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

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.

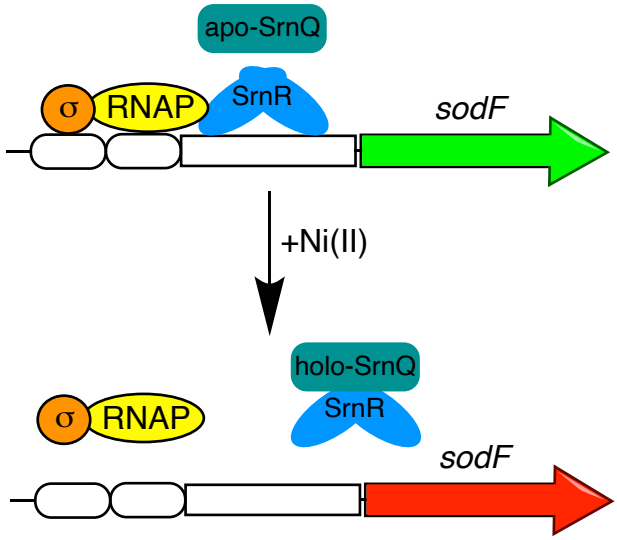
INTRODUCTION

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 shown in Figure 3). 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, we inferred the interacting residues from the proposed model reported for *Staphylococcus aureus* CzrA [9] derived 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.

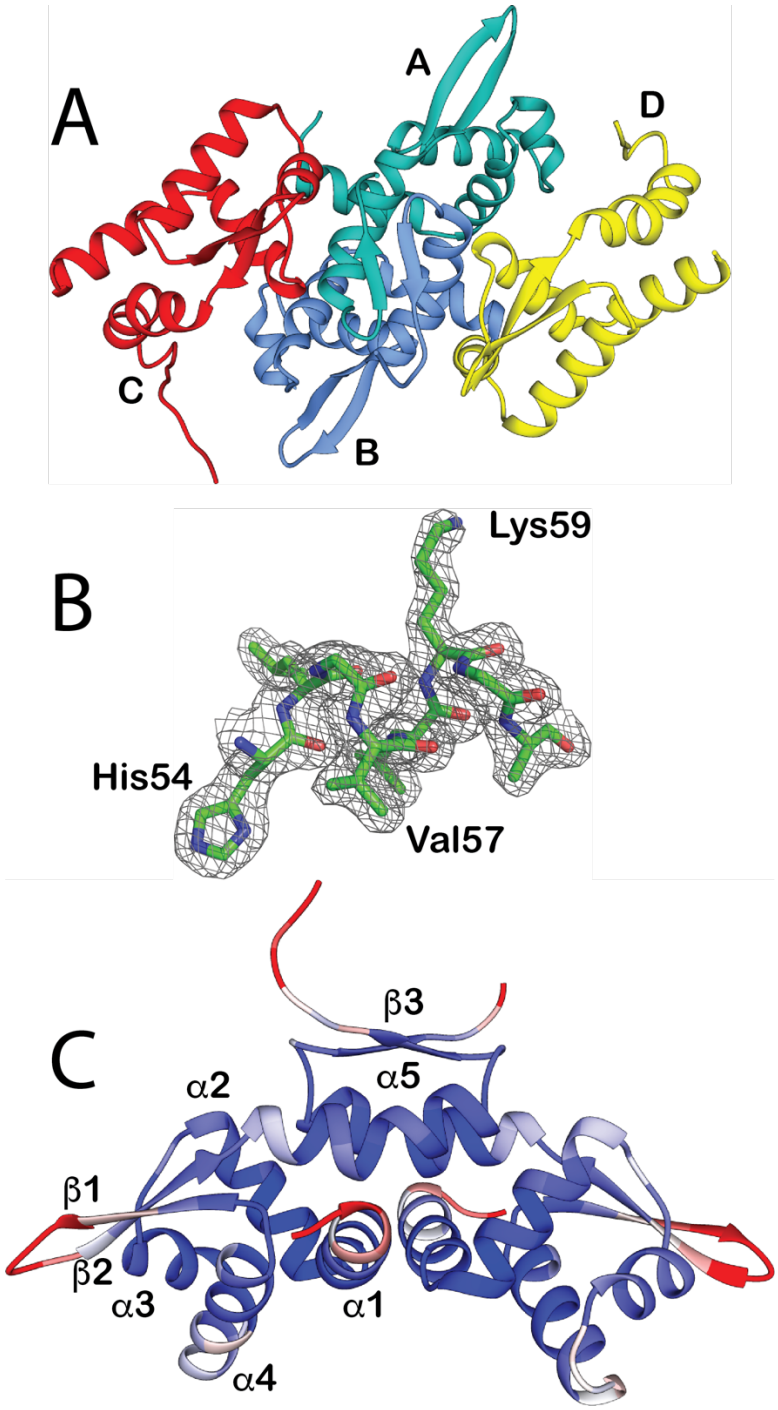


Figure 2. Representation of *Sg*SrnR as (**A**) ribbon coloured by residue conservation determined with ConSurf (maroon = conserved, cyan = variable) and as (**B**) molecular surface coloured by electrostatic potential using DelPhi (red = negative, blue = positive). In panel (C) a 'sausage' representation 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.

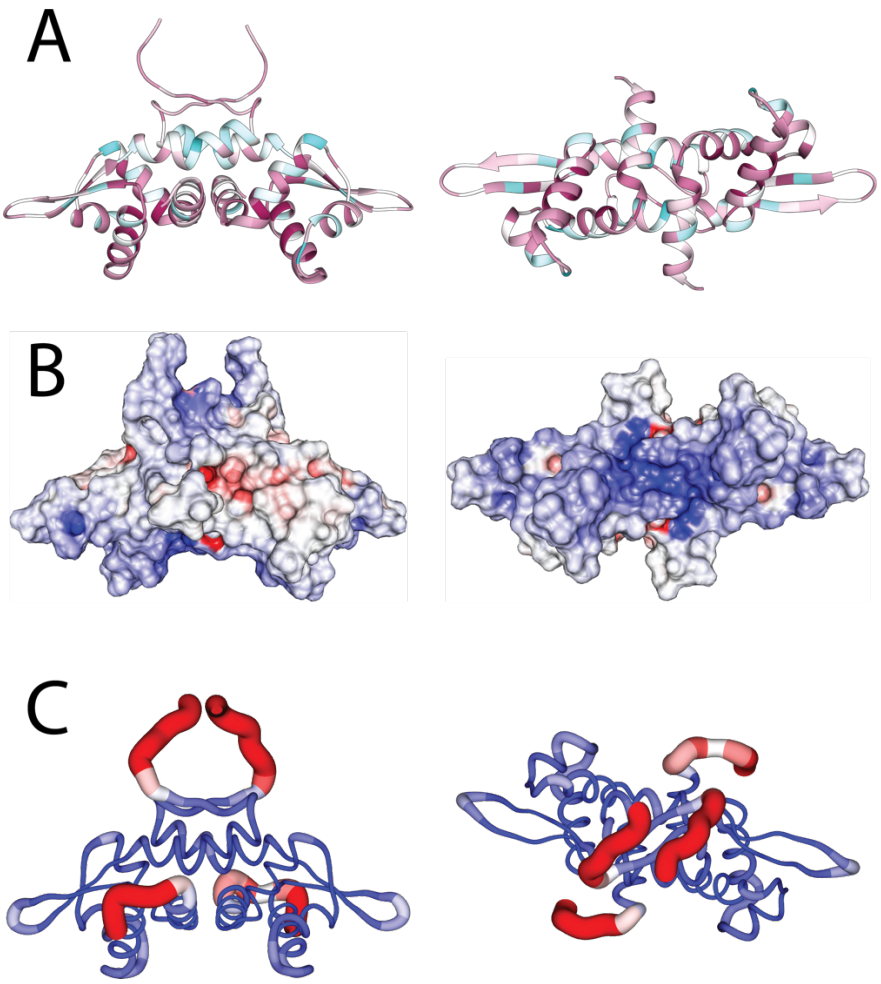


Figure 3. ^{15}N–^{1}H HSQC spectrum of triply labelled *Sg*SrnR at 800 MHz and pH 7.5. The labels indicate the single-letter amino acid code and the corresponding residue number. The peaks around 7/126 ppm on the ^{1}H/^{15}N dimension are folded unassigned peaks from Arg sidechains; two pairs of signals 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.

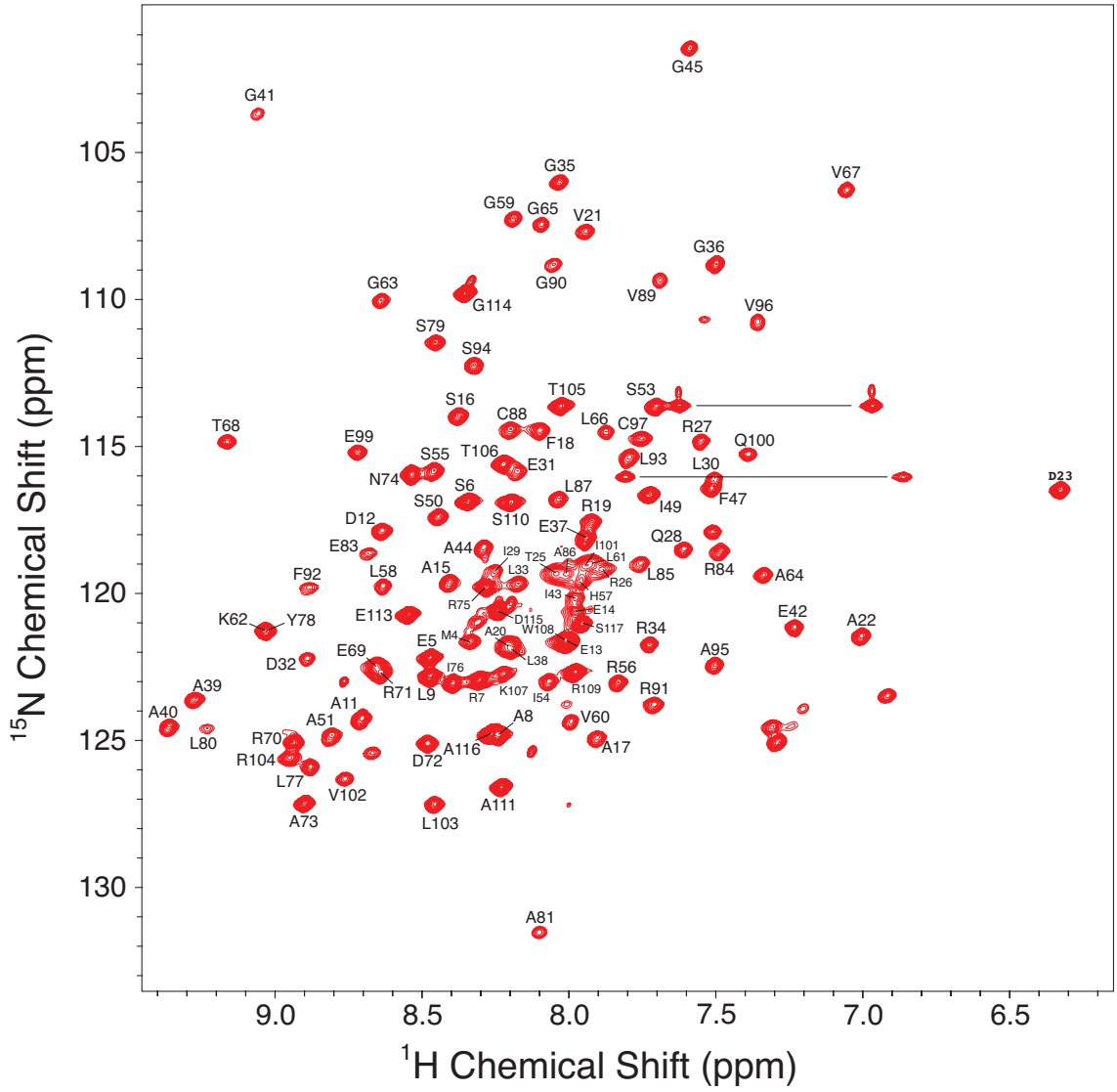


Figure 4. (**A**) Secondary structure analysis based on the *Sg*SrnR NMR chemical shifts assignment. Probability of secondary structure elements distribution along the protein sequence (red: helix; blue: strand) and corresponding order parameters S^2 (dots connected by a line) predicted by TalosN; (**B**)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

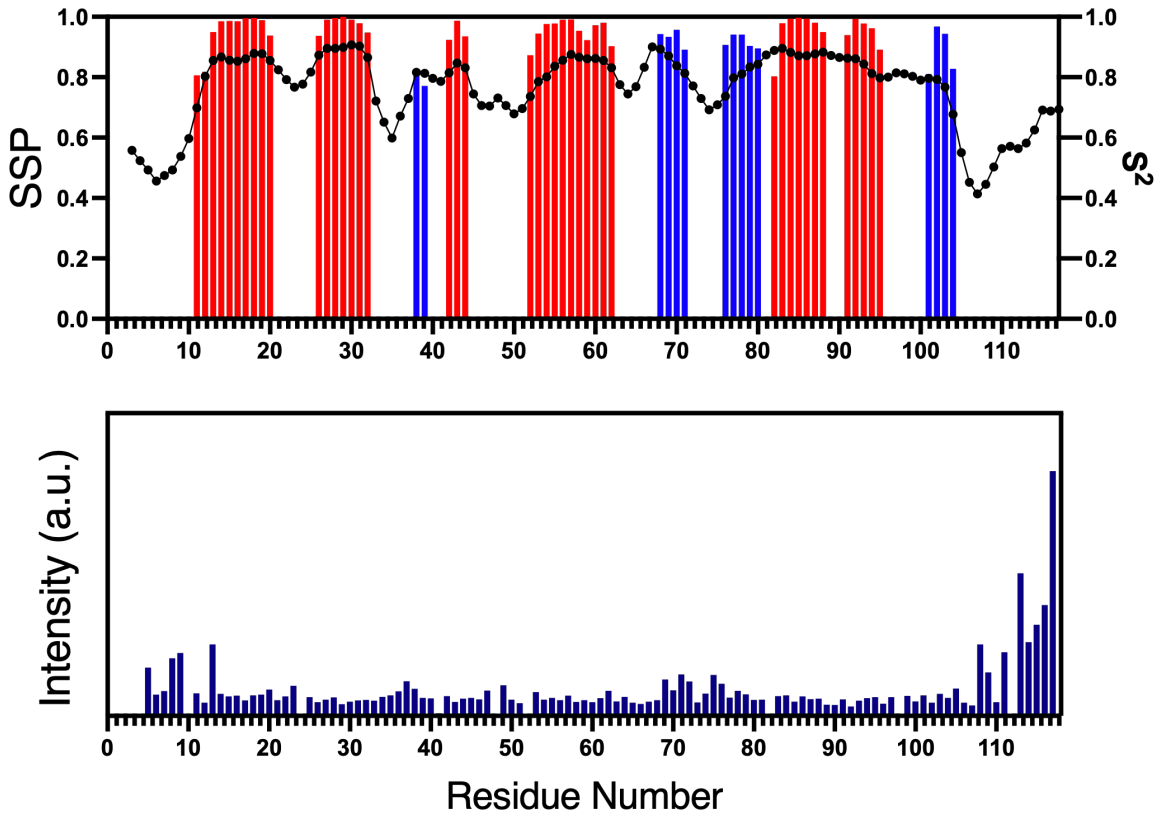


Figure 5. Disordered regions of the sequence of *Sg*SrnR as predicted by the D2P2 server (http://d2p2.pro/). The predicted disordered regions (top), folded domains (middle) and disorder consensus (bottom) are indicated by bars over the residue numbers.

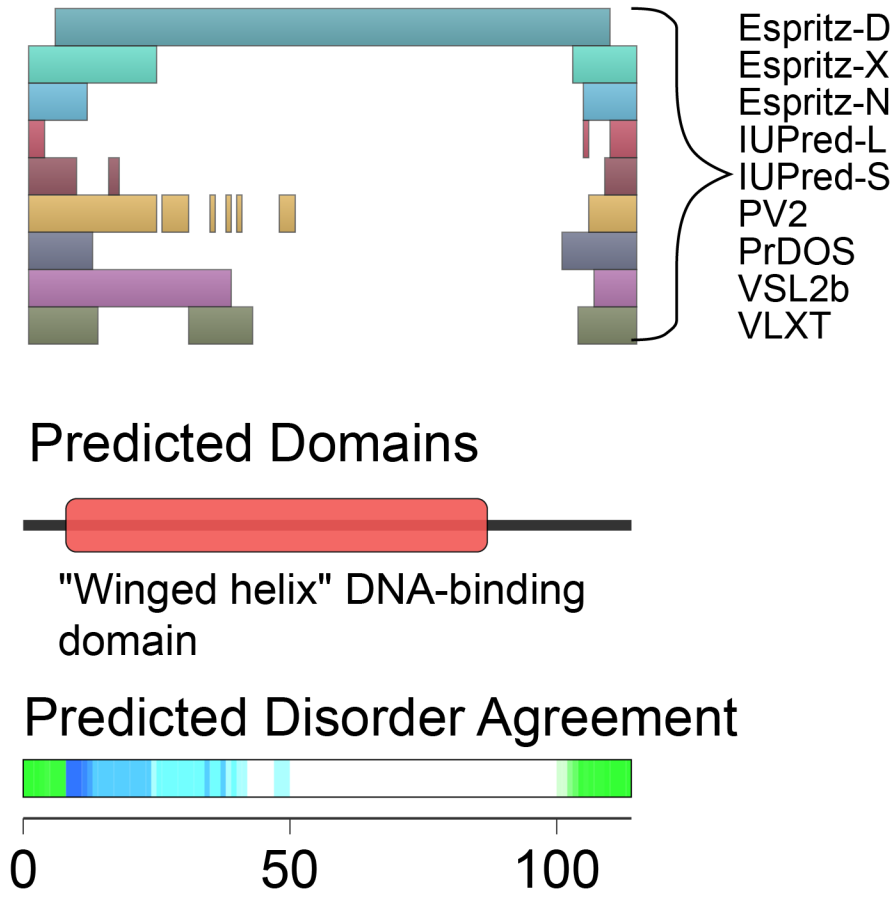


Figure 6. OP_{sodF} binding to *Sg*SrnR by ITC. (**A**) Thermogram obtained by titrating a solution of *Sg*SrnR (13 µM) with a solution of OP_{sodF} DNA sequence (140 µM). (**B**) Integrated heat data (filled dots) fit with a model involving two sets of binding sites (continuous line).

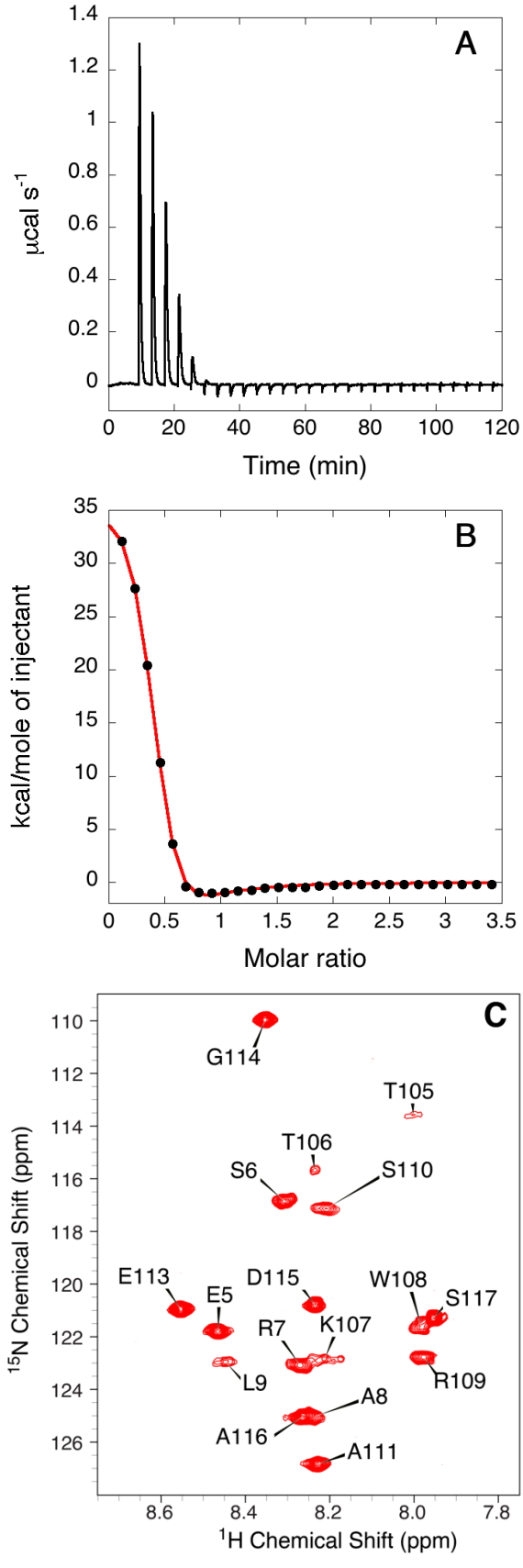
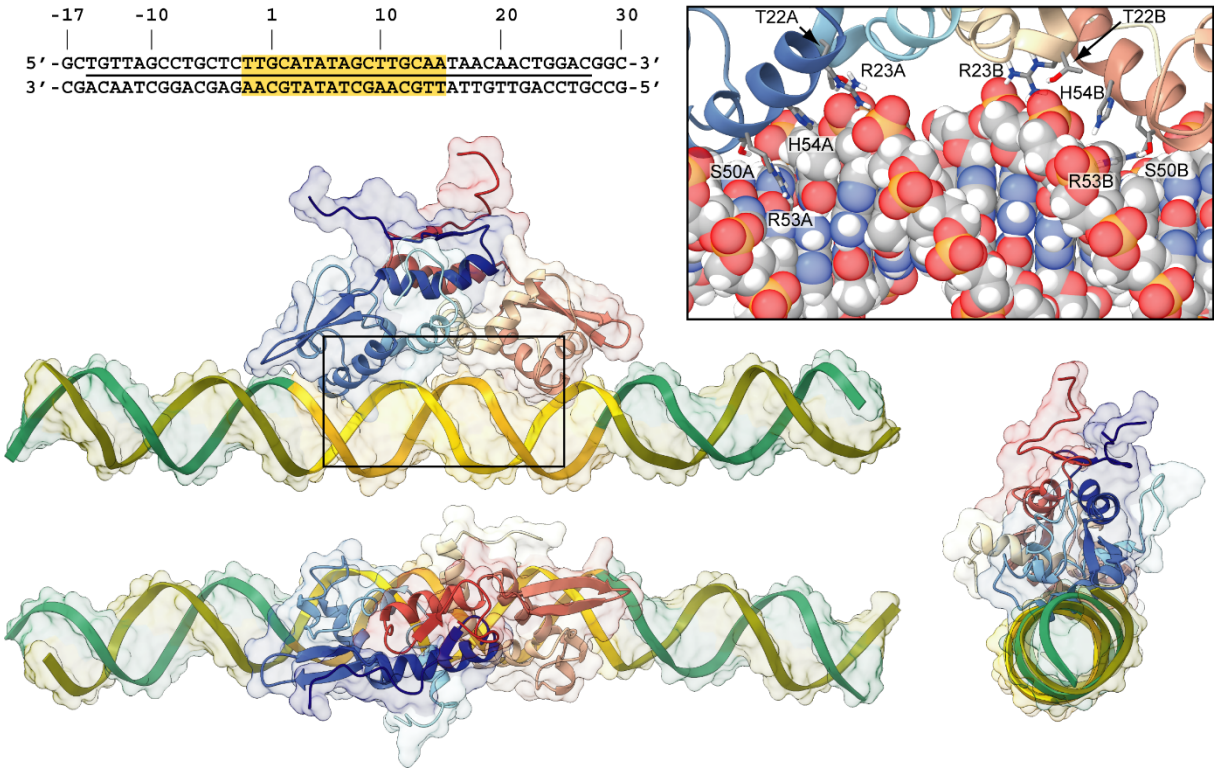


Figure 7. DNA sequence used for the protein-DNA docking (top-left panel) and molecular modelling of the *Sg*SrnR-OP_{sodF} complex (other panels). The OP_{sodF} operator (−15 to +27 with respect to the *sodF* operon transcriptional start site in *S. griseus*) is indicated through a black line between 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.