9760–9772 *Nucleic Acids Research, 2017, Vol. 45, No. 16* *Published online 18 July 2017*
doi: 10.1093/nar/gkx609

The fused SnoaL_2 domain in the *Mycobacterium tuberculosis* sigma factor σ^J modulates promoter recognition

Kapil Goutam, Arvind K. Gupta and Balasubramanian Gopal*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, Karnataka, India

Received April 11, 2017; Revised July 01, 2017; Editorial Decision July 04, 2017; Accepted July 06, 2017

ABSTRACT

Extra-cytoplasmic function (ECF) σ-factors are widespread in bacteria, linking environmental stimuli with changes in gene expression. These transcription factors span several phylogenetically distinct groups and are remarkably diverse in their activation and regulatory mechanisms. Here, we describe the structural and biochemical features of a *Mycobacterium tuberculosis* ECF factor σ^J that suggests that the SnoaL_2 domain at the C-terminus can modulate the activity of this initiation factor in the absence of a cognate regulatory anti-σ factor. *M. tuberculosis* σ^J can bind promoter DNA *in vitro*; this interaction is substantially impaired by the removal of the SnoaL_2 domain. This finding is consistent with assays to evaluate σ^J-mediated gene expression. Structural similarity of the SnoaL_2 domain with epoxide hydrolases also suggests a novel functional role for this domain. The conserved sequence features between *M. tuberculosis* σ^J and other members of the ECF41 family of σ-factors suggest that the regulatory mechanism involving the C-terminal SnoaL_2 domain is likely to be retained in this family of proteins. These studies suggest that the ECF41 family of σ-factors incorporate features of both—the σ^{70} family and bacterial one—component systems thereby providing a direct mechanism to implement environment-mediated transcription changes.

INTRODUCTION

Regulation of gene expression in prokaryotes occurs primarily at the transcription initiation step. σ factors, the specificity defining subunit of RNA polymerase (RNAP), govern gene expression by their reversible association with the RNAP (1). Of the two broad families of σ factors, members of the the σ^{70} family are more diverse due to variations in their activation and regulatory mechanisms. Most characterized σ^{70} members rely on protein–protein interactions to enable DNA promoter binding—a key distinction from the σ^{54} class of σ factors that can form tight DNA complexes but require adenosine triphosphate hydrolysis for open promoter complex formation (2,3). σ^{70} members have been further classified into groups based on domain architecture (4,5). While the housekeeping σ factor contains an N-terminal polypeptide segment (region 1.1) and four DNA binding domains and governs basal expression of genes, the extra-cytoplasmic function (ECF) σ factors are much smaller with only two DNA binding domains (referred to as σ_2 and σ_4) and govern transcription under stress or starvation conditions (6). ECF σ factors are the largest and most divergent group in σ^{70} σ factors and govern transcription in response to various stresses and starvation conditions. The activity of most ECF σ factors (and a few members of other σ factor families) is regulated by interaction with a protein antagonist also referred to as an anti-σ factor (7). The anti-σ factor can be either cytoplasmic or membrane bound. The release of a σ factor from these σ/anti-σ complexes is brought about by diverse mechanisms including proteolysis, phosphorylation and redox-dependent conformational changes (8–10). More recently, ECF group members were classified into 43 sub-groups based on sequence architecture (11). Another study utilizing under-represented genomes extended these subgroups to more than 50 (12). Four of the ECF groups viz ECF41, ECF42, ECF44 and ECF01-Gob contain an additional domain at the carboxyl-terminus. Of these, the ECF44 sub-group σ factors which contain a conserved carboxyl-terminal cysteine rich domain (CRD) have been better characterized (13–15). The activity of two ECF44 σ factors, *Myxococcus xanthus* corE1 and corE2 is directly regulated by metal ions. While corE1 responds to Cu, corE2 binds Cd and Zn and mutation of key cysteine residues in the CRD affects metal ion binding. Indeed, deletion of the ECF44-specific Cysteine-X-Cysteine (CXC) motif from σ_2 results in loss of activity (14,15).

The ECF41 σ factor sub-group, that contains nearly 400 annotated members distributed across 10 phyla, is poorly understood. All ECF41 group members possess a distinct domain organization with σ_2, σ_4 and an additional

*To whom correspondence should be addressed. Tel: +91 80 2293 3219; Fax: +91 80 2360 0535; Email: bgopal@mbu.iisc.ernet.in

carboxy-terminal domain (11,16). ECF41 members lack an apparent anti-σ factor and the regulation of ECF41 σ factor activity remains unclear. In the absence of an antagonist (anti-σ factor), the domain at the C-terminal has been suggested to play a role similar to an anti-σ factor. A recent study of two ECF41 group σ factors from *Bacillus licheniformis* and *Rhodobacter sphaeroides* suggested a regulatory role for this additional domain at the Carboxy-terminus. These studies, performed using deletion analysis and interaction assays, proposed the additional domain to be a fused anti-σ-factor-like domain with a potential role in promoter activation as well as interaction with the RNAP. (16). An interesting observation is that of genomic context—ECF41 genes are often next to those of carboxy muconolactone decarboxylases, oxidoreductases or epimerases (referred to as COE) (11,16). This genomic proximity suggested a possible role for this σ factor in maintaining redox homeostasis.

Mycobacterium tuberculosis σ factors play a critical role in the virulence of this human pathogen (17). *M. tuberculosis* has two ECF41 σ factors σ^I and σ^J (11). The cellular level of σ^J is upregulated in late stationary phase (18). The only known gene under *M. tuberculosis* σ^J regulation is that encoding σ^I (19). Although the *M. tuberculosis sigJ* gene is not flanked by any of the COE genes, the target *sigI* gene has a putative proline dehydrogenase in the proximity. Interestingly, *M. tuberculosis* ECF41 σ factor σ^J has been reported to influence resistance to hydrogen peroxide-mediated oxidative stress (20). Another study on *M. marinum* σ factors under different stress conditions suggests the possible involvement of σ^J in heat stress (21). The lack of an apparent anti-σ factor that is a receptor for redox stimuli made it interesting to explore whether the additional domain in σ^J plays a role akin to an anti-σ factor. The C-terminal domain in σ^J was predicted to be a SnoaL_2 domain that is classified under the nuclear transport factor 2 (NTF2-like) superfamily (22). These proteins do not contain any specific conserved sequence motifs but share a common structural fold (23,24). The NTF-2 like fold is a cone shaped structure with an internal cavity. Proteins with this fold are involved in a wide range of functions—both enzymatic and non-enzymatic. The catalytically active members include SnoaL polyketide cyclases, limonene-1,2-epoxide hydrolase and delta-5-3-ketosteroid isomerases (25–28). In the case of a characterized non-catalytic role, the NTF2 domain of the calcium/calmodulin-dependent protein kinase II is involved in oligomerization (29).

Here, we describe the crystal structure of *M. tuberculosis* σ^J and biochemical studies to understand the role of the SnoaL_2 domain in this protein. These studies reveal a structural role for this C-terminal domain in enabling promoter binding. The rigidity conferred to σ^J by the SnoaL_2 domain was evaluated by Molecular dynamics (MD) simulations. These observations reveal an interesting variation in σ^{70} proteins wherein a pre-formed promoter recognizing conformation can be modulated by a receptor domain that can facilitate protein–protein or protein–ligand interactions. This feature, which is likely to be conserved across ECF41 members, blurs the distinction between σ factors as dissociable subunits of the RNAP and one-component systems that couple cellular or extra-cellular stimuli with transcription.

MATERIALS AND METHODS

Expression and purification of recombinant proteins

The details of expression constructs used to express and purify σ^J (Rv3328c) are summarized in Supplementary Table S1. All the proteins used in crystallization and interaction assays were purified using the same protocol unless otherwise mentioned. Clones were confirmed using single primer based sequencing (Amnion Biosciences Pvt. Ltd.). In each case, the plasmid with the gene of interest was transformed into *Escherichia coli* BL21(DE3) strain (Novagen *Inc.*). A single colony was inoculated in Luria–Bertani medium or minimal medium (for seleno-methionine (Se-Met) derivative) containing an appropriate antibiotic. Cultures were allowed to grow up to OD_{600nm} of 0.5–0.6 at 37°C prior to induction with 0.2 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). For the Se-Met derivative, amino acid supplements and Se-Met were added at an OD_{600nm} of 0.4. Post-induction, cultures were grown for 12 h at 18°C. Cells were then spun at 7000 *g* for 15 min. The pellet was re-suspended in buffer A (50 mM Tris–HCl pH8.0, 300 mM NaCl) containing 2 mM phenylmethanesulfonyl fluoride (PMSF) and ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor tablets (Roche). Cells were lysed by sonication and the cell debris were removed by centrifugation at 30 000 *g* for 45 min at 4°C. The supernatant was incubated with Ni-NTA resin (Sigma–Aldrich) for 1 h. The protein was then eluted using buffer A with a gradient of imidazole (10–200 mM). The protein was further purified by size-exclusion chromatography using Sephacryl S200 Hi-Prep 16/60 column (GE Healthcare, *Inc.*). The fractions containing the purified protein were concentrated to *ca* 10 mg/ml for crystallization trials. The purity and molecular mass of the protein was further verified using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and liquid chromatography–electrospray ionization mass spectrometry (LCESI-MS) (Bruker Daltonics, *Inc.*). Se-Met labeled protein was purified similarly with the exception that 1 mM Tris-(2- carboxyethyl)phosphine or 2 mM dithiothreitol (DTT) was included during the purification.

Crystal structure determination

Crystallization trials for σ^J have been described earlier (30). The native and selenium SAD and MAD datasets were collected at the BM-14 beamline of the European Synchrotron Radiation Facility at 100 K. The structure was solved using SAD diffraction data. The SAD diffraction data were collected at an oscillation set at 0.25° per image at 0.978Å wavelength. The diffraction data were processed using iMosflm (31) and scaled using SCALA (32) in the CCP4 suite (33). Analysis of diffraction data revealed pseudo-merohedral twinning. Initially, the crystal symmetry appeared to be I422 due to significant twinning (twin fraction = 0.478). The space group was later determined to be I222 (Supplementary Table S3). An initial model of the structure was obtained using the Autosol module in Phenix (34,35). The model was further built with the Autobuild module of Phenix and Arp/Warp (36). Subsequent model building and refinement was performed using COOT

(37) and phenix.refine (34) using the twin operator (–h, -l, -k). The fit of the model to the electron density was evaluated using COOT. The final structures were validated using MOLPROBITY (38). The interface areas were analyzed using PDBePISA server (39). All the structure figures were generated using UCSF Chimera (40).

Sequence and structural analysis

The sequences of ECF41 σ factors were selected from a compilation published earlier (16). Sequences shorter than 280 residues and longer than 340 residues were removed from subsequent analysis. Sequences were clustered at 90% sequence identity using BLASTClust (41). The resulting 323 sequences and the crystal structure of σ^J were submitted to the PROMALS3D server (42). The alignment generated from PROMALS3D was submitted to the ConSurf server (43–47) to estimate evolutionary conservation of amino acid residues. ConSurf estimates evolutionary rates using evolutionary relatedness of protein sequences with consideration to the similarity between amino acid residues. The conservation scores are projected onto the structure or sequence alignments. To obtain an insight into the functional role and potential ligands for SnoaL_2 domain, a DALI database search was performed using the structure of σ^{JSnoaL_2} domain (48). All the unique PDB outputs with *Z*-score > 4 were manually analyzed to search for potential ligands of σ^{JSnoaL_2} domain.

Spectroscopic studies of σ^J interaction with Limonene-1,2-epoxide

The λ_{max} of Limonene-1,2-epoxide and Limonene-1,2-diol was determined spectroscopically. The purified σ^J protein was incubated with Limonene-1,2-epoxide for different time intervals. A UV-visible spectrum of the samples from 240 to 340 nm was recorded and analyzed.

Molecular dynamics simulations

Classical MD simulations were performed on σ^J and $\sigma^{J\Delta SnoaL\_2}$ to evaluate the influence of the SnoaL_2 domain on the dynamics of $\sigma^J{}_2$ and $\sigma^J{}_4$ domains. The initial atomic coordinates were obtained from the crystal structure of σ^J. All the missing loop residues in σ^J were modeled using ModLoop (49,50). The system was neutralized using minimal concentration of counter ions (Na^+ and Cl^- ions). Water molecules were added to solvate the system according to TIP3P model (51). This system was subjected to 3000 steps of steepest descent followed by a 1500 step conjugate gradient minimization by keeping the positions of ions and solute heavy atoms fixed using a restraint of 300 kcal/mol-Å^2. In the next stage, a reduced restraint of 100 kcal/mol-Å^2 was retained only on solute heavy atoms and 4000 steps of steepest descent and 2000 steps of conjugate gradient minimization were performed for each system. Finally, full system minimization was performed involving 8000 steps of steepest descent followed by 4000 steps of conjugate gradient. Subsequently, each system was heated progressively from 0 to 300 K in 60 ps (52). Equilibration phase of 1 ns was carried out under N*p*T conditions. Production runs were done with an integration time step of 2 fs. The non-bonded pair list was updated every 10 steps. The SHAKE algorithm was applied to constrain all bonds involving hydrogen atoms (53). Coordinates were retained at 1 ps time intervals. The production run was performed for 500 ns of total simulation time for both σ^J and $\sigma^{J\Delta SnoaL\_2}$. All trajectories were analyzed using cpptraj module of AMBER and structures were visualized in VMD and Chimera (40,54,55).

Electrophoretic mobility shift assays

For electrophoretic mobility shift assay (EMSA), σ^J or $\sigma^{J\Delta SnoaL\_2}$ was incubated with 0.5 nM ^{32}P-labeled *sigI*-promoter (*sigIp*) ds DNA on ice for 20 min (Supplementary Figure S6 and Supplementary Table S4). The final volume of the binding mixture was adjusted to 20 μl by adding the DNA-binding buffer (25 mM Tris–HCl pH 7.5, 1 mM DTT, 100 μg/μl bovine serum albumin, 5 mM $MgCl_2$, 6% Glycerol). The DNA–protein complex was then run on a 8% non-denaturing polyacrylamide gel in 0.5× tris–borate–EDTA (TBE) buffer at 4°C. Finally, the gel was dried and analyzed on a Typhoon FLA9000 phosphorimager.

Surface plasmon resonance measurements

Interaction of σ^J and $\sigma^{J\Delta SnoaL2}$ with *sigIp* was examined using surface plasmon resonance (SPR) (BIACORE 2000; GE Healthcare) (Supplementary Figure S6 and Table S4). Biotinylated *sigIp* (5′) was immobilized on a streptavidin (SA) sensor chip (BIACORE; GE Healthcare) at a surface density of (300+100RU) ng/mm^2. The first flow cell of the SA chip was used as control. The experiments were performed in a buffer containing 20 mM Tris–Cl (pH 7.5) and 50 mM NaCl. σ^J and $\sigma^{J\Delta SnoaL2}$ proteins were used as analytes in this study. The interaction kinetics was evaluated using BIA-evaluation software.

β-galactosidase based reporter assays in *Escherichia coli*

To study the role of σ^{JSnoaL_2} in transcription regulation, a β-galactosidase activity based assay was designed using LacZ fusion constructs of *sigI*-promoter in the pJEM13 vector (56). A modified pBAD33 vector (pBAD33m) was used for overexpression of σ^J and $\sigma^{J\Delta SnoaL2}$ under the control of an araBAD promoter (Figure 5A). The expression from the araBAD promoter can be regulated by L-arabinose thus allowing controlled expression of σ proteins. The target plasmid containing *sigI*-promoter-lacZ fusion and donor plasmid containing σ proteins were transformed in the *E. coli* LMG194 strain. Single colonies were inoculated in Luria Bertani media and were subsequently induced using 13 mM L-arabinose. Cultures were removed 3 h post induction and β-galactosidase assays were performed. A list of the constructs, plasmids and strains used in this study are summarized in Supplementary Table S5. Following the β-galactosidase assay, the Miller units were calculated and normalized to convert into relative β-galactosidase activity (56).

Limited proteolysis to evaluate conformational flexibility

Limited proteolysis assays were performed with trypsin. Briefly, 60 μg of purified proteins were incubated with 1 μg

of trypsin on ice. Samples were taken out at 10, 20, 30 and 60 min time intervals, mixed with sample dye, heated at 95°C and loaded on 15% SDS-PAGE for analysis.

RESULTS

Overall structure of *M. tuberculosis* σ^J

The crystallization and diffraction data statistics of σ^J was reported earlier (30). These crystals appeared in *ca* two weeks but could not be reproduced easily. While different variations in the expression constructs and additives in crystallization conditions were examined, addition of a trace of thrombin to the crystallization drop led to easily reproducible crystals (Supplementary Tables S1 and 2). LCESI-MS analysis of the crystals suggested a *ca* 3 kDa difference in mass when compared to the full-length protein. *In silico* analysis using the ExPASy peptide cutter (57) suggested two thrombin sites in σ^J - one from vector (pET-15b; Novagen, *Inc.*) and one in the protein that removed eight amino acid residues from C-terminus (Supplementary Figure S1). The data collection, refinement and model statistics are presented in Table 1. The structure was solved by the Single wavelength Anomalous Dispersion method (58). σ^J crystals are nearly perfectly twinned (pseudo merohedral twinning; twin fraction $\alpha = 0.478$) with an apparent higher crystal symmetry (I422 instead of I222; Supplementary Table S3). The structure of σ^J was solved in the I222 space group. There are two molecules of σ^J in the asymmetric unit. The two monomers of σ^J superpose well with a RMSD of 1.2 Å over 278 C_α atoms (Supplementary Figure S2). The crystal structure of σ^J revealed two DNA binding domains- $\sigma^J{}_2$, $\sigma^J{}_4$ and the SnoaL_2 domain arranged in a compact manner (Figure 1A and B; Supplementary Figure S3). The structure reveals substantial interactions between both $\sigma^J{}_2$ and $\sigma^J{}_4$ (buried surface area (BSA) of ~1100Å^2) and $\sigma^J{}_4$ and SnoaL_2 domain (BSA of ~800 Å^2). Interaction between $\sigma^J{}_2$ and SnoaL_2 involves the polypeptide linker connecting $\sigma^J{}_2$ with $\sigma^J{}_4$ (Table 2). This interaction appears important in tethering σ^J domains in a compact conformation. While $\sigma^J{}_2$ and $\sigma^J{}_4$ are α helical bundles, the SnoaL_2 domain adopts a α+β architecture consisting of five α-helices and β-strands with a ααβαααββββα arrangement. The β-strands are arranged in an anti-parallel fashion to form the β-sheet surface which stacks against $\sigma^J{}_2$ and $\sigma^J{}_4$ (Figure 1A and B).

Structural insights onto promoter recognition

The σ_4 domain that interacts with the −35 element of the promoter DNA consists of a helix-turn-helix motif and is better conserved amongst σ^{70} members (59). Indeed, $\sigma^J{}_4$ aligns well with other σ_4 domains (RMSD of 1.2–1.6 Å; Supplementary Figure S4a). The σ_2 domain has four helices α1–α4 connected by three loops L1-L3. The L3 loop—between the third and fourth helices of σ_2—initiates promoter melting by flipping the DNA base at the −10 promoter element in ECF σ-dependent promoters (60). The L3 loop also determines the specificity of ECF σ factors for the −10th base of promoter DNA. $\sigma^J{}_2$ differs from other σ_2 structures as it lacks the first helix α1 usually present in σ_2 domains (Supplementary Figure S4b and c). This helix is positioned just behind the L3 loop. Also, the second helix α2 is both short and rigid in comparison to the longer and bent helix α2 in other σ_2 domains (Supplementary Figure S4b and c). These variations in $\sigma^J{}_2$ could potentially alter the DNA melting properties of σ^J.

Most ECF σ factors are held in an inactive conformation by anti-σ factors that inhibit DNA and/or RNAP binding by either physical occlusion or conformational rearrangement of the DNA binding domains. Even otherwise, free σ factors (of the σ^{70} family) are usually unable to bind promoter DNA in the absence of RNAP as the DNA binding surfaces are occluded by inter-domain interactions. These aspects prompted a comparison of the orientation of σ_2 and σ_4 domains in σ^J and accessibility of DNA binding surface in $\sigma^J{}_2$ and $\sigma^J{}_4$ domains vis-a-vis other structures of σ factors (Figure 1B) (61,62). This analysis showed that the orientation of $\sigma^J{}_2$ and $\sigma^J{}_4$ is similar to that of the σ_2 and σ_4 domains in CnrH/CnrY (PDB ID: 4CXF) (61). In this orientation, both the DNA binding surfaces of the σ_2 domain and σ_4 are exposed. To evaluate the feasibility of DNA binding in this conformation, we superposed the structure of σ^J on a previously described σ_2/−10 DNA complex (PDB ID: 4LUP) (60) and a σ_4/−35DNA complex (PDB ID: 2H27) (63) (Figure 2A–C). While the −35 promoter DNA element could be docked without any steric hindrance, binding of the −10 promoter DNA element also appeared feasible, albeit with a few minor clashes. These observations suggest that σ^J is unlikely to rely on conformational changes induced by RNAP interactions to bind promoter DNA and that the C-terminal SnoaL_2 domain is unlikely to be inhibitory.

In vitro binding of σ^J and $\sigma^{J\Delta SnoaL\_2}$ to promoter DNA

σ^J recognizes the promoter of another ECF41 σ factor, σ^I, thereby controlling the expression of this σ factor (19). We used EMSA and SPR techniques to study interactions of different σ^J constructs with the *sigIp* DNA (Supplementary Table S4). The DNA binding studies using EMSA revealed that full-length σ^J can bind DNA whereas deletion of the SnoaL_2 domain resulted in substantially reduced DNA binding ability (Figure 2E and F). This result is consistent with observations on ECF41 σ factor based promoter activation reported earlier (16). The dissociation constant k_D determined using EMSA studies was 4.96 ± 0.94 μM. To assess whether SnoaL_2 domain can bind DNA by itself, EMSA was carried out with purified SnoaL_2 domain (σ^{JSnoaL_2}). The σ^{JSnoaL_2} protein did not show any binding to promoter DNA under the experimental conditions (Supplementary Figure S9). The DNA binding studies were also performed using SPR (Figure 2G and H). σ^J bound well to immobilized *sigIp* DNA whereas the $\sigma^{J\Delta SnoaL\_2}$ showed significantly reduced DNA binding ability. The K_{on} and K_{off} reaction constants calculated using SPR sensograms for σ^J were 1.14 ± 0.08 × 10^5 M^{-1}s^{-1} and 5.69 ± 0.53 × 10^{-3} s^{-1} respectively. EMSA and SPR experiments were also performed with a mutant DNA probe (Supplementary Table S4). These experiments provide an estimate of the specificity in DNA recognition suggesting that these specificity determinants serve to discriminate between cognate promoter

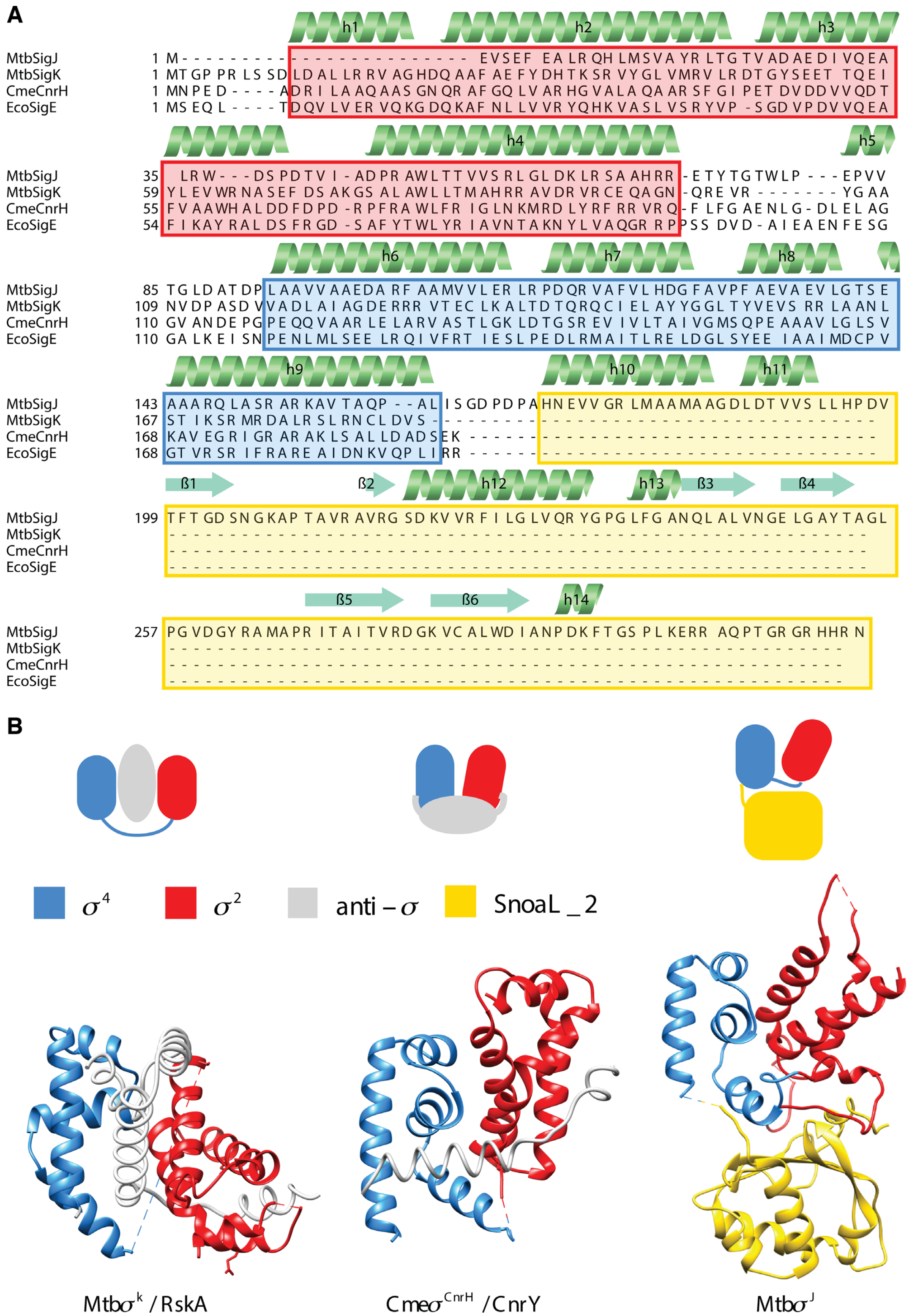

Figure 1. Conformational features of *Mycobacterium tuberculosis* σ^J. (**A**) σ^J contains a structured carboxyl terminal domain that is absent in other ECF σ factors. A structure-based sequence alignment of σ^J with other ECF σ factors reveals that of the two DNA binding domains, σ_4 is more conserved than σ_2. $\sigma^J{}_2$ lacks helix $\alpha1$ present in other σ factors and has a shorter helix $\alpha2$. This variation could alter Pribnow box recognition. (**B**) Comparison of the orientation and accessibility of DNA binding surfaces of σ_2 and σ_4 domains in σ^J with other σ/anti-σ complexes. While σ/anti-σ interactions usually bury DNA binding surfaces of σ factors, in case of σ^J these surfaces are exposed (Mtb σ^K/RskA-4NQW; Cme σ^{CnrH}/CnrY-4CXF) (61,62). The abbreviations in this illustration are—Mtb: *Mycobacterium tuberculosis*; Eco: *Escherichia coli*; Cme: *Cupriavidus metallidurans*.

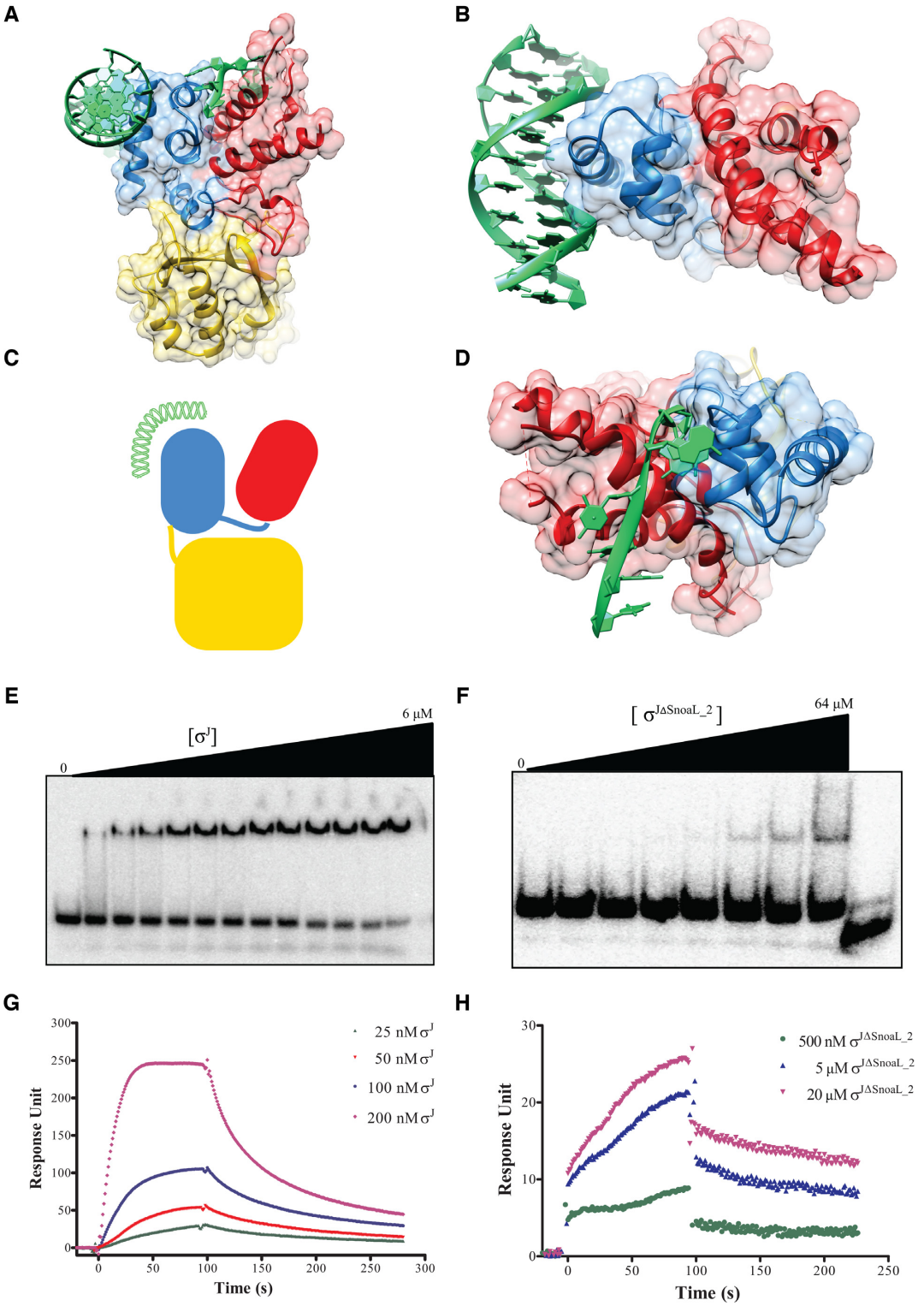

Figure 2. Exposed DNA binding surfaces suggest that σ^J adopts a conformation that can readily interact with the promoter. (**A** and **B**) Mapping of the −10 and −35 promoter DNA on $\sigma^J{}_2$ and $\sigma^J{}_4$ by superposition with previously determined structures of the −10 promoter/σ_2 (PDB ID: 4LUP) and −35 promoter/σ_4 (PDB ID: 2H27) complexes (60,63). (**C**) A schematic representation of DNA binding to σ^J. (**D**) The exposed DNA binding surfaces of $\sigma^J{}_2$ and $\sigma^J{}_4$ reveals that the SnoaI_2 domain tethers the two DNA binding domains in this conformation but is unlikely to interact with the DNA. (**E**) Radiolabeled *sigI*-promoter DNA was incubated with (E) σ^J (0–6 μM) or (**F**) $\sigma^{J\Delta SnoaL\_2}$ (0–64 μM) and separated on a non-denaturing PAGE. DNA binding is significantly reduced in $\sigma^{J\Delta SnoaL\_2}$. The wedge at the top of each panel depicts the gradient of protein concentration. The concentrations of σ^J from left to right are 0, 1, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3, 3.5, 4, 5 and 6 μM respectively. Similarly, the concentration of $\sigma^{J\Delta SnoaL\_2}$ from left to right are 0, 1, 2, 4, 8, 16, 32 and 64 μM respectively. The right-most lane in panel (F) is loaded with 5 μM σ^J as a binding control. (**G**) In surface plasmon resonance experiments varying concentrations of (G) σ^J and (**H**) $\sigma^{J\Delta SnoaL\_2}$ were passed on a chip with an immobilized s*igI*-promoter. Put together, these experiments reveal that deletion of SnoaL_2 significantly impairs σ^J–promoter DNA interaction.

Table 1. Diffraction data, phasing and refinement statistics

PDB ID	5XE7
Data collection	SelMet SAD
Space group	I222
Cell dimensions	
a, *b*, *c* (Å)	a = 74.46, b = 133.60, c = 133.50
Wavelength (Å)	0.978
Resolution (Å)	66.80–2.16 (2.28–2.16)*
R_{sym} or R_{merge}	0.072 (0.574)
I / σ(I)	25.1 (5.9)
CC(1/2)	0.99 (0.94)
Completeness (%)	99.9 (99.8)
Redundancy	22.1 (21.2)
No. mol./asymm. unit	2
Refinement	
Resolution (Å)	66.80–2.16 (2.20–2.16)
Total no. reflections	795 951 (109 453)
Unique reflections	35 959 (5166)
R_{work}/R_{free}	0.26/0.27
No. atoms	
Protein	3805
Ligand/ion	0
Water	180
Average B-factors (Å^2)	
Protein	50.2
Water	42.8
r.m.s d.	
Bond lengths (Å)	0.004
Bond angles (°)	0.852
Ramachandran favoured (%)	91.7
Ramachandran allowed (%)	7.2
Ramachandran outliers (%)	1.1
Twin operator	-h,-l,-k
Twin fraction	0.478
F_o, F_c correlation	0.93

*Values in parentheses are for highest-resolution shell.

Table 2. Interfaces between σ^J domains

Domains	Surface area (Å^2)	Interface area (Å^2)
σ^J_2 and σ^J_4	5908	1140
σ^J_2 and SnoaL_2	6410	354
σ^J_4 and SnoaL_2	6293	790

segments from other DNA sequences (supplementary Figures S7 and 8).

β-galactosidase-based reporter assays

A recent study of the *M. tuberculosis* σ factor network demonstrated the successful use of an *E. coli* model to study *M. tuberculosis* σ factor activity (56). This strategy was utilized to understand σ^J activity. Target plasmids encoding the *sigI*-promoter-*lacZ* fusion protein and donor plasmids containing σ^J and $\sigma^{J\Delta SnoaL\_2}$ were transformed in *E. coli* LMG194 strain (Figure 5). It was envisaged that differences between σ^J- and $\sigma^{J\Delta SnoaL\_2}$-driven transcription initiation would translate into variations in β-galactosidase activity. We note that the β-galactosidase activity of the carboxy terminal deletion construct $\sigma^{J\Delta SnoaL\_2}$ was lower when compared to full-length σ^J (Figure 5). These results suggest the opposite of an anti-σ like role for the SnoaL_2 domain. The SnoaL_2 domain at the C-terminus is thus more likely to adopt the role of a modulator or positive regulator of transcription by conformational re-arrangement of the two DNA binding domains in σ^J.

Molecular dynamics simulations

To study the effect of the SnoaL_2 domain on the structure of σ^J, MD simulations were performed on σ^J and $\sigma^{J\Delta SnoaL\_2}$ for 500 ns of simulation time. σ^J stabilized within 20 ns and shows less variation in the C_α-backbone afterward whereas the $\sigma^{J\Delta SnoaL\_2}$ model showed significant variations in C_α-backbone RMSD even after 200 ns of simulation time (Supplementary Figure S5). To evaluate these differences, a histogram of RMSD was plotted against number of conformers for σ^J and $\sigma^{J\Delta SnoaL\_2}$ (Figure 4). The RMSD-values for σ^J varies from ∼2.5 to 4 Å whereas RMSD for $\sigma^{J\Delta SnoaL\_2}$ extended beyond 5 Å. This analysis suggests that removal of the SnoaL_2 domain leads to substantial conformational heterogeneity. These results were also supported by analytical size exclusion chromatography experiments which suggest $\sigma^{J\Delta SnoaL\_2}$ can adopt an extended conformation (Figure 4). The elution profile of $\sigma^{J\Delta SnoaL\_2}$ suggested a significantly

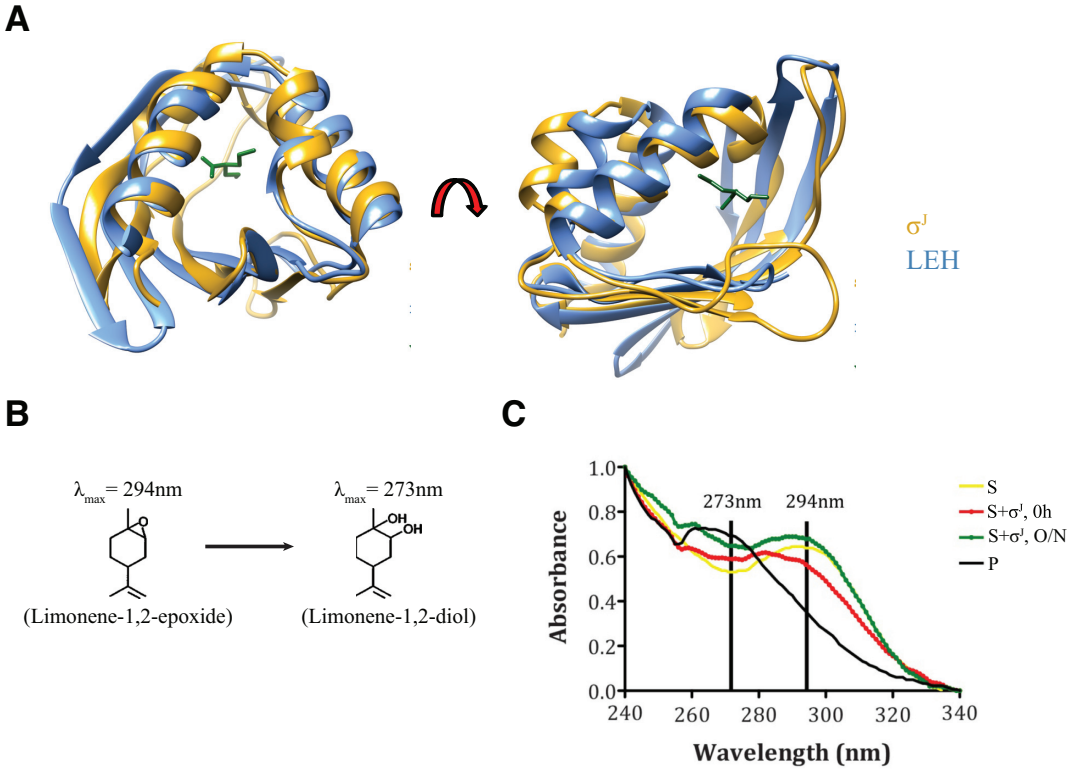

Figure 3. Structural features suggest a ligand-binding role for the SnoaL_2 domain of σ^J. (**A**) Superposition of σ^{JSnoaL_2} with Limolene-1,2-epoxide hydrolase (LEH). The core region of LEH aligned well with σ^{JSnoaL_2} with an RMSD of 1.13Å over 44 atom pairs (LEH: PDB ID: 1NWW) (27). (**B**) LEH catalyzes conversion of Limonene-1,2-epoxide ($\lambda_{max} = 294$ nm) into Limonene-1,2-diol ($\lambda_{max} = 273$ nm). (**C**) Spectroscopic differences upon the incubation of Limonene-1,2-epoxide with σ^J. In this figure, S represents the substrate (Limonene-1,2-epoxide) while P refers to the product (Limonene-1,2-diol).

higher mass than that of a monomer (*ca* 28 kDa instead of 19 kDa) in a size exclusion chromatography experiment performed on a Superdex-200 analytical column. This finding that the SnoaL_2 domain confers structural compactness to σ^J was further evaluated by a limited proteolysis experiment. This experiment, correlating proteolytic susceptibility with conformational flexibility, was performed using trypsin (1:60 mass ratio). A comparison between the proteolytic susceptibility of σ^J and $\sigma^{J\Delta SnoaL\_2}$ revealed that $\sigma^{J\Delta SnoaL\_2}$ is significantly more proteolytically labile when compared to full length σ^J (Figure 4C).

The SnoaL_2 domain in σ^J

A search on the DALI database for proteins that were structurally similar to the SnoaL_2 domain of σ^J revealed several potential matches (48). After pruning for redundancy, a total of *ca* 300 different structures were collated. This list included limonene-1, 2-epoxide hydrolases (LEH), polyketide cyclases, ketosteroid isomerases, epoxide hydrolases/cyclases, Ca^{2+}/calmodulin-dependent protein kinase subunits and the nuclear transport factor 2 (NTF2). Of these, the limonene epoxide hydrolases were the closest experimentally characterized proteins (Table 3).

The crystal structure of the SnoaL_2 domain revealed substantial structural similarity with the *Rhodococcus erythropolis* Limonene-1,2-Epoxide Hydrolase (LEH) (27) (DALI Z-score of 10.8; PDB ID: 1NWW; RMSD of 1.13 Å over 44 atom pairs; Figure 3A). LEH catalyses the conversion of Limonene-1,2-Epoxide ($\lambda_{max} = 294$nm) into Limonene-1,2-diol ($\lambda_{max} = 273$nm) (Figure 3B). In an effort to evaluate if σ^J could perform a similar role, freshly purified σ^J was incubated with Limonene-1,2-Epoxide. We note that incubation of Limonene-1,2-Epoxide with σ^J results in differences in the UV-visible spectra (Figure 3C). While this suggests a potential enzymatic role for the SnoaL_2 domain of σ^J, Limonene-1,2-Epoxide appears unlikely to be a physiologically relevant substrate. This observation, however, suggests a role for the SnoaL_2 domain in ligand binding potentially incorporating the role of a receptor modulating σ^J activity.

DISCUSSION

The activity of ECF σ factors has been demonstrated to be governed at the transcriptional, post-transcriptional and post-translational levels (64,65). This multilayered regulation ensures an appropriate cellular concentration of an ECF σ factor that can compete for RNAP binding and consequently, the expression of the cognate regulon. The ECF41 family of ECF σ factors suggests a variation to this theme. ECF41 σ factors do not have a cognate antagonist protein (anti-σ factor). Indeed, the C-terminal polypeptide was inferred to perform this regulatory function. This hypothesis, however, could not be validated in the absence of structural data on this family of σ factors. Three key findings emerge from the crystal structure

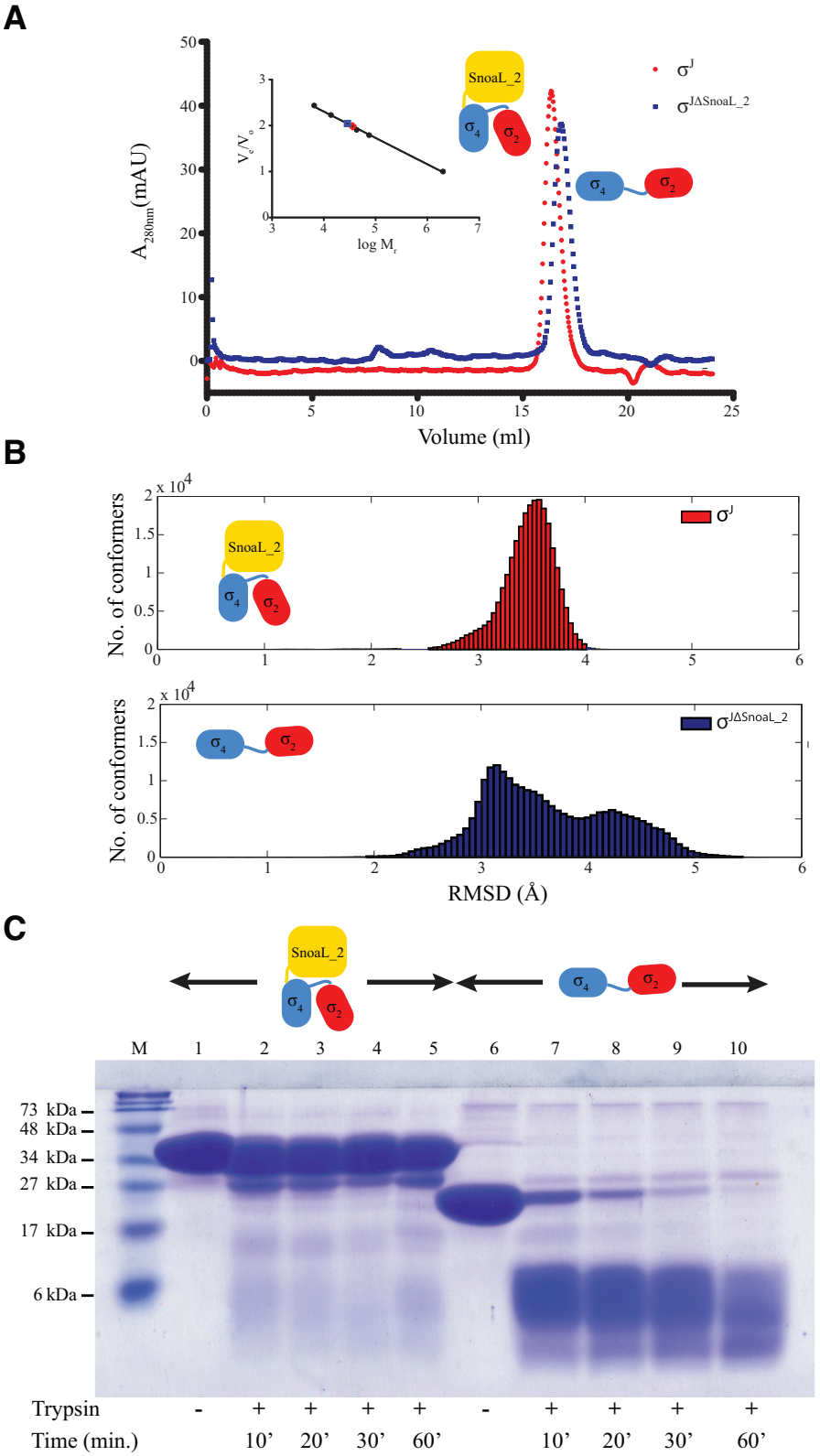

Figure 4. The SnoaL_2 domain contributes to a compact σ^J conformation. (**A**) Comparison of analytical size exclusion chromatogram of σ^J and $\sigma^{J\Delta SnoaL\_2}$. The apparent molar mass (28.5 kDa) of the $\sigma^{J\Delta SnoaL\_2}$ protein is higher than expected (19.1 kDa) for a monomer. (**B**) Molecular dynamics (MD) simulations provide a basis for a comparison between the conformations sampled by σ^J and $\sigma^{J\Delta SnoaL\_2}$. The considerable conformational heterogeneity noted in the MD simulations for $\sigma^{J\Delta SnoaL\_2}$ is consistent with the hypothesis that the SnoaL_2 domain constraints σ^J to a compact structure. (**C**) A comparison between the proteolytic susceptibility of σ^J and $\sigma^{J\Delta SnoaL\_2}$ (using trypsin) revealed that $\sigma^{J\Delta SnoaL\_2}$ is significantly more proteolytically labile when compared to full length σ^J.

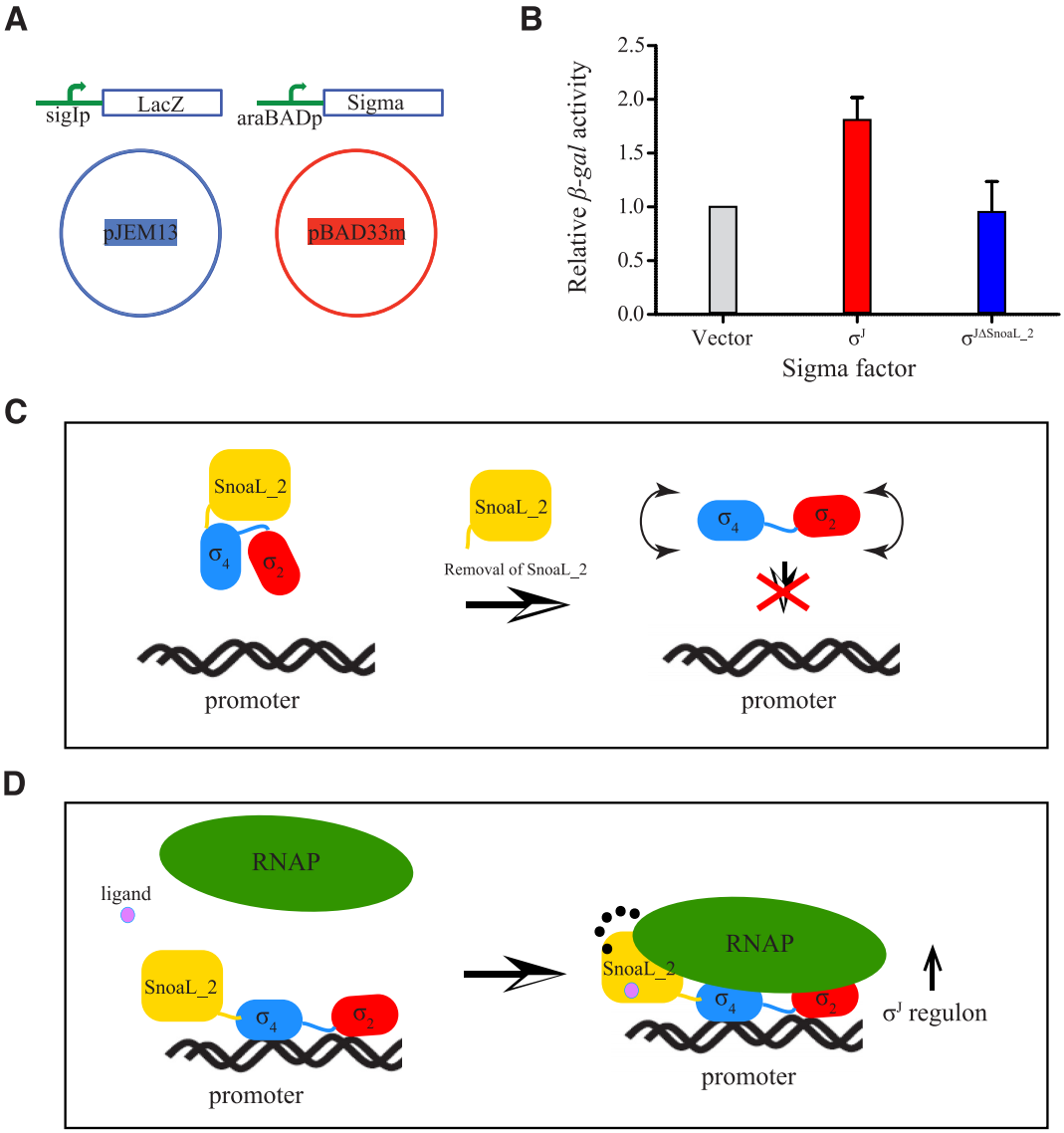

Figure 5. The SnoaL_2 domain modulates σ^J activity. (**A**) A schematic representation of the β-galactosidase assay to evaluate σ^J activity. The target plasmid contained a *sigJ*-promoter fused to a *lacZ* gene cloned in the pJEM13 vector (56). The donor plasmid contained either σ^J or $\sigma^{J\Delta SnoaL\_2}$ cloned in the pBAD33m vector. The pBAD33m vector served as a control in this experiment. (**B**) β-galactosidase activity was significantly reduced upon deleting the SnoaL_2 domain ($\sigma^{J\Delta SnoaL\_2}$) when compared to the full-length σ factor. The SnoaL_2 domain is thus unlikely to play the role of a σ factor antagonist (anti-σ factor). (**C** and **D**) A mechanistic model for σ^J activity. The influence of the Snoal_2 domain on DNA binding and transcription initiation suggests that the C-terminal domain is an essential part of σ^J (C). The SnoaL_2 domain tethers $\sigma^J{}_2$ and $\sigma^J{}_4$ in an orientation compatible with promoter DNA binding and potentially influences σ^J–RNAP interaction (16). These interactions could be modulated by the binding of small molecules to the SnoaL_2 domain (D). The SnoaL_2 domain can thus modulate σ^J activity by inducing conformational changes that regulate promoter DNA or RNAP binding.

Table 3. Functional features of proteins similar to σ^{JSnoaL_2}

PDB ID	Z-score	Function (ligand/substrate/inhibitor)	Organism (source)
1NWW	10.8	Limonene-1, 2-epoxide hydrolase	*Rhodococcus eryhtropolis*
3B4P	8.5	Phenazine biosynthesis protein [2-(cyclohexyl amino) benzoic acid]	*Burkholderia cepacia*
2A15	10.1	Rv0760 (Estradiol-17beta-hemisuccinate)	*Mycobacterium tuberculosis*
4CDL	10.1	Retro-aldolase (inhibitor: 1-(6-methoxy-2-naphthalenyl)-1,3-butanedione)	*Pseudomonas putida*
1BUQ	8.7	Delta-5-3-ketosteroid isomerase (Steroid 19-nortestosterone-hemisuccinate)	*Comamonas testosteroni*
1GY7	8.8	NTF2 (Transport/Nuclear protein)	*Saccharomyces cereviviae*
5IG3	8.8	Ca(2+)/calmodulin-dependent protein kinase II, α hub	*Homo sapiens*

and *in vitro* and *ex vivo* characterization of *M. tuberculosis* σ^J—(i) exposed DNA binding surfaces in σ^J enable this initiation factor to bind DNA (independent of cognate promoter recognition)—other σ factors rely on conformational changes induced by RNAP binding for DNA interactions; (ii) Interactions between the σ^J_2–σ^J_4 loop and the SnoaL_2 domain keeps σ^J structure in a compact conformation and removal of the SnoaL_2 domain leads to a flexible arrangement with substantially impaired DNA binding; (iii) Structural similarities suggest a possible role for the SnoaL_2 domain in protein-ligand interactions potentially modulating σ^J activity. Furthermore, the structures of the DNA binding domains in σ^J themselves suggest functional adaptation. While the σ_4 domain is well conserved, the σ_2 domain has differences when compared to other structures. These variations in σ^J_2—absence of the α1 helix which is present just behind the specificity determining L3 loop and a shorter and rigid α2 helix could potentially alter the DNA melting properties of σ^J.

The compact arrangement of the σ_2, σ_4 and SnoaL_2 domain is likely to be conserved across ECF41 proteins. Most of the conserved residues in the SnoaL_2 domain are positioned in the first two helices (Supplementary Figure S10). Residues in the σ^J_2–σ^J_4 loop which stack against the β-sheet surface of σ^{JSnoaL_2} are also conserved. In a related observation, analytical size exclusion chromatography suggested that although both σ^J and $\sigma^{J\Delta SnoaL\_2}$ are monomers in solution, the observed molar mass of $\sigma^{J\Delta SnoaL\_2}$ is higher than expected. MD simulations also revealed significant differences in the RMSD profile of $\sigma^{J\Delta SnoaL\_2}$ compared to σ^J. The root-mean-square fluctuations revealed increased flexibility in the σ^J_2–σ^J_4 linker, σ^J_4–σ^{JSnoaL_2} linker and the L3 loop (between α2–α3 helices in σ^J_2) in the absence of the SnoaL_2 domain. The σ^J_2–σ^J_4 linker interacts with the β-sheet surface of σ^{JSnoaL_2} and constrains σ^J to a compact conformation. Together, this suggests that σ^J is conformationally more rigid than $\sigma^{J\Delta SnoaL\_2}$. σ^J binds the *sigI*-promoter DNA *in vitro*. While DNA binding was significantly reduced upon deletion of the SnoaL_2 domain, the SnoaL_2 domain alone did not show any DNA binding. Transcription experiments in *E. coli* revealed that σ^J could initiate transcription of reporter genes fused to the *sigI*-promoter. Deletion of the SnoaL_2 domain resulted in a significant decrease in σ^J activity. This suggests that the SnoaL_2 domain is not an antagonist (as an anti-σ factor) but is actively involved in transcription initiation. Although σ^I, another ECF41 σ factor, is the only experimentally validated gene under σ^J regulation so far, σ^J-driven expression of other genes under different environmental conditions cannot be excluded.

ECF σ factors represent the third most abundant mechanism of signal transduction after one component system (OCS) and two component systems (TCS) (11,66,67). While TCS have a separate sensor and effector, OCS possess a simpler design, wherein both the sensor and effector domains are fused together in a single polypeptide chain (66,68,69). ECF σ factors differ from OCS and TCS as they form part of the RNAP holo-enzyme complex. Environment-mediated regulation of ECF σ factors is primarily governed by anti-σ factors that employ diverse mechanisms involving phosphorylation, redox-dependent conformational changes and proteolysis. Thus, the anti-σ factors and other regulatory proteases/kinases together serve as the sensors for environmental stimuli while the DNA binding domains of an ECF σ factor perform the effector function. While we could not trap or identify interacting proteins with the SnoaL_2 domain, transiently interacting protein partners cannot be excluded. Nonetheless, we note that the SnoaL_2 domain is likely to interact with ligands or indeed substrates, suggesting a possible role of ligand-mediated regulation of σ^J activity. Indeed, the possibility of an anti-σ function for the Snoal_2 domain is likely if and when activated by a signal in the presence of a specific, as yet unidentified ligand.

Put together, the structural and biochemical data suggest that the SnoaL_2 domain in σ^J is unlikely to play the role of an antagonist for σ^J-mediated transcription. Rather, the ability of this domain to confer a DNA-binding conformation to σ^J points to a role as modulator of activity. The Snoal_2 domain by itself cannot bind DNA, thus excluding the possibility of this domain to influence transcription bubble formation directly. It thus appears likely that ligand and/or protein-based interactions could alter the conformation of σ^J providing a regulatory mechanism based on inter-domain conformational rearrangement. ECF41 σ factors thus seem to blur the distinction between the traditional definition of a σ factor as a unit of the RNAP and a one-component system with a defined role of a context-dependent transcriptional regulator.

ACCESSION NUMBER

Atomic coordinates and structure factors for the reported crystal structure have been deposited with the Protein Data Bank under accession number 5XE7.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Funding from the Ministry for Human Resource Development for the Indian Institute of Science is gratefully acknowledged. We gratefully acknowledge the help, advice and constructive criticisms from Dr Abhijit Sardesai (Center for DNA Fingerprinting and Diagnostics, Hyderabad), Dr Asmita Gupta (Molecular Biophysics Unit, IISc) and Dr Roshan Singh Thakur (Department of Biochemistry, IISc) at different stages of this study. Prof. Marila Gennaro and colleagues (Public Health Research Institute Center, New Jersey, USA) are gratefully acknowledged for the kind gift of the plasmids used for the β-galactosidase assay.

FUNDING

Department of Biotechnology, Government of India (in part); Department of Science and Technology-Science and Engineering Research Board [IR/SO/LU003/2010-Phase-II]; Department of Science and Technology-Fund for Improvement of S&T Infrastructure in Higher Educational Institutions.(in part) Funding for open access charge: Career Development Allowance (to B.G.).

Conflict of interest statement. None declared.

REFERENCES

1. Burgess,R.R., Travers,A.A., Dunn,J.J. and Bautz,E.K. (1969) Factor stimulating transcription by RNA polymerase. *Nature*, **221**, 43–46.
2. Paget,M.S. and Helmann,J.D. (2003) The sigma70 family of sigma factors. *Genome Biol.*, **4**, 203.
3. Ghosh,T., Bose,D. and Zhang,X. (2010) Mechanisms for activating bacterial RNA polymerase. *FEMS Microbiol. Rev.*, **34**, 611–627.
4. Lonetto,M., Gribskov,M. and Gross,C.A. (1992) The sigma 70 family: sequence conservation and evolutionary relationships. *J. Bacteriol.*, **174**, 3843–3849.
5. Lonetto,M.A., Brown,K.L., Rudd,K.E. and Buttner,M.J. (1994) Analysis of the Streptomyces coelicolor sigE gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 7573–7577.
6. Gruber,T.M. and Gross,C.A. (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.*, **57**, 441–466.
7. Hughes,K.T. and Mathee,K. (1998) The anti-sigma factors. *Annu. Rev. Microbiol.*, **52**, 231–286.
8. Bae,J.B., Park,J.H., Hahn,M.Y., Kim,M.S. and Roe,J.H. (2004) Redox-dependent changes in RsrA, an anti-sigma factor in Streptomyces coelicolor: zinc release and disulfide bond formation. *J. Mol. Biol.*, **335**, 425–435.
9. Park,S.T., Kang,C.M. and Husson,R.N. (2008) Regulation of the SigH stress response regulon by an essential protein kinase in Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 13105–13110.
10. Jaiswal,R.K., Prabha,T.S., Manjeera,G. and Gopal,B. (2013) Mycobacterium tuberculosis RsdA provides a conformational rationale for selective regulation of σ-factor activity by proteolysis. *Nucleic Acids Res.*, **41**, 3414–3423.
11. Staroń,A., Sofia,H.J., Dietrich,S., Ulrich,L.E., Liesegang,H. and Mascher,T. (2009) The third pillar of bacterial signal transduction: Classification of the extracytoplasmic function (ECF) σ factor protein family. *Mol. Microbiol.*, **74**, 557–581.
12. Jogler,C., Waldmann,J., Huang,X., Jogler,M., Glöckner,F.O., Mascher,T. and Kolter,R. (2012) Identification of proteins likely to be involved in morphogenesis, cell division, and signal transduction in Planctomycetes by comparative genomics. *J. Bacteriol.*, **194**, 6419–6430.
13. Gómez-Santos,N., Pérez,J., Sánchez-Sutil,M.C., Moraleda-Muñoz,A. and Muñoz-Dorado,J. (2011) CorE from Myxococcus xanthus is a copper-dependent RNA polymerase sigma factor. *PLoS Genet.*, **7**, e1002106.
14. Marcos-Torres,F.J., Pérez,J., Gómez-Santos,N., Moraleda-Muñoz,A. and Muñoz-Dorado,J. (2016) In depth analysis of the mechanism of action of metal-dependent sigma factors: characterization of CorE2 from Myxococcus xanthus. *Nucleic Acids Res.*, **44**, 5571–5584.
15. Muñoz-Dorado,J., Gómez-Santos,N. and Pérez,J. (2012) A novel mechanism of bacterial adaptation mediated by copper-dependent RNA polymerase σ factors. *Transcription*, **3**, 63–67.
16. Wecke,T., Halang,P., Staroń,A., Dufour,Y.S., Donohue,T.J. and Mascher,T. (2012) Extracytoplasmic function σ factors of the widely distributed group ECF41 contain a fused regulatory domain. *Microbiologyopen*, **1**, 194–213.
17. Manganelli,R., Provvedi,R., Rodrigue,S., Beaucher,J., Gaudreau,L. and Smith,I. (2004) Sigma factors and global gene regulation in Mycobacterium tuberculosis. *J. Bacteriol.*, **186**, 895–902.
18. Hu,Y. and Coates,A.R.M. (2001) Increased levels of sigJ mRNA in late stationary phase cultures of Mycobacterium tuberculosis detected by DNA array hybridisation. *FEMS Microbiol. Lett.*, **202**, 59–65.
19. Homerova,D., Halgasova,L. and Kormanec,J. (2008) Cascade of extracytoplasmic function sigma factors in Mycobacterium tuberculosis: Identification of a σ^J-dependent promoter upstream of sigI. *FEMS Microbiol. Lett.*, **280**, 120–126.
20. Hu,Y., Kendall,S., Stoker,N.G. and Coates,A.R.M. (2004) The Mycobacterium tuberculosis sigJ gene controls sensitivity of the bacterium to hydrogen peroxide. *FEMS Microbiol. Lett.*, **237**, 415–423.
21. Pettersson,B.M., Das,S., Behra,P.R., Jordan,H.R., Ramesh,M., Mallick,A., Root,K.M., Cheramie,M.N., de la Cruz Melara,I., Small,P.L. *et al.* (2015) Comparative sigma factor-mRNA levels in Mycobacterium marinum under stress conditions and during host infection. *PLoS One*, **10**, e0139823.
22. Finn,R.D., Coggill,P., Eberhardt,R.Y., Eddy,S.R., Mistry,J., Mitchell,A.L., Potter,S.C., Punta,M., Qureshi,M., Sangrador-Vegas,A. *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.
23. Bullock,T.L., Clarkson,W.D., Kent,H.M. and Stewart,M. (1996) The 1.6 angstroms resolution crystal structure of nuclear transport factor 2 (NTF2). *J. Mol. Biol.*, **260**, 422–431.
24. Eberhardt,R.Y., Chang,Y., Bateman,A., Murzin,A.G., Axelrod,H.L., Hwang,W.C. and Aravind,L. (2013) Filling out the structural map of the NTF2-like superfamily. *BMC Bioinformatics*, **14**, 327.
25. Sultana,A., Kallio,P., Jansson,A., Wang,J.S., Niemi,J., Mantsala,P. and Schneider,G. (2004) Structure of the polyketide cyclase SnoaL reveals a novel mechanism for enzymatic aldol condensation. *EMBO J.*, **23**, 1911–1921.
26. Nakasako,M., Motoyama,T., Kurahashi,Y. and Yamaguchi,I. (1998) Cryogenic X-ray crystal structure analysis for the complex of scytalone dehydratase of a rice blast fungus and its tight-binding inhibitor, carpropamid: the structural basis of tight-binding inhibition. *Biochemistry*, **37**, 9931–9939.
27. Arand,M., Hallberg,B.M., Zou,J., Bergfors,T., Oesch,F., van der Werf,M.J., de Bont,J.A., Jones,T.A. and Mowbray,S.L. (2003) Structure of Rhodococcus erythropolis limonene-1,2-epoxide hydrolase reveals a novel active site. *EMBO J.*, **22**, 2583–2592.
28. Kim,S.W., Cha,S.S., Cho,H.S., Kim,J.S., Ha,N.C., Cho,M.J., Joo,S., Kim,K.K., Choi,K.Y. and Oh,B.H. (1997) High-resolution crystal structures of delta5-3-ketosteroid isomerase with and without a reaction intermediate analogue. *Biochemistry*, **36**, 14030–14036.
29. Griffith,L.C., Lu,C.S. and Sun,X.X. (2003) CaMKII, an enzyme on the move: regulation of temporospatial localization. *Mol. Interv.*, **3**, 386–403.
30. Goutam,K., Gupta,A.K. and Gopal,B. (2015) Crystallographic studies of the extracytoplasmic function sigma factor sigma(J) from Mycobacterium tuberculosis. *Acta Crystallogr. F Struct. Biol. Commun.*, **71**, 946–950.
31. Battye,T.G., Kontogiannis,L., Johnson,O., Powell,H.R. and Leslie,A.G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 271–281.
32. Evans,P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 72–82.
33. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G., McCoy,A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.
34. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
35. Terwilliger,T.C., Adams,P.D., Read,R.J., McCoy,A.J., Moriarty,N.W., Grosse-Kunstleve,R.W., Afonine,P.V., Zwart,P.H. and Hung,L.W. (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D Biol. Crystallogr.*, **65**, 582–601.
36. Langer,G., Cohen,S.X., Lamzin,V.S. and Perrakis,A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.*, **3**, 1171–1179.
37. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
38. Davis,I.W., Leaver-Fay,A., Chen,V.B., Block,J.N., Kapral,G.J., Wang,X., Murray,L.W., Arendall,W.B. 3rd, Snoeyink,J., Richardson,J.S. *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–W383.
39. Krissinel,E. and Henrick,K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.

40. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
41. Alva,V., Nam,S.Z., Soding,J. and Lupas,A.N. (2016) The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res.*, **44**, W410–W415.
42. Pei,J., Kim,B.H. and Grishin,N.V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.*, **36**, 2295–2300.
43. Celniker,G., Nimrod,G., Ashkenazy,H., Glaser,F., Martz,E., Mayrose,I., Pupko,T. and Ben-Tal,N. (2013) ConSurf: using evolutionary data to raise testable hypotheses about protein function. *Isr. J. Chem.*, **53**, 199–206.
44. Ashkenazy,H., Abadi,S., Martz,E., Chay,O., Mayrose,I., Pupko,T. and Ben-Tal,N. (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.*, **44**, W344–W350.
45. Ashkenazy,H., Erez,E., Martz,E., Pupko,T. and Ben-Tal,N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.*, **38**, W529–W533.
46. Glaser,F., Pupko,T., Paz,I., Bell,R.E., Bechor-Shental,D., Martz,E. and Ben-Tal,N. (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, **19**, 163–164.
47. Landau,M., Mayrose,I., Rosenberg,Y., Glaser,F., Martz,E., Pupko,T. and Ben-Tal,N. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.*, **33**, W299–W302.
48. Holm,L. and Laakso,L.M. (2016) Dali server update. *Nucleic Acids Res.*, **44**, W351–W355.
49. Fiser,A., Do,R.K. and Sali,A. (2000) Modeling of loops in protein structures. *Protein Sci.*, **9**, 1753–1773.
50. Fiser,A. and Sali,A. (2003) ModLoop: automated modeling of loops in protein structures. *Bioinformatics*, **19**, 2500–2501.
51. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
52. Reyes,C.M. and Kollman,P.A. (2000) Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. *J. Mol. Biol.*, **297**, 1145–1158.
53. Ryckaert,J.-P., Ciccotti,G. and Berendsen,H.J.C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*, **23**, 327–341.
54. Roe,D.R. and Cheatham,T.E. III (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.*, **9**, 3084–3095.
55. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
56. Chauhan,R., Ravi,J., Datta,P., Chen,T., Schnappinger,D., Bassler,K.E., Balazsi,G. and Gennaro,M.L. (2016) Reconstruction and topological characterization of the sigma factor regulatory network of Mycobacterium tuberculosis. *Nat. Commun.*, **7**, 11062.
57. Gasteiger,E., Hoogland,C., Gattiker,A., Duvaud,S.e., Wilkins,M.R., Appel,R.D and Bairoch,A. (2005) Protein identification and analysis tools on the ExPASy server. In: Walker,JM (ed). *The Proteomics Protocols Handbook*. Humana Press, Totowa, pp. 571–607.
58. Hendrickson,W.A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**, 51–58.
59. Campbell,E.A., Muzzin,O., Chlenov,M., Sun,J.L., Olson,C.A., Weinman,O., Trester-Zedlitz,M.L. and Darst,S.A. (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol. Cell*, **9**, 527–539.
60. Campagne,S., Marsh,M.E., Capitani,G., Vorholt,J.A. and Allain,F.H.T. (2014) Structural basis for -10 promoter element melting by environmentally induced sigma factors. *Nat. Struct. Mol. Biol.*, **21**, 269–276.
61. Maillard,A.P., Girard,E., Ziani,W., Petit-Härtlein,I., Kahn,R. and Covès,J. (2014) The crystal structure of the anti-σ factor CnrY in complex with the σ factor CnrH shows a new structural class of anti-σ factors targeting extracytoplasmic function σ factors. *J. Mol. Biol.*, **426**, 2313–2327.
62. Shukla,J., Gupta,R., Thakur,K.G., Gokhale,R. and Gopal,B. (2014) Structural basis for the redox sensitivity of the *Mycobacterium tuberculosis* SigK-RskA σ-anti-σ complex. *Acta Crystallogr. D Biol. Crystallogr.*, **70**, 1026–1036.
63. Lane,W.J. and Darst,S.A. (2006) The structural basis for promoter -35 element recognition by the group IV sigma factors. *PLoS Biol.*, **4**, e269.
64. Raman,S., Cascioferro,A., Husson,R. and Manganelli,R. (2008) Mycobacterial sigma factors and surface biology. In: Daffe,M and Reyrat,JM (eds). *The Mycobacterial Cell Envelope*. ASM Press, Washington, DC, pp. 223–234.
65. Sachdeva,P., Misra,R., Tyagi,A.K. and Singh,Y. (2010) The sigma factors of Mycobacterium tuberculosis: regulation of the regulators. *FEBS J.*, **277**, 605–626.
66. Ulrich,L.E., Koonin,E.V. and Zhulin,I.B. (2005) One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol.*, **13**, 52–56.
67. Desai,S.K. and Kenney,L.J. (2017) To approximately P or Not to approximately P? Non-canonical activation by two-component response regulators. *Mol. Microbiol.*, **103**, 203–213.
68. Gebhard,S., Gaballa,A., Helmann,J.D. and Cook,G.M. (2009) Direct stimulus perception and transcription activation by a membrane-bound DNA binding protein. *Mol. Microbiol.*, **73**, 482–491.
69. Iida,T., Waki,T., Nakamura,K., Mukouzaka,Y. and Kudo,T. (2009) The GAF-like-domain-containing transcriptional regulator DfdR is a sensor protein for dibenzofuran and several hydrophobic aromatic compounds. *J. Bacteriol.*, **191**, 123–134.