysed in 0.2% Triton X-100 (Merck Life Science S.r.l., Milan, Italy). A second

centrifugation was needed for cleaning the nuclear extracts. Equal amount of protein, estimated by Bradford colorimetric assay (Thermo-Fisher) were resolved by SDS-PAGE in non-denaturing (native sample buffer: 0.3 M Tris-HCl pH 6.8, 0.03% Bromophenol Blue and 50% Glycerol) or boiled (Laemli sample buffer: 0.3 M Tris-HCl pH 6.8, 0.03% Bromophenol Blue, 9% Sodium Dodecyl Sulfate (SDS), 9% β-Mercapto Ethanol, 50% Glycerol) to obtain denaturing conditions. Samples were transferred to nitrocellulose membrane overnight at 4°C and analyzed by immunoblot with these antibodies: anti-NDRG1 (aminoacid 2-29) (1:500), anti-Actin (1:1000), anti-Lamin A/C (1:200), all from Santa Cruz (Santa Cruz Biotechnology, DBA Italia SRL, Segrate, Italy); anti-α-Catenin (1:3000), β-tubulin (1:2000) and anti-GAPDH (1:8000) from Merck (Merck Life Science S.r.l., Milan, Italy); anti-p21 (1:1000) (Thermo-Fisher Scientific). Final bands recognition was performed with the Amersham ECL detection system and analyzed with ImageJ (National Institute of Health, Bethesda, MD, USA

Materials and methods concerning part 2 are widely discussed in Annex 1 and Annex 2

Acknowledgements

This work benefited from access to the CERMM-CIRMMP infrastructure, the Italian Instruct-ERIC centre in Florence, for the acquisition of the NMR data. The SDSL-EPR results were obtained at the EPR-facility Bioénergétique et Ingénierie des Protéines in Marseille (EPR-MRS), located in the Institute of Microbiology of the Mediterranean through MOSBRI TNA support (project number MOSBRI-2021-11). This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004806. *A*Midex Foundation* of Aix-Marseille University, funded by socio-economic partners, and the *Agence National de la Recherche* gave their financial support (contract ANR-18-CE11-0007-01 Into-the-Cell). All BIP 07 team is thanked for their support and hospitality during may stay as visiting PhD student.

A special thanks to all members of the laboratory of Bioinorganic Chemistry of the Department of Pharmacy and Biotechnology of the University of Bologna. Particular gratitude to Prof. Barbara Zambelli, my research supervisor, for her suggestions and assistance during these years.

References

A.R, Human Carcinogens, Iarc Monographs, O. N. T. H. E. Evaluation, O. F. Carcinogenic Risks, and T. O. Humans. 2012. "Arsenic, Metals, Fibres and Dusts." *Iarc Monographs* 100(Arsenic, metals, fibres, and dusts).

Azatian, Stephan B., Navneet Kaur, and Michael P. Latham. 2019. "Increasing the Buffering Capacity of Minimal Media Leads to Higher Protein Yield." *Journal of Biomolecular NMR* 73(1–2):11–17.

Bae, Dong Hun, Patric J. Jansson, Michael L. Huang, Zaklina Kovacevic, Danuta Kalinowski, C. Soon Lee, Sumit Sahni, and Des R. Richardson. 2013. "The Role of NDRG1 in the Pathology and Potential Treatment of Human Cancers." *Journal of Clinical Pathology* 66(11):911–17.

Barkay, Tamar, Susan M. Miller, and Anne O. Summers. 2003. "Bacterial Mercury Resistance from Atoms to Ecosystems." *FEMS Microbiology Reviews* 27(2–3):355–84.

Belle, Valérie, Sabrina Rouger, Stéphanie Costanzo, Elodie Liquière, Janez Strancar, Bruno Guigliarelli, André Fournel, and Sonia Longhi. 2008. "Mapping α-Helical Induced Folding within the Intrinsically Disordered C-Terminal Domain of the Measles Virus Nucleoprotein by Site-Directed Spin-Labeling EPR Spectroscopy." *Proteins: Structure, Function and Genetics* 73(4):973–88.

Beniamino, Ylenia, Giulia Pesce, Annamaria Zannoni, Davide Roncarati, and Barbara Zambelli. 2020. "SrnR from Streptomyces Griseus Is a Nickel-Binding Transcriptional Activator." *Journal of Biological Inorganic Chemistry* 25(2):187–98.

Berjanskii, Mark V. and David S. Wishart. 2005. "A Simple Method to Predict Protein Flexibility Using Secondary Chemical Shifts." *Journal of the American Chemical Society* 127(43):14970–71.

Bermel, Wolfgang, Ivano Bertini, Veronika Csizmok, Isabella C. Felli, Roberta Pierattelli, and Peter Tompa. 2009. "H-Start for Exclusively Heteronuclear NMR Spectroscopy: The Case of Intrinsically Disordered Proteins." *Journal of Magnetic Resonance* 198(2):275–81.

Bogomolovas, Julius, Bernd Simon, Michael Sattler, and Gunter Stier. 2009. "Screening of Fusion Partners for High Yield Expression and Purification of Bioactive Viscotoxins." *Protein Expression and Purification* 64(1):16–23.

Bonucci, Alessio, Olivier Ouari, Bruno Guigliarelli, Valérie Belle, and Elisabetta Mileo. 2020. "In-Cell EPR: Progress towards Structural Studies Inside Cells." *ChemBioChem* 21(4):451–60.

Bonucci, Alessio, Martina Palomino-Schätzlein, Paula Malo de Molina, Arantxa Arbe, Roberta

Pierattelli, Bruno Rizzuti, Juan L. Iovanna, and José L. Neira. 2021. "Crowding Effects on the Structure and Dynamics of the Intrinsically Disordered Nuclear Chromatin Protein NUPR1." *Frontiers in Molecular Biosciences* 8(July):684622.

Breton, Nolwenn Le, Marlène Martinho, Elisabetta Mileo, Emilien Etienne, Guillaume Gerbaud, Bruno Guigliarelli, and Valérie Belle. 2015. "Exploring Intrinsically Disordered Proteins Using Site-Directed Spin Labeling Electron Paramagnetic Resonance Spectroscopy." *Frontiers in Molecular Biosciences* 2(MAY):1–7.

Brito, Rui M. M. and Winchil L. C. Vaz. 1986. "Determination of the Critical Micelle Concentration of Surfactants Using the Fluorescent Probe N-Phenyl-1-Naphthylamine." *Analytical Biochemistry* 152(2):250–55.

Brown, Nigel L., Jivko V Stoyanov, Stephen P. Kidd, and Jon L. Hobman. 2003. "The MerR Family of Transcriptional Regulators." *FEMS Microbiology Reviews* 27(2–3):145–63.

Browning, Douglas F. and Stephen J. W. Busby. 2004. "The Regulation of Bacterial Transcription Initiation." *Nature Reviews Microbiology* 2(1):57–65.

Busenlehner, Laura S., Mario A. Pennella, and David P. Giedroc. 2003. "The SmtB/ArsR Family of Metalloregulatory Transcriptional Repressors: Structural Insights into Prokaryotic Metal Resistance." *FEMS Microbiology Reviews* 27(2–3):131–43.

Campbell, Duncan R., Kaye E. Chapman, Kevin J. Waldron, Stephen Tottey, Sharon Kendall, Gabriele Cavallaro, Claudia Andreini, Jason Hinds, Neil G. Stoker, Nigel J. Robinson, and Jennifer S. Cavet. 2007. "Mycobacterial Cells Have Dual Nickel-Cobalt Sensors." *Journal of Biological Chemistry* 282(44):32298–310.

Cangul, Hakan. 2004. "Hypoxia Upregulates the Expression of the NDRG1 Gene Leading to Its Overexpression in Various Human Cancers." *BMC Genetics* 5:1–11.

Challener, Cynthia A. 2014. "Coupling Light Scattering with Size-Exclusion Chromatography." *BioPharm International* 27(10):3–5.

Chang, Xiaojing, Xiaoyang Xu, Xiaoying Xue, Jinguo Ma, Zhenhua Li, Peng Deng, Jing Chen, Shuanglong Zhang, Yu Zhi, and Dongqiu Dai. 2016. "NDRG1 Controls Gastric Cancer Migration and Invasion through Regulating MMP-9." *Pathology & Oncology Research* 22(4):789–96.

Changela, Anita, Kui Chen, Yi Xue, Jackie Holschen, Caryn E. Outten, Thomas V O'Halloran, and Alfonso Mondragón. 2003. "Molecular Basis of Metal-Ion Selectivity and Zeptomolar Sensitivity by CueR." *Science* 301(5638):1383–87.

Chen, Haobin, Nitai Charan Giri, Ronghe Zhang, Kenichi Yamane, Yi Zhang, Michael Maroney, and Max Costa. 2010. "Nickel Ions Inhibit Histone Demethylase JMJD1A and DNA Repair

Enzyme ABH2 by Replacing the Ferrous Iron in the Catalytic Centers." *Journal of Biological Chemistry* 285(10):7374–83.

Chen, Haobin, Qingdong Ke, Thomas Kluz, Yan Yan, and Max Costa. 2006. "Nickel Ions Increase Histone H3 Lysine 9 Dimethylation and Induce Transgene Silencing." *Molecular and Cellular Biology* 26(10):3728–37.

Contreras-Martos, Sara, Hung H. Nguyen, Phuong N. Nguyen, Nevena Hristozova, Mauricio Macossay-Castillo, Denes Kovacs, Angela Bekesi, Jesper S. Oemig, Dominique Maes, Kris Pauwels, Peter Tompa, and Pierre Lebrun. 2018. "Quantification of Intrinsically Disordered Proteins: A Problem Not Fully Appreciated." *Frontiers in Molecular Biosciences* 5(SEP).

Costa, Max, Yan Yan, Daoji Zhao, and Konstantin Salnikow. 2003. "Molecular Mechanisms of Nickel Carcinogenesis: Gene Silencing by Nickel Delivery to the Nucleus and Gene Activation/Inactivation by Nickel-Induced Cell Signaling." *Journal of Environmental Monitoring* 5(2):222–23.

Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. "NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes." *Journal of Biomolecular NMR* 6(3):277–93.

Dhar, Apratim, Antonios Samiotakis, Simon Ebbinghaus, Lea Nienhaus, Dirar Homouz, Martin Gruebele, and Margaret S. Cheung. 2010. "Structure, Function, and Folding of Phosphoglycerate Kinase Are Strongly Perturbed by Macromolecular Crowding." *Proceedings of the National Academy of Sciences of the United States of America* 107(41):17586–91.

Drescher, Malte. 2012. "EPR in Protein Science." Pp. 91–119 in *EPR Spectroscopy: Applications in Chemistry and Biology*, edited by M. Drescher and G. Jeschke. Berlin, Heidelberg: Springer Berlin Heidelberg.

Du, Aolin, Yufeng Jiang, and Chuifeng Fan. 2018. "NDRG1 Downregulates ATF3 and Inhibits Cisplatininduced Cytotoxicity in Lung Cancer A549 Cells." *International Journal of Medical Sciences* 15(13):1502–7.

Dyson, H. Jane and Peter E. Wright. 2005. "Intrinsically Unstructured Proteins and Their Functions." *Nature Reviews Molecular Cell Biology* 6(3):197–208.

Etienne, E., N. Le Breton, M. Martinho, E. Mileo, and V. Belle. 2017. "SimLabel: A Graphical User Interface to Simulate Continuous Wave EPR Spectra from Site-Directed Spin Labeling Experiments." *Magnetic Resonance in Chemistry* 55(8):714–19.

Fang, Bernard A., Žaklina Kovačević, Kyung Chan Park, Danuta S. Kalinowski, Patric J. Jansson, Darius J. R. Lane, Sumit Sahni, and Des R. Richardson. 2014. "Molecular Functions of the

Iron-Regulated Metastasis Suppressor, NDRG1, and Its Potential as a Molecular Target for Cancer Therapy." *Biochimica et Biophysica Acta - Reviews on Cancer* 1845(1):1–19.

Felli, Isabella C. and Roberta Pierattelli. 2014. "Novel Methods Based on 13C Detection to Study Intrinsically Disordered Proteins." *Journal of Magnetic Resonance* 241(1):115–25.

Finney, Lydia A. and Thomas V. O'Halloran. 2003. "Transition Metal Speciation in the Cell: Insights from the Chemistry of Metal Ion Receptors." *Science* 300(5621):931–36.

Fridovich, Irwin. 1983. "Superoxide Radical And." *Ann Rev Pharmacol Toxicol* (23):239–57.

Gajiwala, Ketan S. and Stephen K. Burley. 2000. "Winged Helix Proteins." *Current Opinion in Structural Biology* 10(1):110–16.

Girolomoni, Giampiero, Paolo Gisondi, Chiara Ottaviani, and Andrea Cavani. 2004. "Immunoregulation of Allergic Contact Dermatitis." *Journal of Dermatology* 31(4):264–70.

Gnutt, David, Mimi Gao, Oliver Brylski, Matthias Heyden, and Simon Ebbinghaus. 2015. "Excluded-Volume Effects in Living Cells." *Angewandte Chemie - International Edition* 54(8):2548–51.

He, Lang, Kang Liu, Xiaoshan Wang, Hong Chen, Jin Zhou, Xun Wu, Tao Liu, Yongxue Yang, Xuemei Yang, Dandan Cui, Guiqin Song, un Wang, and Jianguo Lei. 2018. "NDRG1 Disruption Alleviates Cisplatin/Sodium Glycididazole-Induced DNA Damage Response and Apoptosis in ERCC1-Defective Lung Cancer Cells." *International Journal of Biochemistry and Cell Biology* 100(33):54–60.

Homouz, Dirar, Michael Perham, Antonios Samiotakis, Margaret S. Cheung, and Pernilla Wittung-Stafshede. 2008. "Crowded, Cell-like Environment Induces Shape Changes in Aspherical Protein." *Proceedings of the National Academy of Sciences of the United States of America* 105(33):11754–59.

Huang, Fei, Christopher J. Oldfield, Bin Xue, Wei Lun Hsu, Jingwei Meng, Xiaowen Liu, Li Shen, Pedro Romero, Vladimir N. Uversky, and A. Keith Dunker. 2014. "Improving Protein Order-Disorder Classification Using Charge-Hydropathy Plots." *BMC Bioinformatics* 15(17):1–13.

Huang, L. Eric and H. Franklin Bunn. 2003. "Hypoxia-Inducible Factor and Its Biomedical Relevance." *Journal of Biological Chemistry* 278(22):19575–78.

Hwang, Jungwon, Yoonjeong Kim, Ho Bum Kang, Lukasz Jaroszewski, Ashley M. Deacon, Hwiseop Lee, Won Chan Choi, Kyung Jin Kim, Cheol Hee Kim, Beom Sik Kang, Jie Oh Lee, Tae Kwang Oh, Jae Wha Kim, Ian A. Wilson, and Myung Hee Kima. 2011. "Crystal Structure of the Human N-Myc Downstream-Regulated Gene 2 Protein Provides Insight into Its Role as a Tumor Suppressor." *Journal of Biological Chemistry* 286(14):12450–60.

Kalaydjieva, Luba, David Gresham, Rebecca Gooding, Lisa Heather, Frank Baas, Rosalein De

Jonge, Karin Blechschmidt, Dora Angelicheva, David Chandler, Penelope Worsley, Andre Rosenthal, Rosalind H. M. King, and P. K. Thomas. 2000. "N-Myc Downstream-Regulated Gene 1 Is Mutated in Hereditary Motor and Sensory Neuropathy-Lom." *American Journal of Human Genetics* 67(1):47–58.

Karaczyn, Aldona A., Filip Golebiowski, and Kazimierz S. Kasprzak. 2006. "Ni(II) Affects Ubiquitination of Core Histones H2B and H2A." *Experimental Cell Research* 312(17):3252–59.

Kim, Eun-Ja, Hye-Jung Chung, Bumsu Suh, Yung Chil Hah, and Jung-Hye Roe. 1998. "Expression and Regulation of the SodF Gene Encoding Iron- and Zinc-Containing Superoxide Dismutase in Streptomyces Coelicolor Müller." *Journal of Bacteriology* 180(8):2014–20.

Kim, Eun Ja, Hye Jung Chung, Bumsu Suh, Yung Chil Hah, and Jung Hye Roe. 1998. "Transcriptional and Post-Transcriptional Regulation by Nickel of SodN Gene Encoding Nickel-Containing Superoxide Dismutase from Streptomyces Coelicolor Muller." *Molecular Microbiology* 27(1):187–95.

Kim, Ju Sim, Sa Ouk Kang, and Jeong K. Lee. 2003. "The Protein Complex Composed of Nickel-Binding SrnQ and DNA Binding Motif-Bearing SrnR of Streptomyces Griseus Represses SodF Transcription in the Presence of Nickel." *Journal of Biological Chemistry* 278(20):18455–63.

Kim, Kyung Rok, Kyung A. Kim, Joon Sung Park, Jun Young Jang, Yuri Choi, Hyung Ho Lee, Dong Chul Lee, Kyung Chan Park, Young Il Yeom, Hyun Jung Kim, and Byung Woo Han. 2020. "Structural and Biophysical Analyses of Human N-MYC Downstream-Regulated Gene 3 (NDRG3) Protein." *Biomolecules* 10(1):1–17.

Kjaergaard, Magnus, Søren Brander, and Flemming M. Poulsen. 2011. "Random Coil Chemical Shift for Intrinsically Disordered Proteins: Effects of Temperature and PH." *Journal of Biomolecular NMR* 49(2):139–49.

Klein, Catherine B., Kathleen Conway, Xin Wei Wang, Rupinder K. Bhamra, Xinhua Lin, Mitchell D. Cohen, Lois Annab, J. Carl Barrett, and Max Costa. 1991. "Senescence of Nickel-Transformed Cells by an X Chromosome: Possible Epigenetic Control." *Science* 251(4995):796–99.

Kosol, Simone, Sara Contreras-Martos, Cesyen Cedeño, and Peter Tompa. 2013. "Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy." *Molecules* 18(9):10802–28.

Kovacevic, Zaklina, Sharleen V. Menezes, Sumit Sahni, Danuta S. Kalinowski, Dong Hun Bae,

Darius J. R. Lane, and Des R. Richardson. 2016. "The Metastasis Suppressor, N-MYC Downstream-Regulated Gene-1 (NDRG1), down-Regulates the ErbB Family of Receptors to Inhibit Downstream Oncogenic Signaling Pathways." *Journal of Biological Chemistry* 291(3):1029–52.

Kumar, Raj. 2009. "Role of Naturally Occurring Osmolytes in Protein Folding and Stability." *Archives of Biochemistry and Biophysics* 491(1–2):1–6.

Kumar Satish and Trivedi A.V. 2016. "A Review on Role of Nickel in the Biological System." *International Journal of Current Microbiology and Applied Sciences* 5(3):719–27.

Kuroda, Makoto, Hideo Hayashi, and Toshiko Ohta. 1999. "Chromosome-Determined Zinc-Responsible Operon Czr in Staphylococcus Aureus Strain 912." *Microbiology and Immunology* 43(2):115–25.

Lachat, Pascale, Philip Shaw, Sandra Gebhard, Nico Van Belzen, Pascal Chaubert, and Fred T. Bosman. 2002. "Expression of NDRG1, a Differentiation-Related Gene, in Human Tissues." *Histochemistry and Cell Biology* 118(5):399–408.

Lande-Diner, L., J. Zhang, T. Hashimshony, A. Goren, I. Keshet, and H. Cedar. 2004. "Gene Repression Paradigms in Animal Cells." *Cold Spring Harbor Symposia on Quantitative Biology* 69:131–38.

Lee, Woonghee, William M. Westler, Arash Bahrami, Hamid R. Eghbalnia, and John I. Markley. 2009. "PINE-SPARKY: Graphical Interface for Evaluating Automated Probabilistic Peak Assignments in Protein NMR Spectroscopy." *Bioinformatics* 25(16):2085–87.

Lee, Y. W., C. B. Klein, B. Kargacin, K. Salnikow, J. Kitahara, K. Dowjat, A. Zhitkovich, N. T. Christie, and M. Costa. 1995. "Carcinogenic Nickel Silences Gene Expression by Chromatin Condensation and DNA Methylation: A New Model for Epigenetic Carcinogens." *Molecular and Cellular Biology* 15(5):2547–57.

Liu, Wen, Fei Xing, Megumi Iiizumi-Gairani, Hiroshi Okuda, Misako Watabe, Sudha K. Pai, Puspa R. Pandey, Shigeru Hirota, Aya Kobayashi, Yin Yuan Mo, Koji Fukuda, Yi Li, and Kounosuke Watabe. 2012. "N-Myc Downstream Regulated Gene 1 Modulates Wnt-β-Catenin Signalling and Pleiotropically Suppresses Metastasis." *EMBO Molecular Medicine* 4(2):93–108.

Lobley, A., L. Whitmore, and B. A. Wallace. 2002. "DICHROWEB: An Interactive Website for the Analysis of Protein Secondary Structure from Circular Dichroism Spectra." *Bioinformatics* 18(1):211–12.

Marley, Jonathan, Min Lu, and Clay Bracken. 2001. "A Method for Efficient Isotopic Labeling of Recombinant Proteins." *Journal of Biomolecular NMR* 20:71–75.

Maroney, Michael J. and Stefano Ciurli. 2021. “Nickel as a Virulence Factor in the Class I Bacterial Carcinogen, Helicobacter Pylori.” *Seminars in Cancer Biology* 76(March):143–55.

Maxwell, Patrick and Konstantin Salnikow. 2004. “HIF-1: An Oxygen and Metal Responsive Transcription Factor.” *Cancer Biology and Therapy* 3(1):29–35.

Mazzei, Luca, Francesco Musiani, Szymon Żerko, Wiktor Koźminski, Michele Cianci, Ylenia Beniamino, Stefano Ciurli, and Barbara Zambelli. 2021. “Structure, Dynamics, and Function of SrnR, a Transcription Factor for Nickel-Dependent Gene Expression.” *Metallomics* 13(12):mfab069.

McCaig, Catherine, Louisa Potter, Olga Abramczyk, and James T. Murray. 2011. “Phosphorylation of NDRG1 Is Temporally and Spatially Controlled during the Cell Cycle.” *Biochemical and Biophysical Research Communications* 411(2):227–34.

Micsonai, András, Frank Wien, Éva Bulyáki, Judit Kun, Éva Moussong, Young Ho Lee, Yuji Goto, Matthieu Réfrégiers, and József Kardos. 2018. “BeStSel: A Web Server for Accurate Protein Secondary Structure Prediction and Fold Recognition from the Circular Dichroism Spectra.” *Nucleic Acids Research* 46(W1):W315–22.

Micsonai, András, Frank Wien, Linda Kernya, Young Ho Lee, Yuji Goto, Matthieu Réfrégiers, and József Kardos. 2015. “Accurate Secondary Structure Prediction and Fold Recognition for Circular Dichroism Spectroscopy.” *Proceedings of the National Academy of Sciences of the United States of America* 112(24):E3095–3103.

Miraula, Manfredi, Stefano Ciurli, and Barbara Zambelli. 2015. “Intrinsic Disorder and Metal Binding in UreG Proteins from Archae Hyperthermophiles: GTPase Enzymes Involved in the Activation of Ni(II) Dependent Urease.” *Journal of Biological Inorganic Chemistry* 20(4):739–55.

Molday, R. S., S. W. Englander, and R. G. Kallen. 1972. “Primary Structure Effects on Peptide Group Hydrogen Exchange.” *Biochemistry* 11(2):150–58.

Morby, Andrew P., Jennifer S. Turner, James W. Huckle, and Nigel J. Robinson. 1993. “SmtB Is a Metal-Dependent Repressor of the Cyanobacterial Metallothionein Gene SmtA : Identification of a Zn Inhibited DNA-Protein Complex.” *Nucleic Acids Research* 21(4):921–25.

Murray, James T., David G. Campbell, Nicholas Morrice, Gillian C. Auld, Natalia Shpiro, Rodolpho Marquez, Mark Peggie, Jenny Bain, Graham B. Bloomberg, Florian Grahammer, Florian Lang, Peer Wulff, Dietmar Kuhl, and Philip Cohen. 2004. “Exploitation of KESTREL to Identify NDRG Family Members as Physiological Substrates for SGK1 and GSK3.” *Biochemical Journal* 384(3):477–88.

Mustonen, Venla, Gopinath Muruganandam, Remy Loris, Petri Kursula, and Salla Ruskamo. 2020. "Crystal and Solution Structure of NDRG1, a Membrane-binding Protein Linked to Myelination and Tumour Suppression." *The FEBS Journal*.

Nagai, Maria Aparecida, Renê Gerhard, José Humberto T. G. Fregnani, Suely Nonogaki, Regina Barbosa Rierger, Mário Mourão Netto, and Fernando A. Soares. 2011. "Prognostic Value of NDRG1 and SPARC Protein Expression in Breast Cancer Patients." *Breast Cancer Research and Treatment* 126(1):1–14.

Novagen. 2011. "Competent Cells." *User Protocol TB009 Rev. H* 5.

Oates, Matt E., Pedro Romero, Takashi Ishida, Mohamed Ghalwash, Marcin J. Mizianty, Bin Xue, Zsuzsanna Dosztányi, Vladimir N. Uversky, Zoran Obradovic, Lukasz Kurgan, A. Keith Dunker, and Julian Gough. 2013. "D2P2: Database of Disordered Protein Predictions." *Nucleic Acids Research* 41(D1):508–16.

Oh, Yu Mi, Hyung Bae Park, Jae Hun Shin, Ji Eun Lee, Ha Young Park, Dhong Hyo Kho, Jun Sung Lee, Heonsik Choi, Tomohiko Okuda, Koichi Kokame, Toshiyuki Miyata, In Hoo Kim, Seung Hoon Lee, Ronald H. Schwartz, and Kyungho Choi. 2015. "Ndrg1 Is a T-Cell Clonal Anergy Factor Negatively Regulated by CD28 Costimulation and Interleukin-2." *Nature Communications* 6(May).

Park, Kyung Chan, Sharleen V. Menezes, Danuta S. Kalinowski, Sumit Sahni, Patric J. Jansson, Zaklina Kovacevic, and Des R. Richardson. 2018. "Identification of Differential Phosphorylation and Sub-Cellular Localization of the Metastasis Suppressor, NDRG1." *Biochimica et Biophysica Acta - Molecular Basis of Disease* 1864(8):2644–63.

Park, Kyung Chan, Jasmina Paluncic, Zaklina Kovacevic, and Des R. Richardson. 2020. "Pharmacological Targeting and the Diverse Functions of the Metastasis Suppressor, NDRG1, in Cancer." *Free Radical Biology and Medicine* 157(May 2019):154–75.

Peana, Massimiliano, Karolina Zdyb, Serenella Medici, Alessio Pelucelli, Giancarlo Simula, Elzbieta Gumienna-Kontecka, and Maria Antonietta Zoroddu. 2017. "Ni(II) Interaction with a Peptide Model of the Human TLR4 Ectodomain." *Journal of Trace Elements in Medicine and Biology* 44(April):151–60.

Pei, Jimin, Bong Hyun Kim, and Nick V. Grishin. 2008. "PROMALS3D: A Tool for Multiple Protein Sequence and Structure Alignments." *Nucleic Acids Research* 36(7):2295–2300.

Pennella, Mario A. and David P. Giedroc. 2005. "Structural Determinants of Metal Selectivity in Prokaryotic Metal-Responsive Transcriptional Regulators." *BioMetals* 18(4):413–28.

Pettersen, Eric F., Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. 2004. "UCSF Chimera - A Visualization

System for Exploratory Research and Analysis." *Journal of Computational Chemistry* 25(13):1605–12.

Ragsdale, Stephen W. 2009. "Nickel-Based Enzyme Systems." *Journal of Biological Chemistry* 284(28):18571–75.

Rothenberg, Marc E. 2010. "Innate Sensing of Nickel." *Nature Immunology* 11(9):781–82.

Sahni, Sumit, Sukriti Krishan, and Des R. Richardson. 2014. "NDRG1 as a Molecular Target to Inhibit the Epithelial–Mesenchymal Transition: The Case for Developing Inhibitors of Metastasis." *Future Medicinal Chemistry* 6(11):1241–44.

Salnikow, Konstantin, Mikhail V. Blagosklonny, Heather Ryan, Randall Johnson, and Max Costa. 2000. "Carcinogenic Nickel Induces Genes Involved with Hypoxic Stress." *Cancer Research* 60(1):38–41.

Salnikow, Konstantin, Steven P. Donald, Richard K. Bruick, Anatoly Zhitkovich, James M. Phang, and Kazimierz S. Kasprzak. 2004. "Depletion of Intracellular Ascorbate by the Carcinogenic Metals Nickel and Cobalt Results in the Induction of Hypoxic Stress." *Journal of Biological Chemistry* 279(39):40337–44.

Salnikow, Konstantin and Kazimierz S. Kasprzak. 2007. "Nickel-Dependent Gene Expression." Pp. 581–618 in *Nickel and Its Surprising Impact in Nature*. John Wiley & Sons, Ltd.

Salnikow, Konstantin and Anatoly Zhitkovich. 2008. *Genetic and Epigenetic Mechanisms in Metal Carcinogenesis and Cocarcinogenesis: Nickel, Arsenic, and Chromium*. Vol. 21.

Schreiter, Eric R. and Catherine L. Drennan. 2007. "Ribbon–Helix–Helix Transcription Factors: Variations on a Theme." *Nature Reviews Microbiology* 5(9):710–20.

Shen, Yang and Ad Bax. 2013. "Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks." *Journal of Biomolecular NMR* 56(3):227–41.

Shi, W., J. Wu, and B. P. Rosen. 1994. "Identification of a Putative Metal Binding Site in a New Family of Metalloregulatory Proteins." *The Journal of Biological Chemistry* 269(31):19826–29.

Silver, Simon and Le T. Phung. 1996. "BACTERIAL HEAVY METAL RESISTANCE: New Surprises." *Annual Review of Microbiology* 50(1):753–89.

Singh, Aditi, Igor Minia, Dorothea Droll, Abeer Fadda, Christine Clayton, and Esteban Erben. 2014. "Trypanosome MKT1 and the RNA-Binding Protein ZC3H11: Interactions and Potential Roles in Post-Transcriptional Regulatory Networks." *Nucleic Acids Research* 42(7):4652–68.

Song, Yan and Lili Cao. 2013. "N-Myc Downstream-Regulated Gene 1: Diverse and Complicated

Functions in Human Hepatocellular Carcinoma (Review)." *Oncology Letters* 6(6):1539–42.

Stark, Chris, Bobby Joe Breitkreutz, Teresa Reguly, Lorrie Boucher, Ashton Breitkreutz, and Mike Tyers. 2006. "BioGRID: A General Repository for Interaction Datasets." *Nucleic Acids Research* 34(Database issue):535–39.

Stein, Susanne, Emily K. Thomas, Birger Herzog, Matthew D. Westfall, Jonathan V. Rocheleau, Roger S. Jackson, Mai Wang, and Peng Liang. 2004. "NDRG1 Is Necessary for P53-Dependent Apoptosis." *Journal of Biological Chemistry* 279(47):48930–40.

Stepanenko, Olesya V., Olga V. Stepanenko, Irina M. Kuznetsova, Vladimir N. Uverskyo, and Konstantin K. Turoverov. 2016. "Peculiarities of the Super-Folder GFP Folding in a Crowded Milieu." *International Journal of Molecular Sciences* 17(11).

Stola, Massimiliano, Francesco Musiani, Stefano Mangani, Paola Turano, Niyaz Safarov, Barbara Zambelli, and Stefano Ciurli. 2006. "The Nickel Site of Bacillus Pasteurii UreE, a Urease Metallo-Chaperone, as Revealed by Metal-Binding Studies and X-Ray Absorption Spectroscopy." *Biochemistry* 45(20):6495–6509.

Stoll, Stefan and Arthur Schweiger. 2006. "EasySpin, a Comprehensive Software Package for Spectral Simulation and Analysis in EPR." *Journal of Magnetic Resonance* 178(1):42–55.

Storz, Gisela, Louis A. Tartaglia, Spencer B. Farr, and Bruce N. Ames. 1990. "Bacterial Defenses against Oxidative Stress." *Trends in Genetics* 6(C):363–68.

Sugiki, Toshihiko, Makoto Murakami, Yoshitaka Taketomi, Rei Kikuchi-Yanoshita, and Ichiro Kudo. 2004. "N-Myc Downregulated Gene 1 Is a Phosphorylated Protein in Mast Cells." *Biological and Pharmaceutical Bulletin* 27(5):624–27.

Sun, Guohui, Lijiao Zhao, Rugang Zhong, and Yongzhen Peng. 2018. "The Specific Role of O 6 -Methylguanine-DNA Methyltransferase Inhibitors in Cancer Chemotherapy." *Future Medicinal Chemistry* 10(16):1971–96.

Szklarczyk, Damian, Annika L. Gable, David Lyon, Alexander Junge, Stefan Wyder, Jaime Huerta-Cepas, Milan Simonovic, Nadezhda T. Doncheva, John H. Morris, Peer Bork, Lars J. Jensen, and Christian Von Mering. 2019. "STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets." *Nucleic Acids Research* 47(D1):D607–13.

Tarazona, M. Pilar and Enrique Saiz. 2003. "Combination of SEC/MALS Experimental Procedures and Theoretical Analysis for Studying the Solution Properties of Macromolecules." *Journal of Biochemical and Biophysical Methods* 56(1–3):95–116.

Theillet, Francois Xavier, Andres Binolfi, Tamara Frembgen-Kesner, Karan Hingorani, Mohona Sarkar, Ciara Kyne, Conggang Li, Peter B. Crowley, Lila Gierasch, Gary J. Pielak, Adrian

H. Elcock, Anne Gershenson, and Philipp Selenko. 2014. “Physicochemical Properties of Cells and Their Effects on Intrinsically Disordered Proteins (IDPs).” *Chemical Reviews* 114(13):6661–6714.

Thelwell, C., N. J. Robinson, and J. S. Turner-Cavet. 1998. “An SmtB-like Repressor from Synechocystis PCC 6803 Regulates a Zinc Exporter.” *Proceedings of the National Academy of Sciences* 95(18):10728–33.

Torricella, F., A. Pierro, E. Mileo, V. Belle, and A. Bonucci. 2021. “Nitroxide Spin Labels and EPR Spectroscopy : A Powerful Association for Protein Dynamics Studies.” *BBA - Proteins and Proteomics* 1869(7):140653.

Utschig, Lisa M., James W. Bryson, and Thomas V O’Halloran. 1995. “Mercury-199 NMR of the Metal Receptor Site in MerR and Its Protein-DNA Complex.” *Science* 268(5209):380–85.

Uversky, Vladimir N. 2009. “Intrinsically Disordered Proteins and Their Environment: Effects of Strong Denaturants, Temperature, PH, Counter Ions, Membranes, Binding Partners, Osmolytes, and Macromolecular Crowding.” *Protein Journal* 28(7–8):305–25.

Uversky, Vladimir N., Joel R. Gillespie, and Anthony L. Fink. 2000. “Why Are ‘natively Unfolded’ Proteins Unstructured under Physiologic Conditions?” *Proteins: Structure, Function and Genetics* 41(3):415–27.

Wang, Bei, Jianli Li, Zhanying Ye, Zhe Li, and Xiaohua Wu. 2014. “N-Myc Downstream Regulated Gene 1 Acts as a Tumor Suppressor in Ovarian Cancer.” *Oncology Reports* 31(5):2279–85.

Wang, Liya, Hamid R. Eghbalnia, Arash Bahrami, and John L. Markley. 2005. “Linear Analysis of Carbon-13 Chemical Shift Differences and Its Application to the Detection and Correction of Errors in Referencing and Spin System Identifications.” *Journal of Biomolecular NMR* 32(1):13–22.

Wang, Qiang, Li Hong Li, Guo Dong Gao, Gang Wang, Liang Qu, Jin Ge Li, and Chun Mei Wang. 2013. “HIF-1α up-Regulates NDRG1 Expression through Binding to NDRG1 Promoter, Leading to Proliferation of Lung Cancer A549 Cells.” *Molecular Biology Reports* 40(5):3723–29.

Wang, Yang and Shaodong Dai. 2013. “Structural Basis of Metal Hypersensitivity.” *Immunologic Research* 55(1–3):83–90.

Weickert, Sabrina, Julia Cattani, and Malte Drescher. 2019. *Intrinsically Disordered Proteins (IDPs) Studied by EPR and in-Cell EPR*. Vol. 26.

Weiler, Markus, Jonas Blaes, Stefan Pusch, Felix Sahm, Marcus Czabanka, Sebastian Luger, Lukas Bunse, Gergely Solecki, Viktoria Eichwald, Manfred Jugold, Sibylle Hodecker,

Matthias Osswald, Christoph Meisner, Thomas Hielscher, P. Rubmann, P. N. Pfenning, Michael Ronellenfitsch, Tore Kempf, M. Schnolzer, Amir Abdollahi, Florian Lang, Martin Bendszus, A. von Deimling, Frank Winkler, Michael Weller, Peter Vajkoczy, Michael Platten, and Wolfgang Wick. 2014. "MTOR Target NDRG1 Confers MGMT-Dependent Resistance to Alkylating Chemotherapy." *Proceedings of the National Academy of Sciences* 111(1):409–14.

Whitmore, Lee and B. A. Wallace. 2008. "Protein Secondary Structure Analyses from Circular Dichroism Spectroscopy: Methods and Reference Databases." *Biopolymers* 89(5):392–400.

Wiedemann, Christoph, Peter Bellstedt, and Matthias Görlach. 2013. "CAPITO - A Web Server-Based Analysis and Plotting Tool for Circular Dichroism Data." *Bioinformatics* 29(14):1750–57.

Wishart, David S. and Brian D. Sykes. 1994. "The 13C Chemical-Shift Index: A Simple Method for the Identification of Protein Secondary Structure Using 13C Chemical-Shift Data." *Journal of Biomolecular NMR* 4(2):171–80.

Wright, Peter E. and H. Jane Dyson. 2015. "Intrinsically Disordered Proteins in Cellular Signalling and Regulation." *Nature Reviews Molecular Cell Biology* 16(1):18–29.

Ying, Jinfa, Frank Delaglio, Dennis A. Torchia, and Ad Bax. 2017. "Sparse Multidimensional Iterative Lineshape-Enhanced (SMILE) Reconstruction of Both Non-Uniformly Sampled and Conventional NMR Data." *Journal of Biomolecular NMR* 68(2):101–18.

Zambelli, Barbara and Stefano Ciurli. 2013. *Nickel and Human Health*. Vol. Metals Ion.

Zambelli, Barbara, Francesco Musiani, and Stefano Ciurli. 2012. "Metal Ion-Mediated DNA-Protein Interactions." Pp. 135–70 in.

Zambelli, Barbara, Vladimir N. Uversky, and Stefano Ciurli. 2016. "Nickel Impact on Human Health: An Intrinsic Disorder Perspective." *Biochimica et Biophysica Acta - Proteins and Proteomics* 1864(12):1714–31.

Zhang, Jian, Suning Chen, Wei Zhang, Jing Zhang, Xinping Liu, Hai Shi, Honglei Che, Weizhong Wang, Fuyang Li, and Libo Yao. 2008. "Human Differentiation-Related Gene NDRG1 Is a Myc Downstream-Regulated Gene That Is Repressed by Myc on the Core Promoter Region." *Gene* 417(1–2):5–12.

Zoroddu, M. A., M. Peana, T. Kowalik-Jankowska, H. Kozlowski, and M. Costa. 2004. "Nickel(II) Binding to Cap43 Protein Fragments." *Journal of Inorganic Biochemistry* 98(6):931–39.

Zoroddu, Maria Antonietta, Massimiliano Peana, Serenella Medici, and Roberto Anedda. 2009. "An NMR Study on Nickel Binding Sites in Cap43 Protein Fragments." *Dalton Transactions (Cambridge, England : 2003)* (28):5523–34.

Annex 1

Accepted manuscript version of "*SrnR from Streptomyces griseus is a nickel-binding transcriptional activator*", published by Journal of Biological Inorganic Chemistry and reproduced in accordance with the rules of the publisher, Springer Nature.

The article is available online at:
DOI: 10.1007/s00775-019-01751-5
https://link.springer.com/article/10.1007%2Fs00775-019-01751-5

SrnR from Streptomyces griseus is a nickel-binding transcriptional activator

Journal:	*Journal of Biological Inorganic Chemistry*
Manuscript ID	JBIC-19-07-00182.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Beniamino, Ylenia; University of Bologna, Pharmacy and Biotechnology Pesce, Giulia; University of Bologna, Pharmacy and Biotechnology Zannoni, Annamaria; University of Bologna, Pharmacy and Biotechnology Roncarati, Davide; University of Bologna, Pharmacy and Biotechnology Zambelli, Barbara; University of Bologna, Pharmacy and Biotechnology
Keywords:	Nickel sensor, nickel trafficking, structure-function relationship, transcriptional regulation, circular dichroism, isothermal titration calorimetry, DNase I footprinting

SrnR from *Streptomyces griseus* is a nickel-binding transcriptional activator

Ylenia Beniamino[#], Giulia Pesce[#], Annamaria Zannoni, Davide Roncarati, Barbara Zambelli*

Dept. of Pharmacy and Biotechnology, University of Bologna, via Fanin 40, 40127 Bologna, Italy

[#]YB and GP equally contributed to this work

* barbara.zambelli@unibo.it;

Tel. +390512096233

Abstract

Nickel ions are crucial components for the catalysis of biological reactions in prokaryotic organisms. As an uncontrolled nickel trafficking is toxic for living organisms, nickel-dependent bacteria have developed tightly regulated strategies to maintain the correct intracellular metal ion quota. These mechanisms require transcriptional regulator proteins that respond to nickel concentration, activating or repressing the expression of specific proteins related to Ni(II) metabolism. In *Streptomyces griseus*, a Gram-positive bacterium used for antibiotic production, *Sg*SrnR and *Sg*SrnQ regulate the nickel-dependent antagonistic expression of two superoxide-dismutase (SOD) enzymes, a Ni-SOD and a FeZn-SOD. According to a previously proposed model, *Sg*SrnR and *Sg*SrnQ form a protein complex in which *Sg*SrnR works as repressor, binding directly to the promoter of the gene coding for FeZn-SOD, while *Sg*SrnQ is the Ni(II)-dependent co-repressor.

The present work focuses on the determination of the biophysical and functional properties of *Sg*SrnR. The protein was heterologously expressed and purified from *E. coli*. The structural and metal-binding analysis, carried out by circular dichroism, light scattering, fluorescence and isothermal titration calorimetry, showed that the protein is a well-structured homodimer, able to bind nickel with moderate affinity. DNase I footprinting and β-galactosidase gene reporter assays revealed that apo-*Sg*SrnR is able to bind its DNA operator and activates a transcriptional response. The structural and functional properties of this protein are discussed relatively to its role as a Ni(II)-dependent sensor.

Keywords: Nickel sensor, nickel trafficking, transcriptional regulation, circular dichroism, isothermal titration calorimetry, DNase I footprinting

Introduction

Transition metal ions are crucial components of life, being estimated to be part of about one third of all proteins [1]. Despite their essentiality, an uncontrolled metal ion intracellular pool is toxic for living organisms, which thus require tightly regulated mechanisms to maintain the correct intracellular metal ion quota. Metal availability is also a key factor in host-pathogen interactions: infected organisms produce metal chelators to withdraw essential micronutrients from pathogenic bacteria as a defense strategy, in a process called *nutritional immunity* [2]. Bacteria, on the other side, have set up complex mechanisms to overcome the competition and satisfy their physiological requirement [3]. In particular, bacteria respond to metal ion availability in their living niche by activating or repressing the expression of specific sets of genes, typically through the action of metal-sensing transcription factors that bind to specific operators and often control metal-specific regulons. Therefore, under conditions of metal excess, the syntheses of metal efflux pumps, metal storage proteins and metallo-enzymes are initiated, while metal limitation results in the enhanced level of metal import systems, mobilization of stored metal ion pools, and alternative metal-independent pathways. For example, in *Escherichia coli* the transcriptional regulators *Ec*NikR and *Ec*RcnR govern the Ni(II)-dependent expression of the Ni(II)-importer system *Ec*NikABCDE and of the Ni(II)-export pump *Ec*RcnA, respectively [4, 5]. Both sensors respond to environmental Ni(II) in the 1 nM - 1 μM range, coherently with the Ni(II) binding affinities of the two proteins measured in solution (K_d = 20-400 nM for *Ec*NikR and K_d = 17-300 nM for *Ec*RcnR) [6, 7]. Similarly, in *Streptomyces coelicolor*, the interplay of two Ni(II)-dependent repressors, the Fur-like *Sc*Nur and the ArsR/SmtB-like *Sc*NmtR, regulates the expression of Ni(II) uptake and Ni(II)-efflux systems, respectively [8, 9]. Similarly to *Ec*NikR, *Sc*Nur responds to environmental Ni(II) concentrations from 1 nM to 100 nM, coherently with its Ni(II) binding affinities measured in solution (K_d = 10-250 nM) [7, 9]. On the other hand, *Sc*NmtR performs its regulation at a higher concentration range, from 0.1 μM to 100 μM of environmental Ni(II), suggesting that this bacterium requires or tolerates higher amounts of cytoplasmic Ni(II) [9].

With similar mechanisms, metal utilization is regulated under conditions of different metal ion availability: an interesting case is represented by superoxide dismutase (SOD), an important component of the protective mechanism against reactive oxygen species produced by aerobic metabolism. This protein is responsible for disproportionating superoxide ion into dioxygen and hydrogen peroxide [10]. The importance of SOD for cellular metabolism is reflected by the multiple types of SODs coded by the genome of many organisms. Based on the metal ions in the active site, four groups of SODs were characterized: CuZn-SOD, Mn-SOD, Fe-SOD, found in prokaryotes and eukaryotes and Ni-SOD, found only in prokaryotes [11, 12]. The expression of a certain type of SOD is switched on or off depending on the availability of the cognate metal ion. For instance, in *E. coli* Fe(II) limitation turns down the Fe-SOD expression, while it activates Mn-SOD production [13]. With a similar mechanism, *S. coelicolor* and *S. griseus* express a Ni-SOD, encoded by the *sodN* gene [14] and a FeZn-SOD, encoded by the *sodF* gene [15]. Their expression is antagonistically regulated by Ni(II): when Ni(II) is available, the expression of the *sodN* gene is increased while the transcription of the *sodF* gene is turned down [16, 17], assuring a constant intracellular SOD activity. In *S. coelicolor*, expression of the *sodF* and *sodN* genes is inversely regulated by *Sc*Nur that binds the *sodF* promoter in the holo-form inhibiting the transcription of FeZn-SOD, as well as of a small anti-*sodN* RNA, indirectly promoting the expression of Ni-SOD [18]. Differently, in *S. griseus*, SOD balance depends on the interplay of two proteins, named *Sg*SrnR and *Sg*SrnQ, coded by genes located downstream the *sodF* locus [19]. According to the currently accepted model, Ni(II)-bound *Sg*SrnQ interacts with *Sg*SrnR, enhancing the DNA binding ability of *Sg*SrnR with a peculiar protein-protein interaction [19]. Preliminary biochemical characterization indicated that *Sg*SrnQ is mostly unfolded in the absence of metal ions, but folds upon Ni(II) binding [19], thus giving the structural basis for the conformational change that modifies the protein ability to interact with *Sg*SrnR. Coherently, disorder prediction analysis based on primary structure suggested that *Sg*SrnQ is highly disordered, with an intrinsic disorder (ID) rate of ca. 80%, and two predicted disorder-based binding sites. Differently, *Sg*SrnR is noticeably more folded, with an ID rate of 25-35 % [20].

In the present work we report the biophysical and functional characterization of *Sg*SrnR. The recombinant protein was purified and its hydrodynamic, structural and functional properties were analyzed combining calorimetric, spectroscopic, light scattering and molecular biology experiments. The structural and functional properties of this protein are discussed relatively to its role as a Ni(II)-dependent sensor.

Materials and methods

Gene cloning

The *srnR* gene from *S. griseus* was commercially synthesized and sub-cloned into a *pEX-A2* vector (Eurofins), with the recognizing sequences of NdeI and BamHI restriction enzymes at 5' and 3' ends, respectively. The resulting construct was double digested with NdeI and BamHI FastDigest (Fermentas). The obtained DNA fragment was ligated (T4 DNA ligase, Promega) into the *pET15b* (5.7 kbp) expression vector (Novagen), previously digested with NcoI e BamHI endonucleases. The resulting construct was analyzed by restriction analysis and sequenced at both strands. Subsequently, the *pET15b-SgsnrR* plasmid was digested with NcoI and BamHI restriction enzymes, and the *His-SgsrnR* gene was cloned into the *pETZZ_1a* expression vector [21], introducing a N-terminal IgG-binding domain ZZ (ZZ-tag) with an additional N-terminal His-rich tag [22], to increase the solubility and the expression of the protein. The thrombin cleavage site present between the His-tag and the cloned *SgsnrR* gene in the *pET15b* vector was maintained in the new construct. The obtained plasmid was transformed into *E. coli* XL10-Gold Ultracompetent Cells (Stratagene), purified, screened by restriction analysis and sequenced on both strands.

The vector for the β-galactosidase assay was constructed as follow: first, the promoter region of the *Helicobacter pylori* cytochrome c553 gene, P_{1227} (HPG27_RS06145) was PCR amplified from genomic DNA with primers *p1227far_HindIII_F* (ATATAAGCTTCGCTTGAGTGTATTCTTCTTTTAACCC) *p1227_TSS_XhoI_R* (ATATctcgaGTCTCCTCTACTTATCAAAATGATCTTAAAATGATAC) and cloned into a *pGEM-T-Easy* vector (Promega). Second, the sequence encoding LacZ was PCR amplified with

oligonucleotides *LacZ_XhoI_F* (ATATCTCGAGACACAGGAAACAGCTATGACCATG) and *LacZ_BamHI_NotI_R* (ATATGCGGCCGCGGATCCTTATTATTTTTGACACCAGACCAACTGG), using the *MiniCTX-LacZ* vector [23] as template, it was then digested with XhoI/NotI restriction endonucleases and cloned into the *pGEM-T-Easy-*P_{1227} construct digested with the same restriction enzymes. The obtained *pGEM-T-Easy-*P_{1227}*-LacZ* construct was double digested with HindIII/BamHI and the obtained P_{1227}*-LacZ* DNA fragment was ligated into the *pKT25* (3.4 kbp) vector (Novagen) [24]. The resulting *pKT25-*P_{1227}*-LacZ* construct was examined by restriction analysis and sequenced on both strands. The 113 bp promoter sequence (P_{sod}) encompassing the *sodF* operator from *S. griseus* between the StyI and BalI recognition sequences [19] was commercially synthesized and sub-cloned into a *pEX-A218* vector (Eurofins), with the recognizing sequences of HindIII and XhoI restriction enzymes at 5' and 3' ends, respectively. The resulting *pEX-A128-*P_{sod} plasmid was double digested with HindIII and XhoI (Fermentas) and the P_{sod} fragment was cloned upstream to a β-galactosidase gene in pKT25-P_{1227}-*lacZ* digested with the same restriction enzymes. The resulting *pKT25-*P_{sod}*-LacZ* construct was examined by restriction analysis and sequenced on both strands.

Protein heterologous expression and purification

E. coli BL21 (DE3) competent cells (Stratagene) were transformed with the recombinant *pETZZ1a-SgsrnR* construct and grown in 2 L of lysogeny broth (LB) medium at 37 °C, supplemented with 30 μg/mL of kanamycin with vigorous stirring. When OD_{600} reached 0.6, the expression of the protein was induced by 0.5 mM IPTG. Protein expression was performed at 25 °C for 18 h. After that, cells were harvested by centrifugation at 10,000 x *g* for 20 min at 4°C, resuspended in 25 mL of 50 mM TrisHCl buffer pH 7.6 containing 500 mM NaCl, 20 mM imidazole, DNaseI 20 μg/mL and 10 mM $MgCl_2$ and disrupted by two passages through a French pressure cell (SLM Aminco) at 20000 pounds/square inch. The insoluble fraction of the bacterial lysate was separated from the supernatant by centrifugation at 76,000 x *g* for 25 min at 4°C. The clarified fraction was loaded onto a His-Trap HP 5 mL column (GE Healthcare), pre-equilibrated with 25 mL of 50 mM TrisHCl pH 7.6, containing

500 mM NaCl and 20 mM imidazole (binding buffer). The column was washed with the same buffer until the baseline was stable. Then, 50 U of thrombin protease (1 U/ mg of protein) were loaded onto the column and incubated at room temperature, to cleave the His-tag/ZZ-tag/His-tag fusion polypeptide from *Sg*SrnR protein. After 4 h, the column was washed with 15 mL of binding buffer which was directly loaded onto a HiPrep 16/60 Desalting column pre-equilibrated with 20 mM Tris-HCl pH 7.5, 1 mM TCEP (buffer A) and connected in tandem with the His-Trap column. *Sg*SrnR was then further purified with a HiTrap Q HP 5 mL column (GE Healthcare), equilibrated with buffer A and eluted by a linear gradient from 0 to 1 M NaCl. Protein elution occurred at 100-150 mM NaCl. Finally, protein polishing was obtained on a Superdex 75 10/300 column equilibrated with 20 mM TrisHCl pH 7.5, 150 mM NaCl, 1 mM TCEP.

The purity of the protein was assessed by electrophoresis on NuPAGE Novex Pre-Cast Gel System (Invitrogen) with NuPAGE 12% Bis-Tris gels and the fractions containing *Sg*SrnR were combined and concentrated, by using 3-kDa-cut off membrane Amicon ultra-filtration units (Millipore). The integrity of purified *Sg*SrnR was checked by mass spectrometry, which indicated a unique major peak (12,613 Da) corresponding to the theoretical molar mass with the N-terminal GSH sequence, derived from the thrombin recognition sequence.

Protein concentration was estimated by absorption spectroscopy using theoretical extinction coefficient ε_{280}= 6990 M^{-1} cm^{-1} calculated according to the amino acid sequence using the ProtParam tool (http://web.expasy.org/protparam/). Absence of Ni(II) ions in the protein solutions, as well as in the buffers used in all subsequent experiments, was verified using inductively coupled plasma emission spectroscopy, with a protocol previously reported [25].

Thermal shift assay

Purified *Sg*SrnR (16 µM) was incubated with 5xSYPRO Orange dye (ThermoFisher) in a 96-well PCR plate. One µL of solutions from the kit Slice pH HR2-070 (Hampton Research) was added to each well and the final volume was adjusted to 20 µL with distilled water. Protein stability under

different buffer conditions was determined by thermal unfolding reactions heating the protein solutions from 25 °C to 95 °C, at a speed of 1 °C per minute. The melting temperature (T_m) was determined by plotting the first derivative of the fluorescence emission at 569 nm (excitation at 470 nm) as a function of temperature. The conditions that corresponded to the highest T_m values were used to screen the effect of ionic strength, repeating the experiment from 100 mM to 1 M NaCl. All subsequent experiments were performed in the buffer identified as the one producing an unfolding reaction with the highest T_m (20 mM TrisHCl, pH 7.5, containing 150 mM NaCl and 1 mM TCEP). The effect of increasing concentration of Ni(II) and Zn(II) on the T_m value were evaluated using the same protocol.

Isothermal titration calorimetry

Ni(II) and Zn(II) binding to *Sg*SrnR were investigated at 25 °C using a high-sensitivity VP-ITC microcalorimeter (MicroCal). The protein sample (30 - 40 μM) was loaded into the sample cell (1.4093 mL) and was titrated with 55 x 5 μL injections of a solution containing 600 μM $NiSO_4$ or $ZnSO_4$, dissolved in the same buffer, using a computer-controlled 310-μL microsyringe. Heat of dilution of the metal ions into the buffer was verified to be negligible by control experiments. Integrated heat data were fitted using a non-linear least-squares minimization algorithm to theoretical curves corresponding to different binding models and processed with the Affinimeter software (www.affinimeter.com). ΔH (reaction enthalpy change in cal mol^{-1}), K_a (binding affinity constant in M^{-1}), n (number of binding sites) and Q_{dil} (dilution heat) were the fitting parameters. The chi-square parameter $\chi 2$ was used to establish the best fit. The reaction entropy was calculated using the equations: $\Delta G = -RT \ln K_a$ (R =1.9872 cal mol^{-1} K^{-1}, T = 298 K) and $\Delta G = \Delta H - T\Delta S$. The values given for ΔH and ΔS are apparent, and include contributions not only from metal binding but also from associated events such as protonation/deprotonation of the amino acid residues involved in the binding and consequent change in the buffer ionization state.

Circular dichroism spectroscopy

The secondary structure of *Sg*SrnR (90 μM) was estimated by circular dichroism (CD) in 20 mM TrisHCl, pH 7.5, containing 150 mM NaCl and 1 mM TCEP. Experiments were conducted in the absence and in the presence of increasing concentrations of Ni(II) (1, 2 and 5 equivalents) or Zn (II) (1 and 2 equivalents). The CD spectra were recorded at 25 °C using a JASCO-810 spectropolarimeter flushed with N_2 and cuvette of 0.01 cm of path length. Ten spectra were registered from 260 to 190 nm, at 0.2 nm intervals. The secondary structure composition was quantitatively evaluated, using CDSSTR program and reference sets 3, 4, 7, available at the Dichroweb server (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) [26]. Unfolding of secondary structure was followed by heating a protein solution (45 μM) from 10 °C to 90 °C and following the decrease in ellipticity at 222 nm. Data analysis was performed as previously reported [27].

The protein conformation involving the side chains of aromatic residues (3 Phe, 1 Tyr, 1 Trp) was evaluated for the apo-protein and in the presence of increasing concentration of Ni(II) or Zn(II) using near UV CD spectroscopy. CD data collection was performed between 250 and 350 nm on a protein sample of 560 μM in a 1 mm path length cuvette.

Fluorescence spectroscopy

Steady-state fluorescence intensities and emission spectra were recorded using a PTI QuantaMaster C60/2000 (Photon Technology International) spectrofluorometer, operating in ‘‘photon counting’’ mode. Excitation and emission band-passes were set to 2.5 nm each. An excitation wavelength of 295 nm was used.

To monitor the effect of Ni(II) and Zn(II) binding on SgSrnR (20 μM), the experiments were performed adding subsequent aliquots of metal ions from a concentrated stock solution (2 mM). The samples were equilibrated at each temperature before data collection. Fluorescence spectra were collected over time until no further change was observed.

Light scattering measurements

The oligomerization state and the hydrodynamic radius of *Sg*SrnR were determined using a combination of size-exclusion chromatography (SEC), multiple-angle light scattering (MALS) and quasi-elastic light scattering (QELS). *Sg*SrnR (100 µL, 560 µM) in 20 mM TrisHCl, pH 7.5, containing 150 mM NaCl and 1 mM TCEP, in the absence and in presence of one equivalent of $NiSO_4$ or $ZnSO_4$, was loaded onto a Superdex 75 10/300 GL column (GE Healthcare) pre-equilibrated with the same buffer. Elution was carried out at room temperature with a flow rate of 0.5 mL /min

The column was connected downstream to a multiple-angle laser light (690 nm) scattering DAWN EOS photometer (Wyatt Technology) and to a quasi-elastic light scattering apparatus (Wyatt QELS). The specific refractive index increment (*dn/dc*) for the protein was taken as 0.185 mL/g [28]. The value of 1.331 for the solvent refractive index and the concentration of the eluted protein were determined using the refractive index detector. The weight average molecular masses, M_w, were determined across the entire elution profile in the intervals of 0.2 s from MALS measurement using the ASTRA software (Wyatt Technology). A Rayleigh–Debye–Gans light scattering model was used to determine M_w, using a Zimm plot. The uncertainties on M_w are a measure of the statistical consistency of the MALS data, obtained combining the standard deviations calculated for each slice in the analyzed peaks. Data analysis was performed using Astra version 5.3.4 following the manufacturer's instructions.

DNase I footprinting assay.

For probe labelling, the *pEX-A128-*P_{sod} plasmid was linearized with NotI, 5'end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase at one extremity and subsequently further digested with BamHI. Following gel purification, ca. 15 fmol of labelled probe were used for each footprinting reaction. Footprinting experiments were performed as described previously [29]. Basically, the P_{sod} probe was incubated with different concentrations of purified *Sg*SrnR in 50 µl of 1 X Binding Buffer (10 mM Tris-HCl pH 7.5, 50 mM KCl, 5 mM $MgCl_2$, 0.01% NP-40 [Igepal], 10% glycerol),

containing 300 ng of sonicated salmon sperm DNA as nonspecific competitor, in the absence or in the presence of 10 μM $NiSO_4$ for 15 min at room temperature. Then, 0.055 U of DNase I (Novagen), freshly diluted in 1 X Binding Buffer containing 5 mM $CaCl_2$, were added to the reaction mixture, and the digestion was incubated for 75 seconds at room temperature. The reaction was arrested by the addition of 140 μl of DNase I stop buffer (192 mM NaOAc pH 5.2, 32 mM EDTA, 0.1 % SDS, 64 mg/ml sonicated salmon sperm DNA). Samples were extracted with phenol-chloroform reaction, precipitated with ethanol, and resuspended in 10 μl of formamide loading buffer (95% formamide, 10 mM EDTA, 0.1% Xylene Cyanol FF, 0.1% Bromophenol Blue). Samples were then denatured at 95°C for 5 min, chilled on ice, subjected to 6% polyacrylamide – 8M urea gel electrophoresis, and autoradiographed.

β-galactosidase gene reporter assays

The *pKT25-*P_{sod}*-LacZ* construct was co-trasformed into *E. coli* BL21 (DE3) competent cells (Stratagene) with *pET15b-SgsrnR*. Negative controls were performed by co-trasforming *pKT25-*P_{sod}*-LacZ* with empty *pET15b* vector, and co-trasforming the *pKT25-*P_{1227}*-LacZ*, containing a *H. pylori* promoter not specific for *Sg*SrnR (see above), with *pET15b-SgsrnR*. Cells were grown in LB medium at 37 °C, in the absence or in the presence of increasing concentrations of $NiSO_4$ and $ZnSO_4$. When OD_{600} was equal to 0.5, the culture was split into two aliquots and, in one of them, expression of *Sg*SrnR was induced with 0.5 mM IPTG at 26 °C. Aliquots of 20 μL were removed immediately before and after induction at 30 mins intervals, and 80 μL of permeabilization solution (100 mM Na_2HPO_4, 20 mM KCl, 2 mM $MgSO_4$, 0.8 mg/mL hexadecyltrimethylammonium bromide, 0.4 mg/mL sodium deoxycholate and 5.4 μL/mL) were added. At the end of the experiment, 600 μL of substrate solution (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 1 mg/mL o-nitrophenyl-β-D-Galactoside (ONPG), 2.7 μL/mL β-mercaptoethanol) were added to each tube. When yellow color had developed in the reaction solution, 700 μL of stop solution (1 M Na_2CO_3) were added, and the time of reaction was recorded. Tubes were centrifuged for 5-10 min at 14,000 rpm and absorbance of the clarified

solutions was measured at 420 nm and at 550 nm. Enzymatic activity at different times was reported as Miller units, calculated as indicated below.

$$Miller\ Units = 1000\ x\ \frac{(Abs\ 420 - (1.75\ x\ \ Abs\ 550)}{(Abs\ 600\ x\ (volume\ [0.02\ mL] x\ reaction\ time)}$$

Results

Protein purification

The sequence coding for *Sg*SrnR was obtained by retro-translating its amino acid sequence (UniProt code: Q8L1Y3), and optimizing it for the expression in *E. coli*. The gene was initially cloned in a *pET15b* vector but the expression of the His-tagged protein was low, thus we moved to a different expression approach. This involved the N-terminal tagging of *Sg*SrnR with the modified immunoglobulin binding domain of protein A of *Staphylococcus aureus* (ZZ-tag), an effective carrier for recombinant protein expression, in terms of yields, solubility and conformational stability [22]. This tag was preceded by a His-rich sequence to facilitate protein purification by affinity chromatography. The obtained construct was used to overproduce His-ZZ-His-*Sg*SrnR (31.1 kDa) in *E. coli* BL21 (DE3) by induction with IPTG. This protocol produced a polypeptide with the correct molar mass that accumulated in the soluble fraction of the cellular extract. The protein was purified with a Ni(II)-based affinity chromatography, in column proteolytic cleavage to remove the tag, and elution of *Sg*SrnR (12.6 kDa). The flow was connected in line with a desalting column to decrease the ionic strength of the buffer for a subsequent anionic exchange chromatographic step. Protein polishing was obtained with a size exclusion chromatography (SEC) step, achieving a final yield of 10-15 mg L^{-1} of initial culture. Protein purity was verified using SDS-PAGE (Figure 1A).

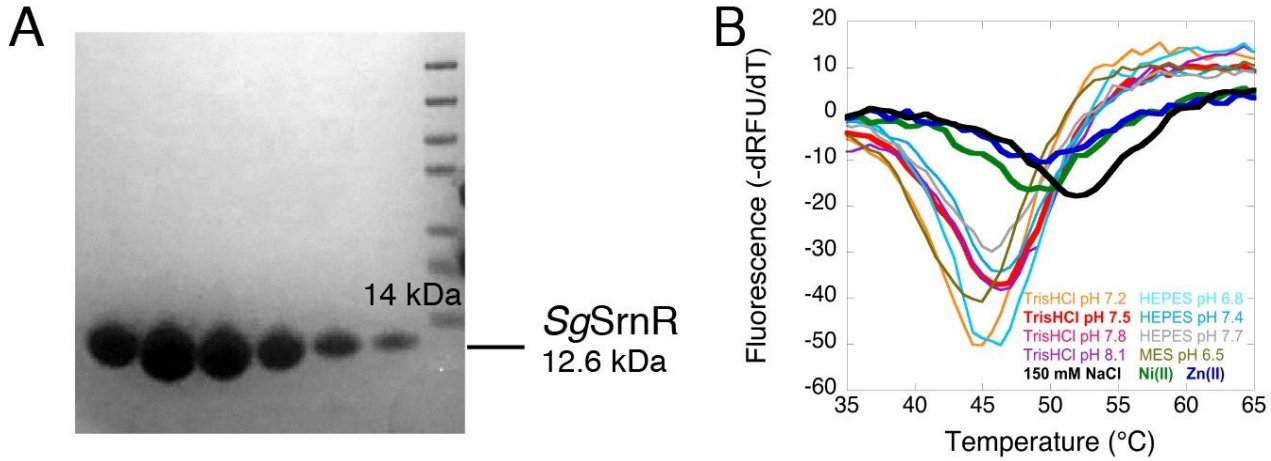


Figure 1: (A) Purity of the isolated *Sg*SrnR analyzed by SDS-PAGE electrophoresis of the fractions from the last chromatographic step. (B) Representative first derivative curves of fluorescence for a thermal shift assay formulated in a series of buffers that range in pH from 6.5-8.1. The minimum of the curves corresponds to the protein melting temperatures. The effect of 1:1 Ni(II)/Zn(II):protein ratio on the protein unfolding in 20 mM TrisHCl pH 7.5, 150 mM NaCl, 1 mM TCEP is shown.

Conditions for protein storage and analysis were selected by testing protein stability under different buffer conditions using thermal shift assay (TSA) experiments. In some cases, normalized fluorescence signals for the protein/dye combinations did not have a distinct transition in the fluorescence signal and were therefore excluded from the analysis. The T_m values for protein unfolding under the other conditions were calculated as the midpoint of the fluorescence transition, reported as a negative peak in the first derivative curve (Figure 1B). The buffer selected for producing the highest T_m (46.4 °C) was 20 mM TrisHCl pH 7.5. Starting from this condition, we tested increasing ionic strengths from 0 to 1 M NaCl. The T_m shifted to slightly higher temperature with 150 mM NaCl (52.0 °C). Thus, the buffer used for the subsequent analyses was 20 mM Tris HCl pH 7.5, 150 mM NaCl, 1 mM TCEP.

Isothermal titration calorimetry

The metal binding properties of *Sg*SrnR were evaluated using isothermal titration calorimetry (ITC). The calorimetric data indicate that the protein binds both Ni(II) and Zn(II), as can be derived from the exothermic peaks in the heat trace of the ITC raw data (Figure 2A,C). In the case of Ni(II), the integrated titration data (Figure 2B) show a single inflection point, thus the data were fitted using a

single set of sites model (χ^2 = 1.5), indicating one Ni(II) ion bound per protein monomer, with a moderate affinity $K_a = 6.2 \pm 0.8 \times 10^4$ ($K_d = 16 \pm 2$ μM). The reaction was driven by negative enthalpy ($\Delta H = -3.3 \pm 0.2$ kcal mol^{-1}) and positive entropy ($\Delta S = +10.9$ cal mol^{-1} K^{-1}). Zn(II) binding to *Sg*SrnR shows a binding isotherm that suggests the presence of two binding events (Figure 2D). Consistently, a fit of the thermodynamic data using a model involving two independent sets of binding sites gives better statistics as compared to the other fitting models ($\chi^2 = 4.79$). This analysis indicates that Zn(II) binds to the *Sg*SrnR monomer in two binding events, with a stoichiometries $N_1 = 0.25$ and $N_2 = 1$ and with $K_{a1} = 7.0 \pm 0.1 \times 10^6$ ($K_d = 0.14 \pm 0.02$ μM), $\Delta H_1 = -11.56 \pm 0.02$ kcal mol^{-1}, ΔS_1 = -7.47 cal mol^{-1} K^{-1}, and $K_{a2} = 2.56 \pm 0.02 \times 10^5$ ($K_{d2} = 3.91 \pm 0.03$ μM), $\Delta H_2 = -8.51 \pm 0.03$ kcal mol^{-1}, ΔS_2 = -3.81 cal mol^{-1} K^{-1}. The stoichiometry of the first binding event is lower than one, suggesting that one Zn(II) binds to four *Sg*SrnR monomers, hinting for protein oligomerization upon Zn(II) binding.

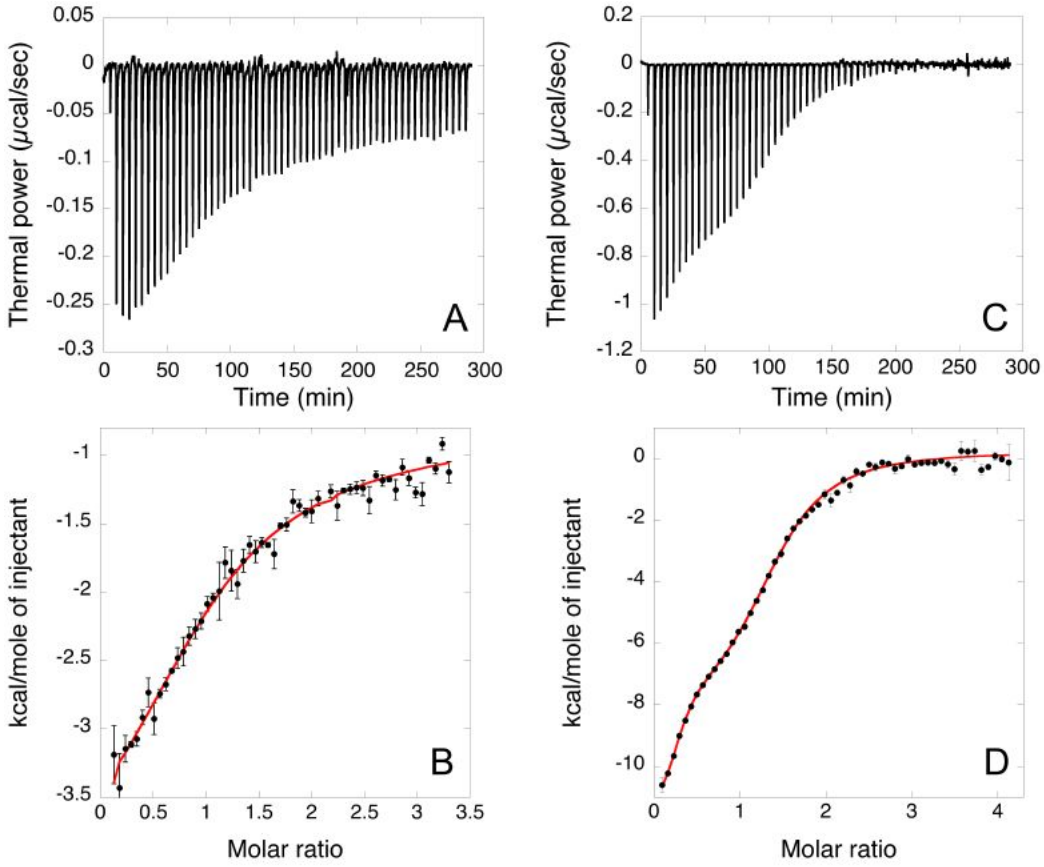


Figure 2: ITC titration data for the binding of Ni(II) (A, B) and Zn(II) (C,D) to *Sg*SrnR. Raw titration data for Ni(II) (A) and Zn(II) (C) represent the thermal effect of 5 μL injections of metal ion solutions (600 μM) into protein solution (30-40 μM). Fits of the normalized heat reaction data for Ni(II), with

a single site model (B) and for Zn(II) with a two sets of independent sites model (D) are represented as continuous red lines.

Circular dichroism and fluorescence spectroscopies

The CD spectrum of apo-*Sg*SrnR is typical of a protein with significant amount of secondary structure, with a positive band at 192 nm and negative signals at 210 and 224 nm. The quantitative analysis performed using Dichroweb [26] hints to the prevalence of α-helices (52 %), while β-strands relative amount is minor (15 %), as well as unordered structures (16 % turns and 16 % random coil). The obtained values are similar to those predicted using the software JPred4 [30], corresponding to 53 % of α-helices and 8.5 % of β-strands. The CD spectrum of the protein is not significantly influenced by the addition of up to five equivalents of Ni(II), indicating that Ni(II) binding does not affect the secondary structure of the protein (Figure 3A). Differently, addition of one equivalent of Zn(II) considerably modifies the secondary structure of *Sg*SrnR, with calculated relative amounts of 18% α-helix, 29% β-strands, 23% turns and 29% random coil, indicating a loss of secondary structure upon Zn(II) addition. Upon addition of two or more Zn(II) equivalents, a partial precipitation of the sample is observed, accompanied by a further decrease in the α-helical content and by an increase of the β-strands amount, with a secondary structure compositions of 8% α-helices, 44% β-strands, 14% turns and 34% random coil in the presence of two Zn(II) equivalents (Fig. 3A).

Upon temperature increase, the absolute value of ellipticity considerably decreases and becomes close to zero at 90 °C (Figure 3B), indicating that *Sg*SrnR loses secondary structure with a sigmoidal profile typical of a two-state denaturation process. The denaturation curve can be fitted with a two-state thermal denaturation function, exhibiting midpoint transition temperature (T_M) of 48.5 ± 0.2 °C. This is slightly lower than the observed midpoint transition registered for the TSA analysis (52.0 °C), indicating that the secondary structure unfolding precedes the loss of hydrophobic clusters in the protein fold. The apparent van't Hoff enthalpy of unfolding is calculated as −12.8 ± 0.5 kcal mol^{-1}.

The CD spectrum in the 250-350 nm wavelength region can be sensitive to certain aspects of protein tertiary structure. At these wavelengths, the chromophores are the aromatic amino acids, and the CD

signals that they produce are sensitive to the tertiary structure of the protein in the vicinity of the aromatic moieties. In the spectrum of the native protein (Figure 3C, D), two bands covering the 250-270 nm range, likely corresponding to the three phenynalanines signals and one band from 270 to 290 nm, possibly referred to the single tyrosine, are clearly visible, indicating that these aromatic residues are located in asymmetric regions of the protein, containing significant tertiary structure. Addition of one Ni(II) equivalent (Figure 3C), produces an additional signal at 280 nm, likely corresponding to the single tryptophan residue, indicating a gain of rigidity in the region surrounding this residue. Further addition of Ni(II), however, decreases this band intensity, indicative of an increased mobility of this region. Ni(II) addition also produces an increase of the phenylalanine band at ca. 260 nm, which is maintained at higher Ni(II) stoichiometries (Figure 3C). Differently, Zn(II) addition does not produce any major change in the near-UV CD spectrum, suggesting a non-significant change of the regions closed to the aromatic residues (Figure 3D).

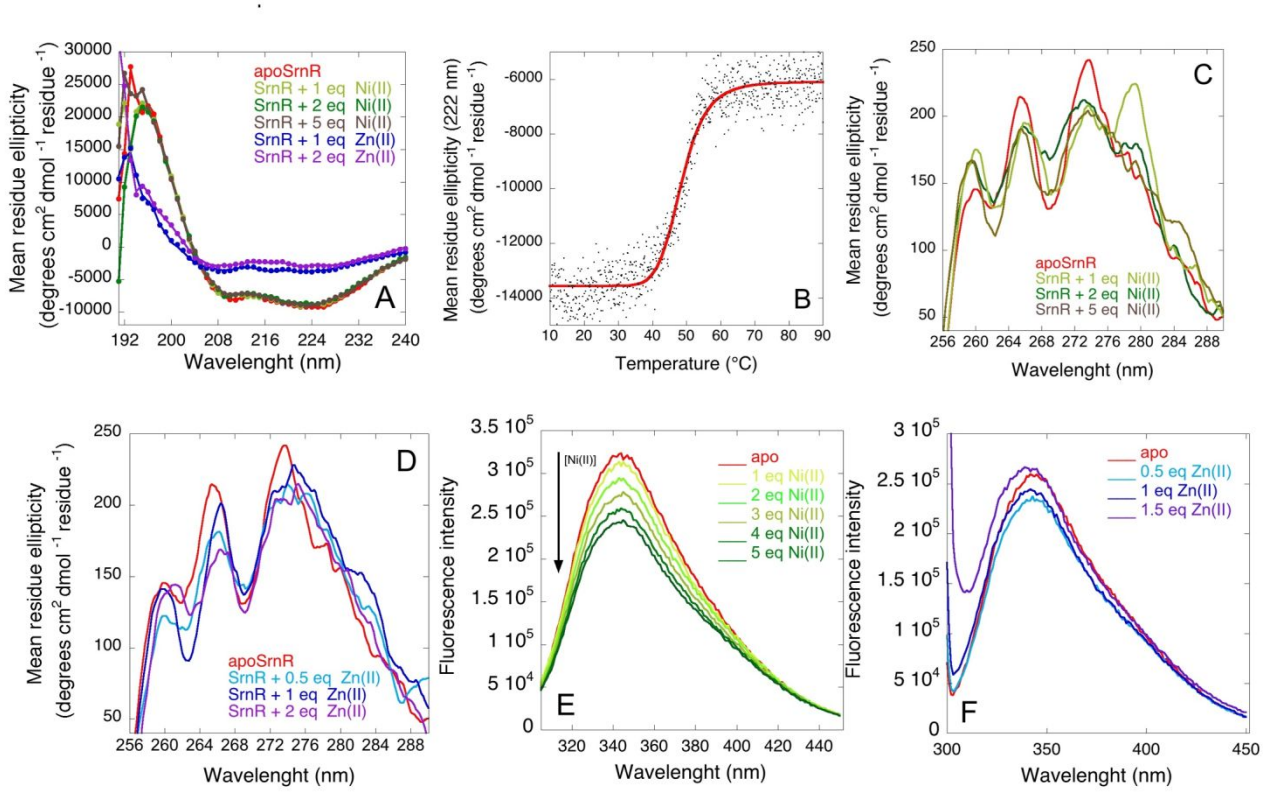


Figure 3. A. Far-UV CD spectra of *Sg*SrnR in the absence (red) and in the presence of one (light green), two (green) or five (brown) equivalents of Ni(II), or one (blue) or two (purple) equivalents of Zn(II). B. Temperature-induced conformational transitions observed as changes in ellipticity at 222

nm in the 10 °C–90 °C temperature range for *Sg*SrnR. C. Near-UV CD spectra of *Sg*SrnR in the absence (red) and in the presence of one (light green), two (green) and five (brown) equivalents of Ni(II). D. Near-UV CD spectra of *Sg*SrnR in the absence (red) and in the presence of half (cyan), one (blue) or two (purple) equivalents of Zn(II). E. Intrinsic fluorescence intensity steady-state measurements of *Sg*SrnR, in the absence and in the presence of increasing concentrations of Ni(II) ions. E. Intrinsic fluorescence intensity steady-state measurements of *Sg*SrnR, in the absence and in the presence of increasing concentrations of Zn(II) ions.

The local conformational transitions affecting the tryptophan region of *Sg*SrnR were monitored by following the changes in the protein intrinsic fluorescence spectra as a function of increasing Ni(II) (Figure 3E) and Zn(II) concentration (Figure 3F). The fluorescence spectrum of the protein presents a maximum at 343-345 nm. Addition of Ni(II) produces a moderate decrease of the fluorescence intensity, with no major modifications of the spectrum profile and position (Figure 3E). These results indicate that the binding of Ni(II) to *Sg*SrnR induces some modifications of the chemical state and photophysics of the tryptophan, likely associated to a moderate increase of the mobility of the residue side chain, without changing the polarity of its molecular environment. Differently, addition of half of the stoichiometric equivalent of Zn(II) produces a minor quenching of the tryptophan fluorescence (Figure 3F), while in the presence of higher Zn(II) concentrations a significant increase of the scattering at wavelengths lower than 300 nm is observed, indicating phenomena of Zn(II)-induced protein aggregations, as already suggested by far-UV CD spectroscopy. The Zn(II)-driven protein aggregation was also coherent with an increased absorbance of the protein solution at wavelengths higher than 300 nm in the UV-vis absorbance spectrum of *Sg*SrnR.

The metal-induced changes of protein tertiary structure were also monitored by extrinsic fluorescence using TSA assays. Addition of both metal ions shifts the T_m to a lower value (from 52.0 °C to 49.3 °C) suggesting a metal-ion driven decrease of protein stability (Figure 1B).

Light scattering measurements

The hydrodynamic and oligomeric properties of *Sg*SrnR in solution, were determined with SEC in line with multiple-angle and quasi-elastic light scattering (MALS and QELS). The results indicate

that *Sg*SrnR elute from the SEC column in a single peak corresponding to the dimeric form of the protein (MW = 26.4 kDa, R_h= 2.9 nm). The results are similar in the presence of one equivalent of Ni(II) (MW = 25.9 kDa, R_h= 3.2 nm) or Zn(II) (MW = 27.0 kDa, R_h= 3.5 nm) per protein monomer, confirming that the addition of these metals does not change the oligomeric state of the protein and causes a slight increase of the hydrodynamic volume (Figure 4). Addition of higher concentrations of Zn(II) causes protein precipitation.

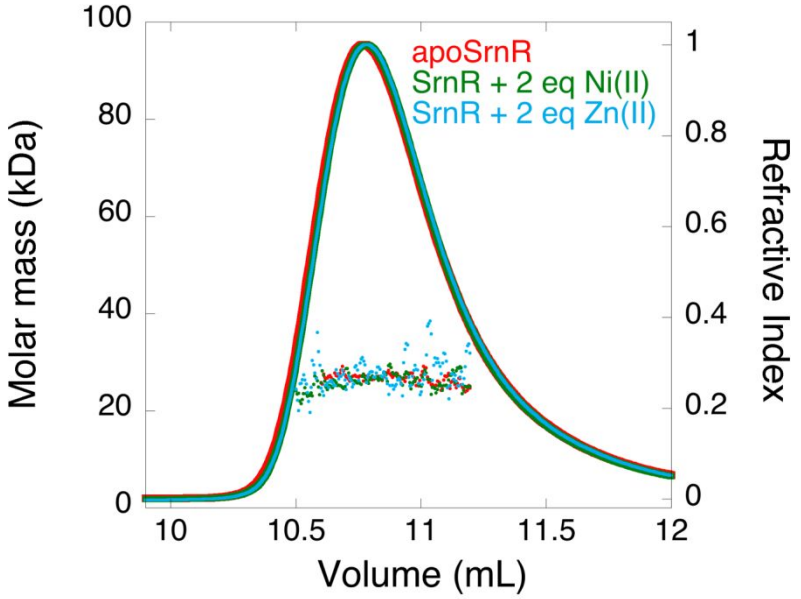


Figure 4: Size exclusion profile monitored by the refractive index detector (lines) and weight-averaged molar mass molar distribution (dots) of *Sg*SrnR obtained with a combination of SEC-MALS-QELS in the absence (red) and in the presence of one equivalents of Ni(II) (green) or Zn(II) (blue).

DNase I footprinting

The nickel-responsive transcriptional regulation of *S. griseus sodF* is exerted through an inverted-repeat sequence (TTGCA-N7-TGCAA) overlapping the transcription start site of the gene [19]. In particular, *Sg*SrnR was found to retard the *sodF* regulatory region in gel-mobility shift assays, when both *Sg*SrnQ and Ni(II) were provided [19]. To further investigate the direct *Sg*SrnR-P_{sod} interaction and map the DNA contacts with high resolution, a DNase I footprinting assay was set up with the purified recombinant protein using the P_{sod} region as radiolabelled probe, both in the absence or in the presence of Ni(II). The addition of at least 480 nM of protein dimer was able to protect the region from -15 to +27 with respect to the transcriptional start site (Figure 5A), indicative of apo-*Sg*SrnR

binding to the DNA probe. The binding site is specific and is overlapping the transcriptional start site encompassing the inverted-repeat sequence (TTGCA-N7-TGCAA), previously suggested to have a role in *sodF* transcriptional regulation (Figure 5B) [19]. The addition 10 μM $NiSO_4$ (Figure 5A) did not change the binding pattern and relative binding affinity of *Sg*SrnR for P_{sod}, denoting that, in the tested experimental conditions, *Sg*SrnR is able to bind directly to the *sodF* promoter in a nickel-independent manner.

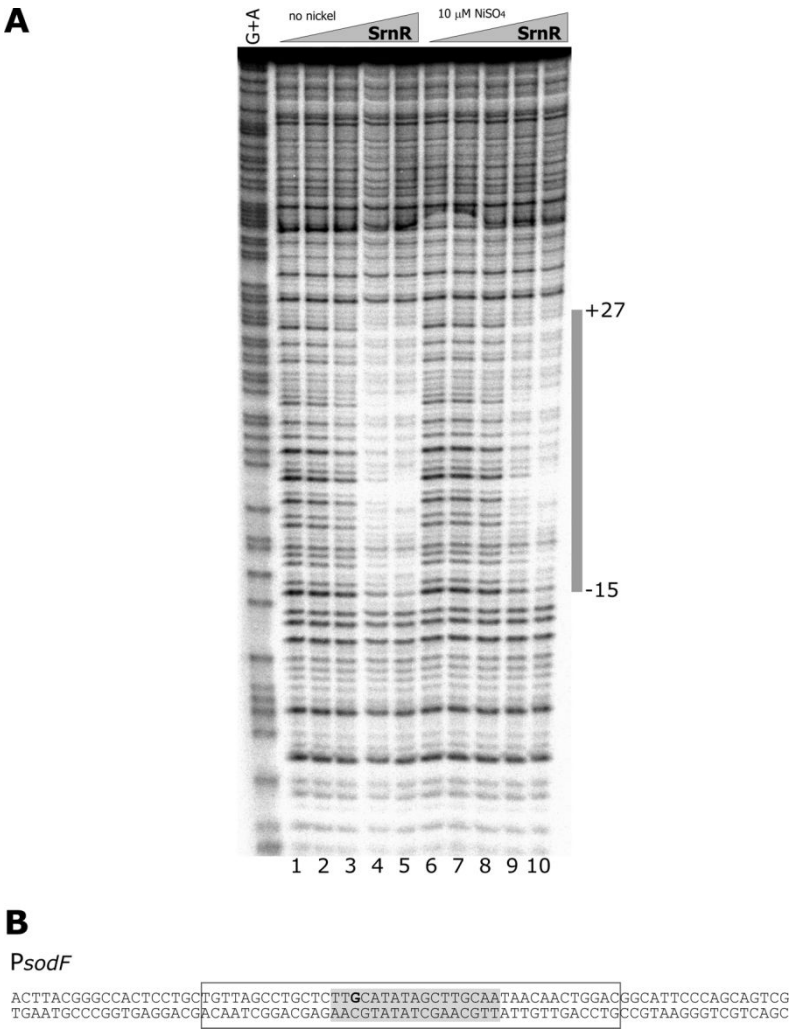


Figure 5. DNase I footprinting analysis of *Sg*SrnR binding to P_{sodF}. (A) A radiolabelled DNA probe, encompassing the *sodF* promotorial region, was incubated with increasing concentrations of purified *Sg*SrnR both in the absence and in the presence of $NiSO_4$ (Lanes 1 to 5: 0, 120, 240, 480, 960 nM dimeric *Sg*SrnR, respectively, absence of $NiSO_4$; lanes 6 to 10: 0, 120, 240, 480, 960 nM dimeric *Sg*SrnR, respectively, in the presence of 10 μM $NiSO_4$) and subjected to DNaseI partial digestion. Purified DNA fragments were separated on a polyacrylamide denaturing gel along with a G+A sequence reaction ladder to map the binding site. The grey box on the right of the autoradiograph represents the DNA region protected from DNase I digestion. (B) Mapping of the *Sg*SrnR binding site on the *sodF* promoter region. The region of protection is included in an open grey box, the

inverted repeat sequence proposed by Kim *et al.* [19] is highlighted in light grey, while the nucleotide that initiates the transcription of *sodF* is representedin bold.

Reporter gene assay

To analyze the role of *Sg*SrnR as trans regulator over *sodF* promoter, an assay was set up using β−galactosidase (β-gal) as reporter gene. The candidate promoter P_{sod} was cloned upstream of *lacZ* encoding β-gal in a plasmid vector that was then transformed into *E. coli* BL21(DE3), together with an expression plasmid containing the His-tagged *Sg*SrnR under the control of the T7 promoter. The obtained strain did not produce β-gal activity as determined by the Miller assay (70 Miller units) [31], suggesting that P_{sod} exhibited weak basal transcriptional activity in *E. coli* in the absence of a specific inductor of transcription. Differently, expression of *Sg*SrnR induced with IPTG produced an increasing color change from transparent to yellow in the reaction mixture over the time after induction (Figure 6). This change is indicative of the up-regulation of the β-gal activity over time. Negative controls were performed by substituting the P_{sod} promoter with a non-specific one or co-transforming the P_{sod} reporter plasmid with an empty expression vector. In both cases the β-gal induced color change was negligible, supporting the role of *Sg*SrnR as a transcriptional activator.

No significant change was observed including up to 200 μM Ni(II) in the reaction mixture, suggesting that, coherently with the DNase I footprinting experiment, Ni(II) binding to *Sg*SrnR has no or little effect on the protein function as a transcriptional activator. The same result was obtained performing an analogous experiment in the presence of increasing concentrations of Zn(II).

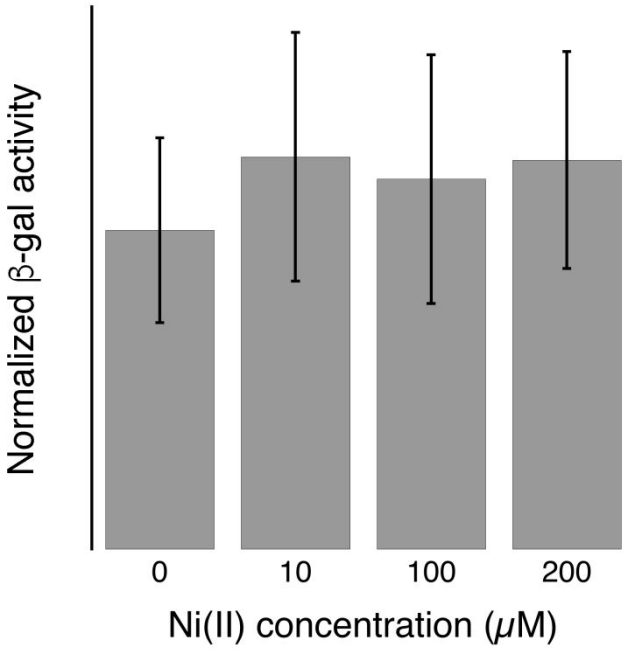


Figure 6. Measurement of β-gal activity, measured as Miller units produced over time by *E. coli* cultures co-transformed with *pKT25-P*$_{sod}$ and *pET15b-SgsrnR* plasmids, after induction of *Sg*SrnR expression by IPTG and in the absence or in the presence of increasing Ni(II) concentrations. The activity measured at the different Ni(II) concentrations is represented as a ratio of the activity produced by the culture without addition of Ni(II) ions.

Discussion

The present work describes the biophysical and functional characterization of *Sg*SrnR, one of the two components of a transcriptional regulation system that, in *S. griseus,* governs the antagonistic expression of FeZn-SOD and Ni-SOD [19]. The protocol for the over-expression of *Sg*SrnR, the DNA binding protein of the system, fused to a N-terminal ZZ-tag and its purification to homogeneity is reported. This protein belongs to the ArsR/SmtB family, the most represented and wide-spread group of metal sensors [19, 32]. Members of this family are homo-dimeric proteins that share a general fold with at least five α-helices and two three-stranded β-sheets. Consistently, SEC-MALS data show that isolated *Sg*SrnR is a stable homodimer in solution. A previous work, which described the expression and purification of the protein using a GST-based approach, found the protein in an apparent monomeric state according to SEC experiments [19]. However, SEC is known to be unreliable for the determination of protein molar mass, as the calibration approach is often inaccurate due to

possible significant structural differences between the sample and the standards. In addition, artifacts can be caused by interactions of the samples with the solid phase, and by the impact of the experimental parameters on the observed outcomes [33]. Differently, SEC-MALS is a more reliable and absolute technique that eliminates the calibration step and directly measures the molar mass of the species in solution [33].

A quantitative analysis of far-UV CD spectral curves indicates that *Sg*SrnR presents a prevalence of α-helical content, coherently with its belonging to the ArsR/SmtB class of transcriptional regulators. Thermal unfolding revealed that the protein behaves as a well-folded polypeptide, undergoing an apparent two-state denaturation process. The protein unfolding follows the loss of α-helical structure, detected by CD (T_m = 48.5 °C), while the disruption of hydrophobic clusters, monitored by extrinsic fluorescence in TSA assays, occurs at slightly higher temperatures (T_m = 52.0 °C).

ArsR/SmtB proteins generally interact with their target DNA operators acting as apo-repressors, using two-winged helix-turn-helix motifs. Three ArsR/SmtB members of Ni(II) sensors were formerly characterized in bacteria: NmtR and KmtR from *Mycobacterium tuberculosis* and NmtR from *S. coelicolor* [9, 34]. In all these cases, metal ion binding induces protein dissociation from DNA resulting in an increased gene transcription (Figure 7A) [32].

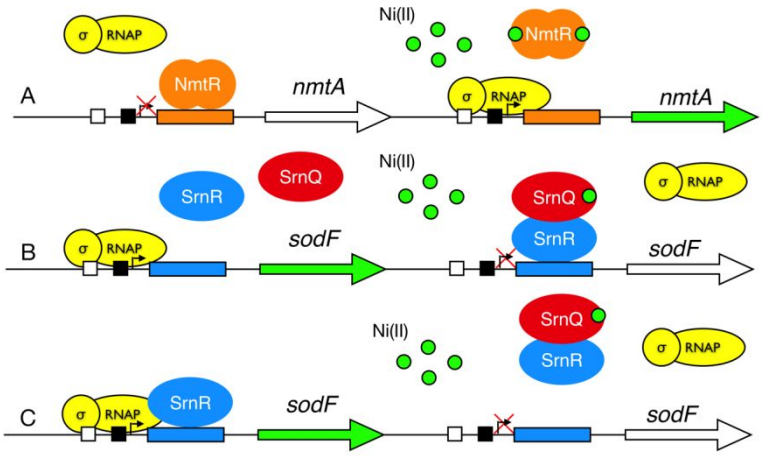


Figure 7. DNA binding models for ArsR/SmtB family members. **A.** General mechanism of functioning for known Ni(II)-responsive ArsR/SmtB proteins, functioning as apo-repressors, whose DNA binding activity is attenuated by Ni(II). **B.** Mechanism of functioning for *Sg*SrnR/*Sg*SrnQ proteins, acting as holo-repressors, as previously proposed[19]: *Sg*SrnR binds its operator in a complex with *Sg*SrnQ and Ni(II) and represses gene transcription **C.** Mechanism of functioning for *Sg*SrnR as results from the present work: apo-*Sg*SrnR binds its operator DNA and activates the transcription. In the presence of a cognate protein, such as *Sg*SrnQ, binding from DNA is relieved and transcription is repressed.

To investigate the ability of *Sg*SrnR to function as a DNA binding protein, we performed DNase I footprinting assays using a radioactively labeled probe encompassing the inverted repeat sequence previously identified as the operator that regulates the *sodF* expression [19]. Under these experimental conditions, the protein is able to protect a region of 42 bp overlapping the inverted repeat sequence and the transcriptional start site. This result clearly indicates that, as other ArsR/SmtB members, apo-*Sg*SrnR is active as operator binder. On the other hand, this observation contrasts with previously reported EMSA experiments, that failed to reveal protein-DNA interaction when Ni(II) and *Sg*SrnQ were absent from the protein solution [19]. This discrepancy can be explained considering the intrinsic limitations of the EMSA methodology: nonappearance of electrophoretic shift is not always indicative of the absence of protein-DNA interactions, as dissociation during electrophoresis can occur because samples are not in equilibrium during the electrophoretic run [35]. The concomitant expression of *Sg*SrnR and *Sg*SrnQ, together with the addition of Ni(II), was reported to significantly inhibit the transcription of the *sodF* mRNA, which became undetectable when 5 μM Ni(II) was added to the reaction solution [19]. This observation, coupled to the EMSA results reported above, was interpreted considering *Sg*SrnRQ a holo-repressor. Ni(II) binding to *Sg*SrnQ was postulated to drive the formation of a Ni(II)-*Sg*SrnRQ complex that, interacting with the *sodF* DNA operator, prevented binding of RNA polymerase and consequent gene transcription (Figure 7B). However, this transcriptional response is compatible with another mechanicistic model, that sees apo-*Sg*SrnR binding to its operator and acting as an activator of transcription, while Ni(II) and/or *Sg*SrnQ, interacting with the regulator, drive its dissociation from DNA. Even though metal sensors more often act as transcriptional repressors, in a limited number of cases, such as the members of the MerR family, they function as activators, facilitating binding of RNA-polymerase to its promoter [36]. The evidence that *Sg*SrnR was actually able to bind DNA in the absence of cofactors and/or protein partner in DNase I footprinting assays prompted us to set up a gene reporter assay, to investigate the role of *Sg*SrnR as transcriptional repressor or activator. *E. coli* was co-transformed with an expression vector

coding for *Sg*SrnR and a reporter vector, with the β-gal gene cloned downstream to the P_{sod} promoter. Measurement of β-gal activity in the obtained strain clearly demonstrates an increase of enzyme catalysis specifically associated to *Sg*SrnR production. Hence, this outcome supports the role of apo-*Sg*SrnR as an activator of transcription (Figure 7C).

Previous data reported the absence of activity by isolated *Sg*SrnR in response to Ni(II) and found the need of *Sg*SrnQ as Ni(II) receptor. On the other hand, ITC experiments indicate that *Sg*SrnR is able to bind one Ni(II) ion per protein monomer with moderate affinity in an enthalpy and entropy-driven reaction. As a comparison, titration of the protein with Zn(II) reveals the occurrence of two binding events: the first binding showed non-integer stoichiometry, negative entropy and higher affinity, suggesting that some different event, not directly related to binding, occurs upon Zn(II) titration onto the protein. Indeed, addition of more than one equivalent of Zn(II) induces protein aggregation and a drop in secondary structure, as observed by CD and fluorescence.

Ni(II) binding does not significantly affect protein secondary structure nor the oligomerization state, as reported by far-UV CD and light scattering experiments. However, Ni(II) addition slightly changes the protein tertiary structure, as revealed near-UV and fluorescence spectroscopy. In particular, Ni(II) quenches the intrinsic fluorescence of the unique Trp residue located in the C-terminal region of the protein, suggesting an increased mobility of the Trp moiety, as previously observed for the Ni(II)-sensors *H. pylori* NikR [37]. In addition, Ni(II), as well as Zn(II), slightly decrease protein stability as revealed by extrinsic fluorescence, with T_m shifting from 52.0 °C to 49.3 °C. This result is unexpected as, usually, ligand binding improves protein stability. We interpret this observation considering that metal binding might change the protein conformation rendering it more labile, with Ni(II) increasing the flexibility of the protein and Zn(II) driving protein aggregation.

Protein activity as a DNA binder was unaffected, as observed by footprinting analysis in the presence of an excess of metal ion. Coherently, the protein-driven activation of transcription remains unchanged in the presence of Ni(II). The measured protein-metal dissociation constant of 16 μM is not compatible with the transcriptional response to Ni(II) observed in vivo, which is already observed

at a Ni(II) concentration of 0.1 μM and is complete at 5 μM [19]. All these considerations support the previously postulated role of *Sg*SrnQ in the transcriptional regulation network. Thus, the physiological function of the Ni(II) binding activity of *Sg*SrnR remains unclear and needs further investigation. We speculate that the Ni(II) binding site of *Sg*SrnR takes part in the interaction with *Sg*SrnQ, possibly bridging the two proteins together. Understanding of this process would require high resolution structural information on *Sg*SrnR, in the absence and in the presence of Ni(II) and/or DNA operator, as well as in the presence of its physiological partner *Sg*SrnQ. These studies are currently underway in our laboratory.

Acknowledgements

The authors thank Prof. Stefano Ciurli for financial support and useful discussion. They also thank Prof. Paolo Neyroz for helpful assistance with fluorescence measurements and examination of the data. This work was supported by the Department of Pharmacy and Biotechnology of the University of Bologna through funds for fundamental research. YB and AZ are recipient of PhD fellowships by the University of Bologna.

References

1 C. Andreini, G. Cavallaro, S. Lorenzini and A. Rosato (2013) Nucleic Acids Res 41:D312-319
2 M. I. Hood and E. P. Skaar (2012) Nat Rev Microbiol 10:525-537
3 P. Chandrangsu, C. Rensing and J. D. Helmann (2017) Nat Rev Microbiol 15:338-350
4 P. T. Chivers and R. T. Sauer (2000) J Biol Chem 275:19735-19741
5 J. S. Iwig, S. Leitch, R. W. Herbst, M. J. Maroney and P. T. Chivers (2008) J Am Chem Soc 130:7592-7606
6 J. S. Iwig, J. L. Rowe and P. T. Chivers (2006) Mol Microbiol 62:252-262
7 F. Musiani, B. Zambelli, M. Bazzani, L. Mazzei and S. Ciurli (2015) Metallomics 7:1305-1318
8 B. E. Ahn, J. Cha, E. J. Lee, A. R. Han, C. J. Thompson and J. H. Roe (2006) Mol Microbiol 59:1848-1858
9 H. M. Kim, B. E. Ahn, J. H. Lee and J. H. Roe (2015) Metallomics 7:702-709
10 I. Fridovich (1995) Annu Rev Biochem 64:97-112
11 F. Johnson and C. Giulivi (2005) Mol Aspects Med 26:340-352
12 H. D. Youn, E. J. Kim, J. H. Roe, Y. C. Hah and S. O. Kang (1996) Biochem J 318 (Pt 3):889-896
13 J. A. Fee (1991) Mol Microbiol 5:2599-2610
14 E. J. Kim, H. J. Chung, B. Suh, Y. C. Hah and J. H. Roe (1998) Mol Microbiol 27:187-195
15 E. J. Kim, H. J. Chung, B. Suh, Y. C. Hah and J. H. Roe (1998) J Bacteriol 180:2014-2020

16 H. J. Chung, J. H. Choi, E. J. Kim, Y. H. Cho and J. H. Roe (1999) J Bacteriol 181:7381-7384
17 J. S. Kim, J. H. Jang, J. W. Lee, S. O. Kang, K. S. Kim and J. K. Lee (2000) Biochim Biophys Acta 1493:200-207
18 H. M. Kim, J. H. Shin, Y. B. Cho and J. H. Roe (2014) Nucleic Acids Res 42:2003-2014
19 J. S. Kim, S. O. Kang and J. K. Lee (2003) J Biol Chem 278:18455-18463
20 B. Zambelli, V. N. Uversky and S. Ciurli (2016) Biochim Biophys Acta 1864:1714-1731
21 J. Bogomolovas, B. Simon, M. Sattler and G. Stier (2009) Protein Expr Purif 64:16-23
22 Y. Zhao, Y. Benita, M. Lok, B. Kuipers, P. van der Ley, W. Jiskoot, W. E. Hennink, D. J. Crommelin and R. S. Oosting (2005) Vaccine 23:5082-5090
23 T. T. Hoang, A. J. Kutchma, A. Becher and H. P. Schweizer (2000) Plasmid 43:59-72
24 G. Karimova, A. Ullmann and D. Ladant (2001) J Mol Microbiol Biotechnol 3:73-82
25 A. D'Urzo, C. Santambrogio, R. Grandori, S. Ciurli and B. Zambelli (2014) J Biol Inorg Chem 19:1341-1354
26 L. Whitmore and B. A. Wallace (2008) Biopolymers 89:392-400
27 M. Miraula, S. Ciurli and B. Zambelli (2015) J Biol Inorg Chem 20:739-755
28 P. A. Charlwood (1957) J Am Chem Soc 79:776-781
29 S. Pelliciari, E. Pinatel, A. Vannini, C. Peano, S. Puccio, G. De Bellis, A. Danielli, V. Scarlato and D. Roncarati (2017) Sci Rep 7:41063
30 A. Drozdetskiy, C. Cole, J. Procter and G. J. Barton (2015) Nucleic Acids Res 43:W389-394
31 J. H. Miller (1972). Cold Spring Harbor Laboratory,
32 B. Zambelli, F. Musiani and S. Ciurli (2012) Met Ions Life Sci 10:135-170
33 S. Podzimek (2014) J Appl Polym Sci 131
34 C. W. Lee, D. K. Chakravorty, F. M. Chang, H. Reyes-Caballero, Y. Ye, K. M. Merz, Jr. and D. P. Giedroc (2012) Biochemistry 51:2619-2629
35 L. M. Hellman and M. G. Fried (2007) Nat Protoc 2:1849-1861
36 N. L. Brown, J. V. Stoyanov, S. P. Kidd and J. L. Hobman (2003) FEMS Microbiol Rev 27:145-163
37 B. Zambelli, A. Danielli, S. Romagnoli, P. Neyroz, S. Ciurli and V. Scarlato (2008) J Mol Biol 383:1129-1143

Annex 2

DOI: 10.1093/mtomcs/mfab069
https://academic.oup.com/metallomics/articleabstract/13/12/mfab069/6445039?redirectedFrom=fulltext

OXFORD UNIVERSITY PRESS | Metallomics

Structure, dynamics and function of SrnR, a transcription factor for nickel-dependent gene expression

Journal:	*Metallomics Journal*
Manuscript ID	MTOMCS-2021-109.R2
Manuscript Type:	Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Mazzei, Luca; University of Bologna, Department of Pharmacy and Biotechnology Musiani, Francesco; University of Bologna, Pharmacy and Biotechnology Żerko, Szymon ; University of Warsaw, Faculty of Chemistry, Biological and Chemical Research Centre Kozminski, Wiktor ; University of Warsaw, Faculty of Chemistry, Biological and Chemical Research Centre Cianci, Michele; Polytechnic University of Marche, Department of Agricultural, Food and Environmental Sciences Beniamino, Ylenia; University of Bologna, Department of Pharmacy and Biotechnology Ciurli, Stefano; University of Bologna, Pharmacy and Biotechnology Zambelli, Barbara; University of Bologna, Department of Pharmacy and Biotechnology
Keywords:	Nickel sensor, Transcriptional regulation, Crystallography, NMR, Molecular docking, Molecular dynamics calculations, isothermal titration calorimetry, Streptomyces griseus, Metal homeostasis

Structure, dynamics, and function of SrnR, a transcription factor for nickel-dependent gene expression

Luca Mazzei,[a] Francesco Musiani,[a] Szymon Żerko,[b] Wiktor Koźminski,[b] Michele Cianci,[c] Ylenia Beniamino,[a] Stefano Ciurli,[*,a] Barbara Zambelli[*,a]

[a] Laboratory of Bioinorganic Chemistry, Department of Pharmacy and Biotechnology (FaBiT), University of Bologna, Via Giuseppe Fanin 40, I-40127 Bologna (Italy)

[b] Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089, Warsaw, Poland

[c] Department of Agricultural, Food and Environmental Sciences, Polytechnic University of Marche, Via Brecce Bianche, I-60131 Ancona (Italy)

[a]: LM and FM contributed equally to this study

* Corresponding authors: stefano.ciurli@unibo.it; barbara.zambelli@unibo.it

ORCID for Luca Mazzei: 0000-0003-1335-9365
ORCID for Francesco Musiani: 0000-0003-0200-1712
ORCID for Szymon Żerko: 0000-0002-0044-1100
ORCID for Wiktor Koźminski: 0000-0003-2319-4525
ORCID for Michele Cianci: 0000-0001-5607-6061
ORCID for Stefano Ciurli: 0000-0001-9557-926X
ORCID for Barbara Zambelli: 0000-0002-3876-0051

KEYWORDS

Nickel homeostasis; Nickel sensing; *Streptomyces griseus*; Protein crystallography; NMR spectroscopy; Calorimetry; Molecular modelling; Molecular dynamics

This article is dedicated to the memory of Deborah Zamble, whose insights in nickel biology were critical for the development of this field of bioinorganic chemistry.

ABSTRACT

Streptomyces griseus, a bacterium producing antibacterial drugs and featuring possible application in phytoremediation, expresses two metal-dependent superoxide dismutase (SOD) enzymes, containing either Fe(II) or Ni(II) in their active site. In particular, the alternative expression of the two proteins occurs in a metal-dependent mode, with the Fe(II)-enzyme gene (*sodF*) repressed at high intracellular Ni(II) concentrations by a two-component system (TCS). This complex involves two proteins, namely *Sg*SrnR and *Sg*SrnQ, which represent the transcriptional regulator and the Ni(II) sensor of the system, respectively. *Sg*SrnR belongs to the ArsR/SmtB family of metal-dependent transcription factors; in the apo-form and in the absence of *Sg*SrnQ, it can bind the DNA operator of *sodF*, up-regulating gene transcription. According to a recently proposed hypothesis, Ni(II) binding to *Sg*SrnQ would promote its interaction with *Sg*SrnR, causing the release of the complex from DNA and the consequent down-regulation of the *sodF* expression. *Sg*SrnQ is predicted to be highly disordered, thus the understanding, at the molecular level, of how the *Sg*SrnR/*Sg*SrnQ TCS specifically responds to Ni(II) requires the knowledge of the structural, dynamic, and functional features of *Sg*SrnR. These were investigated synergistically in this work using X-ray crystallography, NMR spectroscopy, atomistic molecular dynamics calculations, isothermal titration calorimetry and *in silico* molecular docking. The results reveal that the homodimeric apo-*Sg*SrnR binds to its operator in a two-step process that involves the more rigid globular portion of the protein and leaves its largely disordered regions available to possibly interact with the disordered *Sg*SrnQ in a Ni-dependent process.

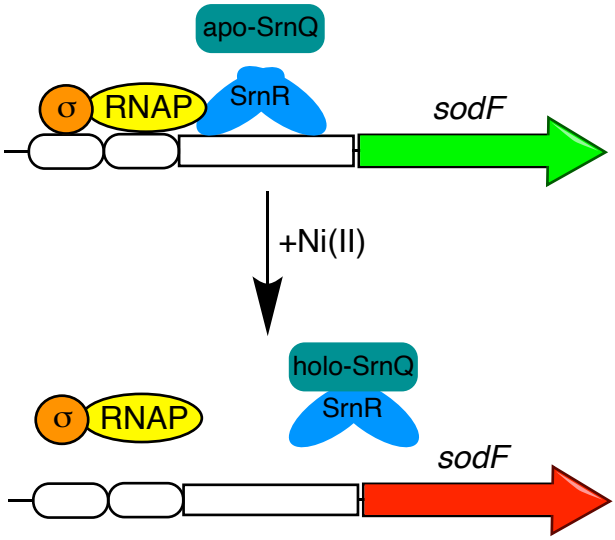
INTRODUCTION

About one fourth of all known proteins require metal ions as cofactors for their physiological function [1-2]. Due to their dual nature as both toxic and essential, the intracellular concentration of these elements is controlled by a tightly regulated homeostasis that involves specific membrane import and efflux pumps, as well as by cytoplasmic metallo-chaperones that deliver metal ions into their final subcellular destination, most often in the active site of enzymes. The expression of proteins involved in metal ion trafficking and utilization is regulated at the level of gene transcription by the coordinated network of specific metal sensors, whose action of repressing or activating genes in response to the concentration of specific cognate metal ions determines the composition of the intracellular metallome [2-4].

Seven main families of metallo-regulators have been described in bacteria [3], and four additional structural families contain some underrepresented metal sensors [4]. Among them, the family of metal-dependent ArsR/SmtB transcription factors is the most frequently found in the prokaryotic world, with members present in all bacterial taxonomy groups and with most bacterial genomes possessing at least one of these sequences [5-6]. The ArsR/SmtB members that have been structurally characterized show a common homodimeric fold, including at least five α-helices and a two-stranded antiparallel β-sheet [3, 6] connected by a β–turn between α4 and α5 (α1–α2–α3–α4–β1–β2–α5). Recognition and binding of an inverted repeated operator on DNA is performed by two symmetric winged helix-turn-helix (HTH, α3-turn-α4) motifs per dimer, with helix α4 directly contacting the DNA major groove. The additional three helices present in the structure are involved in hydrophobic interactions that orient the DNA binding motifs. Helices α1 and α5 form an orthogonal bundle that contributes to the dimerization [3].

The multiplicity of metal ions recognized by this class of proteins is reflected by the structural variety of the metal binding sites, despite the homologous global folds. Thirteen metal sensing motifs have been identified according to their position on the secondary structural elements, and divided into seven different groups according either to the position of the metal binding ligands or to the presence and identity of additional bound ligands, and further sub-divided into subclasses [7-8]. Structures of protein-operator complexes of ArsR/SmtB members indicate that they bind DNA as homodimers, with the HTH motifs placed symmetrically on two major grooves of the double helix to recognize a palindromic sequence [9-10]. Metal ion coordination in the regulatory site of metal sensors is allosterically transduced through the protein backbone, with a conformational change that modulates the protein affinity to DNA. This is well exemplified by the case of *Synechococcus* (*Sy*) SmtB, for which the crystal structures of the apo-protein and the metal-bound forms show that metal binding to the regulatory site compacts the homo-dimer altering the relative position of one subunit with respect to the other and changing the positions of the DNA recognition sites [11]. Analogously, a comparison between the Zn(II)-bound form of *Staphylococcus aureus* (*Sa*) CzrA and its apo-form bound to DNA reveals that in the latter complex, the protein exists in a "closed" state with a lower inter-protomer packing of the C-terminal region that allows the HTH motif to recognize and fasten the DNA operator.

Generally, ArsR/SmtB metal sensors function as transcriptional *repressors*, shielding the binding site of RNA-polymerase on DNA and consequently blocking the initiation of the transcription of genes encoding proteins that expel metal ions, chelate them, or change their oxidation state. Upon cognate metal binding, these regulators dissociate from DNA, de-repressing gene expression, thus reducing metal-derived cellular toxicity [7]. Recently, an exception to this rule was reported for the transcriptional regulator SrnR from *Streptomyces griseus* (*Sg*), which *in vitro* functions as a transcriptional *activator* despite belonging to the ArsR/SmtB family (Scheme 1) [12]. In this case, *Sg*SrnR bound to DNA recruits the RNA polymerase, either by direct interaction with the enzyme or by modifying the structure of the DNA to increase its accessibility for the transcriptional machinery (Scheme 1). A similar effect has been also suggested for *Sinorhizobium fredii* NolR, the global ArsR/SmtB regulator of the nodulation process [10]. *Sg*SrnR appears to operate in association with *Sg*SrnQ, a largely disordered protein that has been proposed to act as the Ni(II) component that modulates the *Sg*SrnR-DNA interaction [13]. *Sg*SrnR and *Sg*SrnQ form a two-component system (TCS) involved in the Ni(II)-dependent expression of *sodF*, a gene encoding a superoxide dismutase (SOD) that requires Fe(II) in its active site (Fe-SOD). This enzyme is antagonistically produced with SodN, a Ni(II)-dependent SOD (Ni-SOD). In the presence of Ni(II), the interplay between *Sg*SrnR and *Sg*SrnQ down-regulates the expression of Fe-SOD, thus promoting the activity of Ni-SOD (Scheme 1).



Scheme 1. Current hypothesis of the mechanism for the transcriptional regulation of *sodF* by the *Sg*SrnR-*Sg*SrnQ two-component system. RNAP = RNA polymerase; σ: sigma factor.

*Sg*SrnR activity as a DNA binder and transcriptional activator occurs independently of the presence of Ni(II) in solution, as observed by DNase footprinting and gene-reporter analysis [12]. Consistently, isothermal titration calorimetry experiments indicated that the protein binds Ni(II) with mild affinity (K_d ca. 16 μM), not compatible with the transcriptional response to Ni(II) observed *in vivo* [12]. In addition, its secondary structure and oligomeric state do not change in the presence of Ni(II), as proven by circular

dichroism and light-scattering [12]. These observations indicate that *Sg*SrnR alone is unlikely to act as a Ni(II) sensor. Instead, this role is likely played by *Sg*SrnQ. According to the most recently proposed hypothesis [12], Ni(II) sensing is performed by the cognate protein *Sg*SrnQ in a regulation network involving two different partners. Metal binding to *Sg*SrnQ would promote its interaction with *Sg*SrnR, causing the release of the complex from DNA, a decreased ability for the RNA polymerase to contact the SodF promoter, and the consequent down-regulation of the operon expression (Scheme 1). Thus, uniquely among all sensors belonging to the ArsR/SmtB family, the transcriptional regulation would not depend on the punctual binding of a metal ion or small molecule to a specific site on the DNA binding protein; rather, it appears to require a more extensive SrnR-SrnQ interaction that modulates the ability of *Sg*SrnR to bind DNA and to recruit the RNA polymerase.

The peculiarities of this system are likely reflected into the structural features of the transcription factor, as well as into its dynamical response to protein-protein interaction with its partner. Full understanding of the transcriptional process orchestrated by this TCS requires highly detailed structural and dynamic information on the two proteins involved. In the present work, a complementary study was carried out to determine the structural and dynamic features of *Sg*SrnR using solid state (X-ray crystallography) and solution (NMR) techniques as well as *in silico* modelling of the dynamics of the protein. The interaction of *Sg*SrnR with the double strand DNA operator of the *sodF* promoter (OP_{sodF}) was investigated using calorimetric techniques and NMR spectroscopy, while the structural determinants of the protein-DNA complex were explored using molecular docking. The results provide crucial information on the molecular framework at the basis of the function of this nickel-dependent expression modulator system.

MATERIALS AND METHODS

Protein preparation

Recombinant apo-SrnR from *Streptomyces griseus* (*Sg*SrnR) containing a GSH tail at the N-terminus (117 residues overall) was prepared as previously described [12]. Protein purity was verified using SDS-PAGE; the purified protein was devoid of metal ions as shown by inductively coupled plasma emission spectrometry (ICP-ES) as previously described [14]. The protein was stored at -80 ºC in 20 mM TrisHCl buffer at pH 7.5, containing 150 mM NaCl and 1 mM TCEP, and thawed prior to use.

Samples for NMR were prepared as single (^{15}N), double (^{13}C,^{15}N) and triple (^{2}H,^{13}C,^{15}N) labelled variants using the following protocol. Cells were grown in 2 L of LB at 37 °C. When the optical cell densities at 600 nm was ~ 0.6, the cells were centrifugated for 20 min at 7000 x *g* at room temperature. The cells were then resuspended in 500 mL of M9 minimal medium, containing ^{13}C or ^{2}H,^{13}C glucose for carbon or carbon/deuterium labelling, ^{15}N ammonium sulphate for nitrogen labelling and 70% of $^{2}D_2O$ for

deuteration. After an additional incubation of 30 min, protein expression was induced with 0.5 mM IPTG for 18 h at 26 °C. The protein was purified as previously reported [12].

Crystallization, X ray data collection and refinement

Protein crystallization was carried out at 293 K by using the micro-batch under oil technique in 96-wells MRC plates (Cambridge, UK) and the Clear Strategy Screen II-HT96 (Molecular Dimensions). Drops of 1 μL of *Sg*SrnR solutions (12.5 mg mL^{-1} in 20 mM TrisHCl PH 7.5, 150 mM NaCl, 1 mM TCEP, corresponding to 0.5 mM dimer) were added to 20 μL of volatile oil (Molecular Dimensions), immediately followed by 1 μL of precipitant. The crystallization wells were protected from drying using adhesive ClearView sheets (Molecular Dimensions). The best crystals of about 0.2 mm^3 appeared within four days in condition G6 (0.2 M calcium acetate hydrate, 0.1 M TrisHCl 8.5, 15 % w/v PEG 4000); crystals were cryoprotected by soaking them in a solution containing equal volumes of G6 crystallization mix and PEG 8000 50%, then fished out from the mother liquor by cryoloops and flash cooled into liquid nitrogen for storage.

Diffraction data were collected at 100 K using synchrotron X-ray radiation were recorded at the EMBL P13 beamline of the Petra III storage ring, c/o DESY, Hamburg (Germany) [15]. Data processing and reduction was carried out using XDS [16] and AIMLESS [17]. The crystal diffracted to 1.93 Å resolution with unit cell dimensions $a = b = 113.4$ Å, and $c = 124.9$ Å and belonged to space group $P6_222$. The asymmetric unit consisted of four *Sg*SrnR molecules giving a solvent content of 53.68%.

The structure of *Sg*SrnR was determined by molecular replacement using the program Phaser [18] and the region comprising residues 26-90 of the crystal structure of the possible transcriptional regulator for arsenical resistance (PDB code: 3F6V) as the search model. Initial model was automatically built using the program PHENIX Phase and Build refined using TLS refinement against experimental data by using REFMAC [19]. Visual inspection, as well as manual model building and addition of solvent molecules, were carried out using COOT [20-21]. The refinement converged to a final R_{factor} and R_{free} were 17.8% and 21.7%. The stereochemistry of the final model was routinely checked using COOT [20-21] and PROCHECK [22]. The final crystallographic model and structure factor amplitudes were deposited in the Protein Data Bank with the accession code 7P6F. Details for data collection and refinement statistics are reported in Table 1-SI. Figures were generated using PyMol (The PyMol Molecular Graphics System, v. 1.8 Schrödinger, LLC.), and Chimera X [23-24].

NMR Backbone Resonance Assignment

NMR experiments were performed using ca. 0.5 mM dimer of triply labelled apo-*Sg*SrnR in 20 mM TrisHCl buffer at pH 7.5, containing 150 mM NaCl and 1 mM TCEP containing 5% D_2O, at 298 K. All experiments were performed on a Bruker AVANCE III spectrometer operating at 18.8 T (799.67 MHz ^1H Larmor frequency), equipped with 5 mm TCI z-gradient cryo-probe. Salt-tolerant susceptibility matched slot NMR tubes (Shigemi Inc.) were used to improve the signal-to-noise ratio during NMR data collection. Proton chemical shifts were referenced to 2,2-dimethyl-2-silapentane-5-sulfonic acid

sodium salt (DSS), while the ^{13}C and ^{15}N chemical shifts were referenced indirectly to DSS, using the ratios of the gyromagnetic constants.

The backbone and side chains Cβ nuclei were assigned using 3D HNCO, HN(CA)CO, HNCA, HN(CO)CA and HNCACB spectra, as well as 4D HNCOCA and HNCACO spectra (Table 2-SI). These spectra were processed using ToASTD [25]. In the case of NUS spectra, cleaner3d and cleaner4d with Signal Separation Algorithm reconstruction were used [26]. Sequence-specific assignment was carried out manually using UCSF Sparky [27]. Overall, 95% of Cα, 84% of Cβ and 91% of CO carbons chemical shifts were successfully assigned. The assignment was deposited in the Biological Magnetic Resonance Bank (BMRB) with the accession code 50753. The interaction of *Sg*SrnR with the double strand operator of *sodF* (OP_{sodF}) was investigated by obtaining ^{1}H,^{15}N TROSY-HSQC spectra of the apo-protein in the presence of one equivalent of the DNA fragment.

Protein dynamics by ^{15}N NMR spectroscopy

The experiments for the determination of ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates, and of the ^{1}H-^{15}N cross-relaxation rate measured via steady-state heteronuclear ^{1}H-^{15}N NOE, were acquired at 298 K on a Bruker AVANCE NEO spectrometer operating at 16.4 T (700.13 MHz ^{1}H Larmor frequency) equipped with a 5 mm TCI z-gradient cryo-probe. Samples of ^{15}N-labelled apo-*Sg*SrnR (0.85 mM) in NMR buffer containing 10% D_2O were utilized. Shaped NMR tubes (Bruker BioSpin AG) were used to improve the signal-to-noise ratio during NMR data collection. Spectra were processed using Topspin 4.0.3 (Bruker BioSpin) and peak intensities were analysed using Dynamics Center 2.7.1 (Bruker BioSpin). The details of spectra acquisition, processing and analysis are provided in the Supplementary Information.

Molecular dynamics simulations

For each of the *Sg*SrnR dimers that can be reconstructed from the crystallographic asymmetric unit (see Results below: namely AB, CC' and DD' hereafter) the residues not visible in the crystal structure were added using the software Modeller 10.0 [28] and using the most complete *Sg*SrnR monomer as template. The first three residues and residues 108-110 at the C-terminus, not visible in the crystal structure, were modelled through a standard loop optimization procedure. The last three residues at the C-term of the *Sg*SrnR sequence were not included in the models. The most probable protonation states of titratable amino acids and the tautomeric state of histidine residues at pH 7.2 were assigned using the H++ 3.2 server [29-31]. The protein was embedded into a truncated octahedron water box using a 10-Å buffer zone of solvent. The resulting systems consisted of ca. 53,700, 56,600 and 54,000 atoms for *Sg*SrnR AB, CC' and DD' dimers, respectively. The Amber ff14SB force field [32] for the protein and the TIP3P model [33] for water were used. The Na^+ ion bound to each monomer and found in the crystal structure was included in the system preparation. Each system was neutralized by adding 4 Cl^- ions using the genion program of the GROMACS 2020.1 package [34-35] Analogously, additional Na^+ and Cl^- ions were placed in the water box to achieve a physiological ionic strength (200 mM). The system was energy-minimized

and then equilibrated at 300 K and 1 atm by performing 1 ns of gradual annealing using GROMACS 2020.1. The geometry optimization was performed in four cycles. In the first two cycles, which comprised 800 steps of steepest descent followed by 200 steps of conjugate gradient, the water molecules were relaxed while the position of the protein heavy atoms were constrained using a harmonic potential with a force constant of 1,000 J mol^{-1} $Å^{-2}$. In the third and in the fourth cycles the procedure was repeated without applying any constraint. During this equilibration phase, positional constraints were applied on the protein heavy atoms (force constant of 1,000 J mol^{-1} $Å^{-2}$). Temperature and pressure were controlled using a Berendsen thermostat and barostat [36], respectively. An integration step of 2 fs was used, and the structures were sampled every 0.1 ps. LINCS constraints [37] were applied on the hydrogen-involved covalent bonds. Periodic boundary conditions were applied. The Particle Mesh Ewald method was used to calculate electrostatic interactions [38]. The cut-off values for the real part of the electrostatic interactions and for the van der Waals interactions were set to 9 Å. During the 100 ns-long molecular dynamics (MD) production runs, the temperature and pressure coupling was made using a v-rescale thermostat [39] and a Parrinello-Raman barostat [40-41], respectively. Clustering analysis was performed using the cluster module of GROMACS, using the Gromos algorithm [42]. A 0.15 nm cut-off for the RMSD was used to include structures in the same cluster.

Isothermal titration calorimetry

Binding of *Sg*SrnR to the double strand DNA operator of *sodF* (OP_{sodF}) was investigated at 25 °C using a high-sensitivity VP-ITC microcalorimeter (MicroCal). The protein (13 μM dimer in 20 mM TrisHCl pH 7.5, 150 mM NaCl, 1 mM TCEP) was loaded into the sample cell (1.4093 mL) and was titrated with 22 × 10 μL injections of a solution containing 140 μM OP_{sodF}, dissolved in the same buffer, using a computer-controlled 310-μL microsyringe. Heat of dilution of DNA into the buffer was verified to be negligible by control experiments. Integrated heat data were fitted using a non-linear least-square minimization algorithm to a theoretical curve corresponding to a two sets of sites model and processed using the Origin 7.0 software provided by the manufacturer. ΔH (reaction enthalpy change in cal mol^{-1}), K_A (binding affinity constant in M^{-1}) and n (number of binding sites) were the fitting parameters. The Chi-square parameter χ^2 was used to establish the best fit. The reaction entropy was calculated using the equations: $\Delta G = -RT \ln K_A$ (R=1.9872 cal mol^{-1} K^{-1}, T=298 K) and $\Delta G = \Delta H - T\Delta S$.

Protein-DNA docking

The most representative structure of each of the four more populated clusters obtained from the MD calculations were used as *Sg*SrnR starting structure for the molecular docking. A starting model for the unbound operator of *sodF* (OP_{sodF}) was generated using the DNA analysis and rebuilding software x3DNA-DSSR (http://x3dna.org/) [43-44]. OP_{sodF} comprises nucleotides from −15 to +27 with respect to the *sodF* operon transcriptional start site in *S. griseus*. To avoid biasing effects due to the highly charged DNA termini, two and three nucleotides were added respectively at the 5' and 3' side of the operator using the *S. griseus* genome. In this way, on each side of the inverted repeat sequence proposed by Kim *et al.* [13] there are fifteen nucleotides. The model was generated in the canonical B-DNA conformation.

*Sg*SrnR was docked onto OP_{sodF} using the data-driven docking program HADDOCK 2.2 [45-46] and a previously described protocol [47-48] that involves a two-stage protein–DNA docking approach [49]. In the first docking round, a rigid body energy minimization was carried out, 1000 structures were calculated, and the 200 best solutions based on the intermolecular energy were used for a semiflexible, simulated annealing step followed by an explicit water refinement on the same docked poses used for the second step. The calculation was guided by selecting *Sg*SrnR residues corresponding to those involved in the interaction with DNA in the homologous protein *Staphylococcus aureus* CzrA [9] (*Sg*SrnR Ser50, Arg53 and His58), as well as the inverted repeat sequence (from –2 to +15 with respect to the *sodF* operon transcriptional start site in *S. griseus*) [12]. The docking algorithm rewards the complexes that have these so-called "active" protein residues or DNA nucleotides at the interaction interface [45-46]. A second set of "passive" protein residues (Asp20, Thr22, Arg23, Iso42, Ser47, Pro49, Ser52, Gly56, and Val57), as well as "passive" DNA nucleotides (from to –15 to +27 with respect to the *sodF* operon transcriptional start site in *S. griseus*), located in the vicinity of the "active" residues or nucleotides, was also included in the calculation. The experimental information is thus translated in the docking process to ambiguous interaction restraints (AIRs) that are used to drive the docking process. An AIR is defined as an ambiguous intermolecular distance with a maximum value of 3 Å between any atom of an active residue of the biomolecule A (*Sg*SrnR in the present case) and any atom of both active and passive residues of the biomolecule B (the DNA in the present case) [45-46]. Additional restraints were introduced for the DNA fragment to maintain base planarity and Watson–Crick bonds. The 200 models thus refined were clustered using a cut-off of 7.5 Å based on the pairwise backbone root mean square deviation matrix. Subsequently, the DNA conformation in the docked resulting structures was analysed using the program 3D-DART [50] to determine trends in DNA bending and twisting, a type of information that was used to generate an ensemble of custom DNA models representing the accessible conformations, using a local version of the program 3D-DART (https://github.com/haddocking/3D-DART). A second HADDOCK docking round was then carried out following the same approach as described for the first round, but this time including the ensemble of DNA models generated above. In this round, the conformational freedom of the DNA molecule was restricted at the semi-flexible refinement stage to prevent helical deformation.

RESULTS

X-ray crystallography

The crystal structure of *Sg*SrnR was obtained and refined at 1.93 Å resolution using synchrotron radiation X-ray diffraction data collected on a single crystal at cryogenic conditions. The structure reveals that the asymmetric unit of the crystal contains four *Sg*SrnR monomers, namely A, B, C, and D, related by non-crystallographic 2-fold axes (Figure 1A). This arrangement is consistent with a dimeric

oligomerization of *Sg*SrnR in the solid state, where monomers A and B form a dimer within the same asymmetric unit, while monomers C and D dimerize with a C' and D' monomers, respectively, belonging to adjacent asymmetric units. The approximate dimensions for the dimer are 70 x 50 x 35 Å and the interface area calculated by PDBePISA server (https://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver) is ~1300-1400 Å^2 per monomer.

The electron density is well defined for residues 5 to 107 for monomers A and C, and 4 to 104 and 7 to 103 for monomers B and D, respectively (Figure 1B for a representative portion of the electron density). *Sg*SrnR shows the typical ArsR repressor folding, encompassing residues 7–93, and containing five α-helices (α1: residues 7-18, α2: residues 21-32, α3: residues 37-43, α4: residues 48-60, and α5: residues 79-93) and two β-strands (β1: residues 64-69, β2: residues 72-77) to give an overall α1-α2-α3-α4-β1-β2-α5 fold, as found for other ArsR folds (Figure 1C). In *Sg*SrnR, the C-terminal consists of a long unstructured and mobile portion, which contains, in the case of monomers A and B, a short β-strand (β3: residues 98-100).

In this topology, helices α3 and α4 from each monomer form the helix-turn-helix (HTH) motif known to be responsible for the DNA binding within the major groove in analogous ArsR/SmtB transcription factors [3]. The correct positioning of the HTH motif is ensured by the compact scaffold provided by helices α1, α2 and α5. Additionally, α1 and α5 helices from one monomer are nearby and anti-parallel to the two-fold symmetry related helices from the other monomer, with these four secondary structure elements providing stabilization of the dimer. Remarkably, the topological orientation of α5 with respect to α4 is significantly different as compared with most other members of the ArsR/SmtB (Figure 1-SI). In the current structure (Figure 1-SI A), an obtuse angle between the two helices is observed. A similar conformation has been previously reported for two ArsR crystal structures (Figure 1-SI B) [51]. Differently, in most of the structures of this protein family deposited in the PDB, represented by the structure of *Sy*SmtB in Figure 1-SI C, the two helices form an acute angle. While the ArsR proteins were modified by the addition of a C-terminal His tag, implying the possibility of an artefact in the protein topology due to the primary structure variation, in the case of *Sg*SrnR the GSH sequence left by the cloning procedure is positioned at the N-terminus, thus excluding that the observed topology is modified by a cloning artefact at the C-terminus.

The β1 and β2 strands of each monomer, which are positioned in the sequence following the HTH motif and that are spaced by a two-residue turn (Ala70 and Asn71), form an intramolecular antiparallel β-sheet showing a hairpin structural motif. Residues comprised in this region have the highest B-factors in the molecule (Figure 1C), indicating considerable mobility (except for chain C, where the hairpin is blocked by crystal packing and therefore the B-factor values for its residues are lower). The additional β3 strand located on monomers A and B form a short intermolecular antiparallel β-sheet also contributing to the dimer association.

A ConSurf analysis (https://consurf.tau.ac.il/) was carried out to estimate the evolutionary conservation of the amino acid sequence of *Sg*SrnR (Figure 2A). The results show an overall high conservation for residues belonging to the first half of the protein (helices α1-α4). Three stretches of highly conserved residues are visible: the first is located on the α1-α2 connecting region covering residues Ala 17-Arg24, with three residues (Ala17, Val18 and Ala19) being hydrophobic and belonging to the terminal portion of helix α1, in a region located at the monomer-monomer interface of the *Sg*SrnR dimer; the two following residues, Asp20 and Pro21, form the connection between helices α1 and α2, while the last three residues of the first conserved stretch are Thr22, Arg23 and Arg24, which point towards the bulk solvent and are possibly involved in DNA-binding. The second conserved amino acid stretch is located on the N-terminal portion of the α4 helix (Ser 47-Leu55) and consists of three hydrophobic residues (Ala48, Ile51 and Leu55) interspersed with three Ser residues (Ser47, Ser50, and Ser52), and the polar residues Arg53 and His54; except for Ser52, the other non-hydrophobic residues point towards the bulk solvent. The third conserved stretch corresponds to the fully hydrophobic region connecting the α4 helix with the β1 strand, and consists of residues Gly60, Ala61, Gly62, Leu63, and Val64. The noticeable hydrophobic-rich environment present in the conserved regions continues intermittently between these three main stretches. Indeed, except for Glu34, the α-helical secondary structure elements show the presence of single highly conserved hydrophobic residues (Ile26, Leu27, Leu30, Ala37, Ile40, Ala41, Leu58) positioned every *ca*. 4 residues along the α-helices backbone. This arrangement gives rise to hydrophobic patches on each α-helix that are involved in the constitution of a hydrophobic core, providing a scaffold to correctly positioning the HTH motif. The second half of the protein moiety displays a global lower residue conservation. Significantly conserved residues are Tyr75, located on the β2 strand, and Pro95, positioned at the end of helix α5 and at the beginning of the unstructured and mobile C-terminal region.

An analysis of the electrostatic potential, performed by DelPhi [52-53], highlights the presence of two positively charged regions (Figure 2B): the first is located in the dimerization cleft originated by the α1-α4 helices, and is due to the presence of positive side chains on Arg16 (α1), Arg23, Arg24 and Arg31 (α2), Arg43 (α3), Arg53 and His54 (α4), with Arg23, Arg24, Arg53 and His54 being highly conserved; the second is located in the C-terminal portion of the protein, where the presence of additional several Arg residues is observed. These regions suggest possible interaction patches involved in DNA recognition.

Additional electron density was found in the vicinity of the protein surface at the end of the β2 strand and the beginning of the α5 helix in the A, B, and D monomers. This density was modelled with an Na^+ ion bound to the carboxylate Oε1 atom of Glu79 and the carbonylic O atom of Leu77 and to water molecules completing a pseudo-octahedral coordination geometry. In monomer C, a weaker electron density was found in the same position and was modelled as a water molecule even though the presence of a less occupied Na^+ ion cannot be ruled out.

NMR spectroscopy

The structural features thus established for *Sg*SrnR were then complemented with solution properties investigated using high-resolution protein NMR spectroscopy. The solution $^{1}H,^{15}N$ TROSY-HSQC spectrum of *Sg*SrnR is shown in Figure 3. The signal assignment of a total of 107 out of 111 cross-peaks of backbone amide groups was carried out using 3D and 4D resonance NMR experiments. The signals of the five proline residues are not observable. The four unassigned resonances include the N-terminal Gly-Ser-His extension and Glu82 (Glu79 in the native sequence): the latter residue is observed binding a Na^{+} ion in the crystal structure, which could undergo exchange phenomena that broaden the signals of this residue beyond detection.

Prediction of the protein secondary structure performed by TalosN [54] using the obtained chemical shifts (Figure 4A) revealed that *Sg*SrnR in solution is largely folded and consists of both α-helices (α1: residues 8-17; α2: 23-29; α3: 39-41; α4: 49-59; α5: 79-85; α6: 89-92) and β-strands (β1: residue 35-36; β2: 65-68; β3: 73-77; β4: 98-101). These regions largely correspond to those identified in the solid state by crystallography, with the addition of a short strand between α2 and α3 that extends the β-sheet comprising also the β hairpin. Amplitudes of motions in the ps–ns time scale detected using the Random Coil Index (RCI) method [55], based on chemical shift analysis and referred by TalosN as the S^{2} order parameter, suggest that both the N–terminus and, more significantly, the C–terminus are subjected to motions in this time scale, as indicated by lower order parameters (Figure 4A). The disordered nature of the C-terminus is further corroborated by the elevated intensities observed for the signals corresponding to residues in this region (Figure 4B). The presence of significant disorder in these protein portions is consistent with the predictions made by disorder predictors using the D2P2 web server (http://d2p2.pro/), which also recognizes the presence of a folded DNA binding domain in the central part of the protein (Figure 5) [56].

The results of the structural analyses of the NMR chemical shifts described above prompted us to investigate the solution protein dynamics of *Sg*SrnR by measuring the ^{15}N relaxation rates R_1 (Figure 2A-SI) and R_2 (Figure 2B-SI) as well as the ^{1}H-^{15}N heteronuclear NOE values (Figure 2C-SI) of all assigned backbone amide groups of *Sg*SrnR (see Supplementary Information for details). The presence of local internal motions in the ps–ns time scales is expected to contribute to the R_1, R_2 and NOE values, with NOEs being more sensitive to ultrafast internal dynamics than R_1 and R_2 [57], while conformational exchange processes occurring on the μs–ms time scale additionally contribute to increase the R_2 rates [58].

A rotational correlation time τ_m = 17.1± 0.9 ns was initially determined on the basis of R_1 and R_2 values; this value corresponds to a molecular mass of 28.5 ± 1.5 kDa estimated using the empirical relationship τ_m (ns) ~ 0.6 kDa for folded proteins [59], supporting the presence of the homodimer of *Sg*SrnR in solution under the experimental conditions used, in agreement with light scattering data [12]. A qualitative analysis of the relaxation data for *Sg*SrnR indicates that relatively large NOE values are

generally observed in all protein regions predicted as helix or strand fragments by the chemical shift analysis (Figure 2C-SI), while smaller NOE values are observed for all other regions, especially in the C-terminal portion of the protein, which features large and negative NOE values indicating greater mobility in the sub-nanosecond time range. This is consistent with the disorder observed also in the solid state. A similar behaviour is observed for R_1 (Figure 2A-SI) which additionally features a peculiar increase in the 100-110 region followed by a decrease in the last portion of the C-terminus, indicative of a further increase in the motion frequency that decreases the efficiency of the longitudinal relaxation while contributing largely to the decrease of the NOE values. The values of R_2 (Figure 2B-SI) are found to be generally more uniform throughout the amino acid sequence, with a pronounced decrease in the C-terminal region, again consistently with the ensuing increase in the motion frequency in this portion of the protein. This is again coherent with the large disorder observed in the solid-state crystal structure. The relaxation data, quantitatively analysed using the reduced spectral density mapping approach [60-66] (see Figure 2-SI and Supplementary Information for details) further corroborates the presence of a stable and relatively rigid protein fold, with the exception of the N- and C-termini, which show internal motions faster than the ns time scale, the absence of slow (ms) exchange phenomena, the presence of internal dynamics in the sub-ns time scale, and even faster dynamics, in the ps time scale, for the final portions of the sequence.

Atomistic molecular dynamics calculations

To gain a deeper understanding of the dynamic behaviour of *Sg*SrnR, the mobility features of *Sg*SrnR determined by NMR spectroscopy were further probed by atomistic molecular dynamics calculations in explicit solvent. Three 100 ns-long MD simulation in explicit solvent were carried out using an atomistic force field and starting from the three dimers derived from the asymmetric unit of the crystal. In all the dimers (AB, CC' and DD'), the three N-terminal residues and the C-terminal residues up to residue 110 – that were not solved in the crystal structure – were added to the structure through homology modelling. The root mean square deviation (RMSD) of the Cα atoms of the whole protein and of both monomers – excluding the N-terminal and C-terminal residues that were not present in the crystal structure – appears to be converged (Figure 3-SI) at values close to 0.2 nm after few ns of simulation time. Only in one case, one monomer shows RMSD values at ca. 0.4 nm after ca. 55 ns of simulation time. This is due to a partial unfolding in the initial portion of helix α1 of monomer A that has not been observed in the case of other monomers. The unstructured N- and C-terminal unstructured regions are extremely mobile, as confirmed by the RMSD of the protein calculated also considering these regions and by the root mean square fluctuations (RMSF) of both monomers (Figure 4-SI). The remaining parts of the protein fluctuate between 0.1 and 0.2 nm with the largest values recorder for residues 32 (C-terminal of α-helix α3), 40-48 (loop between loop α-helices α3 and α4), 68-72 (loop between β-strands β1 and β2) and 88-90 (N-terminal part of α-helix α5) (Figure 2C). In general, the consistency of the RMSD values for the

dimer and both monomers can be ascribed to a structural stability of the dimer in the hundreds of ns time scale.

The three calculated trajectories were then summed to increase the sampling of the conformational space. The clustering of the summed trajectories done on the dimeric *Sg*SrnR with the exclusion of the mobile N- and C-terminals revealed the presence of four clusters accounting for ca. 80% of the total frames (Figures 5-SI and 6-SI). The representative structures of the four most populated clusters were used as input for the subsequent protein-DNA docking stage (see below). Motion correlations between various subparts of the protein can be identified by a calculation of the covariance matrix of the amino acid displacements. Visual inspection of the corresponding map (Figure 7-SI) suggests that the motion of the C-terminal regions (β-strands β2 and β3 separated by α-helix α5) of both monomers is correlated, while the motion of the central part of the protein (α-helices α3 and α4) is anticorrelated with the C-terminal region described before. These correlated/anticorrelated motions are relevant especially for monomers A, C, and D, while are less visible for monomers B, C' and D'. Here, the basis for the functional characterization of *Sg*SrnR were determined by experimentally investigating protein-DNA binding using calorimetry.

Protein-DNA interaction by isothermal titration calorimetry and NMR spectroscopy

Isothermal titration calorimetry (ITC) experiments were carried out by titrating the double strand DNA operator of *sodF* (OP_{sodF}) into a solution containing *Sg*SrnR protein. The sequence used (TGTTAGCCTGCTCTTGCATATAGCTTGCAATAACAACTGGACG), containing an inverted repeat motif (underlined) previously suggested to have a role in *sodF* transcriptional regulation [13], was chosen including the base pairs from -15 to + 27 with respect to transcription start site, protected by *Sg*SrnR in DNase I footprinting experiments [12].

The binding thermogram shows large endothermic peaks following each injection at the beginning of the titration (Figure 6A). As the titration proceeds, exothermic peaks appear, indicating the presence of at least two different events, with opposite enthalpy of binding, occurring upon DNA addition to the protein solution. The best fit of the binding isotherm calculated from peak integrations (Figure 6B) could be obtained using a model involving two sets of binding sites, both showing a half-integer stoichiometry. This can be explained by considering the dimeric nature of *Sg*SrnR, with one monomer that may initially recognize one DNA hemi-operator with higher affinity (K_{D1} = 80 ± 10 nM), followed by a second event (occurring with a lower equilibrium constant, K_{D2} = 1.0 ± 0.2 μM) that completes the formation of the homodimeric protein-DNA complex through the interaction of the second monomer to the other half of inverted-repeated sequence. The thermodynamic parameters obtained from the fit indicated that the first higher affinity event is largely entropy-driven (ΔH_1 = + 38.17 ± 0.06 kcal mol^{-1}, ΔS_1 = + 160 kcal $mol^{-1} K^{-1}$) consistently with the formation of a protein-DNA complex accompanied by release of water molecules into the bulk, while the second lower affinity binding is entropically disfavoured and enthalpy-driven

(ΔH_2 = - 11.72 ± 0.09 kcal mol^{-1}, ΔS_2 = -11.9 kcal $mol^{-1} K^{-1}$), which is compatible with a conformational change that decreases the disorder of the system occurring when the protein completes the DNA binding.

The $^{1}H,^{15}N$ TROSY-HSQC spectrum of the *Sg*SrnR - OP_{sodF} complex (Figure 6C) is characterized by the disappearance of all signals corresponding to residues located in the well-folded portion of the protein (compare with the spectrum of the apo protein, Figure 8-SI). This is ascribed to the formation of a larger protein-DNA complex, with a slower overall rotational correlation time that leads to faster relaxation and decrease of signal intensities beyond detection. A site-specific analysis of the interaction site was therefore impossible. However, the significant presence, in the spectrum of the complex, of the NMR signals of residues belonging to the N-terminus (residues 5-9) and the C-terminus (residues 110-117) clearly indicates that these disordered terminal regions of SrnR maintain their large mobility in the complex and are thus minimally involved in the interaction of the protein with OP_{sodF}.

Protein-DNA docking

The experimental data for the protein-DNA interaction were then validated using an unbiased computational molecular docking study to calculate a model for the interaction between *Sg*SrnR and the OP_{sodF}. The calculations were performed using a two-steps knowledge-based docking approach [47,49,69] that allows both to generate docking poses in agreement with experimental data and bioinformatics predictions as well as to adapt the DNA structure to the protein structure during the docking procedure. In the absence of any direct structural information on the *Sg*SrnR protein-DNA interaction, v
the interacting residues from the proposed model reported for *Staphylococcus aureus* Czr[illegible]
from NMR data of the DNA bound apo-protein. On the DNA side, the inverted repeat sequence (from −2 to +15 with respect to the *sodF* operon transcriptional start site in *S. griseus*) found on OP_{sodF} was used. The results of the docking, as well as the DNA sequence used, are reported in Figure 7.

According to the calculated structural model, *Sg*SrnR interacts with the OP_{sodF} inverted repeat sequence by inserting α-helix α4 in the major groove and by interacting with the DNA backbone though the C-terminal part of α-helix α2 (Thr22 and Arg23). Interestingly, the latter residues were not used to guide the calculation. The DNA major groove appears to be slightly deformed in the *Sg*SrnR interacting region, to allow the insertion of α-helix α4. In particular, the interaction in this region is stabilized by the presence of Arg53, which inserts its positively charged side chain in the major groove and is in contact with the nitrogenous bases at its bottom. In the calculated model, the disordered regions at the N- and C-termini appear not to be involved in the formation of the protein-DNA complex, a conclusion supported by the NMR-based evidence.

DISCUSSION

Streptomyces such as *S. griseus* are the major producers of all known antibacterial drugs, with over two-thirds of the clinically useful antibiotics of natural origin obtained from this source; they are thus considered a promising resource for the war against multi-drug resistant pathogens [70]. In addition, *Streptomyces* have possible applications in bioremediation, especially for phytoextraction processes of metal ions, as they are often associated to hyper-accumulating plants [71-73]. The production of secondary metabolites, as well as the acquisition of a metal resistant phenotype, generally involves specific gene clusters [74-75], that, in these bacteria, are often regulated by two-components systems (TCS) [76]. Therefore, the understanding, at the molecular level, of how the *Sg*SrnR/*Sg*SrnQ TCS specifically responds to its Ni(II) cofactor is crucial, both because it is a regulation system belonging to an important bacterial genus, and because this system is the only known TCS able to regulate Ni(II)-dependent expression [77], representing therefore a paradigmatic example of transcriptional regulation of the intracellular homeostasis of this metal ion. The physiological function of *Sg*SrnR as a transcriptional regulator in the Ni(II)-dependent TCS that controls superoxide dismutase expression requires extensive structural and dynamical information on the protein both in the absence and in the presence of its DNA operator.

In the present work, we have obtained highly detailed structural data on this TCS, using a combination of independent techniques, namely X-ray crystallography, NMR spectroscopy, calorimetry, atomistic molecular dynamics simulations and biocomputational modelling. The results provide a congruent description of the structure of the dimeric protein, confirming that its core adopts an ArsR/SmtB-like fold, with a conserved HTH DNA binding motif and an unusual topology. On the other hand, the N- and the C-termini possess flexible extensions, as consistently derived from disorder predictions, X-ray crystallography, and NMR spectroscopy.

One dimeric unit of *Sg*SrnR appears to form a complex with its operator in a two-step process, as resulted by ITC experiments, in which the initial tight interaction is made with a monomer, followed by a clamping of the DNA using the second monomer, in a less favourable equilibrium. A similar two-steps binding mode, showing an initial protein-DNA interaction followed by protein conformational rearrangement that results in high affinity DNA binding, has been proposed for the Ni(II)-sensor *Helicobacter pylori* (*Hp*) NikR [67]. A two-step DNA-binding event, both enthalpically and entropically driven as measured by ITC, was also observed for the transcriptional regulator *Sa*CzrA, which however presented a different stoichiometry with two protein dimers that bind one DNA molecule; in that case, binding of the first dimer occurs with K_{D1} = 7.7 pM, while the second event occurs with lower affinity (K_{D2} = 1.6 nM) [68]. Modelling calculations indicated the viability of the contact between the α_4 helix that belong to the HTH motif and the inverted repeated sequence, previously identified as having role for *sodF* regulation. The observation that only the structured globular portion of the protein is involved in the formation of its complex with DNA, leaving the unstructured terminal regions free, was also supported by the *in-silico* docking.

A different situation was previously observed for *Mycobacterium tuberculosis (Mt)* NmtR, a Ni(II)-repressor of the ArsR/SmtB family that features both the N-terminal and the C-terminal regions unstructured in solution; in that case, the N-terminal sequence was suggested to be involved in direct DNA binding and allosteric regulation for metal-driven transcriptional de-repression [78-79]. In particular, the His3 residue in the N-terminal disordered region of *Mt*NmtR was shown to be involved in Ni(II) binding, with the N-terminus functioning as an "arm" that opens and closes when the metal ion is bound to the protein. Ni(II) binding to *Mt*NmtR induces dynamic disorder on the μs - ms time scale of key DNA interacting regions, which likely impairs the ability of the protein to bind DNA when bound to the cognate metal ion [78]. Notably, His3 mutation affects *Mt*NmtR Ni(II) selectivity, as the mutated protein becomes responsive to Zn(II) in vitro, suggesting a functional role for the flexible regions of the protein, which includes direct DNA binding and allosteric regulation [79].

The functional dynamics of several ArsR/SmtB proteins has been proven to be the basis for the metal-driven allosteric modulation of conformational changes that lead to the formation (or rupture) of protein-DNA complexes. In the case of *Sa*CzrA, minimal structural rearrangements upon metal binding [7] are contrasted by significant modifications of the fast dynamic motions that perturb the entropic contribution to DNA binding, eventually impairing the ability of the holo-protein to bind DNA; in this case, the allosteric regulation driven by metal binding derives from the ability of the Zn(II) ion to change the conformational equilibria, rendering some conformational states less accessible with an impact on DNA binding [80]. Analogously, solution NMR studies of the apo and metal-bound forms of the Cd(II)-sensor *Mt*CmtR indicate that binding of the metal ion to the regulatory sites reduces conformational heterogeneity, thus decreasing the number of protein conformations available for DNA selective interaction [81]. In the case of *Hp*NikR, a pleiotropic nickel-sensing transcription factor that regulates the bioavailability of this element in the cell, Ni(II) binding induces conformational and dynamic changes associated with nickel-activated DNA complex formation; in particular, higher levels of dynamics are observed for the apo-protein as shown by ^{19}F NMR spectroscopy, while in the holo form of *Hp*NikR the mobility is decreased and the DNA-binding conformation is more favoured, so that the allosteric mechanism of Ni(II)-activated DNA binding by *Hp*NikR is driven by conformational selection [82].

*Sg*SrnR was reported to bind a single Ni(II) ion with moderate affinity (K_d ca. 16 μM) [12] but this event was proven by NMR to involve the non-native GSH tag at the N-terminus (not shown). Consistently, *Sg*SrnR is not regulated by a metallic cofactor binding [12] but rather by the interaction with the cognate protein *Sg*SrnQ [13]. Therefore, we suggest that the intrinsic disorder of the terminal arms is a driver for protein-protein interactions that involve disorder-to-order transitions. *Sg*SrnQ is predicted to be largely disordered, with two expected disorder-based binding sites potentially involved in the interaction with *Sg*SrnR [83]. In addition, the terminal arms of *Sg*SrnR might directly contact the RNA polymerase, driving the enzyme close to the promoter region, thus fostering transcriptional activation.

It is unknown yet how the availability of Ni(II) ions is transduced into the variation of *Sg*SrnR DNA binding properties, as well as how the peculiarity of this transcriptional regulator, which, uniquely

among the family, functions as an activator and is part of a TCS, is reflected in specific structural and dynamical features. The current accepted hypothesis is that, in the absence of high affinity Ni(II) binding for *Sg*SrnR, this function requires the presence of the cognate protein *Sg*SrnQ, which acts as a Ni(II) sensor. Efforts are underway to obtain the *Sg*SrnR-*Sg*SrnQ complex, both in the presence and in the absence of Ni(II), in order to complete the full picture of the regulation by this paradigmatic Ni(II)-dependent TCS.

ACKNOWLEDGEMENTS

This research was supported by CIRMMP (Consorzio Interuniversitario di Risonanze Magnetiche di Metallo Proteine), by the Polish National Science Centre with MAESTRO Grant No. 2015/18/A/ST4/00270, and by the University of Bologna. X-ray diffraction data were collected at the PETRA III storage ring operated by EMBL Hamburg (DESY, Hamburg, Germany; beam time award number MX-720). NMR data for protein dynamics were collected at the Centre for Magnetic Resonance (CERM) of the University of Florence (Italy); Massimo Lucci and Fabio Calogiuri are duly thanked for their support in NMR data collection. Giulia Pesce is thanked for her help in the first attempts for crystallization setup. The calculated model of the SrnR - OP_{sodF} complex is freely available at the address https://site.unibo.it/bioinorgchem/en/downloads.

Figure 1. (**A**) Ribbon representation of the four *Sg*SrnR monomers in the asymmetric unit, coloured according to each monomer; (**B**) representative portion of the $2F_o$-F_c electron density map contoured at 1σ, in the 54-59 residue range; (**C**) ribbon representation of the dimeric biological unit, coloured according to the B-factor (Å^2) (red, B-factor > 75; white, 25 < B-factor < 75; blue, B-factor < 25); the elements of secondary structure are indicated.

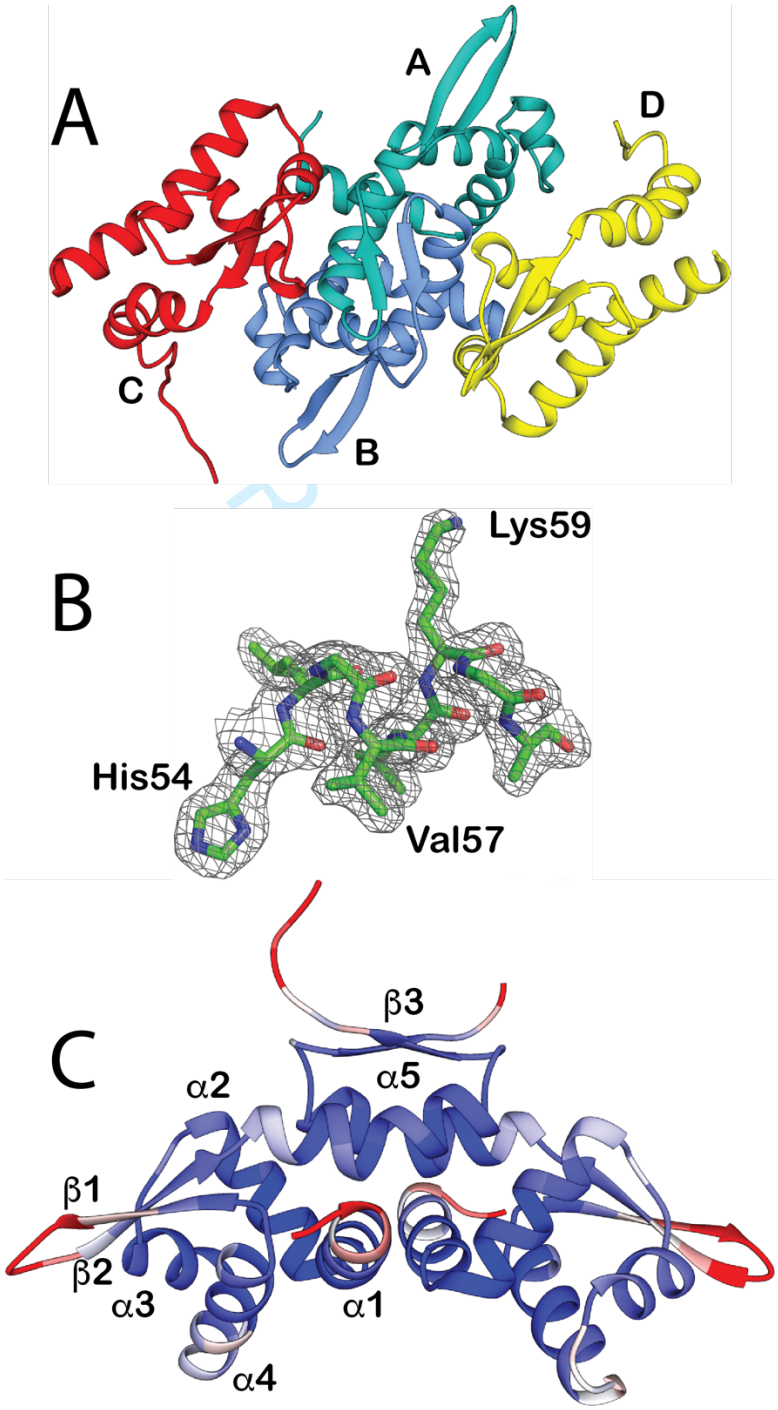


Figure 2. Representation of *Sg*SrnR as (**A**) ribbon coloured by residue conservation determined with ConSurf (maroon = conserved, cyan = variable) and as (**B**) molecular surface coloured by electrostatic potential using DelPhi (red = negative, blue = positive). In panel (C) a 'sausage' representation of the *Sg*SrnR structure is also shown; the diameter of the sausage is proportional to the RMSF of Cα atoms as calculated from the MD simulations. The sausage is coloured from blue to red for RMSF values equal to 0.0 and 0.4 nm, respectively. The right panels of A and B are rotated clockwise, and the right panel of C anticlockwise, by 90° around the horizontal axis *vs*. the left panels.

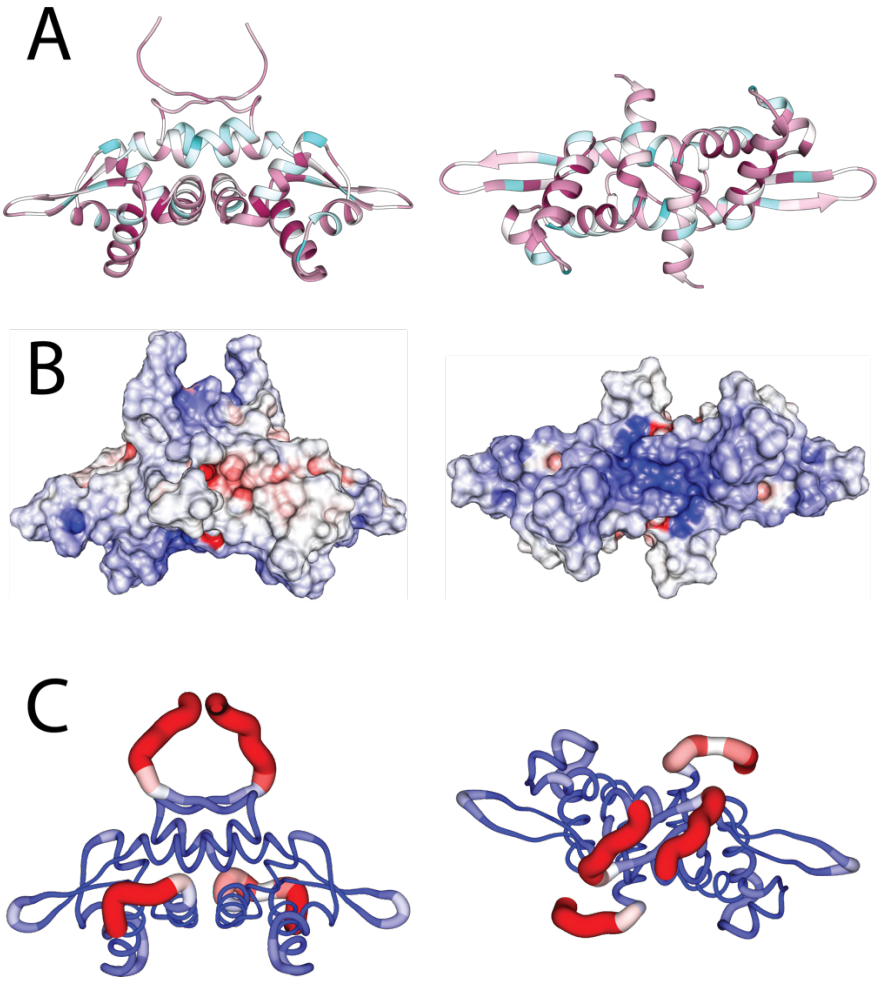


Figure 3. ^{1}H,^{15}N HSQC spectrum of triply labelled *Sg*SrnR at 800 MHz and pH 7.5. The labels indicate the single-letter amino acid code and the corresponding residue number. The peaks around 7/126 ppm on the ^{1}H/^{15}N dimension are folded unassigned peaks from Arg sidechains; two pairs of signals from Asn and Gln sidechains are also visible (joined by a horizontal line); the few remaining unlabelled peaks must originate from the four unassigned residues that gave no signals in 3D or 4D experiments and were left unassigned.

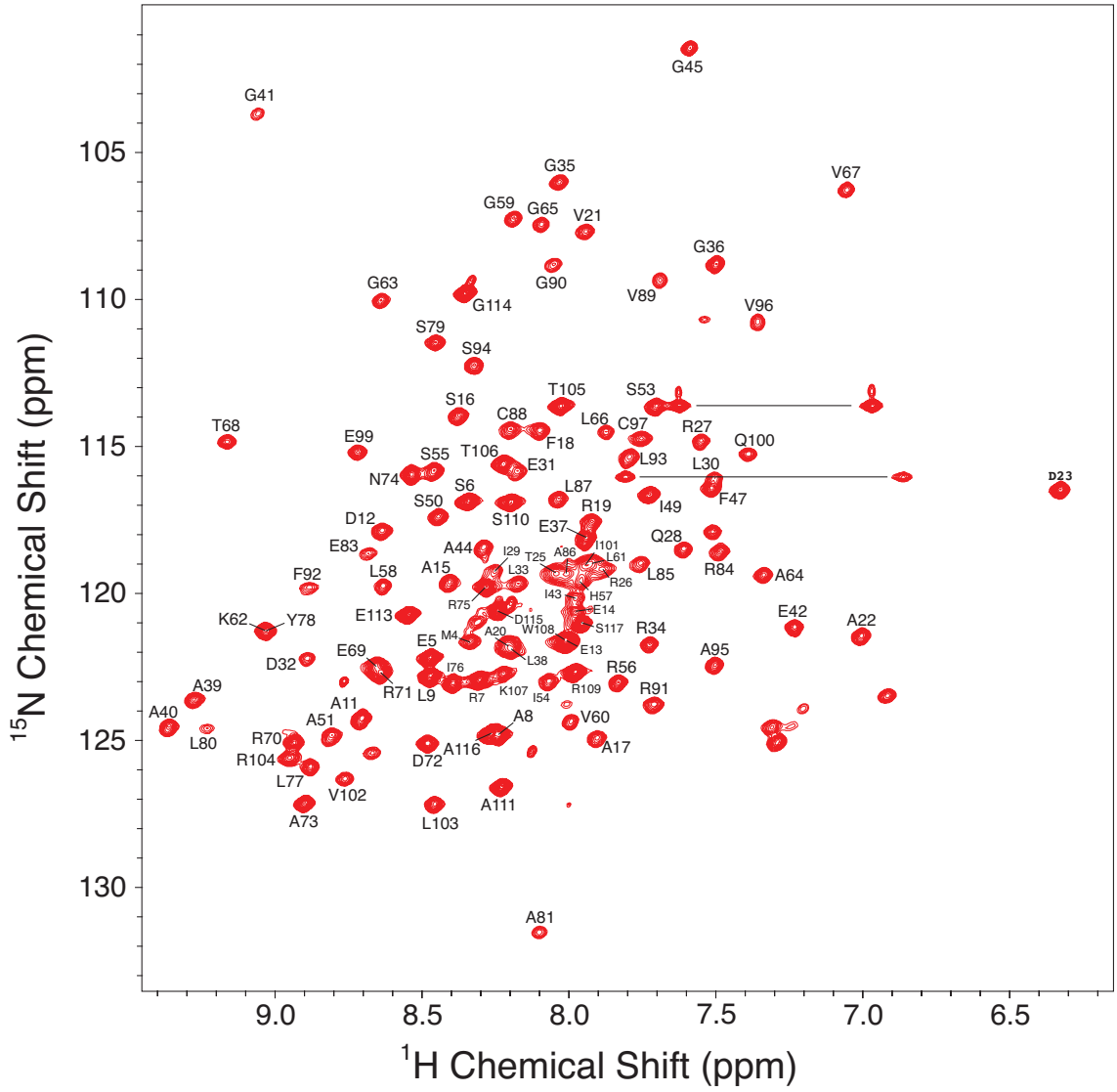


Figure 4. (**A**) Secondary structure analysis based on the *Sg*SrnR NMR chemical shifts assignment. Probability of secondary structure elements distribution along the protein sequence (red: helix; blue: strand) and corresponding order parameters S^2 (dots connected by a line) predicted by TalosN; (**B**) ^{15}N–^{1}H HSQC peaks intensities along the *Sg*SrnR protein sequence (the GSH non-native N-terminal extension is included here)

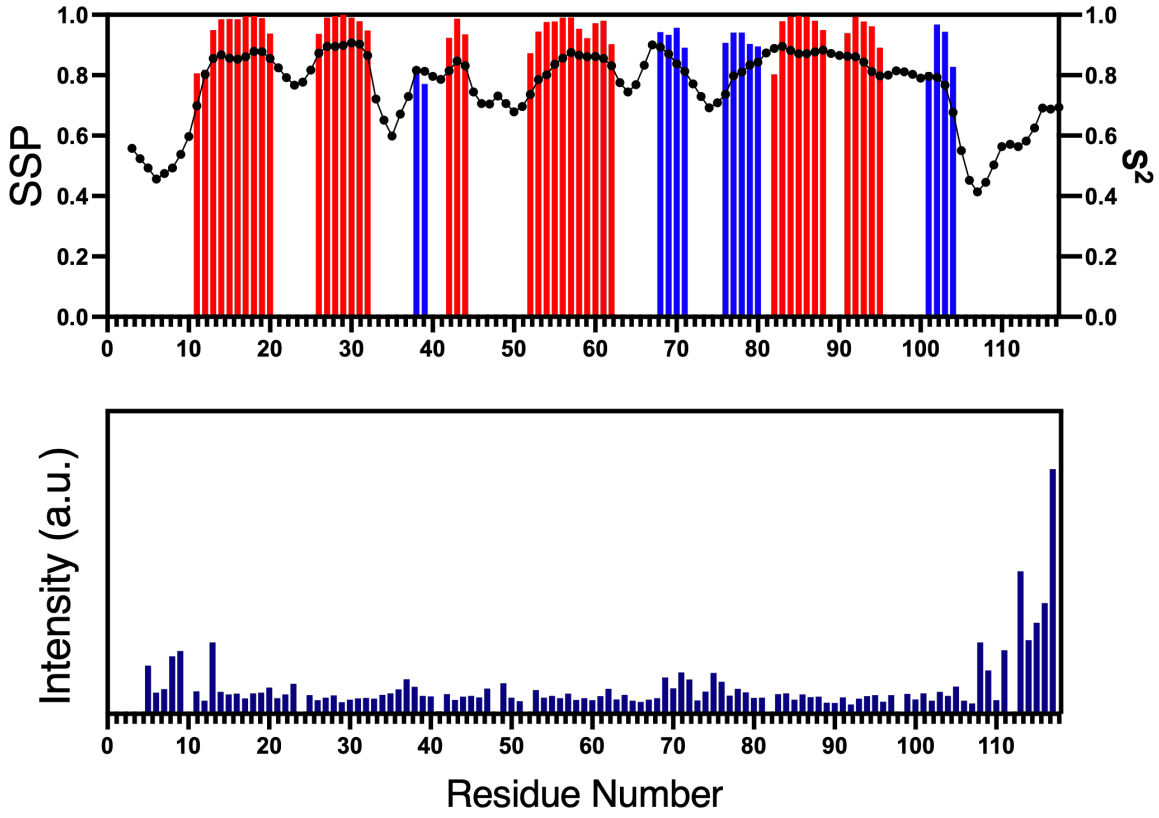


Figure 5. Disordered regions of the sequence of *Sg*SrnR as predicted by the D2P2 server (http://d2p2.pro/). The predicted disordered regions (top), folded domains (middle) and disorder consensus (bottom) are indicated by bars over the residue numbers.

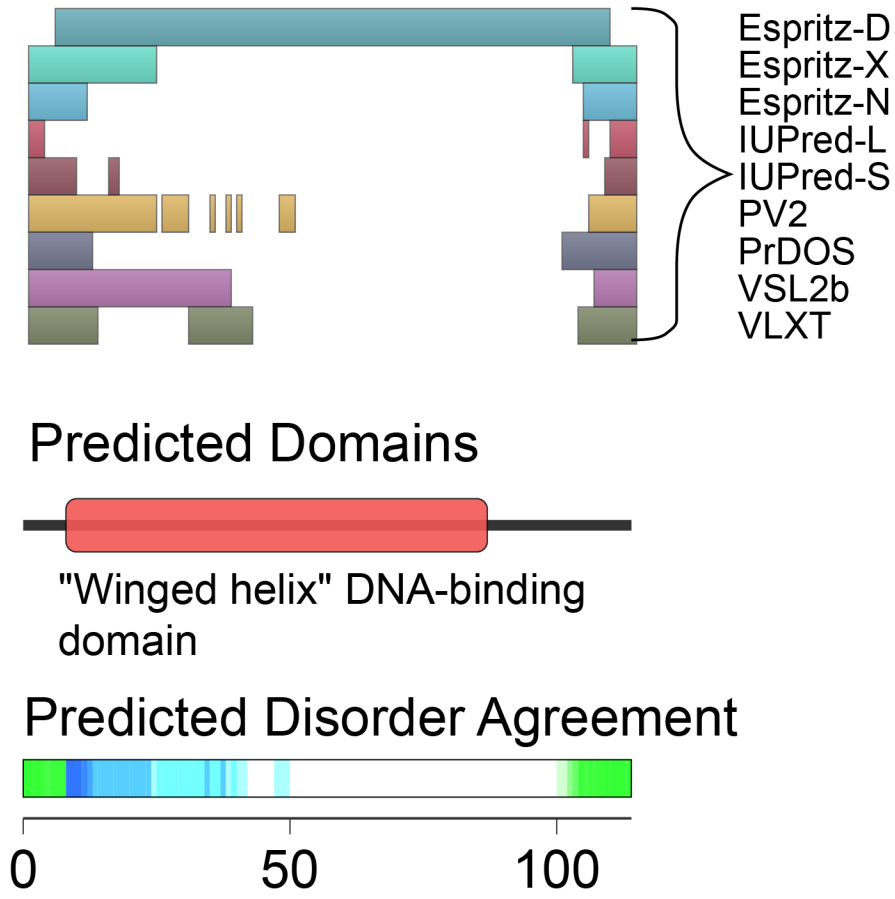


Figure 6. OP_{sodF} binding to *Sg*SrnR by ITC. (**A**) Thermogram obtained by titrating a solution of *Sg*SrnR (13 µM) with a solution of OP_{sodF} DNA sequence (140 µM). (**B**) Integrated heat data (filled dots) fit with a model involving two sets of binding sites (continuous line). (**C**) ^{1}H,^{15}N TROSY-HSQC spectrum of triply labelled *Sg*SrnR at 800 MHz and pH 7.5 in the presence of one equivalent of OP_{sodF}. The labels indicate the single-letter amino acid code and the corresponding residue number.

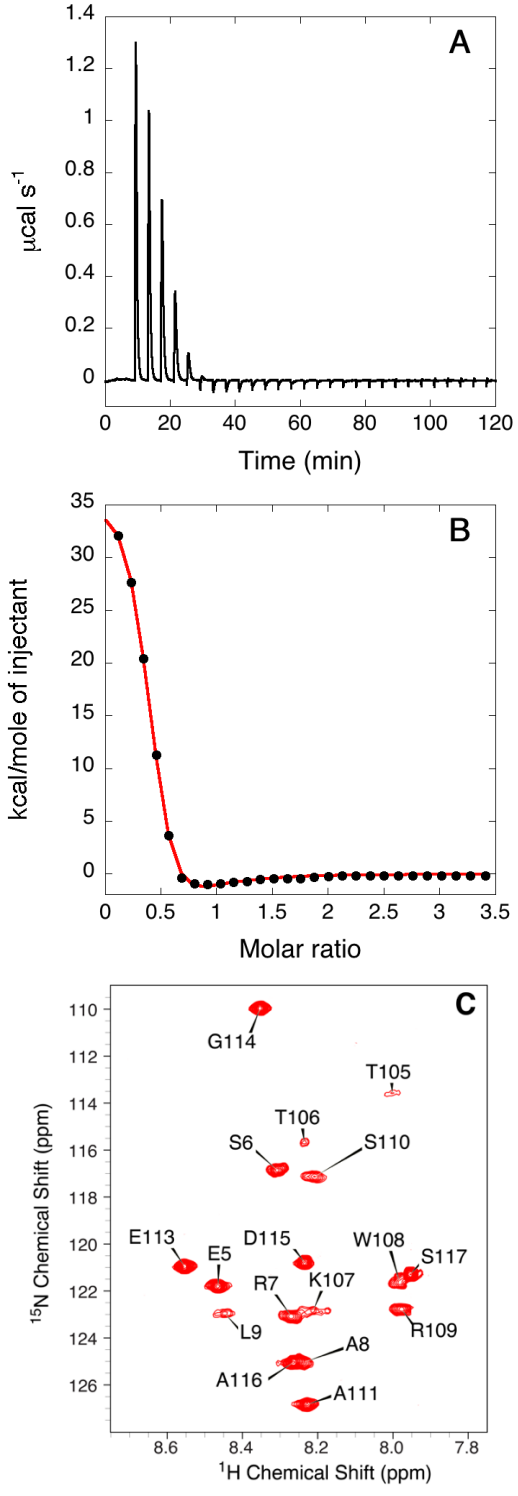
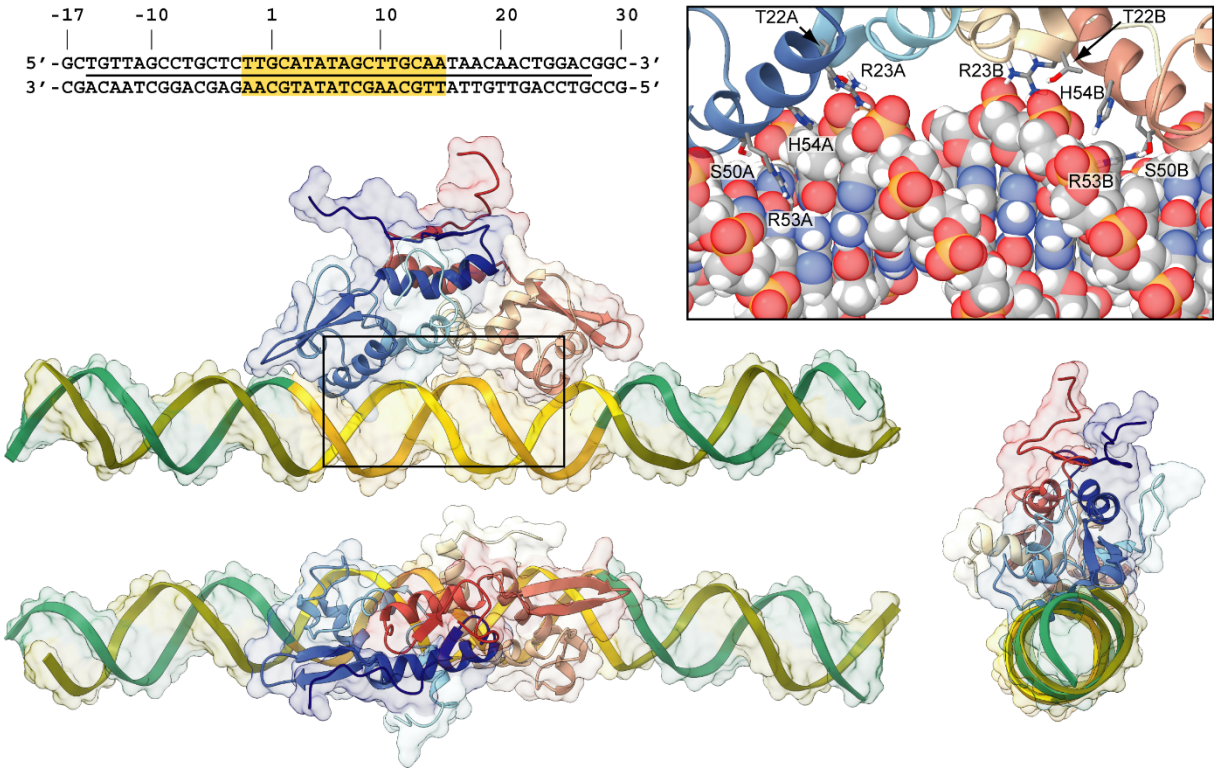


Figure 7. DNA sequence used for the protein-DNA docking (top-left panel) and molecular modelling of the *Sg*SrnR-OP_{sodF} complex (other panels). The OP_{sodF} operator (−15 to +27 with respect to the *sodF* operon transcriptional start site in *S. griseus*) is indicated through a black line between the pairing bases. The inverted repeat sequence used to guide the docking (from −2 to +15 with respect to the *sodF* operon transcriptional start site in *S. griseus*) has been highlighted in yellow. In the panels showing the whole *Sg*SrnR-OP_{sodF} complex, the ribbons of both the protein and of the DNA have been reported together with the molecular surface. *Sg*SrnR ribbons have been coloured from light to dark blue and from yellow to dark red for monomer A and B, respectively. The DNA strands are in lime green and dark green, while the region used to guide the docking is in yellow. In the bottom and in the bottom-right panel, the *Sg*SrnR-OP_{sodF} complex has been rotated by 90° around the horizontal and vertical axis, respectively. In the top-right panel, a detail of the *Sg*SrnR-OP_{sodF} interaction is offered. The DNA is reported using transparent spheres coloured accordingly to the atom type, while *Sg*SrnR residues important for the interaction are in sticks. For clarity, only polar hydrogen atoms have been included in the figure.



REFERENCES

1. Andreini, C.; Bertini, I.; Rosato, A., A hint to search for metalloproteins in gene banks. *Bioinformatics* **2004,** *20* (9), 1373-80.

2. Waldron, K. J.; Robinson, N. J., How do bacterial cells ensure that metalloproteins get the correct metal? *Nat Rev Microbiol* **2009,** *7* (1), 25-35.

3. Zambelli, B.; Musiani, F.; Ciurli, S., Metal ion-mediated DNA-protein interactions. *Metal ions in life sciences* **2012,** *10*, 135-170.

4. Capdevila, D. A.; Edmonds, K. A.; Giedroc, D. P., Metallochaperones and metalloregulation in bacteria. *Essays Biochem* **2017,** *61* (2), 177-200.

5. Campbell, D. R.; Chapman, K. E.; Waldron, K. J.; Tottey, S.; Kendall, S.; Cavallaro, G.; Andreini, C.; Hinds, J.; Stoker, N. G.; Robinson, N. J.; Cavet, J. S., Mycobacterial cells have dual nickel-cobalt sensors: sequence relationships and metal sites of metal-responsive repressors are not congruent. *J Biol Chem* **2007,** *282* (44), 32298-310.

6. Jung, J.; Lee, S. J., Biochemical and Biodiversity Insights into Heavy Metal Ion-Responsive Transcription Regulators for Synthetic Biological Heavy Metal Sensors. *J Microbiol Biotechnol* **2019,** *29* (10), 1522-1542.

7. Saha, R. P.; Samanta, S.; Patra, S.; Sarkar, D.; Saha, A.; Singh, M. K., Metal homeostasis in bacteria: the role of ArsR-SmtB family of transcriptional repressors in combating varying metal concentrations in the environment. *Biometals* **2017,** *30* (4), 459-503.

8. Roy, R.; Samanta, S.; Patra, S.; Mahato, N. K.; Saha, R. P., In silico identification and characterization of sensory motifs in the transcriptional regulators of the ArsR-SmtB family. *Metallomics* **2018,** *10* (10), 1476-1500.

9. Arunkumar, A. I.; Campanello, G. C.; Giedroc, D. P., Solution structure of a paradigm ArsR family zinc sensor in the DNA-bound state. *Proc Natl Acad Sci U S A* **2009,** *106* (43), 18177-82.

10. Lee, S. G.; Krishnan, H. B.; Jez, J. M., Structural basis for regulation of rhizobial nodulation and symbiosis gene expression by the regulatory protein NolR. *Proc Natl Acad Sci U S A* **2014,** *111* (17), 6509-14.

11. Eicken, C.; Pennella, M. A.; Chen, X.; Koshlap, K. M.; VanZile, M. L.; Sacchettini, J. C.; Giedroc, D. P., A metal-ligand-mediated intersubunit allosteric switch in related SmtB/ArsR zinc sensor proteins. *J Mol Biol* **2003,** *333* (4), 683-95.

12. Beniamino, Y.; Pesce, G.; Zannoni, A.; Roncarati, D.; Zambelli, B., SrnR from Streptomyces griseus is a nickel-binding transcriptional activator. *J Biol Inorg Chem* **2020,** *25* (2), 187-198.

13. Kim, J. S.; Kang, S. O.; Lee, J. K., The protein complex composed of nickel-binding SrnQ and DNA binding motif-bearing SrnR of Streptomyces griseus represses sodF transcription in the presence of nickel. *J Biol Chem* **2003,** *278* (20), 18455-63.

14. Stola, M.; Musiani, F.; Mangani, S.; Turano, P.; Safarov, N.; Zambelli, B.; Ciurli, S., The nickel site of *Bacillus pasteurii* UreE, a urease metallo-chaperone, as revealed by metal-binding studies and X-ray absorption spectroscopy. *Biochemistry* **2006,** *45* (20), 6495-509.

15. Cianci, M.; Bourenkov, G.; Pompidor, G.; Karpics, I.; Kallio, J.; Bento, I.; Roessle, M.; Cipriani, F.; Fiedler, S.; Schneider, T. R., P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focusing. *J. Synchrotron Radiat.* **2017,** *24* (1), 323-332.

16. Kabsch, W., Xds. *Acta Crystallogr. D* **2010,** *66*, 125-32.

17. Evans, P., Scaling and assessment of data quality. *Acta Crystallogr. D* **2006,** *62*, 72-82.

18. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40* (Pt 4), 658-674.

19. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **1997,** *53*, 240-255.

20. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **2004,** *60*, 2126-2132.

21. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr. D* **2010,** *66*, 486-501.

22. Laskowski, E. J.; Frankel, R. B.; Gillum, W. O.; Papaefthymiou, G. C.; Renaud, J.; Ibers, J. A.; Holm, R. H., Synthetic analogues of the 4-Fe active sites of reduced ferredoxins. Electronic properties of the tetranuclear trianions $[Fe_4S_4(SR)_4]^{3-}$ of proteins using natural abundance gradient-enhanced ^{13}C-^{1}H correlation spectroscopy. *FEBS Lett.* **1993,** *333*, 251-256.

23. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E., UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **2021,** *30* (1), 70-82.

24. Goddard, T. D.; Huang, C. C.; Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Morris, J. H.; Ferrin, T. E., UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **2018,** *27* (1), 14-25.

25. Kazimierczuk, K.; Zawadzka, A.; Kozminski, W.; Zhukov, I., Random sampling of evolution time space and Fourier transform processing. *J Biomol NMR* **2006,** *36* (3), 157-68.

26. Stanek, J.; Augustyniak, R.; Koźmiński, W., Suppression of sampling artefacts in high-resolution four-dimensional NMR spectra using signal separation algorithm. *Journal of Magnetic Resonance* **2012,** *214*, 91-102.

27. Goddard, T. D.; Kneller, D. G. *SPARKY 3*, University of California: San Francisco.
28. Sali, A.; Blundell, T. L., Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **1993,** *234* (3), 779-815.
29. Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A., H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **2005,** *33* (Web Server issue), W368-W371.
30. Myers, J.; Grothaus, G.; Narayanan, S.; Onufriev, A., A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins* **2006,** *63* (4), 928-938.
31. Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V., H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012,** *40* (Web Server issue), W537-W541.
32. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11* (8), 3696-3713.
33. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926-935.
34. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J., GROMACS: fast, flexible, and free. *J Comput Chem* **2005,** *26* (16), 1701-18.
35. Abraham, M. J.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19.25.
36. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.
37. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997,** *18* (12), 1463-1472.
38. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103* (19), 8577-8593.
39. Bussia, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** *126* (1), 014101.
40. Nosé, S.; Klein, M. L., Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983,** *50* (5), 1055-1076.
41. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.
42. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; an Gunsteren, W. F.; Mark, A. E., Peptide Folding: When Simulation Meets Experiment. *Angewandte Chemie* **2004,** *38* (1-2), 236-240.

43. Zheng, G.; Lu, X. J.; Olson, W. K., Web 3DNA--a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res* **2009,** *37* (Web Server issue), W240-6.

44. Miskiewicz, J.; Sarzynska, J.; Szachniuk, M., How bioinformatics resources work with G4 RNAs. *Brief Bioinform* **2021,** *22* (3).

45. Dominguez, C.; Boelens, R.; Bonvin, A. M., HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **2003,** *125* (7), 1731-7.

46. de Vries, S. J.; van Dijk, A. D.; Krzeminski, M.; van Dijk, M.; Thureau, A.; Hsu, V.; Wassenaar, T.; Bonvin, A. M., HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* **2007,** *69* (4), 726-33.

47. Agriesti, F.; Roncarati, D.; Musiani, F.; Del Campo, C.; Iurlaro, M.; Sparla, F.; Ciurli, S.; Danielli, A.; Scarlato, V., FeON-FeOFF: the Helicobacter pylori Fur regulator commutates iron-responsive transcription by discriminative readout of opposed DNA grooves. *Nucleic Acids Res* **2014,** *42* (5), 3138-51.

48. Mazzei, L.; Dobrovolska, O.; Musiani, F.; Zambelli, B.; Ciurli, S., On the interaction of Helicobacter pylori NikR, a Ni(II)-responsive transcription factor, with the urease operator: In solution and in silico studies. *Journal of Biological Inorganic Chemistry* **2015,** *20* (6), 1021-1037.

49. van Dijk, M.; van Dijk, A. D.; Hsu, V.; Boelens, R.; Bonvin, A. M., Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility. *Nucleic Acids Res* **2006,** *34* (11), 3317-25.

50. van Dijk, M.; Bonvin, A. M., 3D-DART: a DNA structure modelling server. *Nucleic Acids Res* **2009,** *37* (Web Server issue), W235-9.

51. Prabaharan, C.; Kandavelu, P.; Packianathan, C.; Rosen, B. P.; Thiyagarajan, S., Structures of two ArsR As(III)-responsive transcriptional repressors: Implications for the mechanism of derepression. *J Struct Biol* **2019,** *207* (2), 209-217.

52. Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B., Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J Comput Chem* **2002,** *23* (1), 128-37.

53. Rocchia, W.; Alexov, E.; Honig, B., Extending the Applicability of the Nonlinear Poisson−Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *J. Phys. Chem. B* **2001,** *105* (28), 6507-6514.

54. Shen, Y.; Bax, A., Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* **2013,** *56* (3), 227-41.

55. Berjanskii, M. V.; Wishart, D. S., A simple method to predict protein flexibility using secondary chemical shifts. *J Am Chem Soc* **2005,** *127* (43), 14970-1.

56. Oates, M. E.; Romero, P.; Ishida, T.; Ghalwash, M.; Mizianty, M. J.; Xue, B.; Dosztányi, Z.; Uversky, V. N.; Obradovic, Z.; Kurgan, L.; Dunker, A. K.; Gough, J., D2P2: database of disordered protein predictions. *Nucl. Acid. Res.* **2012,** *41* (D1), D508-D516.

57. Kay, L. E.; Torchia, D. A.; Bax, A., Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to *Staphylococcus* nuclease. *Biochemistry* **1989,** *28*, 8972-8979.

58. Palmer, I., A G, Dynamic properties of proteins from NMR spectroscopy. *Curr. Opin. Biotechnol.* **1993,** *4* (4), 385-391.

59. Rossi, P.; Swapna, G. V. T.; Huang, Y. J.; Aramini, J. M.; Anklin, C.; Conover, K.; Hamilton, K.; Xiao, R.; Acton, T.; Ertekin, A.; Everett, J. K.; Montelione, G. T., A microscale protein NMR sample screening pipeline. *J Biomol NMR* **2010,** *46* (1), 11-22.

60. Peng, J. W.; Wagner, G., Mapping of spectral density function using heteronuclear NMR relaxation measurements. *J. Magn. Reson.* **1992,** *98*, 308-332.

61. Farrow, N. A.; Muhandiram, R. R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G. G.; Shoelson, S. E.; Pawson, T. T.; Forman-Kay, J. D.; Kay, L. E. L., Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **1994,** *33* (19), 5984-6003.

62. Ishima, R.; Nagayama, K., Protein backbone dynamics revealed by quasi spectral density function analysis of amide N-15 nuclei. *Biochemistry* **1995,** *34* (10), 3162-71.

63. Farrow, N. A.; Zhang, O.; Szabo, A.; Torchia, D. A.; Kay, L. E., Spectral density function mapping using ^{15}N relaxation data exclusively. *J. Biomol. NMR* **1995,** *6* (2), 153-162.

64. Farrow, N. A.; Zhang, O.; Forman-Kay, J. D.; Kay, L. E., Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer. *Biochemistry* **1995,** *34* (3), 868-78.

65. Peng, J. W.; Wagner, G., Frequency spectrum of NH bonds in eglin c from spectral density mapping at multiple fields. *Biochemistry* **1995,** *34* (51), 16733-52.

66. Lefevre, J. F.; Dayie, K. T.; Peng, J. W.; Wagner, G., Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions. *Biochemistry* **1996,** *35* (8), 2674-86.

67. Fabini, E.; Zambelli, B.; Mazzei, L.; Ciurli, S.; Bertucci, C., Surface plasmon resonance and isothermal titration calorimetry to monitor the Ni(II)-dependent binding of Helicobacter pylori NikR to DNA. *Anal Bioanal Chem* **2016,** *408* (28), 7971-7980.

68. Grossoehme, N. E.; Giedroc, D. P., Energetics of allosteric negative coupling in the zinc sensor S. aureus CzrA. *J Am Chem Soc* **2009,** *131* (49), 17860-70.

69. Mazzei, L.; Dobrovolska, O.; Musiani, F.; Zambelli, B.; Ciurli, S., On the interaction of Helicobacter pylori NikR, a Ni(II)-responsive transcription factor, with the urease operator: in solution and in silico studies. *J Biol Inorg Chem* **2015,** *20* (6), 1021-37.

70. Berdy, J., Bioactive microbial metabolites. *J Antibiot (Tokyo)* **2005,** *58* (1), 1-26.

71. Amoroso, M. J.; Schubert, D.; Mitscherlich, P.; Schumann, P.; Kothe, E., Evidence for high affinity nickel transporter genes in heavy metal resistant Streptomyces spec. *J Basic Microbiol* **2000,** *40* (5-6), 295-301.

72. Mengoni, A.; Barzanti, R.; Gonnelli, C.; Gabbrielli, R.; Bazzicalupo, M., Characterization of nickel-resistant bacteria isolated from serpentine soil. *Environ Microbiol* **2001,** *3* (11), 691-8.

73. Sessitsch, A.; Kuffner, M.; Kidd, P.; Vangronsveld, J.; Wenzel, W. W.; Fallmann, K.; Puschenreiter, M., The role of plant-associated bacteria in the mobilization and phytoextraction of trace elements in contaminated soils. *Soil Biol Biochem* **2013,** *60* (100), 182-194.

74. Nett, M.; Ikeda, H.; Moore, B. S., Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Rep* **2009,** *26* (11), 1362-84.

75. Waksman, S. A.; Schatz, A.; Reynolds, D. M., Production of antibiotic substances by actinomycetes. *Ann N Y Acad Sci* **2010,** *1213*, 112-24.

76. Rodriguez, H.; Rico, S.; Diaz, M.; Santamaria, R. I., Two-component systems in Streptomyces: key regulators of antibiotic complex pathways. *Microb Cell Fact* **2013,** *12*, 127.

77. Musiani, F.; Zambelli, B.; Bazzani, M.; Mazzei, L.; Ciurli, S., Nickel-responsive transcriptional regulators. *Metallomics* **2015,** *7*, 1305-1318.

78. Lee, C. W.; Chakravorty, D. K.; Chang, F. M.; Reyes-Caballero, H.; Ye, Y.; Merz, K. M., Jr.; Giedroc, D. P., Solution structure of Mycobacterium tuberculosis NmtR in the apo state: insights into Ni(II)-mediated allostery. *Biochemistry* **2012,** *51* (12), 2619-29.

79. Reyes-Caballero, H.; Lee, C. W.; Giedroc, D. P., Mycobacterium tuberculosis NmtR harbors a nickel sensing site with parallels to Escherichia coli RcnR. *Biochemistry* **2011,** *50* (37), 7941-52.

80. Capdevila, D. A.; Braymer, J. J.; Edmonds, K. A.; Wu, H.; Giedroc, D. P., Entropy redistribution controls allostery in a metalloregulatory protein. *Proc Natl Acad Sci U S A* **2017,** *114* (17), 4424-4429.

81. Banci, L.; Bertini, I.; Cantini, F.; Ciofi-Baffoni, S.; Cavet, J. S.; Dennison, C.; Graham, A. I.; Harvie, D. R.; Robinson, N. J., NMR structural analysis of cadmium sensing by winged helix repressor CmtR. *J Biol Chem* **2007,** *282* (41), 30181-8.

82. Baksh, K. A.; Pichugin, D.; Prosser, R. S.; Zamble, D. B., Allosteric regulation of the nickel-responsive NikR transcription factor from Helicobacter pylori. *J Biol Chem* **2021,** *296*, 100069.

83. Zambelli, B.; Uversky, V. N.; Ciurli, S., Nickel impact on human health: An intrinsic disorder perspective. *Biochim Biophys Acta* **2016,** *1864* (12), 1714-1731.