Journal of Structural Biology 198 (2017) 134–146

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Structural and dynamics studies of the TetR family protein, CprB from *Streptomyces coelicolor* in complex with its biological operator sequence

Hussain Bhukya [a,b], Asis K. Jana [c,d], Neelanjana Sengupta [e], Ruchi Anand [a,*]

[a] Department of Chemistry, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India
[b] IITB-Monash Research Academy, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India
[c] India Physical Chemistry Division, CSIR-National Chemical Laboratory, Pune 411008, India
[d] Academy of Scientific and Innovative Research (AcSIR), New Delhi, India
[e] Dept. of Biological Sciences, Indian Institute of Science Education and Research (IISER) Kolkata, Mohanpur 741246, W. Bengal, India

ARTICLE INFO

Article history:
Received 28 November 2016
Received in revised form 13 March 2017
Accepted 21 March 2017
Available online 22 March 2017

Keywords:
TetR-FTR
CprB-DNA complex
Quorum sensing
N-terminal
Streptomyces coelicolor
γ-Butyrolactone receptors

ABSTRACT

In *Streptomycetes*, tetracycline repressor family of transcription regulators (TetR-FTRs) controls various biological processes including antibiotic biosynthesis, cellular morphology and innate resistance. Here, we focus on understanding the structural basis of transcription regulation by CprB, a member of TetR-FTRs from *S. coelicolor*. CprB is implicated as a receptor of γ-butyrolactones, a class of quorum sensing molecules, responsible for initiating secondary metabolic pathways. In order to understand the molecular mechanism of DNA recognition, the X-ray structure of CprB in complex with its biological relevant operator sequence was solved to a resolution of 3.95 Å. Furthermore, to refine and compliment the results, atomistic molecular dynamics simulations were carried out using the X-ray structure as the template. The studies reveal that CprB binds to DNA as dimer of dimers with this mode of interaction results in minimal distortion in the DNA, enabling these proteins to recognize multiple sequences with varying affinity. Another crucial finding from our simulation results was that the positively charged N-terminal arm of CprB brings extra stability to the protein-DNA complex by interacting with the minor-groove of the DNA and anchoring itself to the phosphate backbone. Corroborating electrophoretic mobility shift assay and fluorescence anisotropy experiments showed that the mutant ΔN6-CprB exhibited about 7–8 fold reduced DNA binding. Comparison with other TetR-FTRs reveals that this strategy is also employed by over 25% of TetR-FTRs, where N-terminal anchoring mechanism is used to enhance selectivity for a particular DNA sequence.

1. Introduction

The TetR-FTRs are one of the most ubiquitous class of transcription regulators (Cuthbertson and Nodwell, 2013; Ramos et al., 2005). There are over 10,000 TetR-FTRs reported in the non-redundant databases and it is worth noting that almost all bacterial genomes possess at least one such protein. In certain organisms like *Streptomycetes* there is an abundance of TetR-FTRs and some species in this genus are known to have over 100 such regulators (Ahn et al., 2012). TetR-FTRs have attracted a lot of attention because in several pathogenic strains like *Staphylococcus auerus* (Schumacher et al., 2001), *Mycobacterium tuberculosis* (Bolla et al., 2012) etc., they regulate the efflux pathways conferring antibiotic resistance (Cuthbertson and Nodwell, 2013; Ramos et al., 2005). For example, the founding member of this family, TetR, controls the export of tetracycline and thereby aids in acquiring resistance to this antibiotic (Hinrichs et al., 1994; Orth et al., 2000). In actinomycetes, TetR-FTRs are also known to be associated with other important functions like biofilm formation, nitrogen uptake, morphogenesis and density dependent cell-cell communication system known as quorum sensing; here, they help connect the organism with environmental stimuli (Bassler and Losick, 2006; Ramos et al., 2005). *S. griseus*, ArpA (A-factor receptor protein A), a TetR-FTR is one such protein that responds to the small molecule quorum sensing modulator A-factor (2-isocapryloyl-3-R-hydroxyme thyl-γ-butyrolactone), a γ-butyrolactones (GBL) thereby, triggering a host of secondary metabolic pathways that include antibiotic biosynthesis and aerial mycelium formation (Onaka et al., 1995).

CprB (*coelicolor* pigment regulator protein B) from *S. coelicolor* A3(2), a homologue of ArpA is the only structurally characterized

* Corresponding author at: Structural Biochemistry Lab, Department of Chemistry, Indian Institute of Technology Bombay, Mumbai 400076, India.
E-mail address: ruchi@chem.iitb.ac.in (R. Anand).

http://dx.doi.org/10.1016/j.jsb.2017.03.006
1047-8477/

member of this sub-class of TetR-FTRs (Natsume et al., 2004; Onaka et al., 1998; Wasserman et al., 1974). In 2004, Natsume et al. solved the crystal structure of the apo form of CprB which revealed that like other TetR-FTRs, CprB also harbors a modular architecture, consisting of a highly conserved helix-turn-helix (HTH) motif in the N-terminal DNA binding domain (DBD) and a diverse C-terminal ligand binding domain (LBD) (Natsume et al., 2004; Ramos et al., 2005). The origin of the dissimilar nature of LBD is due to the fact that TetR-FTRs control multiple pathways by responding to a spectrum of inducer molecules that operate differentially. Hence, each LBD is tuned to respond to a particular ligand or a class of compounds associated with its downstream function (Ramos et al., 2005). In 2014, the crystal structure of the DNA bound form of CprB was reported, where the choice of DNA used for crystallization was a sequence that was extracted from a random pool of short oligonucleotides capable of binding both ArpA and CprB (Bhukya et al., 2014; Onaka and Horinouchi, 1997). Comparison of the apo and DNA bound crystal structures of CprB revealed that a substantial change in conformation of the protein occurs upon complexation. The DNA bound form of the protein undergoes a pendulum like motion along the dimeric interface facilitating the HTH-motif to dock into the major-groove of the DNA (Bhukya et al., 2014). Moreover, unlike the apo form of CprB that exists as a dimer, the DNA bound form exists as a dimer of dimers (Bhukya et al., 2014). This DNA induced oligomerization of the dimer is also observed earlier like in the case of structurally characterized *Staphalococus auerus* QacR, an efflux pump regulator that binds to a broad spectrum of quaternary ammonium cationic compounds regulating the transcription of the multidrug transporter, *qacA* (Grkovic et al., 1998, 2001; Schumacher et al., 2001). Structural comparison of various TetR-FTRs revealed that they fall in two major classes. One class of proteins such as TetR and SimR are highly specific towards their ligands and bind DNA as a dimer, inducing high degree of distortion (around 15–17°) in their respective operator sequences (Le et al., 2011). The other sub-class binds as dimer of dimers and relies on an extended surface area to ensure effective contact (Bhukya et al., 2014; Le et al., 2011; Schumacher et al., 2002). This dimer of dimer sub-class has over six structurally characterized members and they encompass functions ranging from scanning of DNA like in the case of master regulator Ms6564 (Yang et al., 2013), serving as broad spectrum efflux pumps like in the case of QacR (Grkovic et al., 1998, 2001; Schumacher et al., 2001, 2002) and CgmR (Itou et al., 2010) to quorum sensing as in the case of CprB (Bhukya et al., 2014).

Using the non-biological structurally characterized consensus DNA sequence (CS), a genome wide search revealed that the sequence [−47 to −21] base pair (bp), named as OPB (operator for CprB)] upstream of its own gene, exhibited high sequence similarity to CS. During the course of analysis, it was observed that OPB sequence is only 60% identical to CS and it exhibits an asymmetry in its sequence. Furthermore, a series of electrophoretic mobility shift assay (EMSA) as well as isothermal titration calorimetry (ITC) experiments have previously demonstrated that CprB binds 2–3 folds tighter to the OPB sequence, when compared to CS. Since CprB binds to its own upstream sequence it was proposed that like many other TetR-FTRs (QacR, TetR, ActR etc.), CprB is also autoregulatory in nature (Bhukya et al., 2014) (Bertrand et al., 1983; Grkovic et al., 1998; Tahlan et al., 2008). Here, to understand the similarities and differences in the mode of CprB binding with the two sequences CS and OPB, we report the crystal structure of the CprB-OPB complex to a resolution of 3.95 Å. Moreover, to achieve deeper insights into its mode of binding, comparative, fully atomistic molecular dynamics (MD) studies were also carried out. MD results revealed that the arginine rich N-terminal finger region of CprB is important for stabilizing the protein-DNA complex and in conferring its selectivity. Hence, corroborating experimental studies were undertaken to establish the role of this region. Overall, these studies provide important insights into the underlining mechanism of DNA recognition by the GBL binding sub-class of TetR-FTRs.

2. Material and methods

2.1. Cloning, overexpression and purification of native and mutant CprB

CprB encoding 215 amino acids was overexpressed and purified in the *E. coli* system. The vector pET26b(+) containing *cprB* was obtained from Ryo Natsume. Five amino acids (2–6) from the N-terminal of CprB were truncated to obtain deletion mutant and the cloning was carried out by amplifying the desired gene product using 5′-CGCCATATGGCGGAACGGACCCGCGC-3′ and 5′-CCGCTCG AGTCAGGTCGTGCCCGTCT-3′ primers. The amplified gene was inserted into the NdeI and XhoI restriction sites of pET26b(+). Native and mutant versions of CprB were overexpressed by introduced pET26b(+) containing gene of interest into *E. coli* expression cells, BL21(DE3)pLysS. Induction of native and mutant CprB was done by adding isopropyl-β-thio galactopyranoside to a final concentration of 1 mM (Natsume et al., 2003) and grown for 3 h at 37 °C. The culture was then cooled to 25 °C and grown at the same temperature for 3 h. Cells were harvested and debris was removed by high speed centrifugation at 20,000 rpm for 50 min. The supernatant was then mixed with SP Sepharose beads, which were pre-equilibrated in buffer A (50 mM phosphate buffer pH 7.0) (Natsume et al., 2003). Protein elution was performed with a linear gradient of sodium chloride (100–400 mM) in buffer A. The eluted fractions of protein were desalted using an Econo-Pac 10DG column pre-equilibrated with buffer A. The desalted fractions of CprB were rebound to the SP Sepharose beads and eluted with 1 M NaCl in buffer A. Finally, the protein was desalted with buffer A. The purity of the protein was verified by running an SDS-PAGE gel analysis with 15% polyacrylamide gel followed by Coomassie Blue staining. Detailed purification protocol is given in earlier report (Bhukya et al., 2014).

2.2. Synthesis of oligonucleotides

DNA oligonucleotides sequences used for EMSA and crystallization studies were synthesized on MerMade4. The synthesized oligonucleotides were deprotected and purified by denaturing PAGE (20%, 7 M urea) employing standard protocols. The complementary strands (1:1 ratio in concentration) were annealed by heating at 95 °C for 5 min (in a buffer containing 5 mM Tris-HCl, pH 7.5, 15 mM NaCl and 0.1 mM EDTA, pH 8.0) and allowed to cool down slowly to room temperature, after which, they were stored at −20 °C. Details of synthesis and purification is given in earlier report (Bhukya et al., 2014).

2.3. Radiolabeling of oligonucleotides

Oligonucleotide (OPB) was 5′-end labeled to carry out EMSA studies. 10 pmol of unlabelled OPB was mixed with 1× polynucleotide kinase (PNK) buffer [50 mM Tris–HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM dithiothreitol (DTT) and 0.1 mM spermidine]. T4 polynucleotide kinase enzyme, 5 U and [γ-^{32}P] ATP (3300 Ci/mmol) were further added and the reaction volume was adjusted to 10 μL. After incubating the reaction mixture at 37 °C for 1 h, the enzyme was then deactivated by heating the reaction mixture to 70 °C for 3 min. The labeled product was then purified using the QIAquick nucleotide removal kit protocol provided by Qiagen. Details of labelling is given in earlier report (Bhukya et al., 2014).

2.4. Electrophoretic mobility shift assay (EMSA)

CprB-DNA binding assays were carried out using 5′-end radiolabeled oligonucleotide. Approximately 1 nM of annealed DNA (~5000 cpm) was incubated with a 2-fold serially diluted protein (starting from 6 μM to 23 nM) at 20 °C for 30 min in a buffer containing 10 mM Tris–HCl (pH 7.8), 50 mM KCl, 1 mM EDTA, 1 mM DTT, 5% (vol/vol) glycerol and 10 mg hemoglobin in a total volume of 20–40 mL (Sugiyama et al., 1998). After the incubation, the samples were run on 6% non-denaturing polyacrylamide gel with 1× TBE as a running buffer (89 mM of each Tris and Boric acid and 2 mM of EDTA, pH 8.3) at 4 °C and 100 V for 1 h. EMSA results were collected and analyzed on a Storm825 and autoradiograms were generated using the ImageQuantTL software provided by GE Healthcare.

2.5. Steady state fluorescence measurement

The steady state fluorescence anisotropy measurements of 5′ – carboxyfluorescein (FAM) labeled OPB DNA sequence with native and ΔN6-CprB were performed. The experiments were carried out on a Varian Cary Eclipse spectrofluorimeter with manually installed polarizer at 23 °C. The samples were excited at 495 nm and fluorescence emission was recorded at 520 nm for excitation at parallel ($I_{||}$) and perpendicular ($I_{\perp}$) to the incident light, the path length was 1 cm and the signal recorded for 0.15 s. For all the experiments the 20 μM of the OPB DNA was taken as the titrant in the cuvette and titrated against the solution of the protein in buffer A. After the addition of the protein, the samples were incubated for 3–4 min before recording the emission data. All experiments have been repeated two times. The G-factor was determined from a solution of free fluorescence. The anisotropy was calculated by the following equation and plotted against the concentration of the protein:

$$r = \frac{I_{||} - G(\lambda)I_{\perp}}{I_{||} + 2G(\lambda)I_{\perp}} \quad (1)$$

2.6. Co-crystallization of CprB with OPB

The concentrated CprB (8 mg/mL) in buffer A was mixed with annealed 27 bp blunt OPB DNA sequence (5′-AGGCAGGCGG CACGGT CTGTTGAGTTC-3′) in the ratio of 1:1.2 (CprB:OPB) and was incubated at 20 °C for 30 min. Co-crystallization trails of CprB-OPB were performed using hanging-drop vapor diffusion method with Natrix HR116 and Natrix2 HR117 crystallization screens from Hampton Research. Each drop contained 2.0 μL of CprB-OPB and 1.5 μL of 200 μL well solution. Plates were stored at 20 °C and the crystals were obtained after 2 weeks in the condition with 0.1 M ammonium acetate, 0.02 M $MgCl_2{\cdot}6H_2O$, 0.05 M HEPES sodium pH 7.0 and 5% PEG 8000. The CprB-OPB crystals were cryoprotected with 20% ethylene glycol

2.7. Data collection and refinement

All the crystals were flash cooled using liquid nitrogen and mounted onto the goniostat at the BM-14 of European synchrotron radiation facility (ESRF, Grenoble, France). Data was collected for 8 s of exposure at every 1° oscillation on MAR CCD detector. The resultant data was integrated and scaled using XDS (Kabsch, 1988). The data from CprB-OPB complex was collected to 3.95 Å resolution. The coordinates of CprB-CS complex (PDB entry: 4PXI) were used for molecular replacement and the initial search was performed with Auto-Rikshaw (Panjikar et al., 2005). The asymmetric unit contained two homodimers of CprB and a double stranded OPB DNA. The OPB DNA was fit in the electron density manually using COOT (Emsley et al., 2010), the structure was then refined using CNS (Brunger et al., 1998), REFMAC5 and Phenix (Adams et al., 2010; Murshudov et al., 2011). To establish the validity of the structure sample density for select amino acids involved in DNA specific contacts is given in Fig. S1. Figures were rendered using PyMol (DeLano, 2002) and the helical parameters of OPB DNA were calculated using X3-DNA (Zheng et al., 2009).

2.8. Preparing protein-DNA systems for MD simulations

Initial coordinates were obtained from crystal structure of CprB-CS and CprB-OPB complexes. The missing residues [1–4, 113, 114, 165–175 and 212–215 in monomer A (mA); 1–4, 166–169 and 213–215 in monomer B (mB); 1–4, 115, 168–174 and 213–215 in monomer C (mC); 1–7, 77–79, 118–119, 167–173; and 213–215 in monomer D (mD)] and the DNA bases (1–3 and 24–27 in both the DNA strands) were added using COOT (Emsley et al., 2010). The addition was performed by superposing the apo CprB structure using LsqAB program available in the CCP4i suite of programs and the coordinates for the missing residues were taken and included in the final PDB. All the dihedrals and torsion angles for the added residues were corrected using the refinement procedure in COOT. Energy minimization was performed for CprB-CS and CprB-OPB complexes and then the minimized structures were adopted as starting structures for further atomistic MD simulations. The $NH_3^+$and COO^- were added to N- and C-terminal of each protein molecule. CprB-DNA complexes were solvated with TIP3P water model (Jorgensen et al., 1983) such that there was a minimum distance of 13.0 Å between any solute atom and box edge. Sodium ions were added to neutralize each system.

2.9. MD simulations

All simulations in this study were performed with the NAMD2.9 (Kalé et al., 1999) simulation package, using the CHARMM22 all-atom force field (MacKerell et al., 1998) with the CMAP corrections (Mackerell et al., 2004). The VMD toolkit was used for analyses (Humphrey et al., 1996). The SHAKE algorithm was used to constrain the bond-lengths involving hydrogen atoms (Ryckaert et al., 1977). Energy minimization was initially performed for 10,000 steps on all systems using the conjugate gradient technique, followed by simulations using a time step of 2 fs in the isothermal–isobaric ensemble. Orthorhombic periodic boundary conditions were applied, and the electrostatic interactions were calculated using the particle-mesh Ewald (PME) method (Essmann et al., 1995). The cutoff distances for non-bonded interactions were set to 12 Å, with smoothing started at 10.5 Å. A constant temperature of 310 K was maintained using Langevin dynamics, using a collision frequency of 1 ps^{-1}, and a constant pressure of 1 atmosphere was maintained via the Langevin piston Nose-Hoover algorithm (Feller et al., 1995; Martyna et al., 1994).

3. Results

3.1. Crystal structure of CprB-OPB complex

This study was carried out to understand the mode of recognition of CprB with its biologically relevant operator sequence. In order to do so, the X-ray structure of CprB was solved in complex with OPB sequence. The X-ray crystallographic data processing and structure refinement statistics for CprB-OPB complex are presented in Table 1. The alignment of the 27 bp non-palindromic OPB sequence used for co-crystallization with CS is depicted in Fig. 1A. Due to unavailability of highly diffracting crystals, the structure

Table 1
Crystallographic data processing statistics.

Crystallographic data processing and refinement statistics	
Space group	$P3_2$
Unit cell dimensions (Å)	a = 149.31, b = 149.31, c = 70.35
Resolution (Å)	60.9–3.95 (4.02–3.95)*
Wavelength (Å)	0.9737
Total no. of reflections	53,615
No. of reflections in working set	14,158
No. of reflections in test set	1188
Total no. of unique reflections	15,346 (778)*
Redundancy	3.5 (3.1)*
Completeness	99.7 (100.0)*
R_{merge} (%)	7.9 (67.4)*
I/σ	22.4 (1.8)*
R_{work}/R_{free}	23.84/29.90
Ramachandran favoured (%)	83.90
Ramachandran allowed (%)	13.8
Ramachandran outliers (%)	2.3
RMSD bond distance (Å)	0.01
RMSD bond angle (°)	1.82
PDB entry	5H58

* The figures in brackets signify the values for highest resolution shell. $R_{\text{merge}} = \Sigma_{hkl}\,\Sigma_i\,|I_i(hkl) - \{I(hkl)\}|/\Sigma_{hkl}\,\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the i^{th} observation of reflection hkl and $\{I(hkl)\}$ is the weighted average intensity for all observations i of reflection hkl. The R-factors R_{work} and R_{free} are calculated as follows: $R = \Sigma(|F_{obs} - F_{calc}|)/\Sigma|F_{obs}| \times 100$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

could only be solved to 3.95 Å. A monomer of CprB consists of 215 amino acids, however, due to the weak and/or no observed electron density for the residues 1–4, 113, 114, 165–175 and 212–215 in mA, 1–4, 166–169 and 213–215 in mB, 1–4, 115, 168–174 and 213–215 in mC and 1–7, 77–79, 118–119, 167–173 and 213–215 in mD, they were not included in the final refined structure. The DNA bases (1–3, 24–27 and their complementary bases) were also omitted from the model during the structural solution. The asymmetric unit consists of a pair of CprB dimers and a double stranded OPB DNA sequence. Similar to the previously reported CprB-CS structure (Bhukya et al., 2014), CprB-OPB complex also forms a dimer of dimer with the OPB sequence being sandwiched in between the two dimeric units of CprB (Fig. 1B). Briefly, like other TetRs, CprB is α-helical in nature and possesses a "Ω" shaped scaffold (Fig. 1B). Out of the 10 α-helices in CprB, the first three are engaged in DNA binding whereas, the C-terminal (α4 – α10) of the protein forms LBD that nests the ligand binding pocket. It is here that the γ-butyrolactone is hypothesized to bind via interactions with a conserved tryptophan W127 residue and other hydrophobic amino acids. Helices α8 and α9 are involved in dimerization and it is along this interface conformational changes occur upon DNA binding. CprB undergoes a pendulum like motion and these changes are commensurate with alterations in hydrogen bonding and hydrophobic contacts along the dimerization interface in apo verses the DNA bound forms. The connector helix α4 that interfaces the LBD and the DBD is also involved in this motion.

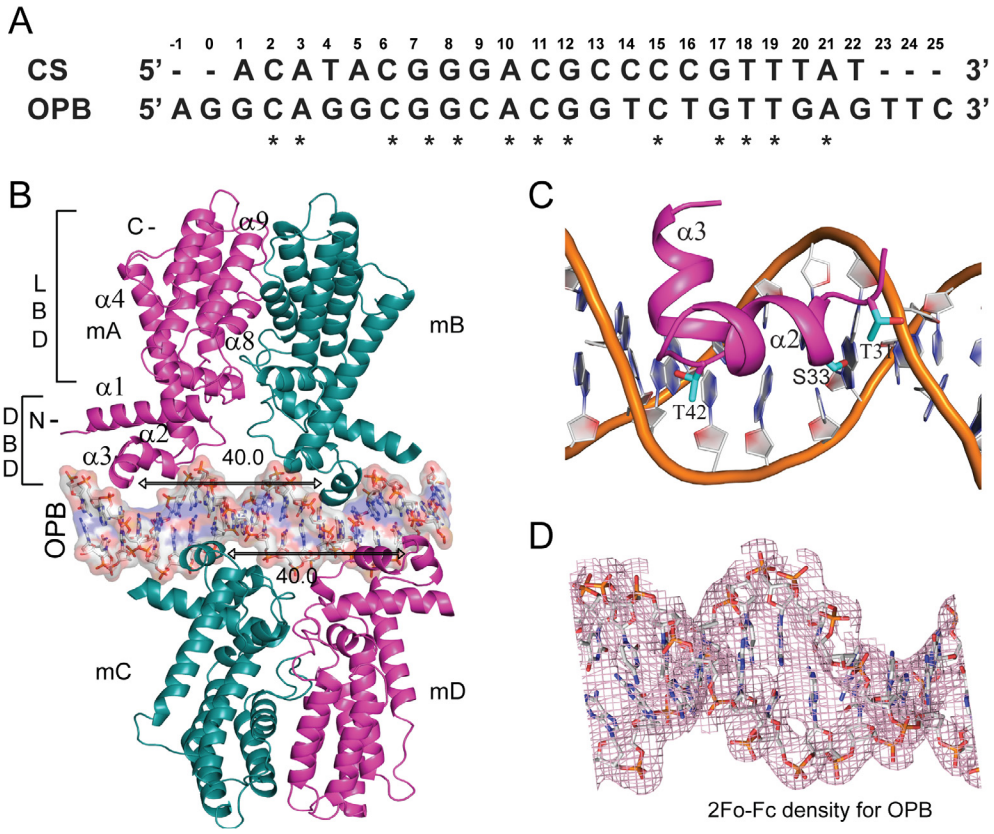

Fig. 1. Overall X-ray crystallographic structure of CprB-OPB complex. A) Alignment of CS and OPB DNA sequences and asterisk to show the identity between two sequences. B) Two dimers of CprB bound to a double stranded OPB sequence. CprB is presented in a cartoon representation and the OPB in stick model with the surface representation. All four monomers are labeled as mA, mB, mC and mD. The monomers mA and mB forms one dimer and mC and mD forms the other. The black arrow showing the distance between two recognition helices α3 of a dimer and the distance was measured to be 40.0 Å from the amide nitrogen atom of G44 from α3 of each monomer. The DNA binding domain and the ligand binding domain of the monomer are represented as DBD and LBD respectively. The N-terminal and the C-terminal of CprB are labeled as N- and C- respectively. C) L-shaped HTH-motif in the DBD of CprB and OPB DNA are given in cartoon representation. The residues (shown in sticks representation) T31, S33 and T42 anchor the phosphate backbone resulting in the best docking of α3 into the major-groove of OPB DNA. D) Electron density map (2Fo-Fc contoured at 1.1 σ) for OPB density, OPB DNA is presented in sticks representation.

3.2. Comparison of CS and OPB CprB complexes

Overall, the architecture of the CprB-OPB complex is similar to that of the CprB-CS complex; the induced fit mechanism is followed in both the complexes. To allow entry of the DBD of CprB, the "L-shaped" HTH-motif residues (ie. T31, L32, S33 from helix α2 and T42 in helix α3) interact with the phosphate backbone of the DNA and widen the major-groove (Fig. 1C). Helix α2 serves as a spacer and helps the recognition helix α3 to dock perfectly into the DNA major-groove (Fig. 1C and Fig. 1D shows the electron density for the DNA). Moreover, the side chain hydroxyl moiety of residues T31, S33 and the amide nitrogen atom from L32 at the start of the helix α2 forms hydrogen bonding interactions with the oxygen (O1P and/or O2P) atom of the phosphate backbone and result in deforming the DNA structure by inducing kinks (Fig. 2A, B) in it. Similar to CprB-CS complex, the residues K43, G44, Y47 and F48 from the helix α3 are involved in base specific contacts in the major-groove of the DNA. The aliphatic side chain of K43 and the aromatic ring of F48 partake in hydrophobic interactions with the hydrophobic core of DNA bases, while the charged amine head group of K43 is involved in hydrogen bonding interactions with them. Additionally, the side chain benzene moiety of Y47 forms T-shaped stacking interactions with DNA bases and the side chain hydroxyl group hydrogen bonds with the oxygen atoms of the phosphate group in the DNA backbone. A comparison of CS and OPB reveals that the presence of residue G44 seems to be very essential in this sub-family of TetR-FTRs. Even in the case of QacR, it was observed that for tight docking of the recognition helix α3, hydrogen bond between amide nitrogen atom of G44 and DNA bases was crucial. This tight docking of helix α3 expels water molecules from the major-groove favoring the entropy component in the binding event (Schumacher et al., 2002).

Although the global architecture of CprB-OPB verses CprB-CS complex seemed to be maintained, a detailed structural comparison disclosed that there are subtle differences both at the DNA sequence level as well as at the protein conformation level. To highlight the structural changes in the DNA binding region, the LBDs (residues 54–210) domains of both CprB-CS and CprB-OPB complexes were superimposed (Fig. 2C, D). The RMSD for the HTH-motif of mA in CprB-OPB complex is around 2.0 Å, where the maximum deviation is observed in helix α3 (2.2 Å) and in helix α2 it is 1.8 Å. For mC, the RMSD of the HTH-motif is around 1.7 Å (deviation in helix α3 is 1.9 Å and in helix α2 is 1.6 Å). In the case of the HTH-motifs in mB and mD, the deviation is observed to be 1.0 Å and 1.5 Å, respectively. Thus, this global analysis shows that the perturbation in the structure maximally occurs in subunits mA and mC. Both these subunits encounter bases that are different between CS and OPB DNA sequences, with the OPB sequence being GC rich in this region (5′-end). A detailed residue wise comparison shows that several specific interactions have readjusted. For example, it was observed that the interaction of residue Y47 (present in helix α3) with DNA, conserved in several TetR-FTR structures, adopts a variable conformation in CprB-CS and the CprB-OPB complex. It encounters different bases in both the DNA sequences and hence Y47 adapts its conformation to provide maximum stability to each complex. This scenario is further complicated by the non-palindromic nature of the OPB DNA sequence. Y47 from mC stacks with the complementary bases of dC9 and dA10 in the CprB-OPB complex, whereas in CprB-CS complex, it stacks with the complementary bases of dG9 and dA10. Here too, there is a change in the recognition base for Y47, where the pyrimidine base (complementary to dA10 in CprB-CS) was replaced with a purine base (complementary to dC9); this is shown in Fig. 3A and B. Similarly, in mA, Y47 stacks with the complementary base for dG7 in CprB-OPB complex (Fig. 3C) and in the CprB-CS complex, it stacks with the complementary base for dC6 and dG7 (Fig. 3D). Thus, both the interactions of Y47 within the OPB-complex itself are different between mA and mC and the overall proximity of Y47 to DNA is also altered. This interplay of differential interactions results in an overall lateral shift in α3 causing an increased splaying of the two dimers in OPB bound form. The center to center distance from each recognition helix α3, as measured from the amide nitrogen atoms of G44 was found to be 40.0 Å (Fig. 1B) for CprB-OPB complex and 38.2 Å in CprB-CS complex. This sideways shift of α3 results in altering the contact profile of the second most important interacting residue F48. In mA of CprB-CS complex F48 was seen to stack with dA3 at a distance of ~5.0 Å (Fig. 3E). However, the same stacking interaction is lost in CprB-OPB complex where the distance between F48 and dA3 was found to be ~8.5 Å, Fig. 3F. Similarly, due to alteration of sequence in mD, F48 stacks with dC2 in CprB-OPB complex and dA1 in CprB-CS complex. The interactions of F48 in mB and mC nevertheless, are not much disturbed. Another residue that shows variation is K43, in mB, mC and mD subunits it forms hydrogen bonding contacts with the bases dT16, complementary base to dC13 and dT20 respectively in CprB-OPB complex. However, K43 from mB, mC and mD subunits in CprB-CS complex forms hydrogen bonding contacts with complementary base to dC14, dC11 and dT18 respectively. Overall, in this DNA-binding event, both the hydrophobic region and the hydrophilic side chain of the K43, Y47 and F48 interact with various DNA bases via hydrogen bonding and hydrophilic interactions that differ in the two sequences.

3.3. DNA conformation analysis

Since OPB and CS sequences are not identical, to understand how the overall architecture of DNA is altered on tweaking the specific bases, the structural changes that occurred in CS and OPB were compared and contrasted in presence and absence of CprB. The DNA base-pair parameters for both CS and OPB were analyzed using Web 3DNA (Zheng et al., 2009). Analysis revealed that there is a symmetric appearance of A-T base pairs in the CS, whereas in the case of OPB it is asymmetric with OPB possessing an A-T rich base pairing at the 3′-end and a G-C base rich pairing at the 5′-end (Fig. 1A). It was observed that binding of protein induces kinks in both the CS (Fig. 2A) and OPB (Fig. 2B) sequences which resulted in positive roll values at bases positions like the bases dG12 (2.8°), dC16 (6.1°), and complementary bases to dC13 (7.7°) and dG9 (9.3°) respectively in CS. Similarly, the positive rolls at the bases dG13 (3.4°), dT16 (13.4°), complementary bases to dG13 (5.6°) and dC8 (3.6°) respectively in OPB DNA sequence. The average roll in a standard B-DNA is −2.8° and these positive rolls as observed in both CS and OPB DNA sequences leads to the global bend in the DNA. A unique feature found in proteins that bind as dimer of dimers to DNA is that the kinks occur in both the strands as dimers are interacting on either sides of the DNA. This leads to a compensation of the net global bend and results in minimal distortion of the DNA. A similar scenario is also observed in other TetR-FTRs like QacR-DNA complex, where the two dimers bind the QacR DNA and sandwich it on either side leading to a net global bend of only around 3.0° (Schumacher et al., 2002). However, in cases like TetR and SimR, which only exist as dimers with no compensating protein attacking on the other side, large positive roll values are observed leading to the global bend of 15–17° in the DNA.

In order to further understand the implication of architectural differences, the analysis of the DNA helical parameters like the width of major- and minor-grooves were also performed. CS DNA follows a symmetrical pattern in the minor-groove width, whereas for OPB DNA this pattern is disrupted (Fig. S2A). The minor-groove

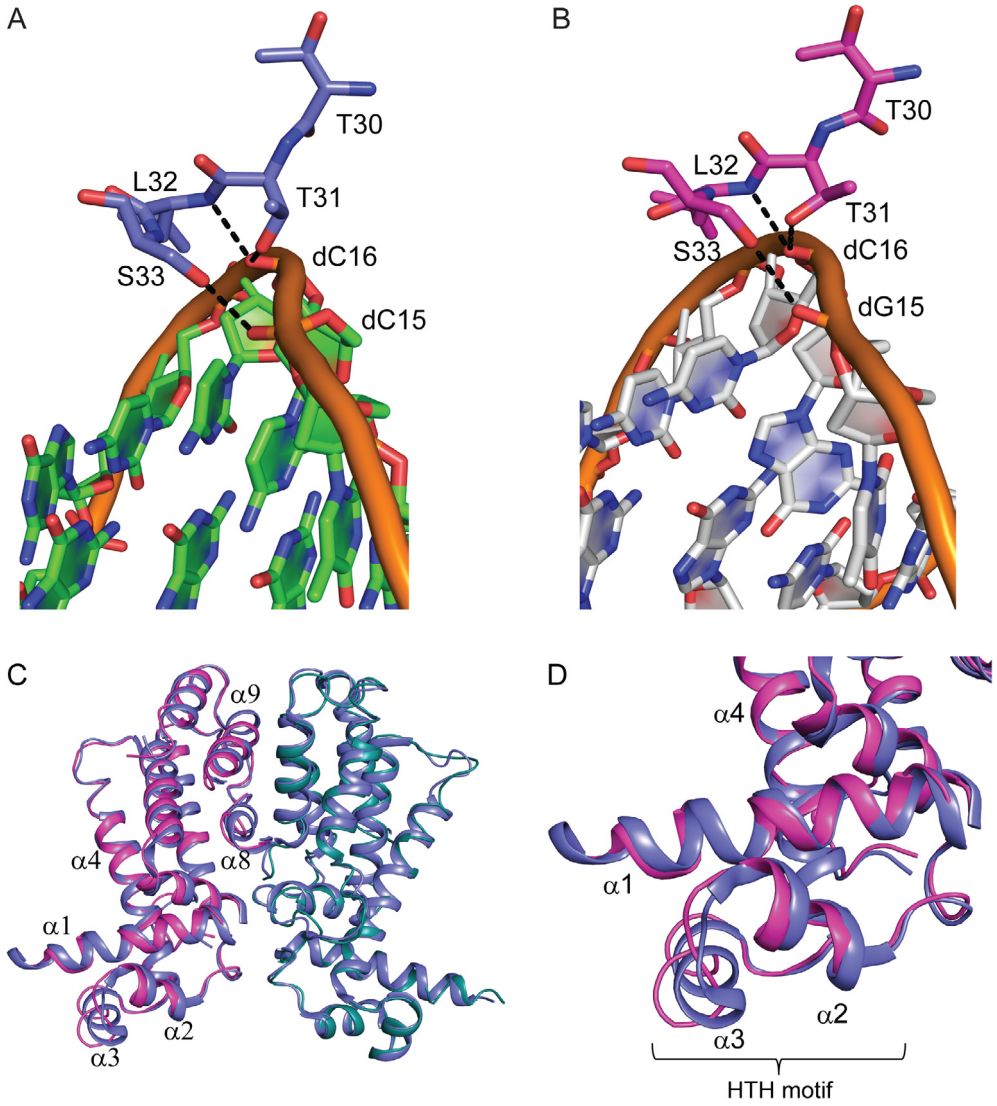

Fig. 2. Comparison of CprB-OPB and CprB-CS complex structures. CprB induced kink in the CS (A) and OPB (B) DNA sequences. The interacting residues are shown in sticks model and the DNA in cartoon representation. The black dotted lines are showing the hydrogen bonding interactions of CprB and DNA backbone. C) Superposition of LBDs (residues 54–210) of mA from both the complexes. CprB represented in cartoon, magenta and green are two monomers of homodimer bound to OPB and blue dimer bound to CS. D) Highlighting the changes in the DBD of CprB when it is bound to OPB and CS. The helices α1 – α4 are labeled and the HTH-motif indicated.

width for CS towards both the 5′ and 3′-ends is around 9.5–12.0 Å (Fig. S2A) and in the middle, it is around 13.5 Å. In the case of OPB sequence, the width of minor-groove towards 5′-end is around 10.5–13.5 Å and towards the 3′-end, it is around 12.5–13.5 Å. The minor-groove width in the middle for OPB is around 13.5 Å, whereas for B-form of OPB sequence, the minor-groove width is uniformly 11.5 Å throughout the sequence. This suggests in both the cases the minor-groove width is widened when compared to the B-form of DNA, however the effect is more prominent for OPB sequence in the 5′-end and major differences in protein conformation are also observed here. Similar trend was also observed in the major-groove width of CS and OPB sequences, Fig. S2B. The widening of major-groove for OPB sequence towards 5′-end is complemented by decreased minor-groove width and the decreased major-groove width is complemented with an increased minor-groove size (Fig. S2A and B). These observations strengthen our belief that asymmetric G-C content as found in the OPB sequence has a marked affect on DNA architecture and structural asymmetry is enhanced when it encounters protein.

3.4. Insights with MD data

Fully atomistic MD simulations were exploited in order to enhance the information content of the lower resolution X-ray data of the CprB-OPB complex, and to gain insights into the molecular origins of the experimentally observed differences between the CprB-CS and CprB-OPB complexes. The MD protocol was first validated by comparing the simulated conformations of the CprB-CS complex with the corresponding features in the high resolution X-ray crystallographic data reported previously (Bhukya et al., 2014). Fig. S3 shows the RMSD of protein and DNA backbone atoms of the CprB-CS complex obtained over the simulation trajectories. The traces of RMSD of all four monomeric subunits, mA, mB, mC and mD, as well as that of the DNA backbone, are shown individually. It is observed form the evolution of the RMSDs that the CprB-CS complexes attain equilibration within about 50 ns. The mean RMSD value of the backbone atoms of the subunits mA, mB, mC and mD over the last 20 ns are 2.9 (±0.1) Å, 3.6 (±0.1) Å, 2.8 (±0.2) Å and 2.9 (±0.1) Å, respectively. We further calculated the

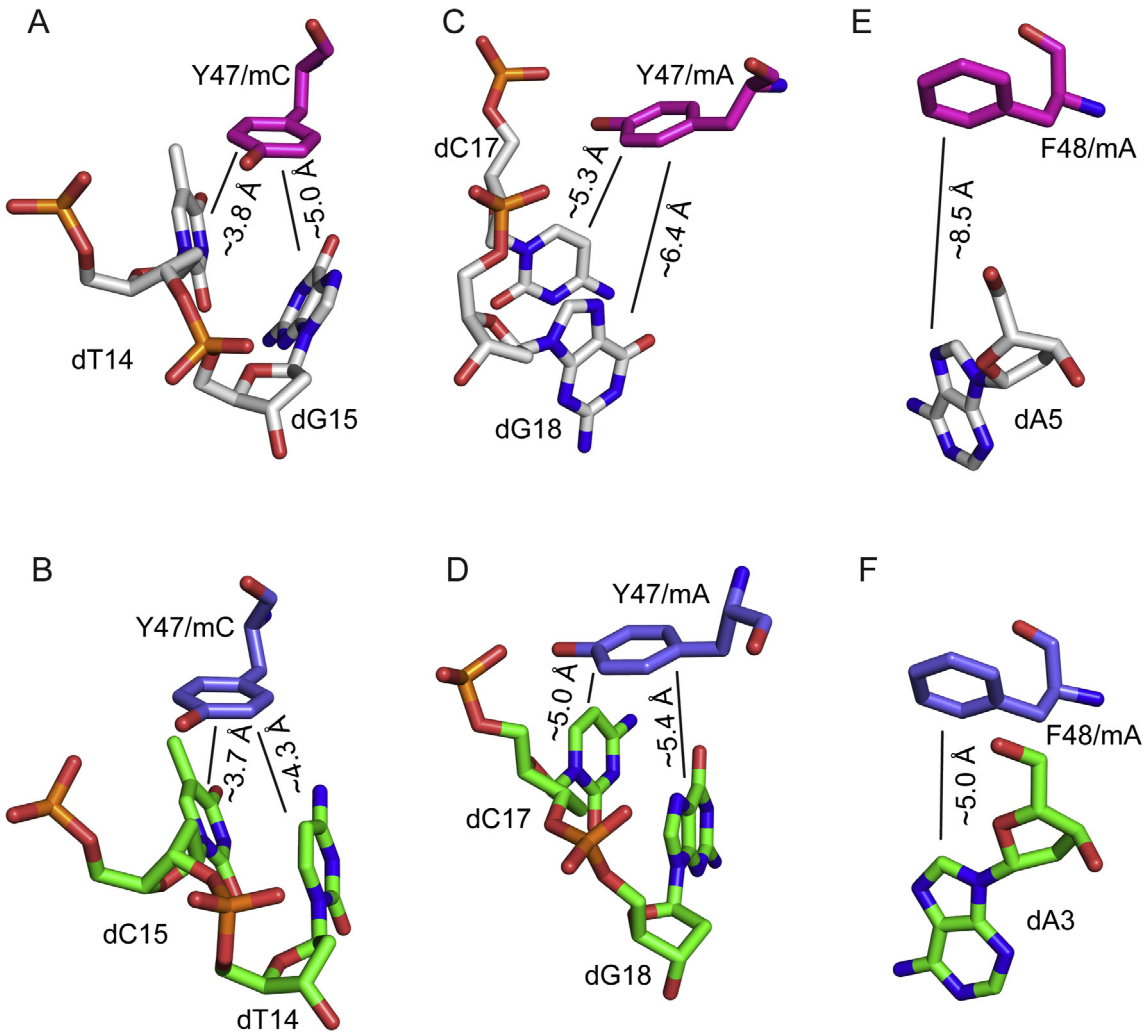

Fig. 3. Comparison of the CprB-DNA interactions in CprB-OPB and CrpB-CS complex. The protein-DNA interactions (A, C and E) shown in the top panel are for CprB-OPB complex and in the bottom panel (B, D and F) are for CprB-CS complex. The solid black line showing the distance in Å. The monomers A, B, C and D are labeled as mA, mB, mC and mD respectively.

residue-wise backbone mean squared fluctuation (MSF) of the protein units over the early (0–30 ns) and latter (80–140 ns) segments of the trajectories; these data are shown in Fig. S4. The MSF value for the whole protein except for the disordered N-terminal region (residues 1–7) and C-terminal region (residues 210–215) in the CprB-CS complex is significantly low at both time regimes. The large spike in the MSF (3–4.5 Å) as seen in the Fig. S4 for the residues (75–79, 112–118, 163–175) in the initial 30 ns simulation, and 1.5–2.0 Å in the latter segment, are due the missing residues in the X-ray structure that were added to the system prior the simulations. Similarly, the disordered N- and C-terminal residues were also added, and hence the MSF are initially high during the course of simulation; later, the MSF drops to around 6.0 Å for the N-terminal residues. This drop is because the N-terminal finger region finds a stable conformation that interacts with the phosphate backbone of the DNA. Two-dimensional (2D) RMSD maps were further calculated for the individual subunits and averaged over the independent trajectories; these data are shown in Fig. S5. It is evident from the data that there are no significant deviations in the conformations of the CprB-CS complex during the course of simulation relative to the initial conformation. This minimal change in the conformation can be attributed to the coordinates of the starting structure that were obtained from the high resolution X-ray crystal structure.

A detailed contact map describing the intermolecular interaction between CprB and CS are given in a schematic representation in Fig. S6. This data can be directly compared to the experimentally obtained contact map published previously. A comparison between the experimental and simulated maps indicates that a large majority of the protein-DNA contacts are maintained during the course of the simulation. For example, the important stacking interactions of Y47 and F48 with the respective bases are maintained, and the major groove widening residues as well as the kink inducing conformations are preserved. However, the base stacking interaction of F48 with complementary base to dG20 and dG7, and DNA backbone interactions of H49 and K53 in mC and mD are missing. Interestingly, the N-terminal region of CprB, which could not be seen in the crystal structure, showed additional interactions in mB and mC subunits, with the N-terminal residues interacting with the minor-groove of the CS DNA. Particularly, residue M1, R3 and R6 in mA and mD appear to be involved in forming hydrogen bonding and hydrophobic interactions with DNA backbone. The overall evidence shows that the X-ray structure and theoretical simulation results are in good agreement, and therefore, the all-atom CHARMM force field is a good choice for studying CprB-DNA complexes.

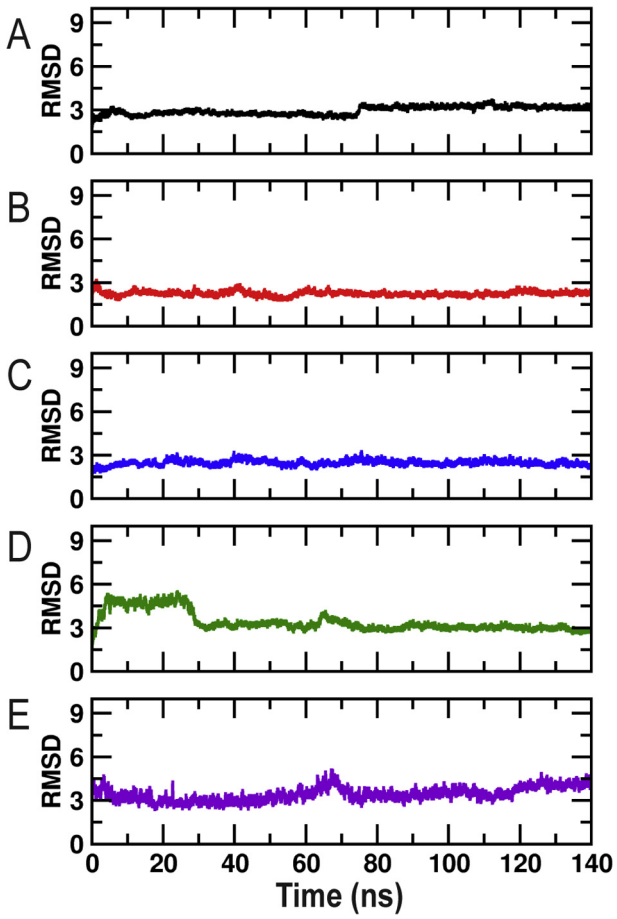

Fig. 4. The backbone root mean squared deviation (RMSD, in Å) of all four monomers and DNA as a function of simulation time for CprB-OPB complex with respect to initial structure. RMSD of all four monomers mA, mB, mC and mD and DNA are shown in A, B, C, D and E respectively.

3.5. MD studies of CprB-OPB complex

The low resolution crystal structure of the CprB-OPB complex was used as a starting model for the simulation. The aim was to use a combination of experiment and theoretical analysis to develop a high confidence interaction map of the protein with the biologically relevant DNA sequence. Three trajectories of the CprB-OPB complex were simulated independently, and the results analyzed as previously described. The mean RMSD value of the backbone atoms of the four monomers, mA, mB, mC and mD over the last 20 ns are 3.2 (±0.1) Å, 2.3 (±0.1) Å, 2.4 (±0.1) Å and 2.9 (±0.1) Å, respectively. The RMSDs for all four monomers of protein and DNA backbone were plotted and given in Fig. 4.

Similar to the CprB-CS complex, the reduction in the magnitudes of the residue-wise MSF for the protein backbone, from the initial 30 ns to the final 80 ns indicates stabilization of the trajectories over the course of the simulation. Again, large spikes in the MSF (2.0–4.0 Å) due to the added residues that are disordered in the X-ray structure as seen in the Fig. 5 for the residues (75–79, 112–118, 163–175) in CprB-OPB complex at the initial 30 ns simulation reduce to 1.5–2.0 Å during the later simulation period. The MSF for the disordered N- and C-terminal residues is dropped down 6.0 Å. The drop in the MSF is attributed to the interaction of N-terminal finger region with the phosphate backbone of the DNA. A comparative account of specific interactions which are mostly in accord with the obtained low resolution structure are provided in Supplementary section.

The 2D-RMSDs over simulation time for the CprB-OPB complex is shown in Fig. 6, and describe interesting differences with the CprB-CS complex. Unlike the latter, the individual monomeric subunits in CprB-OPB show non-identical correlations with time evolution. Monomers A and D show markedly higher RMSD at later times over the starting structure compared to the monomers B and C. While this asymmetry in the subunit behavior may be related to the lower resolution of the initial crystallographic data, it could also arise due to a more complex, step-wise association of the sequence with the subunits. Such an association appears to be suggested in earlier observations, where the ITC profile of

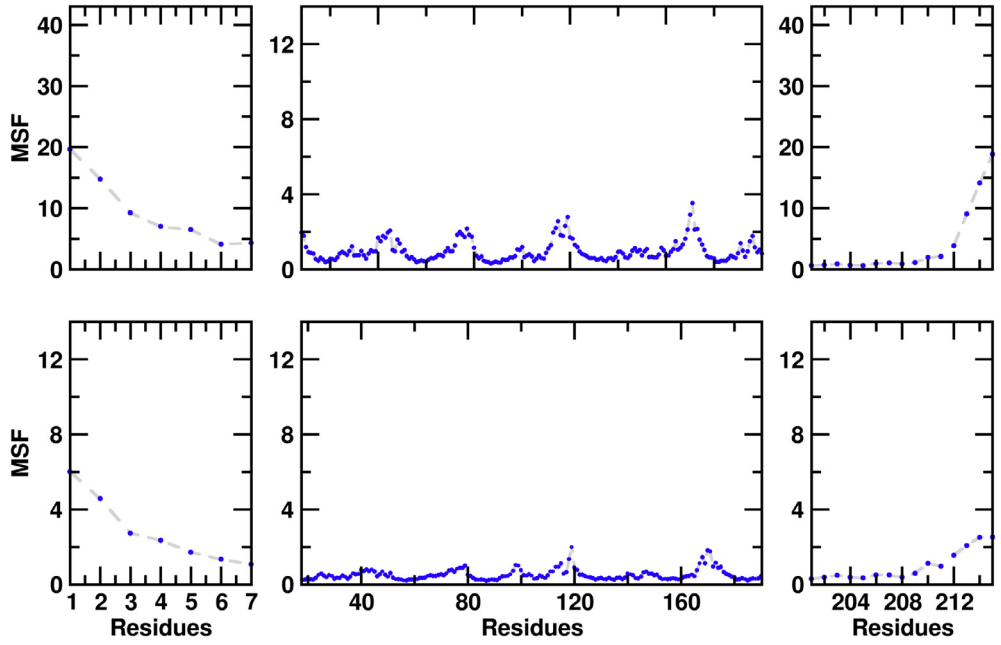

Fig. 5. Comparison of the Cα mean square fluctuation (MSF, in Å^2) of the CprB-OPB complex (blue). The data for initial 30 ns are shown in upper column and the data for final 50 ns are shown in lower column.

the OPB-CprB binding was found to be markedly different from that of the CprB-CS binding, with the heat profile of the former only fitting to a two-step binding model (Bhukya et al., 2014).

It is to be noted here that the OPB sequence is longer and consists of 27 bases as opposed to 22 bases in the CS sequence. Therefore, OPB is able to utilize the extra region to make stabilizing contacts with the N-terminal arm of CprB. The shorter length of CS sequence results in only two subunits were partial contacts of the N-terminal region of CprB protein could be seen in MD simulations. Whereas, simulations with the CprB-OPB complex shows all four monomers of CprB interacting with the DNA. It was noticed that these ordered N-terminal residues primarily network with the minor-groove of the OPB DNA. (Fig. 7A–E). The interactions were primarily of the hydrophobic side chain of the first methionine residue with the sugar moiety of the DNA bases, and of the phosphate backbone with the arginine residues R3 and R6 in all the subunits. Thus, the extended, N-terminal positively charged fingers appear to play a role in stabilizing the protein-DNA interactions. Other residue such as Q4, where the nitrogen atom (NE2) also makes hydrogen bonding interactions with the O2 oxygen atom of the complementary base to dG20 also strengthens this interface. In CprB-CS complex also vestige of these interactions begins to be visible. As mentioned earlier, N-terminal residues Q4, L5 and T10 from mB appear to form hydrogen bonds with the DNA backbone. However, due to the shorter size of CS the contacts are not completely manifested. For example, variations occur in the binding pattern, the residue F48 is sandwiched between the aliphatic side chains of residue R3 and R6 of mA in the CprB-CS complex, whereas in the CprB-OPB complex, the residue R3 is involved in hydrogen bonding interaction with phosphate backbone. The analysis however, reasserts that N-terminal region of CprB is indeed involved in stabilization.

In Fig. S7, we have compared distributions of the protein-DNA contact area (Fig. S7A), as well as the configuration entropy per heavy atom (Fig. S7B) calculated from the trajectories of the CprB-CS and CprB-OPB complexes. Additionally, evidence that the N-terminal region of the OPB complex is heavily involved in stabilization of the complex, comes from the calculation of contact area which, is in the range of 1200–1400 Å^2. In comparison, in the CprB-CS complex, as mentioned earlier, the N-terminal fingers are unable to effectively latch onto the DNA ends, leading to a contact area in the range of 800–1000 Å^2. We further find that there is a small but distinct lowering of the configurational entropy from 25.1(±0.2) JK^{-1} mol^{-1} in the CprB-CS complex to 24.5(±0.2) JK^{-1} mol^{-1} in the CprB-OPB complex, indicating a relative tightening of the latter complex. In both cases, however, the simulation results indicate that the protein-DNA complex is conformationally stable, as a large proportion of the interactions is *via* the phosphate backbone are hydrophobic in nature, and therefore, generic changes in the sequence do not affect the overall architecture.

3.6. EMSA and fluorescence anisotropy studies with native and ΔN6-CprB

The MD studies performed on CprB-OPB complex revealed that the N-terminal residues are crucial and interact with the minor-

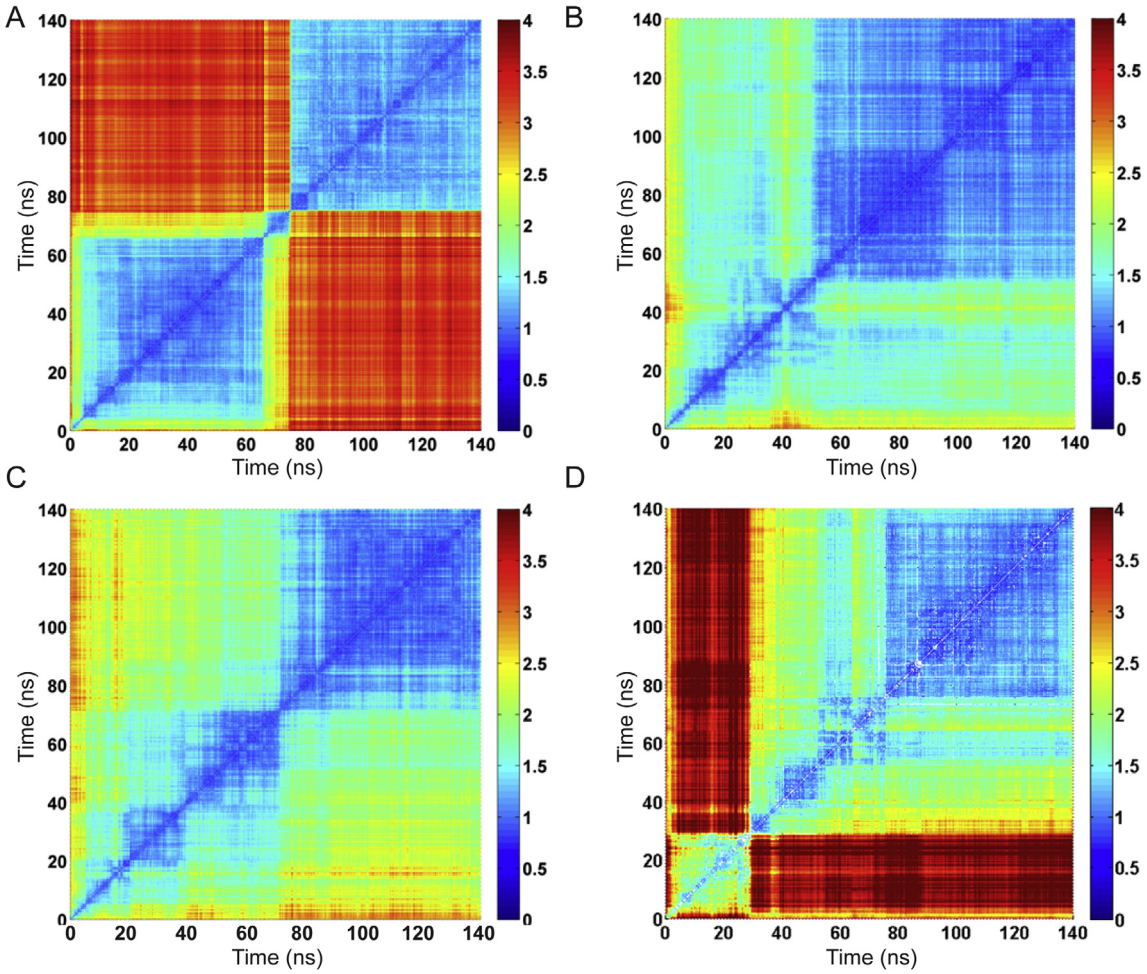

Fig. 6. 2D-RMSD maps of CprB-OPB complex during 140 ns of MD simulation. All four monomers mA, mB, mC and mD are shown in A, B, C, and D respectively.

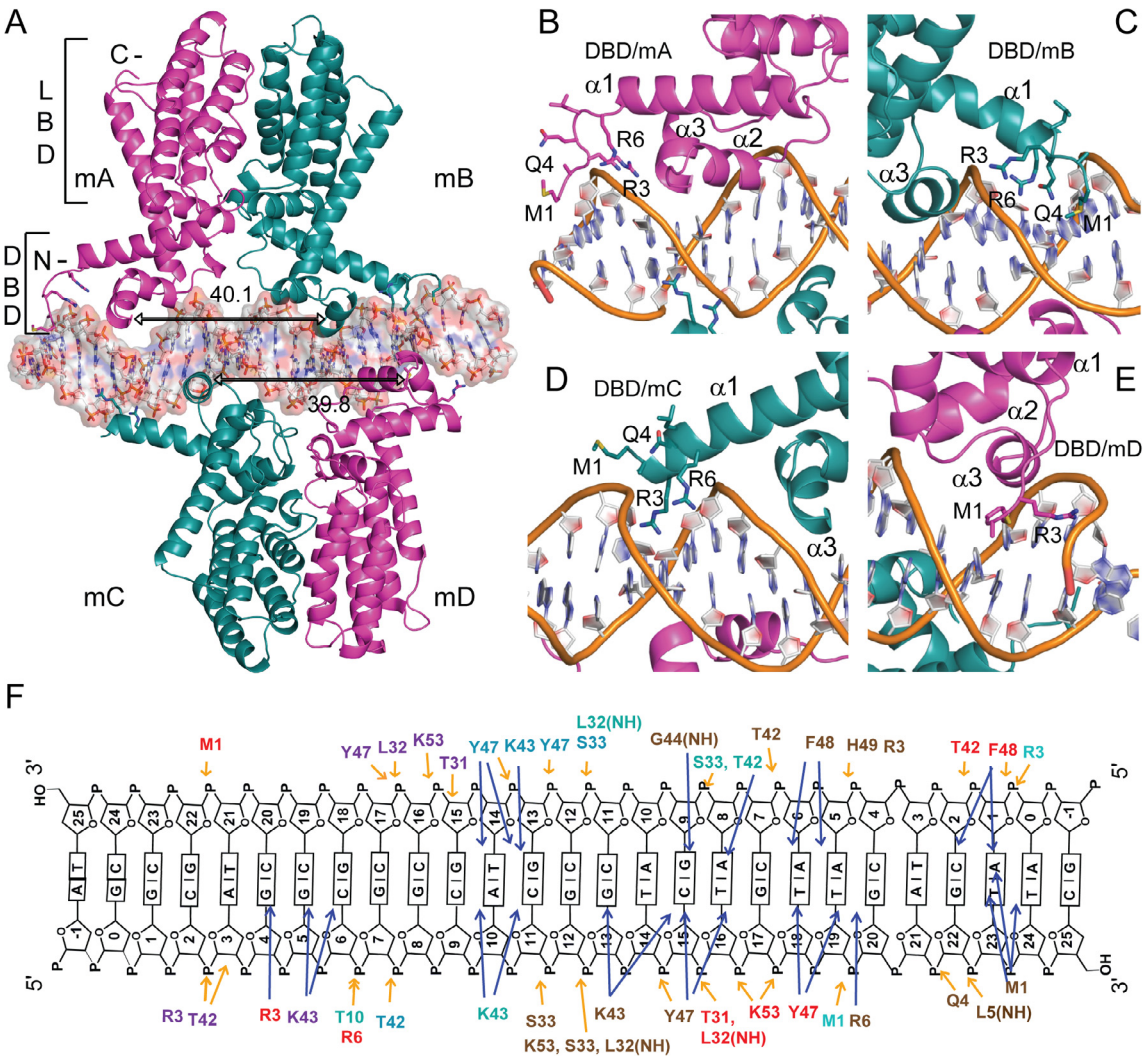

Fig. 7. Snapshot of overall structure of CprB-OPB complex from the molecular dynamics studies and also highlighting the interactions of N-terminal arginine rich region of CprB with the phosphate backbone and minor-groove of OPB DNA. A) Two dimers of CprB bound to a double stranded OPB sequence. CprB is presented in a cartoon representation and the OPB in stick model with the surface representation. All four monomers are labeled as mA, mB, mC and mD. The black arrow showing the distance between two recognition helix α3 of a dimer and the distance was measured to be 40.1 Å from amide nitrogen atom of G44 from helices α3 of each monomer for mA and mB whereas for mC and mD it is 39.8 Å. The DNA binding domain and the ligand binding domain of the monomer are represented as DBD and LBD respectively. The N-terminal and the C-terminal of CprB are labeled as N- and C- respectively. B) The interaction of N-terminal residues M1, R3 and R6 from DBD of monomer A (DBD/mA) of CprB with OPB DNA sequence (shown in cartoon representation). Similarly, C), D) and E) are showing the interactions of N-terminal residues from monomer B (DBD/mB), monomer C (DBD/mC) and monomer D (DBD/mD) respectively with the OPB DNA sequence. F) Schematic representation of the OPB DNA sequence highlighting the CprB-OPB interactions. Residues coloured in purple are from mA, brown are from mB, cyan are from mC and red are from mD. Note that P = PO_3^{2-}. The blue and orange coloured arrows are indicating the base and backbone specific interactions of CprB respectively with OPB. The helices α1, α2 and α3 are labeled and the residues M1, R3, Q4 and R6 interacting with OPB are shown in sticks representation.

groove and the phosphate backbone of OPB sequence. Hints of these interactions also appeared in the CprB-CS simulation. To experimentally confirm these findings ΔN6-CprB deletion mutant that lacks the N-terminal finger was designed, cloned and purified using standard procedures. The ΔN6-CprB protein was subsequently subjected to EMSA using OPB as the target sequence. It was observed that the binding affinity was substantially reduced from 100–200 μM in native CprB-OPB to around 0.8–1.2 μM in ΔN6-CprB-OPB which is about 7–8 folds lower. The native CprB-OPB shown in Fig. 8A and ΔN6-CprB-OPB is shown in Fig. 8B.

Furthermore, to get an analytical value for these observations the binding affinities of native and ΔN6-CprB with the OPB DNA sequence were also pursued by performing steady-state anisotropy experiments using 5′-FAM (fluorescein) labeled DNA sequence. From the fluorescence anisotropy experiments (Fig. 8C), it is evident that the trend of reduced target binding ability of ΔN6-CprB toward the OPB is maintained both in the EMSA as well as in the anisotropy studies. Moreover, these studies also highlight the cooperative nature of DNA binding as the data can be fit to a Hill 1 equation with the Hill coefficient ranging from 1.5–2.0, hinting towards positive cooperativity between the two dimers. Overall, this result corroborates the aspects of differential binding suggested by the simulations.

4. Discussion

The structure of the CprB protein in complex with its operator sequence helps to elucidate various features of the dimer of dimers sub-class of TetR-FTRs. It is clear from the structure and the comparative analysis with other members that dimer of dimers only induce mild distortions of only 3–4° to the B-DNA. However, they

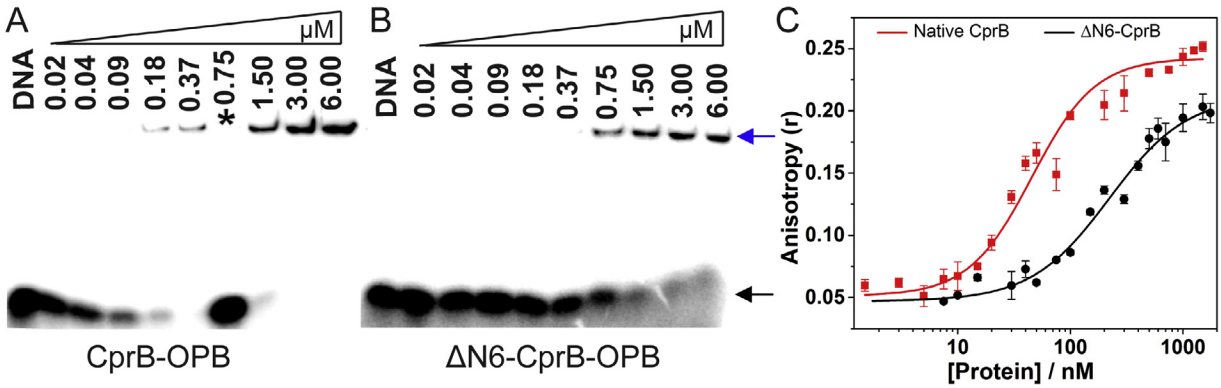

Fig. 8. Results of the EMSA and fluorescence anisotropy studies for native and N-terminal truncated CprB with OPB DNA sequence. A) Native CprB-OPB complex formation, B) N-terminal truncated CprB-OPB. Concentrations mentioned are in micromolar and the lane with asterisk has excess cold OPB DNA. The free DNA and complex are indicated. C) Binding affinities of the native and ΔN6-CprB with 5′-FAM labeled OPB DNA sequence using steady state anisotropy. The red squares and black circles and are the data points for the native and ΔN6-CprB respectively.

are still able to maintain high affinity as they cover a rather extended DNA interface by locking on to it from either direction (Bhukya et al., 2014; Itou et al., 2010; Schumacher et al., 2002). This is because both the strands of the DNA participate equally towards maintaining contact with the protein. Nonetheless, to effectively execute this scenario, the center of each dimer is shifted by around 7–10 Å. The effect of the shift is that both the dimers encounter different bases, resulting in heterogeneity in the binding profile of the two dimers. The scenario is further complicated in the binding of the CprB as the sequence is non-palindromic therefore, the diversity of bases encountered by the two dimers is further enhanced. The heterogeneity induced, determines the order of binding of each monomer in the dimer and is likely to be the factor responsible for cooperative binding effect observed in most of the TetR-FTRs within this sub-class. Based on ITC data and structural information, Brennan and co-workers have concluded that the mechanism of sequential binding in QacR is facilitated by the monomer of the first dimer unwinding the DNA to increase the major-groove width, thereby priming the monomer of the second dimer to fit easily into the other binding site of the DNA, thus inducing cooperative behavior (Schumacher et al., 2002). An analogous mechanism seems to be operating in CprB. A similar ITC profile is observed in case of the OPB sequence and the structural information also points to the fact that each monomer responds specifically to its environment. Moreover, the simulation analysis suggests a pattern where one monomer in each dimer is more mobile than the other, indicating differential binding of each unit. Further, as the DNA base interacting residues in CprB (K43, Y47 and F48) and in QacR (Y40, Y41 and K43) are similar, we believe that like QacR, CprB also follows a click and clamp model of DNA recognition.

The question as to how the same protein like CprB can recognize different sequences CS and OPB is an interesting one. There are several examples of such binding, and it is generally believed that a large percentage of the contacts, either *via* backbone or hydrophobic interactions, are preserved if the overall architecture of the DNA is not dramatically distorted. For example, the regulator of restriction modification enzyme, Esp13961 absorbs the variation in the DNA sequence by slightly altering the conformation adopted by its HTH-motif (Ball et al., 2012). This is similar to the scenario observed in CprB where the HTH-motif changes its conformation to adapt for the sequence changes in CS and OPB DNA sequences. The structural plasticity shown by CprB allows the readjustment of its conformation, thereby allowing it to adapt to the asymmetric nature of the OPB sequence.

One of the most crucial insights obtained in this study pertains to the involvement of the N-terminal residues of CprB that anchor the DNA, thereby consolidating the stability of the CprB-OPB complex. A similar scenario was observed in case of other TetR-FTRs where the extended N-terminal finger region that is enriched with the positively charged residues forms hydrogen bonding as well the salt-bridge interactions with the phosphate backbone and the minor-groove of their respective operator sequences (Le et al., 2011) (Fig. S8). For example, SimR (PDB entry: 3ZQL) contains 28 amino acids extension in its N-terminal arm, and the positively charged residues R18, R19, R22 and R25 are involved in binding to the DNA minor-groove in the adjacent complex (see Fig. 9A and B). Deletion of this extended N-terminal arm has been reported to result in reduced affinity by 120 folds toward its operator sequence (Le et al., 2011). Similarity, the crystal structures of a global regulator Ms6564 (Fig. 9C and D, PDB entry: 4JL3) and the global nitrogen uptake regulator AmtR (Fig. 9E and F, PDB entry: 5DY0) from *Corynebacterium glutamicum* also show that the arginine rich N-terminal region interacts *via* hydrogen bonding and hydrophobic interactions with the phosphate backbone and fits into the DNA minor-groove, thereby enhancing the stability of their respective complexes (Palanca and Rubio, 2016). There are other examples of TetR-FTRs like DesT from *Pseudomonas aeruginosa* (Itou et al., 2010) and CgmR from *C. glutamicum* that also possess N-terminal extensions that are involved in this kind of stabilization (Miller et al., 2010). Interestingly, in the case of the closest structural homology of CprB, QacR, the positive dipole at helix α1 interacts with the phosphate backbone and strengthens the protein-DNA complex (Schumacher et al., 2002).

From the observations, it appears these N-terminal extensions serve the role of anchors. Analysis *via* the RONN server indicates that in most cases the N-terminal extensions are conformationally malleable and only adopt a stable structure upon binding with the appropriate length of DNA. It is possible that the N-terminal extensions also help in attaining selectivity and in guiding the proteins like CprB to the biological OPB sequence. This region could also be responsible for variable affinity that the CprB possess for different sequences. One added advantage of the N-terminal binding is that it results in expansion of the overall volume around CprB, generously exceeding the size of a CprB dimer. Therefore, the presence of this N-terminal minor-groove interacting structure results in an overall enhancement in contact surface area by 30%.

In conclusion, the structure of CprB-OPB complex was able to highlight the nuances in interactions along the protein-DNA interface that determine aspects of its function. The study sheds light on how CprB is able to adjust its interactions and global conformation to accommodate different DNA sequences. Furthermore, it brings out the importance of the flexible N-terminal tail region that anchors into the minor-groove of the DNA, thereby enhancing

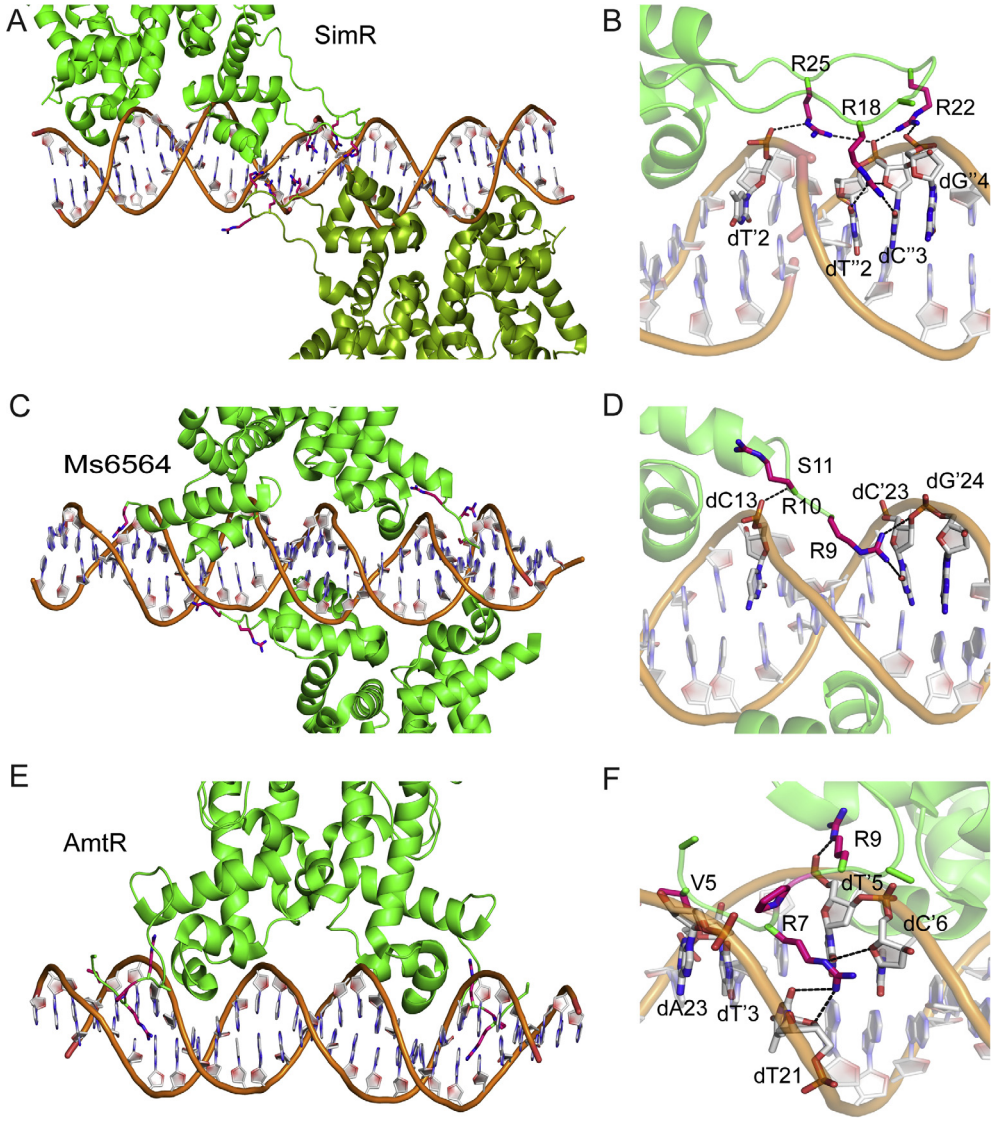

Fig. 9. Structures of protein-DNA complex from TetR-FTRs where the N-terminus of the protein interacts with the minor-groove and the phosphate backbone of their respective operator sequences. The protein-DNA complexes are representation in cartoon model and the N-terminal residues interacting with DNA are shown in sticks representation and coloured in pink. The protein-DNA complexes of SimR, Ms6564 and AmtR are shown in A, C and E respectively. The N-terminal residues interacting with DNA are zoomed as shown in B, D and F for the protein-DNA complexes of SimR, Ms6564 and AmtR respectively. The hydrogen bonding interactions are shown with black dotted lines. The DNA bases in the complementary strand are denoted with a prime (′) and the bases from the DNA of adjacent SimR-DNA complex are denoted with double prime (″).

the stability of the protein-DNA complex. A question that still remains unaddressed is regarding the inherent allostery in the TetR-FTR system. What is the sequence of events and conformational trajectory that results in breaking of this extended DNA binding interface upon associating with ligand? These problems are under active investigations in our laboratories.

Funding

This work was financially supported by IIT, Mumbai, India and Board of Research in Nuclear Sciences (BRNS), India, Grant Nos. 20150237B02RP00614-BRNS.

Acknowledgements

We thank the Ryo Natsume, Japan Biological Informatics Consortium, Tokyo, Japan for gifting us the plasmid containing *cprB* gene. Dr. Pradeepkumar, P. I., for providing access to their laboratory facilities. Prof. K. V. R. Chary and Ms. Geetanjali A. Dhotre for the help in characterization of oligonucleotides MALDI facility at TIFR. Dr. Jessy M. John for helping in the fluorescence anisotropy experiments. BM14-DBT initiative for X-ray crystallographic data collection at European synchrotron radiation facility (ESRF) France. H. B. thanks IITB-Monash Research Academy for the fellowship. N.S and A.K.J acknowledge the CSIR 12th Five Year Plan 'Multi-Scale Simulation and Modeling' project (MSM; grant number CSC0129) for the computational resources.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2017.03.006.

References

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., McCoy, A.J., Moriarty, N.W.,

Oeffner, Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., . PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D 66, 213–221.
Ahn, S.K., Cuthbertson, L., Nodwell, J.R., 2012. Genome context as a predictive tool for identifying regulatory targets of the TetR family transcriptional regulators. PLoS One 7, e50562.
Ball, N.J., McGeehan, J.E., Streeter, S.D., Thresh, S.J., Kneale, G.G., 2012. The structural basis of differential DNA sequence recognition by restriction–modification controller proteins. Nucleic Acids Res. 40, 10532–10542.
Bassler, B.L., Losick, R., 2006. Bacterially speaking. Cell 125, 237–246.
Bertrand, K.P., Postle, K., Wray Jr, L.V., Reznikoff, W.S., 1983. Overlapping divergent promoters control expression of Tn10 tetracycline resistance. Gene 23, 149–156.
Bhukya, H., Bhujbalrao, R., Bitra, A., Anand, R., 2014. Structural and functional basis of transcriptional regulation by TetR family protein CprB from *S. coelicolor A3(2)*. Nucleic Acids Res. 42, 10122–10133.
Bolla, J.R., Do, S.V., Long, F., Dai, L., Su, C.-C., Lei, H.-T., Chen, X., Gerkey, J.E., Murphy, D.C., Rajashankar, K.R., Zhang, Q., Yu, E.W., 2012. Structural and functional analysis of the transcriptional regulator Rv3066 of *Mycobacterium tuberculosis*. Nucleic Acids Res. 40, 9340–9355.
Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L., 1998. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. 54, 905–921.
Cuthbertson, L., Nodwell, J.R., 2013. The TetR family of regulators. Microbiol. Mol. Biol. Rev. 77, 440–475.
DeLano, W.L., 2002. The PyMol User's Manual.
Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501.
Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L.G., 1995. A smooth particle mesh Ewald method. J. Chem. Phys. 103, 8577–8593.
Feller, S.E., Zhang, Y., Pastor, R.W., Brooks, B.R., 1995. Constant pressure molecular dynamics simulation: the Langevin piston method. J. Chem. Phys. 103, 4613–4621.
Grkovic, S., Brown, M.H., Roberts, N.J., Paulsen, I.T., Skurray, R.A., 1998. QacR is a repressor protein that regulates expression of the *Staphylococcus aureus* multidrug efflux pump QacA. J. Biol. Chem. 273, 18665–18673.
Grkovic, S., Brown, M.H., Schumacher, M.A., Brennan, R.G., Skurray, R.A., 2001. The *Staphylococcal* QacR multidrug regulator binds a correctly spaced operator as a pair of dimers. J. Bacteriol. 183, 7102–7109.
Hinrichs, W., Kisker, C., Duvel, M., Muller, A., Tovar, K., Hillen, W., Saenger, W., 1994. Structure of the tet repressor-tetracycline complex and regulation of antibiotic resistance. Science 264, 418–420.
Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14, 33–38.
Itou, H., Watanabe, N., Yao, M., Shirakihara, Y., Tanaka, I., 2010. Crystal structures of the multidrug binding repressor *Corynebacterium glutamicum* CgmR in complex with inducers and with an operator. J. Mol. Biol. 403, 174–184.
Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L., 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926–935.
Kabsch, W., 1988. Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. J. Appl. Cryst. 21, 916–924.
Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., Schulten, K., 1999. NAMD2: greater scalability for parallel molecular dynamics. J. Comput. Phys. 151, 283–312.
Le, T.B.K., Schumacher, M.A., Lawson, D.M., Brennan, R.G., Buttner, M.J., 2011. The crystal structure of the TetR family transcriptional repressor SimR bound to DNA and the role of a flexible N-terminal extension in minor groove binding. Nucleic Acids Res. 39, 9433–9447.
Mackerell, A.D., Feig, M., Brooks, C.L., 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J. Comput. Chem. 25, 1400–1415.
MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., Karplus, M., 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B 102, 3586–3616.
Martyna, G.J., Tobias, D.J., Klein, M.L., 1994. Constant pressure molecular dynamics algorithms. J. Chem. Phys. 101, 4177–4189.
Miller, D.J., Zhang, Y.-M., Subramanian, C., Rock, C.O., White, S.W., 2010. Structural basis for the transcriptional regulation of membrane lipid homeostasis. Nat. Struct. Mol. Biol. 17, 971–975.
Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A., 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. 67, 355–367.
Natsume, R., Ohnishi, Y., Senda, T., Horinouchi, S., 2004. Crystal structure of a γ-butyrolactone autoregulator receptor protein in *Streptomyces coelicolor* A3(2). J. Mol. Biol. 336, 409–419.
Natsume, R., Takeshita, R., Sugiyama, M., Ohnishi, Y., Senda, T., Horinouchi, S., 2003. Crystallization of CprB, an autoregulator-receptor protein from *Streptomyces coelicolor* A3(2). Acta Crystallogr. D Biol. Crystallogr. 59, 2313–2315.
Onaka, H., Horinouchi, S., 1997. DNA-binding activity of the A-factor receptor protein and its recognition DNA sequences. Mol. Microbiol. 24, 991–1000.
Onaka, H., Nakagawa, T., Horinouchi, S., 1998. Involvement of two A-factor receptor homologues in *Streptomyces coelicolor* A3(2) in the regulation of secondary metabolism and morphogenesis. Mol. Microbiol. 28, 743–753.
Onaka, H., Ando, N., Nihira, T., Yamada, Y., Beppu, T., Horinouchi, S., 1995. Cloning and characterization of the A-factor receptor gene from *Streptomyces griseus*. J. Bacteriol. 177, 6083–6092.
Orth, P., Schnappinger, D., Hillen, W., Saenger, W., Hinrichs, W., 2000. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. Nat. Struct. Mol. Biol. 7, 215–219.
Palanca, C., Rubio, V., 2016. Structure of AmtR, the global nitrogen regulator of *Corynebacterium glutamicum*, in free and DNA-bound forms. FEBS J., 1–21
Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S., Tucker, P.A., 2005. Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. Acta Crystallogr. D Biol. Crystallogr. 61, 449–457.
Ramos, J.L., Martinez-Bueno, M., Molina-Henares, A.J., Teran, W., Watanabe, K., Zhang, X., Gallegos, M.T., Brennan, R., Tobes, R., 2005. The TetR family of transcriptional repressors. Microbiol. Mol. Biol. Rev. 69, 326–356.
Ryckaert, J.-P., Ciccotti, G., 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.
Schumacher, M.A., Miller, M.C., Grkovic, S., Brown, M.H., Skurray, R.A., Brennan, R.G., 2001. Structural mechanisms of QacR induction and multidrug recognition. Science 294, 2158–2163.
Schumacher, M.A., Miller, M.C., Grkovic, S., Brown, M.H., Skurray, R.A., Brennan, R.G., 2002. Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. EMBO J. 21, 1210–1218.
Sugiyama, M., Onaka, H., Nakagawa, T., Horinouchi, S., 1998. Site-directed mutagenesis of the A-factor receptor protein: Val-41 important for DNA-binding and Trp-119 important for ligand-binding. Gene 222, 133–144.
Tahlan, K., Yu, Z., Xu, Y., Davidson, A.R., Nodwell, J.R., 2008. Ligand recognition by ActR, a TetR-like regulator of actinorhodin export. J. Mol. Biol. 383, 753–761.
Wasserman, H.H., Shaw, C.K., Sykes, R.J., Cushley, R.J., 1974. The biosynthesis of metacycloprodigiosin and undecylprodigiosin. Tetrahedron Lett. 15, 2787–2790.
Yang, S., Gao, Z., Li, T., Yang, M., Zhang, T., Dong, Y., He, Z.-G., 2013. Structural basis for interaction between *Mycobacterium smegmatis* Ms6564, a TetR family master regulator, and its target DNA. J. Biol. Chem. 288, 23687–23695.
Zheng, G., Lu, X.-J., Olson, W.K., 2009. Web 3DNA—a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. Nucleic Acids Res. 37, W240–W246.