Structural analysis of the role of the two conserved motifs of the ECF41 family sigma factor in the autoregulation of its own promoter in Azospirillum brasilense Sp245

Ekta Pathak[1], AshutoshPrakash Dubey[1], Vijay Shankar Singh[1], Rajeev Mishra[1], and Anil Kumar Tripathi[1]

[1]Banaras Hindu University Faculty of Science

September 25, 2021

Abstract

In *Azospirillum brasilense,* an extra-cytoplasmic function sigma factor (RpoE10) shows the characteristic 119 amino acid long C-terminal extension found in ECF41-type sigma factors, which possesses three conserved motifs (WLPEP, DGGGR, and NPDKV), one in the linker region between the sigma ^{2} and sigma ^{4}, and the other two in the SnoaL_2 domain of the C-terminal extension. Here, we have described the role of the two conserved motifs in the SnoaL_2 domain of RpoE10 in the inhibition and activation of its activity, respectively. Truncation of the distal part of the C-terminal sequence of the RpoE10 (including NPDKV but excluding the DGGGR motif) results in its promoter's activation suggesting autoregulation. Further truncation of the C-terminal sequence up to its proximal part, including NPDKV and DGGGR motif, abolished promoter activation. Replacement of NPDKV motif with NAAAV in RpoE10 increased its ability to activate its promoter, whereas replacement of DGGGR motif led to reduced promoter activation. We have explored the dynamic modulation of sigma2 – sigma4 domains and the relevant molecular interactions mediated by the two conserved motifs of the SnoaL2 domain using molecular dynamics simulation. The analysis enabled us to explain that the NPDKV motif located distally in the C-terminus negatively impacts transcriptional activation. In contrast, the DGGGR motif found proximally of the C-terminal extension is required to activate RpoE1

Στρυστυραλ αναλψσις οφ τηε ρολε οφ τηε τωο ϛονσερεδ μοτιφς οφ τηε E͡Φ41 φαμιλψ σ φαϛτορ ιν τηε αυτορεγυλατιον οφ ιτς οων προμοτερ ιν*Αζοσπιριλλυμ βρασιλενσε* Σπ245

Ekta Pathak[2]+ , AshutoshPrakash Dubey[1]+ , Vijay Shankar Singh[1], Rajeev Mishra[2]*, and Anil Kumar Tripathi[1]*

[1]School of Biotechnology, Institute of Science, Banaras Hindu University, Varanasi-221005, India.

[2]Bioinformatics, MMV, Institute of Science, Banaras Hindu University, Varanasi-221005, India.

+ Joint first authors

* To whom correspondence should be addressed:

Anil Kumar Tripathi,

School of Biotechnology, Institute of Science, Banaras Hindu University, Varanasi-221005, India.

Tel:+91 9451525811

Fax: +542-368693/2368174

Email:tripathianil@rediffmail.com

Rajeev Mishra,

Bioinformatics, MMV, Institute of Science, Banaras Hindu University, Varanasi-221005, India.

Tel: +91 9935338891

Email:rajeev17@bhu.ac.in

Στρυστυραλ αναλψσις οφ τηε ρολε οφ τηε τωο ςονσερεδ μοτιφς οφ τηε Eˆ Φ41 φαμιλψ σ φαςτορ ιν τηε αυτορεγυλατιον οφ ιτς οων προμοτερ ιν*Αζοσπιριλλυμ βρασιλενσε* Σπ245

ABSTRACT

In *Azospirillum brasilense,* an extra-cytoplasmic function σ factor (RpoE10) shows the characteristic 119 amino acid long C-terminal extension found in ECF41-type σ factors, which possesses three conserved motifs (WLPEP, DGGGR, and NPDKV), one in the linker region between the σ^2 and σ^4, and the other two in the SnoaL_2 domain of the C-terminal extension. Here, we have described the role of the two conserved motifs in the SnoaL_2 domain of RpoE10 in the inhibition and activation of its activity, respectively. Truncation of the distal part of the C-terminal sequence of the RpoE10 (including NPDKV but excluding the DGGGR motif) results in its promoter's activation suggesting autoregulation. Further truncation of the C-terminal sequence up to its proximal part, including NPDKV and DGGGR motif, abolished promoter activation. Replacement of NPDKV motif with NAAAV in RpoE10 increased its ability to activate its promoter, whereas replacement of DGGGR motif led to reduced promoter activation. We have explored the dynamic modulation of ?2 – ?4 domains and the relevant molecular interactions mediated by the two conserved motifs of the SnoaL2 domain using molecular dynamics simulation. The analysis enabled us to explain that the NPDKV motif located distally in the C-terminus negatively impacts transcriptional activation. In contrast, the DGGGR motif found proximally of the C-terminal extension is required to activate RpoE10.

Keywords: ECF41 family, Promoter regulation, Molecular dynamics simulation

INTRODUCTION

Bacteria sense the fluctuations in their external environment and respond by expressing genes required for adapting to the altered environmental conditions. Expression of new sets of genes is initiated at promoter sequences recognized explicitly by RNA polymerase with specific sigma (σ) factors. While a primary housekeeping σ factor initiates gene expression in exponentially growing cells, alternative σ factors are activated under specific conditions to control the expression of a specific set of genes by recognizing alternative promoter sequences [1,2]. Based on their sequence, domain architecture, and function, σ factors of the σ^{70} family are divided into four groups[3,4]. The primary σ factor belongs to Group 1 and contains four highly conserved domains (designated σ^1 through σ^4) along with a non-conserved region [3]. Group 2 σ factors are closely related to Group 1 but are not essential for growth. However, the Group 3 σ factors lack the σ^1 domain and control cellular processes such as sporulation, flagella biosynthesis, or heat shock response. Group 4 constitutes the largest and most diverse group of σ factors that regulate the cellular response to extracellular stimuli, hence known as extra-cytoplasmic function (ECF) σ factors[5-7]. In contrast to the other σ^{70} family members, the ECF σ factors contain only two of the four conserved domains, σ^2 and σ^4, which are enough for promoter recognition and interaction with core enzyme.

Based on sequence similarity and conservation of genomic context, ECF σ factors have been subdivided into 40 phylogenetically distinct groups[8]. An ECF σ factor is usually co-transcribed with a gene encoding its cognate anti-σ factor, regulating the σ factor activity [6,8,9]. ECF σ factors are also characterized by auto-regulation of their promoter. Genes encoding ECF σ factor and anti- σ factor are often organized as part of the same operon[6,8-10]. After their expression, anti-σ factors sequester their cognate ECF σ factors to occlude their binding to the core enzyme and their cognate promoters. Specific intracellular or extracellular stimuli inactivate the anti-σ factor either by changing its conformation or by proteolytical degradation [9-11]. This sets the ECF σ factor free to associate with the core enzyme to initiate transcription from its target promoters. Many ECF σ factors, however, are not associated with anti-σ factors. Instead, they harbor a

C-terminal extension fused to the σ^4 domain with the help of a flexible linker [8,9,12]. A conserved SnoaL_2 like domain (Pfam: PF12680) in the C-terminal extension of the ECF41 family of σ factors was thought to play a dual role as an activator and inhibitor of the ECF σ activity [13] by interacting with the core regions of the ECF41 σ factor [14].

Azospirillumbrasilense is a plant growth-promoting rhizobacterium, which colonizes many grasses' roots and promotes their growth by producing phytohormones and fixing atmospheric nitrogen. The genome of *A. brasilense* encodes a primary σ factor and 22 alternative σ factors. Out of its 10 ECF σ factors, two are co-transcribed with and translationally coupled to their cognate Zinc-binding anti-σ factors [15,16]. These two pairs of ECF σ- and anti-σ factors constitute two different regulatory cascades, which control the biosynthesis of carotenoids in *A. brasilense*[17]. One of the ECF σ factors encoded in the *A. brasilense* genome was not accompanied with an anti-σ factor and contained an extension of 119 amino acids at its C-terminus, suggesting its similarity to the ECF41 type of σ factors. Crystal structure of SigJ of *Mycobacterium tuberculosis* (Mtb-SigJ), which belongs to the ECF41 family of σ factors, sheds some light on the role of the C-terminal SnoaL_2 domain on the structure and function of ECF41 type of σ factors [18]. Direct coupling analysis combined with mutational analysis of the conserved residues of the C-terminal region of the ECF41 σ factor of *Bacillus licheniformis* identified the contact interface between the C-terminal extension and the core σ factor regions required for controlling ECF activity[14].

In this study, we investigated the role of the two conserved motifs in the SnoaL_2 -like an extension of the C-terminal domain of ECF41 σ factor of *A. brasilense* Sp245: a proximal DGGGR motif and a distal NPDKV motif. Despite the increasing attention to the role of SnoaL_2 domain in modulating ECF41 σ activity and function[13,14,18], the physiological role of ECF41 family σ factors is not known yet [13,14]. We have recently shown the ECF41 σ factor (RpoE10) role in controlling the motility and biogenesis of lateral flagella in *A. brasilense* Sp245[19]. Here, we describe the role of the two conserved motifs in the SnoaL_2 domain of RpoE10 in the inhibition and activation of its activity, respectively. An all-atom Molecular Dynamics (MD) simulations and principal component analysis (PCA) of RpoE10 with*in silico* mutations at conserved motifs structural study was carried out along with the experimental validation of the consequence.

MATERIAL AND METHODS

Bacterial strains, plasmid, chemicals, and growth conditions :

The *E. coli* DH5αand *E. coli* S17.1 were grown in Luria Bertani (LB) medium at 37°C. *A.brasilense* Sp245 was grown in minimal malate (MM) medium or LB medium at 30°C. All the chemicals used for culturing bacteria were from Hi-media (Mumbai, India), and enzymes used for DNA manipulation and cloning were from New England Biolabs. All strains and plasmids used in this work are listed in Table1.

Cloning of *rpoE10* and its deletion derivatives in a low copy, broad host range expression vector, pMMB206:

To examine the effect of expression of a wild copy of the *rpoE10*gene in *rpoE10::* km mutant, a wild-type copy of the *rpoE10*gene was supplied to the *rpoE10::* km by cloning the entire coding region of *rpoE10* in an expression vector. The gene encoding RpoE10 was amplified by PCR using DreamTaq DNA polymerase (Fermentas), primer pairs *RpoE10* -F' and *RpoE10*- R' having *PstI*and *HindIII* restriction overhangs in their 5'ends, respectively. The gel-purified PCR product was digested using *PstI* and*HindIII* , purified again, and ligated with compatible ends downstream of IPTG inducible *lacUV* 5 promoter region in a broad host range expression vector, pMMB206. The resulting plasmid (pAPD7) was conjugatively mobilized into *A. brasilense* , and exconjugants selected on plates containing chloramphenicol. The deletion derivatives RpoE10(Del1) (pAPD8) and RpoE10(Del2) (pAPD9) were constructed as described earlier [19].

Construction of *abm:gfp* fusion:

The promoter region of *rpoE10* up to ATG was amplified with the help of forward (ProRpoE10GFP_FP) and reverse (ProRpoE10GFP_RP) primers containing *KpnI* and *NdeI* site, respectively. The purified insert was digested and ligated into a broad host range vector, pBBR1MCS-3 (designated as pAPD10). The PCR

amplified E-GFP was digested with *NdeI* and *XhoI* restriction enzymes and ligated into pAPD10, and the resulting construct was designated as pAPD11.

Site-directed mutagenesis:

The role of the conserved amino acids present in the NPDKV motif and DGGGR motif of RpoE10was validated by PCR-based site-directed mutagenesis. The native *rpoE10* gene was PCR-amplified with gene-specific primers (RpoE10F/RpoE10R, Supplemental Table S1) containing *Bam* HI and *Hind* III restriction enzyme sites in the forward and reverse primers, respectively. After purification, the amplicon of *rpoE10* was cloned in pGEMT easy vector (Promega) by TA cloning method, and the recombinant plasmid (pAKVSS1, Supplemental Table S2) harboring*rpoE10* gene was directly used for PCR-based mutagenesis. Restriction sites (*Bam* HI and *Hind* III) were required to generate mutated versions of *rpoE10* after PCR-based mutagenesis and in-frame cloning in the expression vector, pMMB206. In the C-terminal, NPDKV motif was replaced by NAAAV motif in RpoE10(Mut1), and the DGGGR motif was replaced by AAAGR motif in RpoE10(Mut-2) using sets of complementary primers (RpoE10MOT1F/ RpoE10MOT1R for NPDKV motif and RpoE10MOT2F/ RpoE10MOT2R for DGGGR motif listed in Supplemental Table S2) having mutations flanked by unmodified nucleotides for the described amino acids. Site-directed mutagenesis was carried out by using QuikChange II site-directed mutagenesis kit (Agilent), and the PCR cycling conditions consisted of initial denaturation step, at 95^0C for 3 min followed by amplification step of 18 cycles, which was further composed of three sub-steps including denaturation at 94^0C for 1min, annealing at 46^0C for 1 min and extension at 68^0C for 12 min. The final extension step was at 68^0C for 30 min. Mutagenesis-PCR was performed in a 50 µl reaction mixture containing 1X Pfu Ultra HF DNA polymerase buffer, 1 mMdNTPs, 0.5 µM forward and reverse primers, 1% DMSO, and 2-unit Pfu Ultra HF DNA polymerase. After amplification 1 µl of the *Dpn* I restriction enzyme (10 U/µl) was directly added to each amplification reaction and kept at 37^0C for 3 h to digest the parental DNA. After that, 5-10 µl of each *Dpn* I-digested DNA sample was directly transformed in XL1-Blue super-competent cells. Recombinant plasmids having mutations in the NPDKV motif (pAKVSS2) and DGGGR motif (pAKVSS3) were confirmed by DNA sequencing using gene-specific primers (RpoE10F/RpoE10R, Supplemental Table S2). After mutations, modified inserts of *rpoE10* were generated by restriction digestion with*Bam* HI and *Hind* III and cloned in the similarly digested and eluted vector backbone of pMMB206.

Measurement of E-GFP:

The modified over-expression plasmids (pAPD8, pAPDp, AKVSS1, and pAKVSS2) along with wild type (pAPD7) were conjugatively mobilized in*A. brasilense* Sp245 already having pAPD11 plasmid. The effect of deletions and mutations of two conserved motifs of RpoE10 on*abm:gfp* were monitored. For this, overnight grown cultures of recombinants and control were inoculated in the same media to maintain OD_{600} to ˜ 0.1 for all cultures. When bacterial growth reached up to ˜0.6 OD_{600} then each culture was divided into two equal parts. One part of the culture was induced by IPTG (0.5 mM) and allowed to grow for another 6 h. After sample collection and washing with phosphate buffer saline, GFP intensity was measured by exciting at 405 nm and emission 567 at 485 nm to 525 nm in Varian spectrofluorimeter. The second part of the culture was used for plate assay in which cells with equal OD were kept on a MM agar plate as a drop for 48 h in the presence and absence of IPTG. GFP expression was observed under UV light and photographs taken by digital camera.

Homology Modelling of RpoE10:

The homology models of RpoE10 and its truncated versions RpoE10(Del1) and RpoE10(Del2) were generated using the X-ray crystal structure of Mtb-SigJ (PDB ID: 5XE7, chain B) as template [20]. We first constructed a model of the RpoE10 protein containing the SnoaL_2 domain and then modeled the deleted part of the loop. The model of Mtb-SigJ was built with the Modeler module using the HHPRED server[21]. The quality and reliability of the built model were evaluated using PROCHECK and Verify 3D programs. Structural superposition and structure-based sequence alignment were performed using the MatchMaker and MatchAlign tool of the UCSF Chimera package[22]. In the RpoE10 model, *in silico* substitution of DGG

to AAA in RpoE10(Mut1) and PDK to AAA in RpoE10(Mut2) was generated using the rotamer tool of UCSF-Chimera tool[22].

Molecular Dynamics (MD) simulation of RpoE10 and its mutants

To provide atomistic details of bimolecular motions and an additional perspective relative to experimental results, we performed "computational microscopy" using molecular dynamics (MD) simulations for 200 ns for each of structural models of (i) RpoE10 native; (ii) RpoE10(Mut1) (*in silico* substitution of NPDKV with NAAAV) and (iii) RpoE10(Mut2) (*in silico* substitution of DGGGR to AAAGR) (iv) RpoE10 (Del1). All simulations were performed with the GROMACS 2019.2 simulation suite (http://www.gromacs.org/) using the CHARMM 27 force field parameter set [23-25]. In the first step of MD simulation, the selected homology model of RpoE10, native, and the two mutants, Mut1 and Mut2, were solvated in a cubic water box. The system was solvated with TIP3P water molecules. Periodic boundary conditions were applied in all directions, and Na^+Cl^- counter ions were added to make the system electrically neutral. Then, the systems' energy minimization was done by 50,000 steps of steepest descents to relax any steric conflicts. NVT and NPT equilibration then followed energy minimization for 300ps. Once equilibrated regarding pressure and density, these systems were then subjected to MD simulation at 300K for 200 ns, with no restrictions on the residues. During the simulation, the temperature of the system was maintained at 300K using a Berendsen thermostat. The particle mesh Ewald method [26] was employed to account for the long-range electrostatic interactions, and the LINCS algorithm [27] was used, with a time step of 2 fs to restrain bond lengths. The constant temperature and pressure (300K and 1bar) were maintained using a V-rescale thermostat[28] and Parrinello-Rahmanbarostat [29]. The production run was performed for 200ns with leapfrog integrator[30], and the coordinates were saved every 10picoseconds. A total of 20,000 frames were generated. MD trajectories were analyzed using the tools of GROMACS 2019.2. The backbone root-mean-square deviations (RMSDs) of the RpoE10, RpoE10 (Mut-1 and Mut-2), and RpoE10 (Del1) derivatives were calculated following the structural alignment of the core structure. The RMSD and RMSF values depict the regions that fluctuate differentially regarding other regions of the protein of interest. The GROMACS analysis plots were generated by XMGRACE utility from Grace software (http://plasma-gate.weizmann.ac.il/Grace/). UCFC Chimera[22] and VMD [31] software were employed to analyze and visualize the molecular interactions and MD trajectories to create molecular graphics.

Essential dynamics

Principal component analysis (PCA) or Essential dynamics (ED) was performed to understand the impact of the NPDKV and DGGGR motifs on the global motion of the atomic coordinates of RpoE10 and RpoE10(Mut1, Mut2, and Del1) forms during 200ns MD simulation. The backbone atoms of the protein molecules were considered for this analysis. The principal components analysis is done by building a covariance matrix of the atomic fluctuations. Diagonalisation of this matrix yields a set of eigenvectors and eigenvalues that describe collective modes of fluctuations of the protein. The largest-amplitude collective motions are represented by the eigenvectors corresponding to the largest eigenvalues and are called "principal components"[32-34]{Campagne, 2014 #80;Lane, 2006 #81;Campbell, 2002 #89}. The covariance matrix was constructed and diagonalized using g_covar and g_anaeig programs of Gromacs 2019.2, respectively, to generate the eigenvectors and eigenvalues. The prominent mobile regions of RpoE10 were carefully inspected and interpreted from the PCA and covariance matrix data.

Molecular dynamics simulation analysis

To understand the effect of mutations in^{277}NPDKV^{281}and^{200}DGGGR^{204} motifs on the structure and function of RpoE10, the RpoE10 model was subjected to*in silico* substitution of^{200}DGG^{202} to AAA and^{278}PDK^{280} to AAA to generate RpoE10(Mut1) RpoE10(Mut2) models, respectively. Furthermore, we carried out an 800 ns MD simulation study to evaluate the effect of these mutations and partial deletion of SnoaL_2 domain on the overall conformational dynamics, structural stability, compactness of domains, and structural features conferring the enhancement/elimination of promoter activation. According to our simulation studies, the two most common measures of structural fluctuations, root-mean-square deviation (RMSD) and root-mean-

square fluctuation (RMSF), and a measure of compactness (radius of gyration, i.e., Rg) explain the structural effects caused due to mutations in the NPDKV and DGGGR motifs, and its impact on promoter activation.

Backbone conformational (RMSD) analysis

RESULTS

RpoE10 model Structure of *A. brasilense*:

The structural model of the RpoE10 of *A. brasilense* exhibited similar spatial arrangements to its template crystal structure of SigJ of *M. tuberculosis* (Mtb-SigJ; PDB ID: 5XE7) (Figure 1A). The overall architecture of RpoE10 can be structurally categorized into three discrete sub-domains, viz., the N-terminal $\sigma^2$and σ^4 domain, and the C-terminal SnoaL_2 domain. Global sequence alignment using the Needleman Wunsch tool of NCBI showed 36% sequence identity between RpoE10 and Mtb-SigJ. The target-template structures showed 44% sequence identity in the σ^2-σ^4domain sequences, excluding the C-terminal extension. The C-terminal extension of Mtb-SigJ and RpoE10 revealed 27% sequence identity. The σ^2 and σ^4 and SnoaL 2 are threaded to each other through a polypeptide linker strand connecting σ^2 with σ^4. The two conserved motifs^{200}DGGGR^{204} and^{277}NPDKV^{281} of Snoal_2 domain were in the vicinity (5Å zone) of the inter-domain linker strand, containing^{83}WLPEP^{87} motif (Figure 1A; inset view).

Effect of truncation and mutation of C-terminal extension of RpoE10 on the activation of *abm-rpoE10* promoter:

In *A. brasilense* , the gene encoding RpoE10 is preceded by a gene encoding an antibiotic biosynthesis monooxygenase (*abm*) in a dicistronic operon (*abm-rpoE10*). To compare the regulation of the*abm-rpoE10* promoter with that of the other ECF41 σ dependent promoters of *R. sphaeroides* and *B. licheniformis*[13], we constructed an *abm:gfp* fusion (pAPD5) and mobilized it into *A. brasilense* Sp245. Since ECF41 σ factors have been shown to activate their own promoters, the full length*rpoE10* was also cloned downstream of an IPTG inducible promoter in a broad host range expression vector, pMMB206 (designated as pAPD7) and mobilized in *A. brasilense* Sp245 harbouring pAPD5. We did not find any significant increase in the fluorescence from *abm*promoter even after inducing it with IPTG. Because of the suggested negative role of the C-terminal extension of ECF41 σ factors[8], we constructed two recombinants by cloning genes encoding two different truncated derivatives of RpoE10 in pMMB206. While construct RpoE10(Del1) had a deletion up to the DGGGR motif (pAPD8), the other construct RpoE10(Del2) had deletion excluding DGGGR residues (Figure 1B). When we mobilized these recombinant plasmids in *A. brasilense* Sp245 harboring*abm:gfp* fusion, we found that RpoE10(Del1), lacking ^{277}NPDKV^{281}motif with deletion up to^{200}DGGGR^{204} motif, increased GFP fluorescence by almost 15-fold after IPTG induction (Figure 1C). But, RpoE10(Del2), which lacked the entire C-terminal region,including^{200}DGGGR^{204}and^{277}NPDKV^{281} motifs of SnoaL, did not show any promoter activity. To analyze the role of the conserved motifs NPDKV and DGGGR, we also constructed two site-directed mutants, RpoE10(Mut1) and RpoE10(Mut2), and mobilized them in *A. brasilense* Sp245 harboring*abm:gfp* fusion. In the RpoE10(Mut1), the ^{277}NPDKV^{281} motif was replaced with the NAAAV motif, and in RpoE10(Mut2), the^{200}DGGGR^{204} motif was replaced with the AAAGR motif. The effect of RpoE10(Mut1) and RpoE10(Mut2) on the expression of *abm:gfp* was similar to that observed with RpoE10(Del1) and RpoE10(Del2), respectively (Figure 1C).

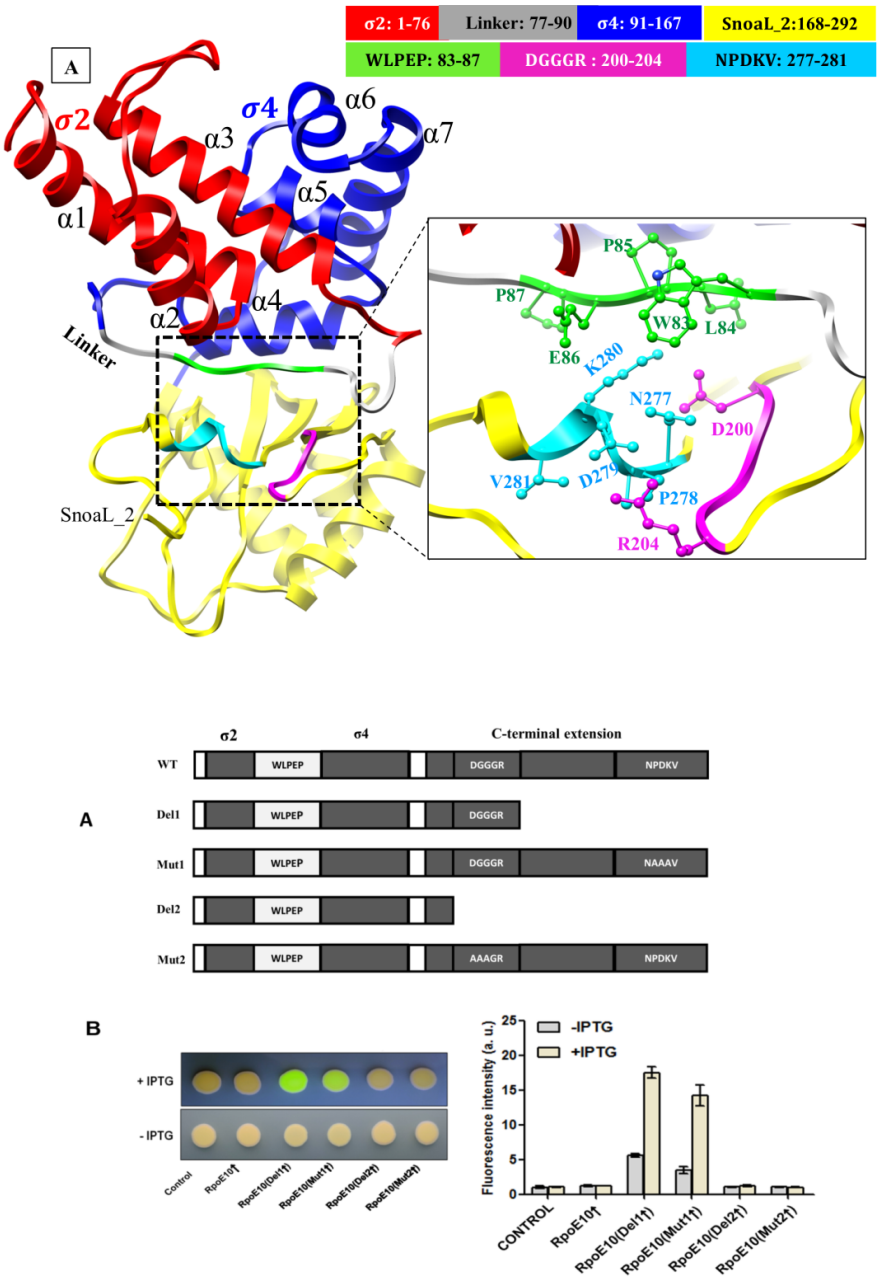
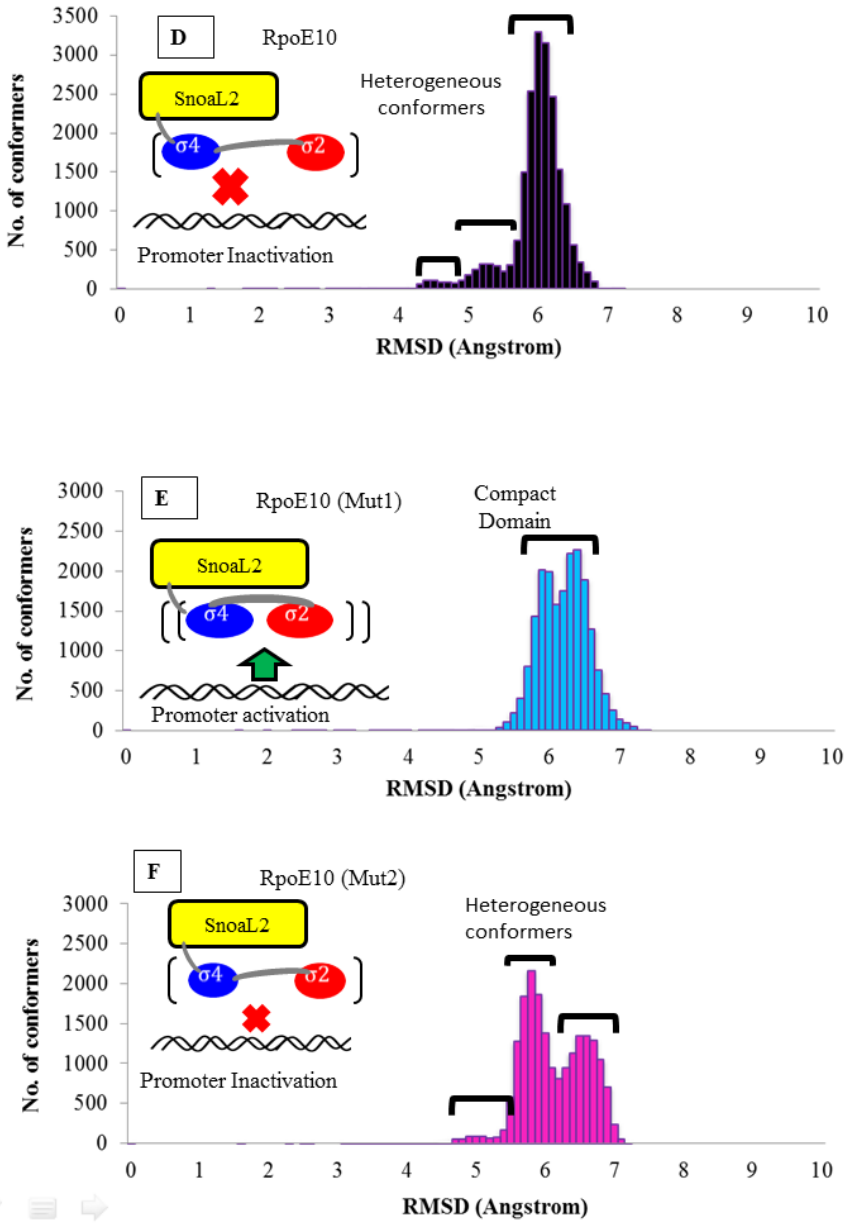


Figure 1 A) ?2, ?4, and SnoaL2 domain of RpoE10 model structure. The residues-residue interaction between WLPEP motif of ?2- ?4 linker, DGGGR and NPDKV motifs of SnoaL2 extension are shown in green, magenta, and cyan (ball and stick), respectively. **B)**Schematic representation of the deletion and mutant derivatives of RpoE10. The characteristic motifs of ECF41 type σ factor, WLPEP motif located between σ2 and σ4 regions, and other motifs (DGGGR and NPDKV) in the C-terminal extension. The first deletion derivative, RpoE10 (Del1) contained the RpoE10 sequence up to DGGGR, while the second deletion derivative RpoE10(Del2) excluded the DGGGR motif. Also, two site-directed mutants of RpoE10 were constructed: in RpoE10(Mut1) NPDKV motif was replaced with NAAAVmotif. InRpoE10(Mut2),

DGGGR motif was replaced with AAAGR motif, and **(C)** : Effect of deletion and mutation ofDGGGR and NPDKV motifs of at the C-terminal of RpoE10 on the expression of *abm:gfp* fusion in *A. brasilense* Sp245 expressing full RpoE10, its two deletion derivatives, RpoE10(Del1) and RpoE10(Del2) and two mutant derivatives RpoE10(Mut1) and RpoE10(Mut2). Each bar represents the mean of triplicates in three independent experiments.

The analysis of 200ns MD simulation trajectories for RpoE10, RpoE10(Mut1), and RpoE(Mut2) enabled us to demonstrate the effect of mutations in the motifs ^{277}NPDKV^{281}and ^{200}DGGGR^{204} of SnoaL_2 domain. The Root Mean Square Deviation (RMSD) was used for evaluating the stability and differences between the backbone trajectories of proteins from its initial structural conformation to its final snapshot. The smaller the RMSD, the more stable or rigid the conformation is. The backbone RMSD analysis of RpoE10, RpoE10(Mut1) and RpoE10(Mut2) system was carried out for full length, individual domains $?^2$(1-76 amino acids), $?^4$ (91-167 amino residues) and combined $?^2$-$?^4$domain (1-167amino acids) (Figure 2). We observed a significant degree of conformational change in the Snoal_2 domain of RpoE10(Mut1) compared to its initial snapshot. The RpoE10 showed a stable and rigid conformation in this domain throughout the simulation. In RpoE10 (Mut1), the SnoaL_2 domain showed an upsurge in RMSD value from ˜4Å to ˜6Å after 20ns, as compared to RpoE10 and RpoE10 (Mut2). The RMSD value for RpoE10(Mut2) gradually increased from 4 Å to 5Å in the first 150ns and then raised from ˜5 Å to 6 Å in the remaining 50ns. The RMSD analysis (with variations in RMSDs) suggested flexible yet stable backbone conformations for RpoE10(Mut1 and Mut2) as compared to the RpoE10 models (Figure 2B and C). Furthermore, to trace the impact of NPDKV and DGGGR motif on conformational dynamics of $?^2$, $?^4$ and SnoaL_2 domain, the RMSD profile of the NPDKV, DGGGR, and WLPEP motif of linker strand segments were also analyzed (Figure 2A, B, C). The WLPEP and NPDKV showed stable RMSD patterns comparable to each other in the case of RpoE10 and RpoE10 (Mut2). Noticeably, a mutation in the NPDKV fragment could result in a sharp rise in its RMSD profile over the 200ns trajectory and impacted the conformational dynamics of the $?^2$, $?^4$, and SnoaL_2 domains in RpoE(Mut1). Compared to RpoE10 and RpoE10 (Mut2), there is a significant RMSD variation in the backbone trajectory of the Snoal_2 domain, NPDKV, and DGGGR segment for RpoE10 (Mut1) indicated the influence of mutations at NPDKV over the ?-domains. (Figure 2 A and C).

Additionally, to evaluate the differences in the backbone trajectory of^2-$?^4$ domain, we plotted a histogram of RMSD against the number of conformers for RpoE10, RpoE10(Mut1), and RpoE10(Mut2) (Figure 2 D, E, and F). In RpoE10(Mut1), the RMSD of the $?^2$-$?^4$ domain conformers were restricted to 5.8-6.5A whereas RMSDs were heterogeneously distributed and varied from 4-7 A in RpoE10 and 5-7 A in RpoE10 (Mut2). This analysis suggests that the RpoE10 and RpoE10 (Mut2) forms acquire substantial conformational heterogeneity leading to an unstable system. However, a mutation in NPDKV leads to conformational stability. Therefore, the observations that SnoaL_2 domain constraints σ domain to a compact structure [18] are consistent in our RpoE10 models. The NPDKV and DGGGR motifs of the Snoal_2 domain can be attributed to significantly alter the conformations of -domains.



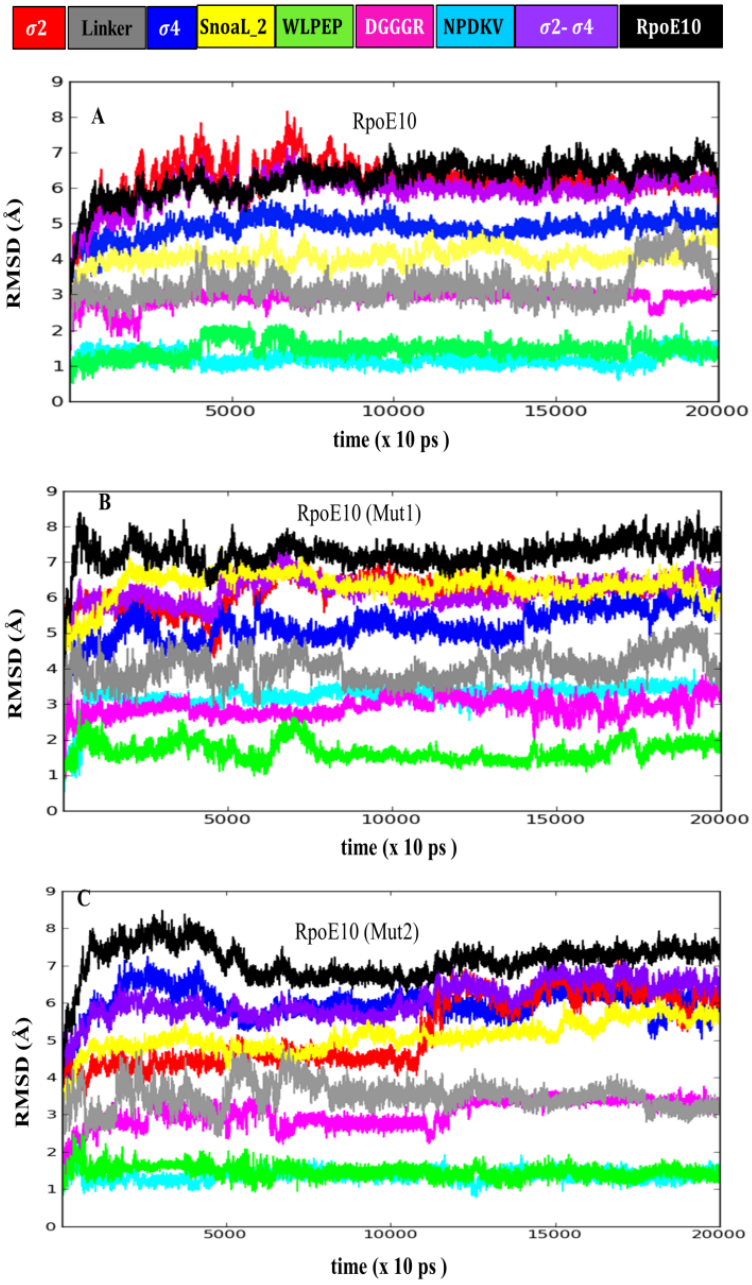


Figure 2. Comparative plots of RMSD profile of (A) RpoE10 (B) RpoE10 (Mut1) and (C) RpoE10 (Mut2). RMSD plot was computed through the least square fitting of the backbone atom. Plots were calculated for ?2(red), ?4 (blue), SnoaL_2(yellow), ?2-?4 (violet), ?2-?4- SnoaL_2 (black), NPDKV (cyan), DGGGR (magenta), and WLPEP (green) motifs of linker strand (grey) segments for RpoE10,

Rp0E10 (Mut1) and Rp0E10 (Mut2) models**. Figure 2 D, E, and F (Right Panel): showing the Impact of SnoaL_2 domain mutations on the conformational dynamics** .The RpoE10 (Mut1) and RpoE10 (Mut2) influence the backbone conformations of σ2 -σ4 domains. RpoE10 (Mut1) showed a more stable conformation of the σ2- σ4 domain than RpoE10 (Mut2) and RpoE10 across many conformers. Conformational heterogeneity can be seen in RpoE10 and RpoE10 (Mut2). The domain fluctuations and its impact in regulating the promoter activation is depicted in the schematic diagram.

Rg analysis (compactness) of ?-domains

The observation that NPDKV and DGGGR motifs of Snoal_2 domain significantly alter the conformations of$^{?2}$-$^{?4}$ domain was further evaluated by measuring the compactness of domains using Radius of gyration (Rg) analysis of MD simulations. A comparison between the Rg profile and average Rg value (indicated at the right margin, in Figure 3 A, B, C) of individual $^{?2}$, $^{?4}$, combined $^{?2}$-$^{?4}$ domain, and full-length RpoE10, RpoE10(Mut1) and RpoE10(Mut2) indicated that the mutation in the NPDKV motif of SnoaL_2 domain constrains the $^{?2}$, $^{?4}$ and $^{?2}$-$^{?4}$domains to a more compact domain structure. A reduced average Rg value of 1.71 nm for $^{?2}$-$^{?4}$ domain of RpoE10(Mut1) as compared to both RpoE10 (1.75 nm) and RpoE10(Mut2) (1.73) was observed. In RpoE10(Mut1), although, a sharp fall in the Rg value from ˜1.5 to ˜1.25 after 50 ns of simulation time indicated a rise in the compactness of the $^{?2}$ domain (Figure 3B), yet a steady Rg value for $^{?2}$-$^{?4}$ and SnoaL_2 domain suggest a stable and compact protein. We also determined the minimum distance between the $^{?2}$ and $^{?4}$ domains as a function of simulation time using the *g_mindist* utility of GROMACS (Figure 3 D, E, F). A varying degree (0.15-0.22 nm) of inter-domain distance motions between $^{?2}$ and $^{?4}$ domains in RpoE10 (Mut1) as compared to RpoE10 (0.15-0.17 nm) and RpoE10(Mut2) (0.15-0.17 nm) proteins suggests a favorable motion for the recognition and activation of promoter.

Furthermore, to zoom in on the differences in the compactness of$^{?2}$-$^{?4}$ domain, a histogram of RMSD was plotted against the number of conformers for RpoE10, RpoE10(Mut1), and RpoE10(Mut2). In RpoE10 (Mut1), the Rg value for $^{?2}$-$^{?4}$ domain conformers was mainly restricted to ˜1.69 nm. However, an increase in the Rg (˜1.73 nm) was noticed for $^{?2}$-$^{?4}$ conformers of both RpoE10 and RpoE10(Mut2) (Figure 3 G, H and I). The reduced Rg value across a large number of σ^2-σ^4 conformers in RpoE10(Mut1) suggests that the mutations in the NPDKV motif of SnoaL_2 domain constrain RpoE10 to a compact and stable structure favoring an enhanced activation of its promoter.

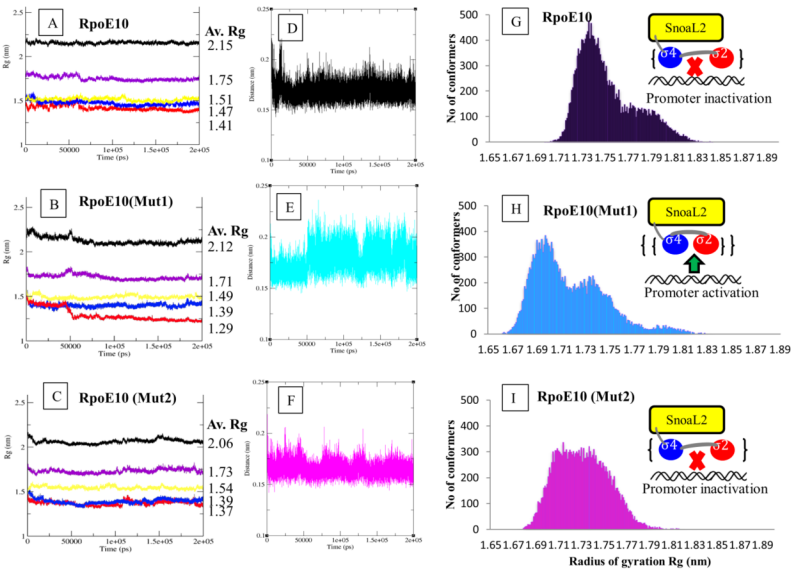


Figure 3. The radius of gyration (Rg) analysis.

Rg trajectory of A) RpoE10, B) RpoE10 (Mut1), and C) RpoE10 (Mut2) shown as σ2 (red), σ4(blue), SnoaL_2(yellow), σ2-σ4 domain (purple), and entire protein chain (black). The Average Rg value is written at the right margin of the corresponding graph. RpoE10 (Mut1) showed a more compact σ2- σ4 domain than RpoE10 (Mut2) and RpoE10. A plot of minimum distance calculated between σ2 and σ4 domain for D) RpoE10, E) RpoE10 (Mut1), and F) RpoE10 (Mut2) from 200ns trajectory. **Figure G, H, and I** depict the analysis of the compactness of the σ2-σ4 domain for G) RpoE10, H) RpoE10 (Mut1), and I) RpoE10(Mut2. RpoE10 (Mut1) showed a more compact σ2- σ4 domain than RpoE10 (Mut2) and RpoE10.

Mutations in the NPDKV and DGGGR alter the structural features of the promoter recognition sites

Structural superimposition of $?^2$ and $?^4$ domains in RpoE10 with other ?/anti- ? complexes[31,32] revealed that in RpoE10, orientation and accessibility of DNA binding surfaces are exposed (Figure 4A). Like its template Mtb-SigJ [18], RpoE10 lacks the first helix α1, usually present in the σ^2 domain, which has three helices α2–α3- α4 connected by two loops L2 and L3. Intriguingly, a comparative analysis of the average fluctuations of all the backbone atoms of the amino acid residues (RMSF profile) of the RpoE10, RpoE10 (Mut1), and RpoE10 (Mut2) showed a contrasting pattern at key positions (Figure 4A). The α1-helix showed an increased RMSF fluctuation in RpoE10 compared to both RpoE10 (Mut1 and Mut2) (marked as 1 in Figures 4A and B). Remarkably, we noticed a prominent fluctuation in the L3 loop (residues 46-52) between second and third helices (α2 andα3) of RpoE10 (Mut1) (marked as 2 in Figure 4A and B). It has been shown that the flexible "specificity loop" initiates promoter recognition and determines ECF σ factors' specificity at the -10 promoter element in ECF σ-dependent promoters [32]. Therefore, a reduced fluctuation in the α1 helix and a simultaneously increased fluctuation in the L3 loop of the $?^2$ domain could favor recognizing -10 promoter and DNA melting by RpoE10 (Mut1), leading to enhanced promoter activation.

Also, in RpoE10, increased fluctuation at the linker-loop junction connecting the $?^2$ and $?^4$ domains (peaks are marked as 3 and 4 in Figure 4B), suggests the possibility of conformational instability of $?^2$ and $?^4$ domains. As compared to RpoE10(Mut2), the junction region (marked as 4 in Figure 4B) towards the $?^4$domain is stabilized in RpoE10 (Mut1). Therefore, similar to Mtb-SigJ (18), the decreased fluctuations at the junction of $?^2$-linker-$?^4$ region depicts the stabilized and tethered $?^2$ and $?^4$domains essential for acquiring a productive conformation of RpoE10(Mut1) for enhanced activation of its promoter. Next, we focused on another important segment (135-151)of the σ^4domain, a helix-turn-helix motif known to interact with the -35 element of the promoter [34]. Both, RpoE10 and RpoE10(Mut2) showed an enhanced average fluctuation of backbone atoms of residues 135-151 segment as compared to RpoE10(Mut1) (peaks are marked as 5 in Figure 4 A and B). Unlike RpoE10 and RpoE10(Mut2), the stabilized helix-turn-helix motif of σ^4 domain of RpoE10(Mut1) reasonably favors interaction with the -35 element of the promoter and, therefore, justifies its enhanced activation.

Interestingly, we notice that the RMSF profile of the RpoE10 (Mut1 and Mut2) showed increased conformational flexibility at both NPDKVand DGGGR motifs (marked as 6 and 9 respectively, in Figure 4B), as compared to that in RpoE10. Since the DGGGR motif is essential for promoter activation (Figure 1), the stabilized conformation of backbone atoms of the NPDKV motif in RpoE10 raises the possibility to obstruct the DGGGR motif away from the $?^2$-$?^4$ linker site. In this situation, the stabilized orientation of the NPDKV motif may lead to the formation of unproductive conformations of^2-$?^4$ domain in RpoE10, and therefore a possible reason for the elimination of its activity. Notably, in RpoE10, another segment of 230-238 residues (marked as 8 in Figures 4A and B) showed a sharp and distinct rise in the fluctuations of backbone atoms compared to RpoE10(Mut1 and Mut2). This segment forms the core of the SnoaL_2 domain, and therefore the backbone fluctuations may act as a trigger signal for eliminating the promoter activation of RpoE10.

RMSF analysis showed that the NPDKV mutant form of RpoE10 containing an intact DGGGR motif showed stable conformations at the -35 promoter binding site (residues 135-151), enhanced flexibility of L3 "specificity loop" (residues 45-52) around the -10 promoter recognition site, and stabilized linker region

connecting the $?^2$-$?^4$ domain which are essential features to attain a productive conformation for enhanced promoter activation.

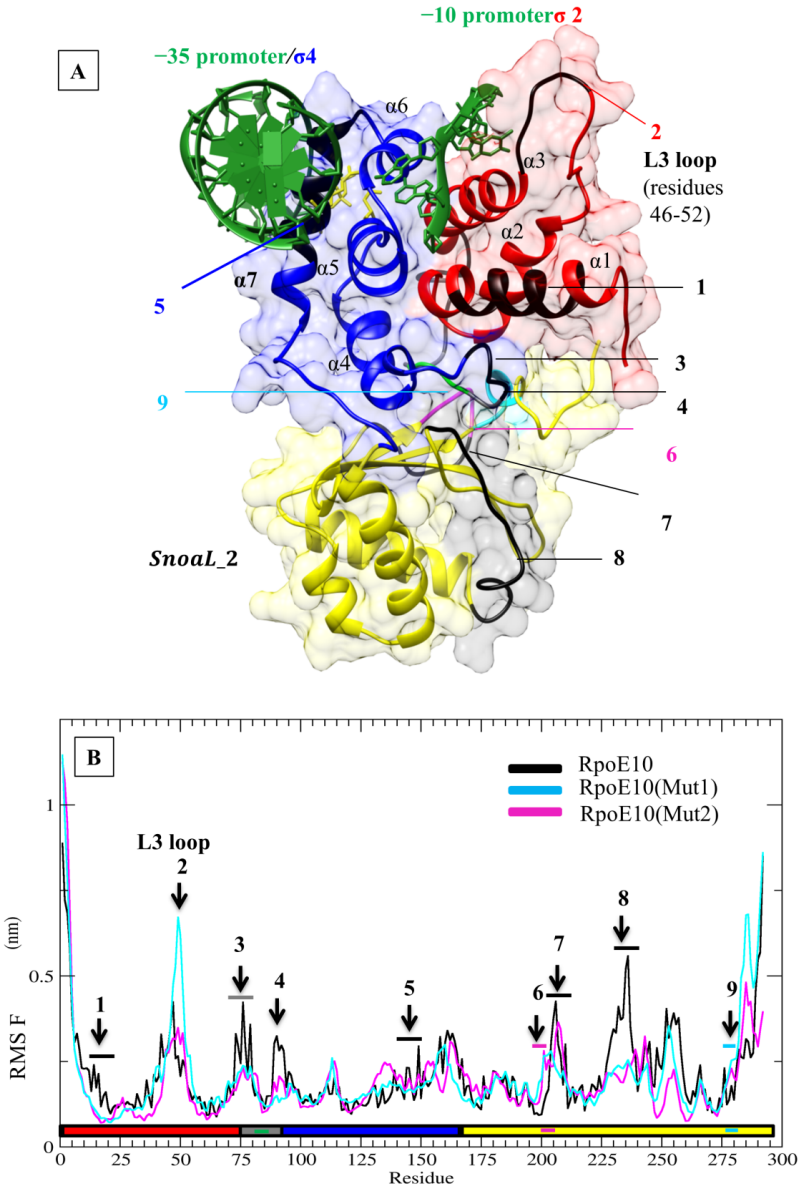


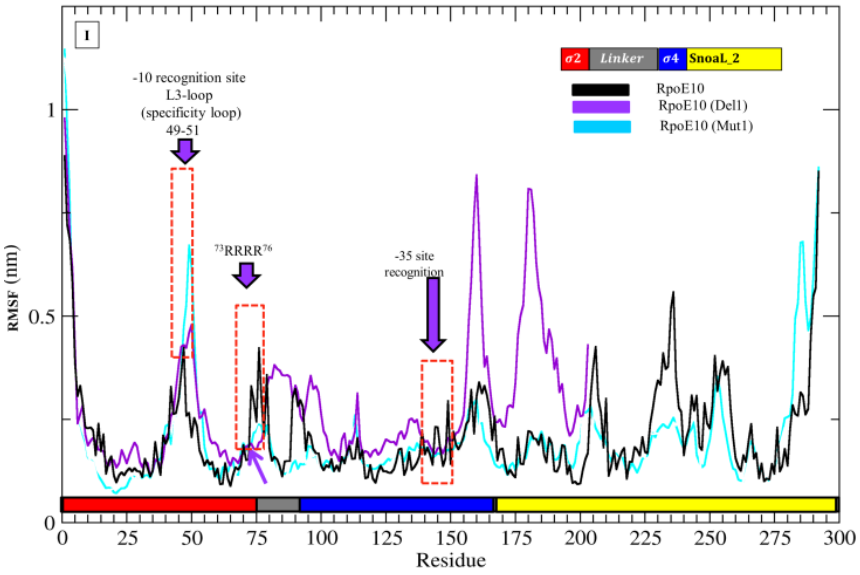
Figure 4. Exposed DNA binding surfaces of RpoE10 adopt a conformation similar to Mtb-SigJ (PDB ID: 5XE7) that can readily interact with the promoter. A) Structural superposition of σ2 and σ4 with previously determined structures of the -10 promoter/ σ 2 (PDB ID: 4LUP) and -35 promoter/σ4 (PDB ID: 2H27) complexes (36, 38).**B)** The comparative Root Mean Square Fluctuation (RMSF) plot of RpoE10 and Its mutant forms. The critical residue positions showing distinct RMSF peaks are indicated by numbers 1-9, and these residues are mapped (in black color) onto the structure of the RpoE10 model.

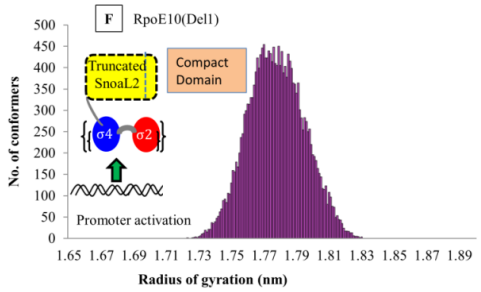
Conformational dynamics of C-terminal partially truncated RpoE10

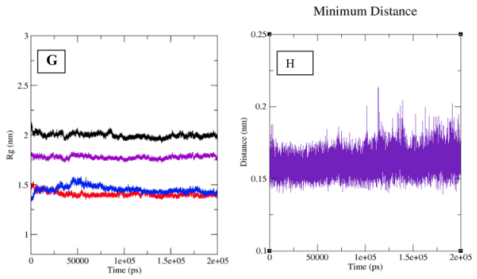
The truncated SnoaL_2 domain model of the RpoE10 (Del1) was subjected to MD simulations for 200 ns trajectories and analyzed to demonstrate the experimentally observed effect of the DGGGR motif on promoter activation without NPDKV motif (Figure 1). Strikingly, the truncated SnoaL_2 domain containing the DGGG motif showed consistent and stable interactions with the σ^2-σ^4 domain across the snapshots of the 200ns simulation time (Figure 5 A, B, C; inset view). Furthermore, we observed stable conformation in the backbone RMSD of ^{2}, $?^4$ combined $?^2$-$?^4$ domain, DGGGR, and WLPEP motifs throughout the simulation (Figure. 5D). In RpoE10 (Del1), after initial fluctuations in the RMSD, both the $?^2$domain and combined $?^2$-$?^4$ domain showed an upsurge in RMSD value at ˜125 ns and then stabilized in remaining ˜75 ns (Figure 5D). The RMSD analysis for $?^2$,^{4} and $?^2$-$?^4$domain along with and linker strand segments suggests that backbone conformations for RpoE10(Del1) were stable even after partial truncation of the SnoaL_2 domain. Furthermore, to ensure the stabilized and homogenous conformers, the histogram of RMSD was plotted against the number of conformers for RpoE10(Del1). In RpoE10(Mut1), the RMSD for most of the $?^2$-$?^4$ domain conformers was restricted to 6-7Å (Figure 5E). This analysis suggests that the truncated SnoaL_2 derivative of RpoE10 containing the DGGG motif adopts stable conformations of $?^2$-$?^4$domain.

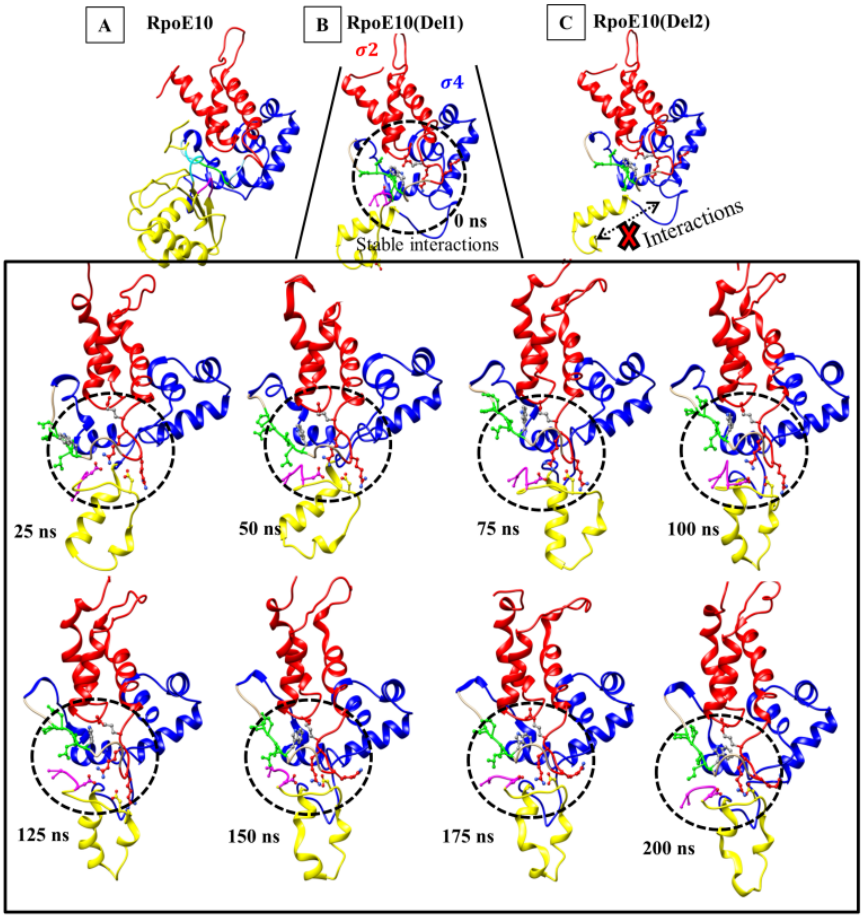
We also evaluated the compactness (Rg value) of the individual $?^2$,^{4}, combined $?^2$-$?^4$ domain, and full-length RpoE10(Del1) model. A steady Rg value for $?^2$,^{4} and combined $?^2$-$?^4$ domain of RpoE10(Del1) was obtained (Figure 5G) Furthermore, the inter-domain distance motions between $?^2$ and $?^4$ domains further mirrors the stable and compact $?^2$-$?^4$domain (Figure 5H). A histogram plot of RMSD against the number of conformers for RpoE10 (Del1) showed that most of the $?^2$-$?^4$ domain conformers were restricted to ˜1.76-1.80 nm (Figure. 5F). This analysis suggests that the truncated SnoaL_2 domain containing the DGGG motif may constraint the $?^2$-$?^4$ domains to a stable and compact protein structure required for promoter recognition and activation. Therefore, the observations that SnoaL_2 domain constraints σ domain to a stable and compact structure[18] can be attributed to the DGGG motifs of Snoal_2 domain.

We constructed and compared the RMSF plot of RpoE10 (Del1) with RpoE10 and RpoE10 (Mut1) to assess the impact of the DGGGR motif onto the key residue positions of the truncated SnoaL_2 domain model (Figure 5I). We focussed on key residue segments essential in initiating the promoter recognition and determining the ECF σ factors' specificity at the -10 and -35 promoter elements in ECF σ-dependent promoters[31,32]. Intriguingly, as shown and marked in Figure 5I, these key residue positions of the RpoE10 showed contrasting fluctuations. As compared to RpoE10, an enhanced fluctuation was noticed in the L3-loop (residues 46-52) of σ2 in both RpoE10(Del1) and RpoE10(Mut1). The enhanced flexibility of the L3-loop "specificity loop" is known to favor the -10 promoter recognition[32]. We also noticed a stable peak in the residues segment 135-151 of σ4 domain of RpoE10(Del1) and RpoE10(Mut1), both. This segment is known to constitute a helix-turn-helix motif that interacts with the -35 element of the promoter DNA [31](Figure 5I). Therefore, our MD simulation analysis of truncated SnoaL_2 domain, without NPDKV motif, containing the DGGG motif suggests that promoter recognition and enhanced activation is due to the stable interactions of DGGG motif with the compact conformations of $?^2$-$?^4$domain, enhanced flexibility of the -10 recognizing L3-loop, and stability of the helix-turn-helix motif that interacts with the -35 element of the promoter DNA.









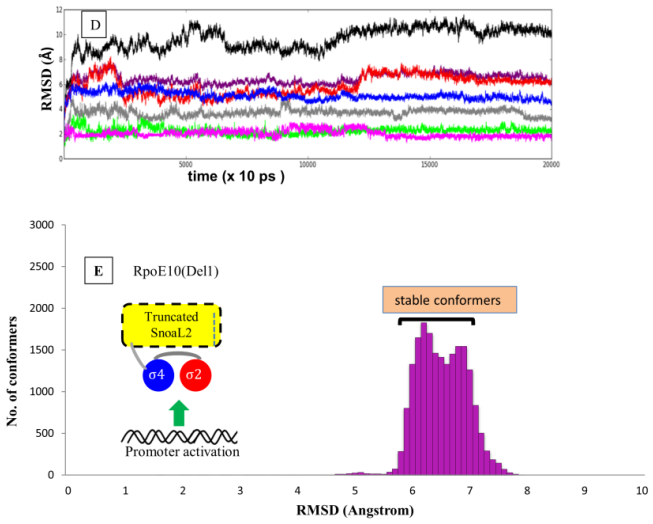


Figure 5. Analysis of RpoE10 (Del 1) model. Structural models A) RpoE10 and its deletion derivatives B) RpoE10 (Del1) and C) RpoE10(Del2). B inset view: Snapshots of the RpoE10(Del1) model structure at various time intervals illustrating the stabilization of the DGGG motif and tethering of the σ2-σ4 domain is partially truncated c-terminal SnoaL_2 domain. The ball and stick model of residues depicts the 200DGGG2G03 (magenta) interactions with linker region residues (83WLPEP87, green color). The RpoE10 structure with intact SnoaL_2 domain showing the interaction between the WLPEP, DGGGR, and NPDKV motifs (interactions are shown in Figure 1A). (C) Model of the RpoE10(Del2) illustrating the loss of interactions between the DGGGR and WLPEP motifs in the C-terminal truncated SnoaL_2 domain. D) Root-mean-square deviation profile of RpoE10(Del1) generated from 200 ns MD simulations trajectory. Truncated SnoaL_2 domain influence the compactness of σ2 -σ4 conformation and influence the backbone conformations of σ2 -σ4 domains. E) RpoE10 Del1) showed the stable conformations of the σ2- σ4 domain across a large number of conformers. F) The schematic diagram depicts the impact of compactness on promoter activation. Compact and homogenous σ2-σ4 domain conformers in RpoE10(Del1) suggest that the truncated SnoaL_2 domain-containing DGGG motif alone may constraint RpoE10 to a compact and stable structure. G) Compactness analysis using the radius of gyration calculations of individual σ2, σ4, SnoaL_2, and combined σ2-σ4 domain for truncated Snoal_2 domain of RpoE10(Del1).H) A plot of minimum distance calculated between σ2 and σ4 domain for RpoE10 (Del1). I) A comparative RMSF plot of RpoE10(Del1) with RpoE10 and RpoE10 (Mut1). An arrow marks the differential and key residue positions showing distinct RMSF peaks. These residues are mapped (in black color) onto the structure of the RpoE10 model (Figure 4A).

Molecular interaction network analyses of NPDKV, DGGGR, and WLPEP motifs

We were keen to understand how, at the molecular level, the DGGGR motif contributes to the enhanced activation of the promoter in both RpoE10(Mut1) and RpoE10(Del1). In contrast, the presence of an intact NPDKV motif has a contrary effect as it eliminates promoter activation. We examined the impact of NPDKV, and DGGGR mutations on the promoter inactivation/activation ability of RpoE10 in the light of molecular interactions, particularly salt-bridge interactions contributed by charged residues N277, D279, and K280 residues of NPDKV motif, and D200 and R204 residues of DGGGR motif across the 200ns MD simulation. These interactions were further mapped onto the representative frame (snapshot) obtained from

a most populated cluster from the 200 ns MD simulation data (Figure 6). The cluster was based on the pairwise best-fit RMSDs. The critical point that we focused upon was the molecular interactions that firmly hooked the flexible truncated fragment of Snoal_2 domain of RpoE (Del1) to the linker strand (Figure 5B (inset view) and Figure 6D). Intriguingly, a positively charged Arg74 anchored the negatively charged Asp200 of the DGGG motif in the vicinity of the linker loop. Additionally, an Arg110 strengthens this interaction by holding D200 through the salt bridge (Figure 6D and Supplemental Table S3). The spatial position of R74 is further stabilized by a network of salt-bridges between R74-E77, R74-E109, R74-D194. The aspartic acid dyad, D193-D194, of the remaining part of the SnoaL_2 domain after truncation stabilizes the spatial conformation of R74 via a network of salt-bridges. D193 interacts with R76, R188, H16, while D194 interacts with R74, R76, andR110 through salt-bridges (Supplemental Table S3). Our analysis revealed that this network of salt-bridge interactions pulls and stabilizes Arg73 to form a salt-bridge with highly conserved Asp30 of ^{30}DEAD^{33}motif of helix α2 of RpoE10-?^{2} domain. The interaction between R73-D30 was further strengthened by the stacking interaction between R73 and a highly conserved W83, which stays firmly between the σ^2-σ^4 domains (Figure 6D). In the NPDKV mutant of RpoE10, a scenario similar to RpoE10 (Del1) was observed with a slight variation on the theme. Here, the D30 is anchored on R76 in the vicinity of the W83, as seen in RpoE10(Del1). The R204 of the DGGGR motif forms a salt-bridge interaction with D279 of the NPDKV motif of the wild type RpoE10 (Figure 6A and Supplemental Table S3). However, due to the elimination of N277-R204 interaction in the mutant RpoE10 containing the NAAAV motif, R204 moves towards the linker region and interacts with E77 through salt-bridges (Supplemental Table S3). The DGGGR motif of the wild type RpoE10 contributes to two salt bridges, i.e., D200-K280 and R204-D279 (Figure 6A). However, due to D279A and K280A substitution in RpoE10(Mut1), the D200-K280 and D279-R204 interactions were eliminated (Figure 6C). Consequently, the two positively charged R289 and R292 residues from the SnoaL_2 terminal occupy the spatial location of R204 to form the salt bridge network with D200. Furthermore, D285 strengthens the position of R204 followed by transient salt-bridge interaction of E286 and E290. This interaction network pushes R204 towards the arginine tetrad (R73-R76) near the ?^{2}-linker junction to form the R204-E77 salt bridge. Interestingly, this network of interactions facilitated the R76-D30 salt bridge formation, similar to the case of RpoE10(Del1) (marked by C1, in Figure 6C). Notably, the residues of the Arginine tetrad (R73-R76), W83, and D200 are highly conserved. This suggests that W83 probably serves as an anchor to stabilize the ?^{2}-?^{4}domain of ECF41 bacterial σ factors to activate the promoter. Our analysis indicates that the stable interactions between the DGGGR motif, W83, D30, and (arginine-tetrad) thread the σ^2-σ^4 domain in a productive orientation and conformation leading to promoter activation.

The next question we pondered was how the NPDKV motif impacts promoter inactivation in the wild type RpoE10 and RpoE10 (Mut2). As evident from Figures 6A and B, the highly conserved D30 and its interaction with R76 of tetrad arginine were eliminated if NPDKV was intact, i.e., RpoE10 and RpoE10 (Mut2). The elimination of the "D30-R76" interaction enabled the entire loop of arginine tetrad (R73-R74-R75-R76) to stay away from the W83 containing linker site, thereby attaining a "switch-off" conformation leading to substantial loss of transcription activity (Figure 1). In RpoE10, the K280 and D279 participate in the salt-bridge interaction with D200 and R204, respectively, to restrict the DGGGR motif to attain an unproductive conformation (Figure 6A). Moreover, the side chain of W83 flipped out of the linker towards the NPDKV and DGGGR motif making it no more available for the stacking interaction of D30 and R76.

We were also interested in finding unique interactions in RpoE10 and RpoE10(Mut1), if any, that could shed some light on the observed inactivation and elimination of promoter activity. Indeed, we found a unique salt-bridge, D32-R282, between D32 of ?^{2}-domain and R282 of SnoaL domain that stayed for a longer duration of simulation time (Figure S1) and was typical to RpoE10 and RpoE10(Mut2). Out of 160 possible salt-bridges (Supplemental Table S4), we could not find D32-R282 salt-bridge interaction in RpoE10 (Mut1 and Del1). The formation of this unique D32-R282 salt bridge interaction could be attributed to the existence of K280-D200 and K280-E86 salt bridge in RpoE10 (Figure S1 A and B). Notably, as shown in Figure S1, D32 is located on the outer surface of the helix α-2, just diagonal to D30, suggesting its impact on the ?^{2}-domain rendering RpoE10 into an unproductive conformation in D32-R282 configuration.

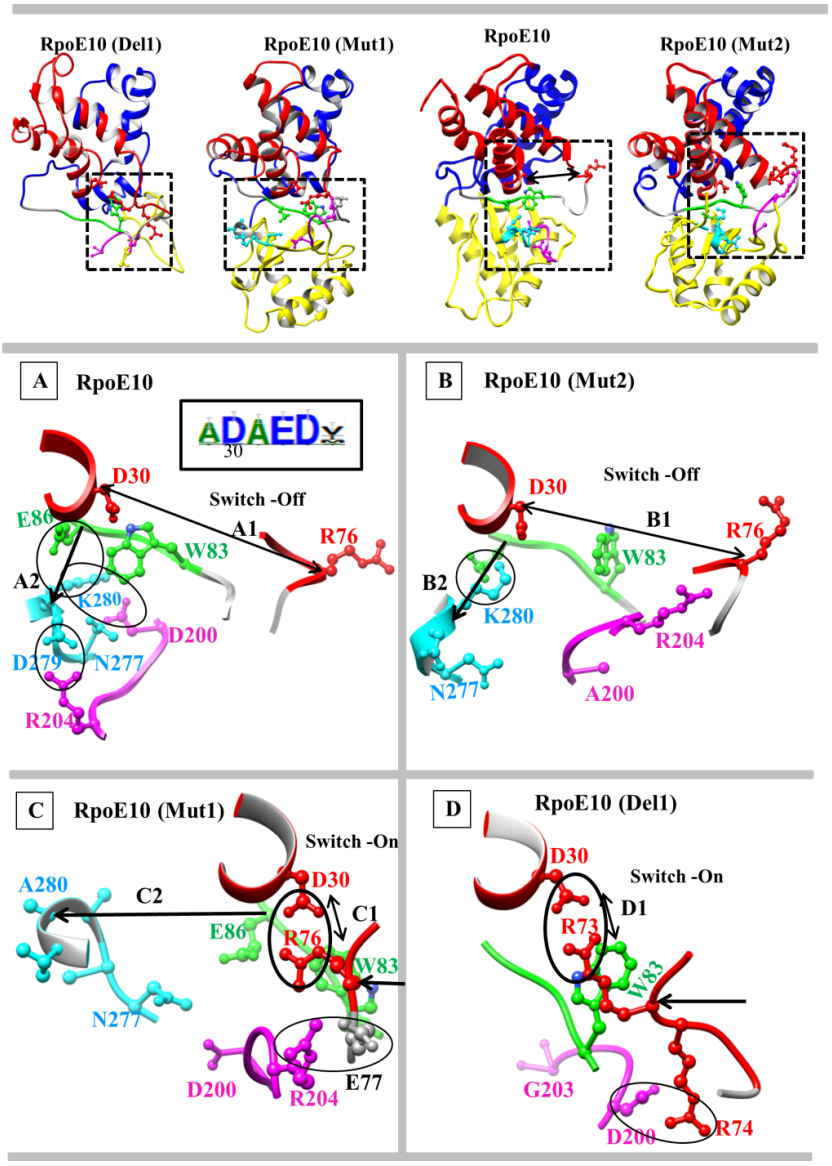


Figure 6. Molecular interaction of NPDKV, DGGGR, and WLPEP motif. The representative structural frame obtained from a most populated cluster from the 200 ns MD simulation data is presented in the top panel. The cluster was based on the pairwise best-fit RMSDs. The stable and consistent interactions were mapped on these structures. A circle shows salt bridge interactions—the double arrow marked by À1 and B1 depicts the increase in distance between the D30 and R76. The double arrow marked by C1 and D1 depicts the close contact of the D30 and R76 residues. Arrow marked by A2 and B2 indicates a close distance between the Cα of the E86 and K280. Arrow marked by C2 indicates a large distance between Cα of the E86 and K280 residues compared to A2 and B2. The colors of the figures are similar to the previous figures.

Principal component analysis of ($?^2$-$?^4$) - SnoaL2 correlated motions

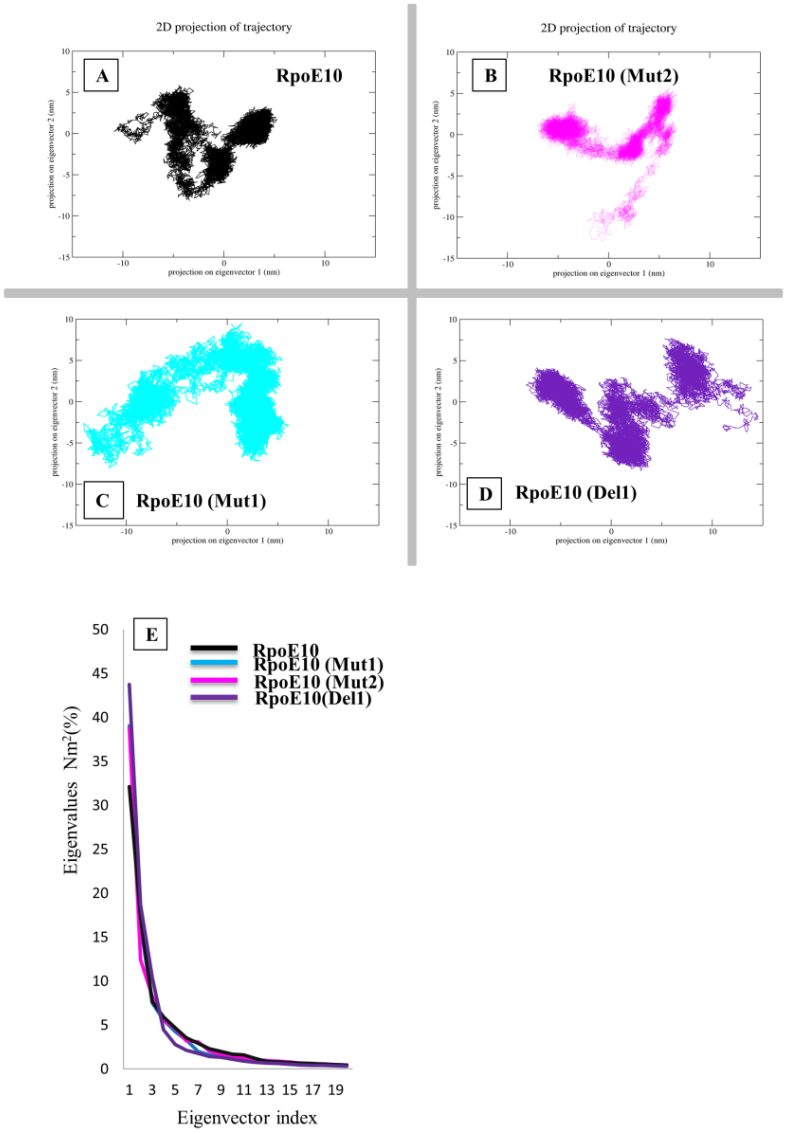
We also carried out the Principal Component Analysis (PCA), i.e., essential dynamics of the covariance matrix resulting from the 200ns MD trajectories, to investigate the underlying interactions at the NPDKV

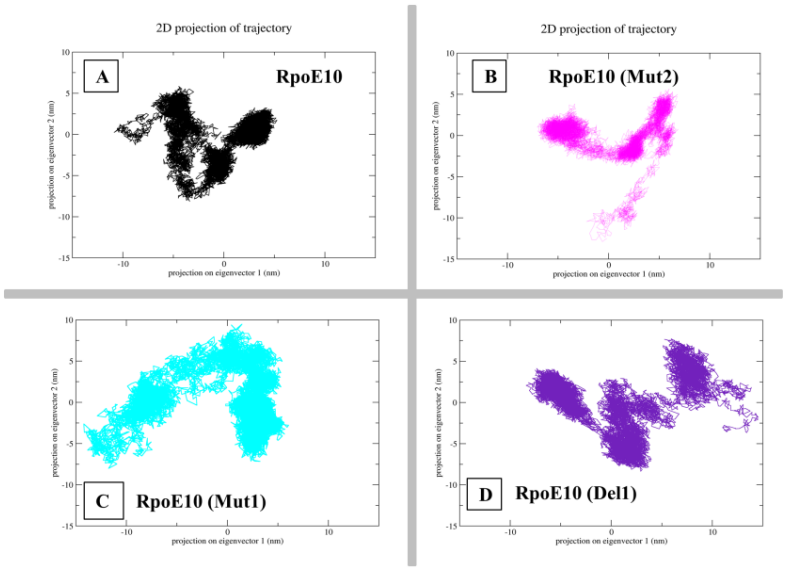
and DGGGR motifs and their impact on the overall domain motions. Principal components (PC) analysis was applied to the backbone atoms of the RpoE10 and RpoE10 (Mut1, Mut2, and Del1) forms. PC analysis is used to reveal the functionally relevant and most dominant internal modes of motion of a MD simulation [35,36]. We chose PC1, the most crucial component, accounting for maximum variability in protein conformation and computed eigenvectors and associated eigenvalues (variance). Covariance values provide information about the correlated motion. All diagonal elements of the covariance matrix were summed and termed as trace values, which provide information about the measure of the total variance. As shown in the scatter plot of PC1 vs. PC2 (Figure. 7A-D), the RpoE10 (Del1) and RpoE10 (Mut1) occupy larger subspaces corresponding to their higher trace values of the covariance matrix 63.03 (nm^2) and 58.51 (nm^2), respectively, as compared to RpoE10 (37.16 nm^2) and RpoE10(Mut-2) (43.58 nm^2) (Figure. 7(A-D)). The higher trace values of RpoE(Del1) and RpoE10(Mut1) relative to the wild-type RpoE10 and RpoE10(Mut2) suggested an association with an enhanced flexible behavior upon mutation in NPDKV or the deletion of the proximal SnoaL2 domain. We observed that the cumulative variance captured by the first 20 eigenvectors of RpoE10 (wild type) is lower as compared to the RpoE10(Mut1) and RpoE10(Del1) (Figure 7E). The RpoE10 (Mut1) and RpoE (Del1) showed 72% and 80% variance, respectively, of the cumulative proportion of the total variance captured by the first five eigenvectors (Supplemental Table S4). This suggests that the mutation at the NPDKV motif has impacted the RpoE10-SnoaL2 correlated motions. The first two principal components PC1 and PC2 account for ˜55% and ˜64% of the total variance of all the motions for RpoE10 (Mut1) and RpoE10(Del1), respectively, as compared to ˜ 49% in RpoE10 and 50% in RpoE10(Mut2) (Supplemental Table S4). The high eigenvalues, i.e., the variance of the covariance matrix in both RpoE10 (Mut1) and RpoE10 (Del1), indicate the signals for critical transitions in the conformational changes lead to enhanced activation. Percentages of variance against eigenvalues of the covariance matrix resulting from simulations are shown in Figure 7E.

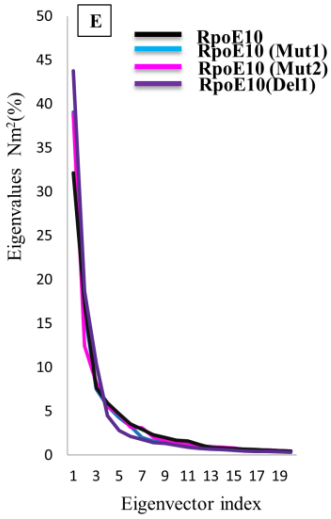
Furthermore, we correlated the elimination or activation of the promoter with the global protein motion in RpoE10 and RpoE10 (Mut1 and Mut2) and RpoE10 (Del1) using PCA analysis (Figure 7F, G, H, and I). The mutation at NPDKV and DGGGR greatly influences the SnoaL_2 domain and overall dynamics of ^{2}-?4 domain, and magnifies the significant conformational movements (Figure 7H and I). The PCA indicated that the essential motion of RpoE10(Mut1) and RpoE10(Del1) was dominated by fluctuations of the critical L3 loop (residues 46-52), also termed as "flexibility loop" and Arginine Tetrad (residues 73-76) (marked by a dotted circle and double arrow and an in Figure 7F, G, H and I) of the $?^{2}$ domain. Another noticeable and differential key fluctuation was observed in the orientation of helix-α7 of $?^{4}$ domain (marked by a dotted square in Figure 7 F, G, H, and I). The helix-α7 constitutes a helix-turn-helix motif that recognizes the -35 element of the promoter [34] (Figure 4A). The principal differential movements of the backbone atoms at -10 and -35 recognition and residues involved in promoter recognition was further recorded from the average structures of RpoE10 and its variant obtained from the PC1 (Figure S2).

A close inspection at -10 promoter binding cleft revealed correlated motions in RpoE10(Mut1) and RpoE10 (Del1); however, RpoE10 and RpoE10 (Mut2) forms exhibited a significant anti-correlated motion (Figure 7 (bottom panel) and Figure S2A and B). More minor fluctuations in RpoE10, as well as RpoE10(Mut2) in comparison to RpoE10 (Mut1) and RpoE10(Del1) were observed in the specificity loop, L3- connecting α2 and α3 (marked with a circle in Figure 7 (F,G,H, and I)). This differential flexibility of the L3-loop and orientation of the helix may directly affect the binding to the promoter region and thus impact the promoter activity. Notably, the wide-open cleft of -10 recognition site with reduced flexibility of L3-loop (shown by the blue double arrow in Figure 7) in both RpoE10 and RpoE10 (Mut2) may constrain the $?^{2}$-$?^{4}$ domain in an "open" and unproductive conformation leading to the elimination of promoter activity (Figure 7). Another contrasting movement both in the direction and distance was observed at the WLPEP motif of RpoE10 (Mut1) and RpoE10(Del1) as compared to RpoE10 and RpoE10(Mut2) (Figure 7, bottom panel). We noticed an increase in the distance between the initial and final conformation of Cα for W83 and P87 in RpoE10 (Mut1) and RpoE10 (Del1) as compared to RpoE10 and RpoE10(Mut2) (Figure 7 (bottom panel) and Supplemental Table S5.) Intriguingly, we find that WLPEP motif moves inward-up in both RpoE10 and RpoE10(Mut2), whereas in RpoE10(Mut1) and RpoE10(Del1), the movements are outward-up and outward-

down, respectively. This differential inward and outward trajectories of the WLPEP motif suggest that it may act as a sensor in transmitting the conformational signal from the Snoal_2 domain to^{2}-?^{4}domains.







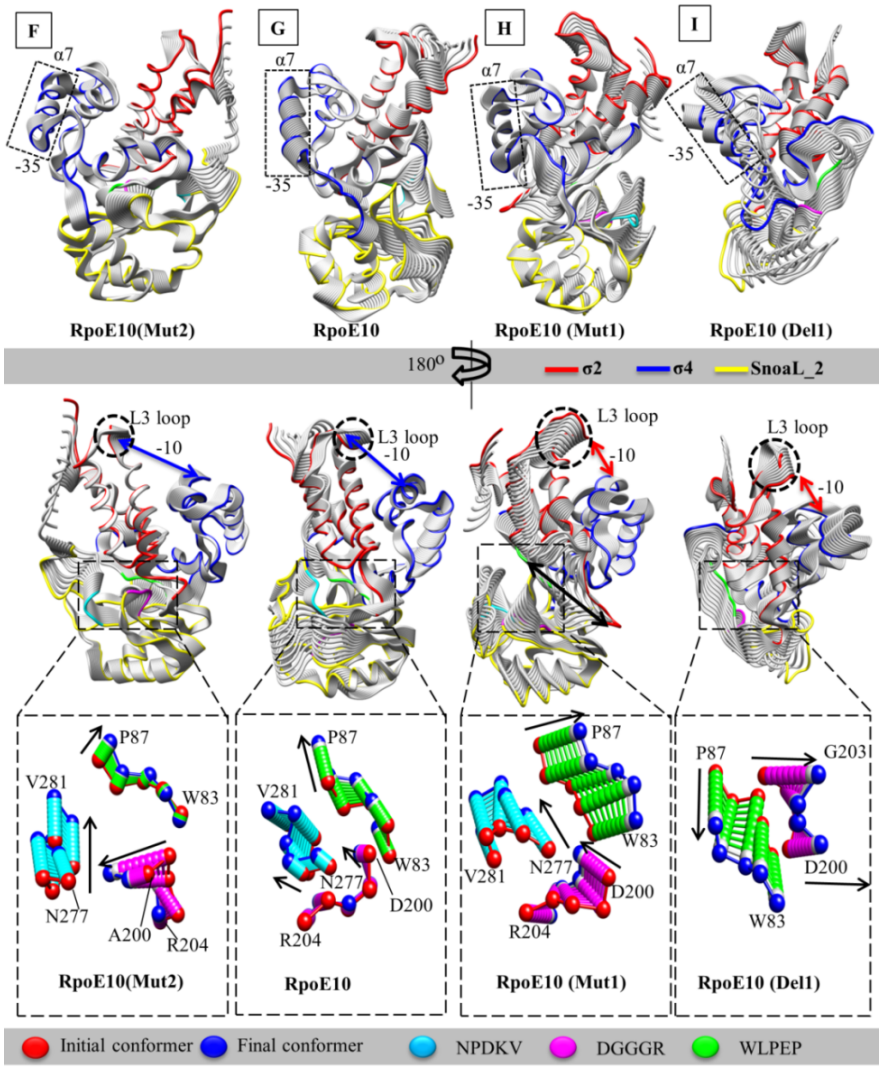


Figure 7. PCA Analysis of (?2-?4) - SnoaL2 correlated motions.PCA scatter plots (PC1 vs. PC2) representing the projections of the Cα displacements along the trajectory onto the first principal eigenvector, PC1 (x-axis), vs. the projections onto the second principal eigenvector, PC2 (y-axis), as derived from MD replicas of RpoE10 (A) wild-type, (B) Mut1, (C) Mut2(D) Del1 models. A larger subspace of B) RpoE10(Mut1) and D) RpoE10(Mut1) relative to the wild-type A) RpoE10 and B) RpoE10 (Mut2) suggested an association with the enhanced flexible behavior upon mutation in NPDKV or the deletion of the proximal SnoaL2 domain. E) Eigenvalues (variance) plotted against eigenvector indices constructed from PCA of the 200ns MD trajectories of backbone atoms for RpoE10 (a) wild-type, (b) Mut1, (c) Mut2 (d) Del1. The line represents the eigenvalues of the covariance matrix of the first 20 eigenvectors. The index with value 1 indicates the largest eigenvalue. An altered Eigenvalues shows the signals for critical

transitions in the conformational changes of the protein. **F, G, H and I: Essential dynamics analysis of Rpo10 and its mutant (Mut1 and Mut2.** The superimposition of structural coordinates associated with the principal component 1 (PC1) of F) RpoE10 and mutant (RpoE10 (G) Mut1, H) Mut2 and I) Del1)) conformers are displaying global motion. The initial conformations are colored (σ2- red, σ4-blue, and SnoaL2- yellow), and the final and the intermediate ones are shown in grey. A circle marks the fluctuations in L3-loop connecting the α2 and α3. The orientations of correlated and anti-correlated motions are indicated with a red and blue double arrow at -10 promoter binding groove, respectively. The differential movement of motif residues was recorded with divergent displacements. The direction of displacement of the NPDKV (Cyan), DGGGR(Magenta), and WLPEP(Green) motifs are labeled and shown below the respective conformers (Magnitude is given in Supplemental Table S5). Red and Blue color balls represent initial and final conformation, respectively. The RpoE10 and its mutant forms are described in the tube model.

DISCUSSION

The characteristic C-terminal extension of the ECF41 σ was thought to act as an anti- σ factor [13]. Deletion of the C-terminal domain, therefore, was assumed to activate the ECF41 σ factor. In an earlier study, the deletion of the SnoaL_2 domain of SigJ in *M. tuberculosis* led to a complete loss of the ability of SigJ to activate its target promoter, indicating a positive role of the C-terminal SnoaL_2 domain of the SigJ in *M. tuberculosis* [18]. I*n-silico* analysis of the MtbSigJ structure also suggested that the C-terminal Snoal_2 domain may not be inhibitory [18]. Based on the site-directed mutagenesis of the conserved residues of the Snoal_2 domain of RpoE10 of *A. brasilense* , we have shown that the DGGGR motif is located at the proximal end of the Snoal_2 domain, is required for the functionality of RpoE10.However, the NPDKV motif located at the distal end of the Snoal_2 domain is responsible for inhibiting the activity of RpoE10, indicating its negative role or anti-sigma factor-like activity.

Earlier experiments with the deletion of NPDKL and DGGGK motifs of ECF41 σ factors in *B. licheniformis* and *R. sphaeroides* have shown that NPDKL motif inhibited the activity of ECF41 σ factor, and deletion of the C-terminal Snoal_2 domain including NPDKV and DGGGK motif led to the complete loss in the ECF41 σ factor activity[13]. A study using Direct Coupling Analysis (DCA) identified the essential residues of $ECF41_{Bli}$ of*Bacillus subtilis* involved in the interaction between conserved residues of the flexible linker region and the conserved residues of the Snoal_2 domain. Out of the ten critical residues identified, N276 and K279 of the NPDKL were predicted to interact with the Y73 and G75 residues of the conserved consensus YVGPWLPEP motif in the linker region of the $ECF41_{Bli}$. The N276A and K279A mutations at the C-terminus of $ECF41_{Bli}$ increased its functionality. This study clearly showed that the contact between the NPDKL and the YVGPWLPEP exerts a negative regulatory effect on the activity of ECF41 σ factor [14]. One possible reason for the increased activity of RpoE10 due to the deletion of the NPDKV motif is the elimination of contact between the distal part of the C-terminal extension and the linker, which occludes binding of RpoE10 to the RNA polymerase or to the promoter. Similar to these observations, we have also shown in this study that replacing NPDKV with NAAAV in RpoE10 increased its functionality.

In $ECF41_{Bli}$, G202 of the DGGGK motif was also found to interact with the Y73 of the YVGPWLPEP motif. The G202A mutation also led to an increase in the functionality of the $ECF41_{Bli}$ [14]. However, in our study of RpoE10, we found that the replacement of DGGGR with the AAAGR led to a complete loss in its functionality. This suggests that DGGG residues in RpoE10 are required for its functionality. Because of the above, we hypothesize that the ECF41 σ factor can assume two alternative conformational possibilities: in one conformation NPDKV motif inhibits the activity of RpoE10, but in the other conformation, inhibitory interaction of NPDKV motif is prevented, leading to the activity of the ECF41 σ factor.

Here, we have demonstrated an elucidation of the role of the two conserved motifs (NPDKV and DGGGR) of the SnoaL2 domain in mediating promoter activation and elimination in RpoE10 sigma factors. A comparative Molecular Dynamics (MD) simulations, Principal Component Analysis (PCA), i.e., essential dynamics, and molecular interaction network analysis, were carried out for RpoE10 and its mutant derivatives RpoE10(Mut1), RpoE10 (Mut2), and RpoE10(Del1) for 200 ns trajectories. The structural model of Abr-RpoE10 based on the crystal structure of the Mtb-SigJ [18] provided us the opportunity to gain insight into

the mechanism of promoter activation. The structural models of RpoE10were well superposed to the crystal structures, -10 promoter/ σ^2 (PDB ID: 4LUP)[32] and -35 promoter/ σ^4 (PDB ID: 2H27) complexes [34] and adopt a conformation that can readily interact with the promoter. The central question that we addressed using MD and PCA analysis is how the NPDKV and DGGGR motifs of the SnoaL_2 domain impact the promoter activation. Essential insights into the responsible interactions at the molecular and structural level have been obtained from our MD simulation experiments. The substantial conformational changes were observed in the NPDKV and DGGGR motif in RpoE(Mut1) and RpoE10(Mut2) proteins, affecting promoter-recognition binding affinity, which is well supported by RMSD and RMSF analysis and PCA analysis. It is worth mentioning that unlike RpoE10 (Mut1) and RpoE10 (Del1), the intact NPDKV motif of wild-type RpoE10 and RpoE10(Mut2) significantly influenced the dynamic behavior of $?^2$-$?^4$and SnoaL_2 domain. We noticed a significant disruption in the salt-bridge interaction network of highly conserved D30 of the DEAD motif of $?^2$ domain and a conserved R76 of "Arginine tetrad^{73}RRRR^{76}" present near the $?^2$linker junction. The stacking interaction between W83-R76 was also eliminated. Strikingly, the presence of a unique salt-bridge interaction between E32-R282 in RpoE10 and RpoE10(Mut2) was suggested to directly impact the dynamics of $?^2$domain. An opposite scenario was observed for NPDKV mutant and truncated SnoaL_2 domain derivative of RpoE10. A schematic diagram of the molecular interaction network between the critical residues of NPDKV, DGGGR, and WLPEP impacting the activation/elimination of promoter activity is presented in Figure 8. We further examined and showed that differential salt-bridge interaction networks could influence the -10 and -35 promoter recognition site through correlated molecular motions. The salt bridge interaction network from N277, D279, K280 and D200, and R204 impacted the conformational alternation in the helix $\alpha 7$ of $\sigma 4$ domain, which forms the helix-turn-helix motif and interacts with the -35 promoter site, and the "specificity loop" L3 connecting the $\alpha 2$-$\alpha 3$ and forming the -10 recognition cleft. These promoter recognition sites were significantly altered upon mutations in NPDKV and DGGGR. Based on these findings, we suggest that NPDKV and DGGGR motifs are the conformational switches that trigger the productive and unproductive conformations responsible for the activation or elimination of promoter. Also, the significant and differential inward and outward movement of the WLPEP motif makes it a sensor in transmitting the conformational signal from the Snoal_2 domain to^2-$?^4$domains. These results suggest that NPDKV and DGGGR, together with WLPEP, may play key roles in modulating correlated motions of the RpoE10 domains. Mutation at the NPDKV motif induces a conformational "switch" of the NPDKV motif to eliminate its inhibitory effect and activate its target promoter.

Overall, our analysis clearly indicates that the NPDKV motif at the C-terminal extension acts as an inhibitory "switch" on RpoE10 activity, and interaction of DGGGR motif with the linker WLPEP motif along with "Arginine tetrad" located between σ^2 and σ^4, may be required for its activity. In the RpoE10(Del2), however, removal of the C-terminal part of the SnoaL_2 domain, including NPDKV motif as well as DGGGR motif, fails to provide the necessary interaction between WLPEP and DGGGR required for the RpoE10 to acquire the conformation needed for activating *abm*promoter (Figure. 2). The above analysis showed that the residues of DGGGR motif interact and stabilize the linker region residues R76 withoutthe ^{277}NPDKV^{281} motif. The amino acid residues of the^{277}NPDKV^{281} motif pull the residues of the DGGGR motif away from the linker region and destabilize the interactions of residues of the^{83}WLPEP^{87} motif. Our simulation results clearly explain the reason for the strong link between NPDKV, DGGGR, and WLPEP motif. Our findings indicated that the conformational transitions associated with the residual motions of NPDKV and DGGGR motif of SnoaL2 domain and WLPEP motif of linker strand connecting the $?^2$ and $?^4$ domains could be used to understand the structural alternations and allosteric regulations in ECF41 regulons. Altogether, the results reported in this study will provide a greater understanding of ECF41-associated promoter regulation and will pave a path for understanding the functional role of C-terminal Snoal_2 domain-containing ECF41-? factors.

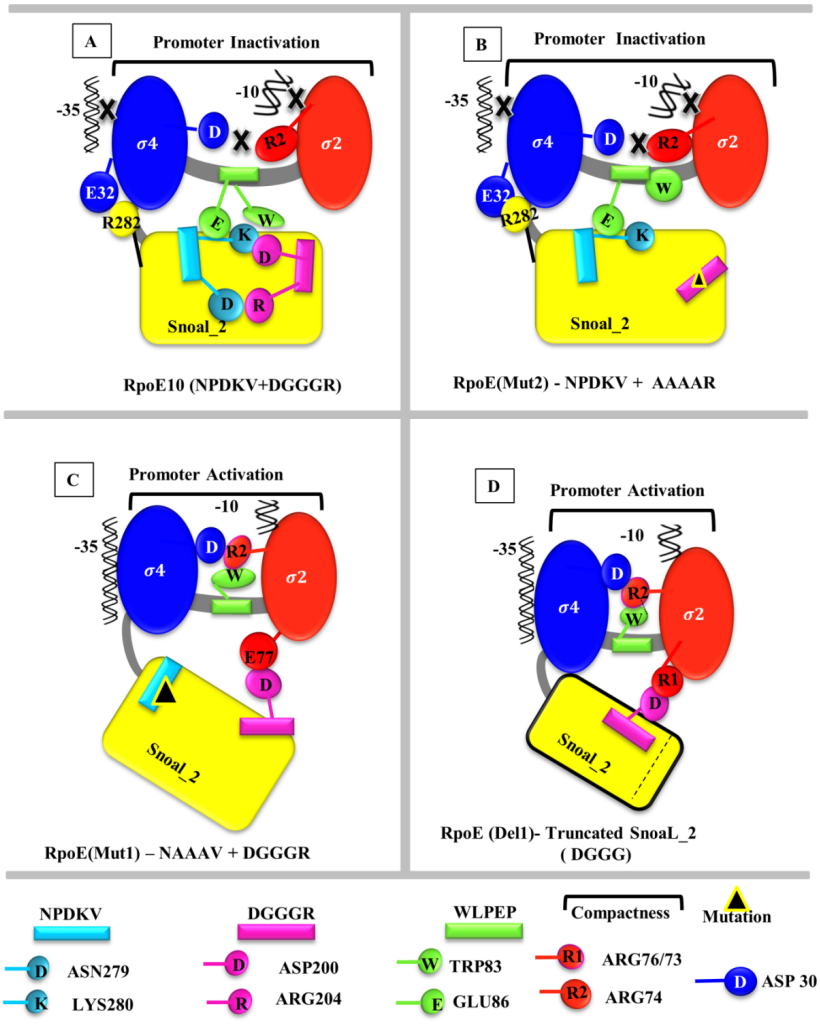


Figure 8. Schematic diagram of the promoter activation regulation using NPDKV, DGGGR, and WLPEP switches of SnoaL_2 domain in RpoE10.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

ACKNOWLEDGEMENT

This work was supported by a research grant to AKT from Council for Scientific and Industrial Research (CSIR), New Delhi. VSS and APD were also supported with a Junior/Senior Research Fellowship from CSIR. Financial support to AKT from the Institution of Eminence (IoE) scheme of BHU is also gratefully acknowledged.

Credit authorship contribution statement

AKT conceptualized and finalized the manuscript's text, APD and VSS performed wet-lab experiments, and EP and RM performed and described all the *in silico* studies. All authors read and approved the final manuscript.

Declaration of competing interest

The authors declare that there are no conflicts of interest with the contents of this article.

REFERENCES

1. Helmann JD, Chamberlin MJJArob. Structure and function of bacterial sigma factors. 1988;57(1):839-872.

2. Helmann JDJBsr. Regulation by alternative sigma factors. 2010:31-43.

3. Gruber TM, Gross CAJARiM. Multiple sigma subunits and the partitioning of bacterial transcription space. 2003;57(1):441-466.

4. Paget MS, Helmann JDJGb. The σ 70 family of sigma factors. 2003;4(1):1-6.

5. Lonetto MA, Brown KL, Rudd KE, Buttner MJJPotNAoS. 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. 1994;91(16):7573-7577.

6. Helmann JD. The extracytoplasmic function (ECF) sigma factors. 2002.

7. Butcher BG, Mascher T, Helmann JD. Environmental sensing and the role of extracytoplasmic function sigma factors. In: *Bacterial Physiology.* Springer; 2008:233-261.

8. Staroń A, Sofia HJ, Dietrich S, Ulrich LE, Liesegang H, Mascher TJMm. The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family. 2009;74(3):557-581.

9. Mascher TJCoim. Signaling diversity and evolution of extracytoplasmic function (ECF) σ factors. 2013;16(2):148-155.

10. Sineva E, Savkina M, Ades SEJCoim. Themes and variations in gene regulation by extracytoplasmic function (ECF) sigma factors. 2017;36:128-137.

11. Ho TD, Ellermeier CDJCoim. Extra cytoplasmic function σ factor activation. 2012;15(2):182-188.

12. Pinto D, Liu Q, Mascher TJMm. ECF σ factors with regulatory extensions: the one-component systems of the σ universe. 2019;112(2):399-409.

13. Wecke T, Halang P, Staroń A, Dufour YS, Donohue TJ, Mascher TJM. Extracytoplasmic function σ factors of the widely distributed group ECF41 contain a fused regulatory domain. 2012;1(2):194-213.

14. Wu H, Liu Q, Casas-Pastor D, Dürr F, Mascher T, Fritz GJMm. The role of C-terminal extensions in controlling ECF σ factor activity in the widely conserved groups ECF41 and ECF42. 2019;112(2):498-514.

15. Mishra MN, Kumar S, Gupta N, Kaur S, Gupta A, Tripathi AKJM. An extracytoplasmic function sigma factor cotranscribed with its cognate anti-sigma factor confers tolerance to NaCl, ethanol and methylene blue in Azospirillum brasilense Sp7. 2011;157(4):988-999.

16. Gupta N, Kumar S, Mishra MN, Tripathi AKJM. A constitutively expressed pair of rpoE2–chrR2 in Azospirillum brasilense Sp7 is required for survival under antibiotic and oxidative stress. 2013;159(Pt_2):205-218.

17. Rai AK, Dubey AP, Kumar S, et al. Carotenoid biosynthetic pathways are regulated by a network of multiple cascades of alternative sigma factors in Azospirillum brasilense Sp7. 2016;198(21):2955-2964.

18. Goutam K, Gupta AK, Gopal BJNar. The fused SnoaL_2 domain in the Mycobacterium tuberculosis sigma factor σJ modulates promoter recognition. 2017;45(16):9760-9772.

19. Dubey AP, Pandey P, Singh VS, et al. An ECF41 family σ factor controls motility and biogenesis of lateral flagella in Azospirillum brasilense Sp245. 2020;202(16):e00231-00220.

20. Goutam K, Gupta AK, Gopal BJACSFSBC. Crystallographic studies of the extracytoplasmic function σ factor σJ from Mycobacterium tuberculosis. 2015;71(8):946-950.

21. Zimmermann L, Stephens A, Nam S-Z, et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. 2018;430(15):2237-2243.

22. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. 2004;25(13):1605-1612.

23. Hess B, Kutzner C, Van Der Spoel D, Lindahl EJJoct, computation. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. 2008;4(3):435-447.

24. Bjelkmar P, Larsson P, Cuendet MA, Hess B, Lindahl EJJoct, computation. Implementation of the CHARMM force field in GROMACS: analysis of protein stability effects from correction maps, virtual interaction sites, and water models. 2010;6(2):459-466.

25. MacKerell Jr AD, Bashford D, Bellott M, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. 1998;102(18):3586-3616.

26. Darden T, York D, Pedersen LJTJocp. Particle mesh Ewald: An N[?] log (N) method for Ewald sums in large systems. 1993;98(12):10089-10092.

27. Hess B, Bekker H, Berendsen HJ, Fraaije JGJJocc. LINCS: a linear constraint solver for molecular simulations. 1997;18(12):1463-1472.

28. Bussi G, Donadio D, Parrinello MJTJocp. Canonical sampling through velocity rescaling. 2007;126(1):014101.

29. Parrinello M, Rahman AJJoAp. Polymorphic transitions in single crystals: A new molecular dynamics method. 1981;52(12):7182-7190.

30. Van Gunsteren WF, Berendsen HJJMS. A leap-frog algorithm for stochastic dynamics. 1988;1(3):173-185.

31. Humphrey W, Dalke A, Schulten KJJomg. VMD: visual molecular dynamics. 1996;14(1):33-38.

32. Campagne S, Marsh ME, Capitani G, Vorholt JA, Allain FHJNs, biology m. Structural basis for- 10 promoter element melting by environmentally induced sigma factors. 2014;21(3):269-276.

33. Campbell EA, Muzzin O, Chlenov M, et al. Structure of the bacterial RNA polymerase promoter specificity σ subunit. 2002;9(3):527-539.

34. Lane WJ, Darst SAJPb. The structural basis for promoter- 35 element recognition by the group IV σ factors. 2006;4(9):e269.

35. Amadei A, Linssen AB, Berendsen HJJPS, Function,, Bioinformatics. Essential dynamics of proteins. 1993;17(4):412-425.

36. David CC, Jacobs DJ. Principal component analysis: a method for determining the essential dynamics of proteins. In: *Protein dynamics.* Springer; 2014:193-226.