Article

Toxicology and Industrial Health

Toxicology and Industrial Health
2018, Vol. 34(12) 819–833

DOI: 10.1177/0748233718795934
journals.sagepub.com/home/tih
SAGE

# Identification of promoter $P_{cadR}$, *in silico* characterization of cadmium resistant gene *cadR* and molecular cloning of promoter $P_{cadR}$ from *Pseudomonas aeruginosa* BC15

**Rajkumar Prabhakaran[1], Sebastin Nirmal Rajkumar[2], Tharmarajan Ramprasath[3] and Govindan Sadasivam Selvam[1]**

## Abstract

Cadmium (Cd) remediation in *Pseudomonas aeruginosa* is achieved through the function of two vital genes, *cadA* and *cadR*, that code for P-type ATPase (CadA) and transcription regulatory protein (CadR), respectively. Although numerous studies are available on these metal-sensing and regulatory proteins, the promoter of these genes, metal sensing and binding ability, are poorly understood. The present work is aimed at the characterization of the CadR protein, identification of the $P_{cadR}$ promoter and protein–promoter–metal binding affinity using bioinformatics and to validate the results by cloning the $P_{cadR}$ promoter in *Escherichia coli* DH5$\alpha$. The promoter regions and its curvature were identified and analysed using PePPER software (University of Groningen, The Netherland) and the Bendit program (Version: v.1.0), respectively. Using Phyre, the three-dimensional structure of CadR was modelled, and the structure was validated by Ramachandran plots. The DNA-binding domain was present in the N-terminal region of CadR. A dimeric interface was observed in helix-turn-helix and metal ion-binding sites at the C-terminal. Docking studies showed higher affinity of Cd to both CadR (Atomic contact energy $= -15.04$ kcal/Mol) and $P_{cadR}$ (Atomic contact energy $= -40.18$ kcal/Mol) when compared to other metal ions. CadR with $P_{cadR}$ showed the highest binding affinity (Atomic contact energy$= -250.40$ kcal/Mol) when compared with $P_{cadA}$. *In vitro* studies using green fluorescent protein tagged with $P_{cadR}$ (*gfp*-$P_{cadR}$) cloned in *E. coli*-expressed *gfp* protein in a concentration-dependent manner upon Cd exposure. Based on our *in silico* studies and *in vitro* molecular cloning analysis, we conclude that $P_{cadR}$ and CadR are active only in the presence of Cd. The CadR protein has the highest binding affinity with $P_{cadR.}$ As it became apparent that the *cadR* gene regulates the $P_{cadR}$ activity in the presence of Cd with high specificity, and the *cadR* and $P_{cadR}$ can be used as a biological tool for development of a microbial biosensor.

## Keywords

$P_{cadR}$, *Pseudomonas aeruginosa*, *cadR* gene, cadmium, cadmium-binding transcriptional regulator, *Escherichia coli*

Received 19 January 2018; Revised 23 June 2018; Accepted 26 July 2018

## Introduction

Heavy metals pollution is an important environmental risk factor for both lower and higher organisms (Rizwan et al., 2017). Among the heavy metals, cadmium (Cd) is a major environmental problem because of its rapid accumulation in soil and water through anthropogenic wastes (Li et al., 2015; Liu et al., 2016). The Cd accumulation in the environment causes severe

[1]Department of Biochemistry, School of Biological Sciences, Madurai Kamaraj University, Madurai, Tamil Nadu, India
[2]Department of Zoology, Thiyagarajar College, Madurai, Tamil Nadu, India
[3]Center for Molecular and Translational Medicine, Petit Science Center, Georgia State University, Atlanta, GA, USA

**Corresponding author:**
Govindan Sadasivam Selvam, Department of Biochemistry, School of Biological Sciences, Madurai Kamaraj University, Madurai, Tamil Nadu, India.
Email: drselvamgsbiochem@rediffmail.com

health problems, such as cancer, hyperglycaemia, anaemia and immune deficiency in humans (Rizwan et al., 2017). Therefore, the detection of environmental Cd is essential and needs more attention. Existing traditional and physico-chemical methods, such as atomic absorption spectroscopy (AAS), inductively coupled plasma mass spectrometry (ICP-MS), have some disadvantages, such as sample manipulation, high cost and it produces the harmful byproducts (Bereza-Malcolm et al., 2015; Kim et al., 2016). Researchers have identified the role of microorganism in Cd resistance and removal (Albano and Macfie, 2016). Microbial Cd resistance is mainly based on energy-dependent mechanisms and is generally controlled by metal-recognizing regulatory proteins. The metal responding regulatory proteins fall under two family types, such as MerR (Philips et al., 2015) and ArsR family (Kang et al., 2016).

The metal regulatory proteins, for instance, CueR, MerR, ZntR, PbR, CoaR and CadR, falls under the MerR family, and ArsR, which is responsible for arsenic (As) resistance falls under the ArsR family (Kang et al., 2016; Wang et al., 2016). These MerR family proteins respond to a range of divalent metal ions, including Cu(II) (copper), Hg(II) (mercury), Zn(II) (zinc), lead (Pb), Co(II) (cobalt) and Cd(II) (cadmium) (Furnholm and Tisa, 2014; Waldron et al., 2009). In gram-negative bacteria, these MerR family proteins are responsible for the activation of sigma 70 factor ($\sigma^{70}$)-targeted promoters; on the other hand, in gram-positive bacteria, MerR family members activate $\sigma^{A}/\sigma^{54}$-targeted promoters (Brocklehurst et al., 2003).

The MerR family transcriptional regulatory proteins activate suboptimal $\sigma^{70}$-dependent metal-sensing promoters, in which the spacing between the $-35$ and $-10$ elements recognized by $\sigma$ factor is greater than the optimal $17 \pm 1$ bp in gram-negative bacteria. The activation of these MerR family promoter transcriptions is through protein-dependent DNA distortion. The MerR regulatory protein CadR has specific high-affinity binding to Cd in *Pseudomonas* species. Especially, *Pseudomonas aeruginosa* is a heavy-metal resistant bacteria isolated from industry effluent wastewater, and which possesses high resistance to Cd. This Cd resistance is mainly conferred by the *cadR* and *cadA* genes, where the *cadA* gene encodes P-type ATPase and *cadR* gene encodes transcriptional regulatory protein (Albano and Macfie, 2016; Li et al., 2016; Raja and Selvam, 2012).

In addition, several reports have demonstrated the function of CadR (Brocklehurst et al., 2003; Li et al., 2016); however, CadR and $P_{cadR}$ metal sensing and binding ability are poorly understood. There have been no previous reports that depict the characterization of CadR, protein–metal binding, interaction with promoter and protein–metal–promoter complex. In the present study, the heavy metal resistance and transcriptional regulation mechanism of proteins as well as physicochemical properties, such as isoelectric point (pI), molecular weight, number of atoms present, aliphatic index, subcellular localization, conserve domains, motif and grand average of hydropathicity (GRAVY), are well decribed by bioinformatics studies. The present study aimed to characterize CadR from *P. aeruginosa* BC15 and to identify the $P_{cadR}$ promoter as well as the protein-promoter-metal binding affinity by *in silico* studies and to validate the results by molecular cloning of $P_{cadR}$ in *Escherichia coli* DH5$\alpha$.

## Materials and methods

### Retrieval of the target sequence and phylogenetic analysis

The nucleotide sequence of the *cadR* gene was retrieved from the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/) using accession number KU302252 in which the CadR ANF04588 (Uniprot ID) protein sequences were obtained from SWISS-PROT data bank for structural and functional analysis. Sequences with significant identity were aligned with the ClustalW algorithm implemented in Molecular Evolutionary Genetic Analysis (MEGA 6; http://www.megasoftware.net) by distance matrix and it was trimmed to consensus (Tamura et al., 2013). Neighbour-joining trees were constructed with 1000 bootstraps at uniform divergence rates (Felsenstein, 1985). Posterior probability and conserved regions among the closely related sequences were performed with MEGA 6.

### Physicochemical characterization

Physicochemical properties of CadR in *P. aeruginosa* BC15, such as pI, molecular weight, number of atoms present, aliphatic index and GRAVY, were computed using the ProtParam tool (http://web.expasy.org/protparam/) (Sigrist et al., 2013). Prediction of subcellular localization of the protein was carried out by CELLOv.2.5 (Yu et al., 2004). The secondary structure of the protein was studied using PSIPRED server and

BLAST from NCBI and compared the query sequence with the database homologues sequences. The query protein domain study was performed using Interproscan analysis (Zdobnov and Apweiler, 2001). CadR-binding residues were analysed by NsitePred – web server. The fingerprints were analysed by PRINTS server, in order to study the tentative functional assignments (Attwood et al., 2002).

## Conserved domains and motif analysis

The conserved pattern of *cadR* gene was identified using motif search (http://www.genome.jp/tools/motif/). Target sequence domain annotation, precise locations of domain boundaries and functional sites were analysed by conserved domain database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The target protein secondary structure was predicted by PDBsum (www.ebi.ac.uk/pdbsum), Phyre and SOPMA server. The protein–protein interaction was evaluated using STRING web server (Snel et al., 2000).

## Three-dimensional structure modelling, energy minimization and validation of the model

Comparative three-dimensional (3D) modelling has been performed in the following stages: *cadR* gene encoded target CadR sequences of *P. aeruginosa* BC15 were identified and taken from the *P. aeruginosa* ANF04588 (Uniprot ID) NCBI protein database. The target protein sequence was retrieved in FASTA, formatted, and subjected to BLASTp against a PDB database for obtaining the most similar crystal structure and used as a template for the target protein modelling, which resulted as the identification of 1Q06B PDB id as a suitable template. The Swiss-Model tool, Phyre2 and RaptorX server were used for the CadR 3D structure prediction (Kallberg et al., 2012) (http://raptorx.uchicago.edu/). The modelled protein structures were viewed using the Swiss-Model tool (http://www.expasy.org/spdbv/). The value of the predicted CadR model was analysed by PyMOL (Schrodinger, 2010). Energy-minimized models were assessed for analysing the stereochemical quality and residue by residue geometry. RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage) was used for validating the CadR modelled structure by Ramachandran Plot (Lovell et al., 2003). The QMEAN Z scores were calculated for the complete assessment and refinement of the predicted model.

## *In silico* identification of $P_{cadR}$ promoter

The promoter sequences were confirmed by PePPER software (http://pepper.molgenrug.nl/index.php; http://genome2d.molgenrug.nl/genome2dresults/promoterprediction/210.212.252.186aacho7715uv4jqnfkkp7la4up1457/results.html). Both BLASTn (Altschul et al., 1990) search and manual sequence analysis of the putative promoter region were performed to the identification of significant matches and conserved motifs of GenBank sequences. ClustalW and BLAST were used to analyse and align the electropherograms. The Bendit online tool was used for the identification of promoter curvature. The $P_{cadR}$ and $P_{cadA}$ promoter sequences are identified by the PePPER database as follows:

$P_{cadA}$ promoter region from *P. aeruginosa* PA01
Motif = TTGACTCTGTAGTTGCTACAGG GTGTGCCAAT
$P_{cadR}$ promoter region from *P. aeruginosa* PA01
Motif = GCCAACCCTCCTCCAATCGCCG ACGCGGCTACCCGAATTGGCAGTAC CGC

## Molecular docking simulation

The protein docking study was performed to identify the interaction between metal–protein complexes. The two-dimensional structure of heavy metals (arsenate, Cd, chromium, cobalt, copper, nickel, lead, silver and zinc ion) was obtained from the PUBCHEM database in SDF file format and converted to PDB file format using an online molecular converter tool (http://cactus.nci.nih.gov/translate/). The docking analysis of the 3D structure of CadR, $P_{cadR}$, $P_{cadA}$, sigma factor (1SIG = crystal structure of a sigma 70 subunit fragment from *E. coli* RNA polymerase), and heavy metal ions was done using PatchDock molecular modelling simulation software, effective for protein–ligand docking (Schneidman-Duhovny et al., 2005).

### Molecular cloning of $P_{cadR}$ in *E. coli* DH5$\alpha$

*Bacterial strains, media and growth conditions.* The strain *P. aeruginosa* BC15 used in this study was previously isolated in our laboratory from oil-mill-treated waste water (Raja et al. 2006). The bacterial strains and plasmids used in this study are listed in Online Supplementary Table S1. *E. coli* DH5$\alpha$ (Invitrogen) was grown in Luria-Bertani (LB) broth, and *P. aeruginosa* BC15 was grown in Kings B media at

37°C in an orbital shaker at 140 r/min. The antibiotic ampicillin (Amp) was added at a final concentration of 100 μg/mL for *E. coli* DH5α.

*DNA techniques and transformation.* All the experiments were carried out according to Sambrook and Russell (2001). Primers used in this study were synthesized and procured from Sigma Aldrich, St Louis, Missouri, USA (see Online Supplementary Table S2). The polymerase chain reaction (PCR)-amplified products were verified by DNA sequencing (Eurofins Genomics India Pvt., Ltd, Chennai, Tamilnadu, India). *E. coli* DH5α was used as the host strain for maintaining plasmids. The plasmid DNA was used to transform *E. coli* DH5α by the calcium chloride transformation method (Ausubel et al., 1999).

*Amplification of $P_{cadR}$ and phylogenetic analysis.* Genomic DNA was extracted from *P. aeruginosa* BC15 by the method of Chen and Kuo (1993). The *cadR* gene upstream region was obtained from the NCBI database, and the promoter region was identified using PePPER and *in silico* prokaryotic promoter analysis. The *cadR* gene promoter sequence amplification primers were designed by online primer designing software using identified sequences. Appropriate primers were designed with flanking restriction sites for $P_{cadR}$ and amplified by PCR. The size of the PCR products was analysed by 2% agarose gel. The $P_{cadR}$ promoter was amplified from *P. aeruginosa* BC15 genomic DNA by primers, and the PCR product was purified and sequenced. The $P_{cadR}$ sequence was obtained from Xcelris labs Pvt. Ltd, Ahmedabad, Gujarat, India Sequences that matched previously published sequences in the NCBI databases using ADVANCED BLAST. Based on the scoring index, the most similar sequences were aligned. Multiple alignments of $P_{cadR}$ sequences were also constructed. The neighbour-joining method was used to construct phylogenetic trees using MEGA 6. Bootstrap confidence values were obtained with 1000 resamplings.

*Expression of green fluorescent protein.* The $P_{cadR}$ was amplified from *P. aeruginosa* genomic DNA using PCR primers (see Online Supplementary Table S2), which contain *Pci*I and *Kpn*I sites, respectively. The resulting fragment was introduced into pUCP18 containing *gfp* and cloned into *E. coli*. The *E. coli* harbouring pUCP-$P_{cadR}$-$GFP^{mut2}$ was grown overnight in LB medium supplemented with 100 μg/mL of ampicillin at 37°C. The overnight culture was a subculture in fresh LB medium supplemented with 100 μg/mL ampicillin and incubated at 37°C in an orbital shaker at 225 r/min to a 0.6 optical density at 600 nm (OD600). The 0.6 OD containing 2-mL aliquots of cells harbouring $P_{cadR}$-$GFP^{mut2}$ were mixed with cadmium chloride ($CdCl_2$; 0.5 mM). Similarly, the control cells harbouring $P_{lac}$-$GFP^{mut2}$ were induced by the addition of 1 mM isopropyl-β-D-thiogalactoside. Bacterial growth was monitored spectrophotometrically at an optical density of 600 nm (OD600), and the fluorescent intensity produced by the bacteria was measured using a high-content screening system (live cell imaging; Operetta, PerkinElmer, School of biological sciences, Madurai Kamaraj University, Madurai, and Tamilnadu, India) (Liao and Ou, 2005).

*Identification of promoter specificity.* To investigate the $P_{cadR}$ sensitivity, *E. coli* harbouring pUCP-$P_{cadR}$-$gfp^{mut2}$ was grown overnight in LB medium supplemented with 50-μM zinc chloride ($ZnCl_2$), 10-μM lead(II) acetate ($Pb(C_2H_3O_2)_2$), 25-μM manganese(II) chloride ($MnCl_2$), 25-μM nickel(II) sulphate ($NiSO_4$), 25-μM Copper(II) sulphate ($CuSO_4$), 0.5-μM mercury(II) chloride ($HgCl_2$) and 0.4-μM $CdCl_2$ at 37°C, respectively. The growth condition was the same as for the generation of the fluorescence (Wu et al., 2009).

*Nucleotide sequence accession numbers.* The *cadR* and $P_{cadR}$ sequences were deposited in GenBank under accession numbers ANF04588 and KU302252.

## Results and discussion

### Physicochemical characterization and secondary structure prediction of CadR

Heavy metal resistance in bacteria is conferred by metal resistant genes, and its expression is orchestrated through regulatory proteins (Li et al., 2016; Liu et al., 2015; Schirawski et al., 2002). In particular, *P. aeruginosa* Cd resistance was regulated by CadR (Lee et al., 2001; Raja and Selvam, 2012). Although there are several reports on Cd resistance mechanisms and regulation in *P. aeruginosa* using CadR protein, the sensitivity, protein–metal interaction and protein-promoter binding are still unclear. Therefore, in the present study, the investigations were aimed to identify and characterize the *P. aeruginosa* BC15 CadR, $P_{cadR}$ promoter, 3D structures of CadR, metal binding ability and probing the regulatory mechanism by studying the interaction of CadR protein with the $P_{cadR}$ promoter, and understanding the structural dynamics associated with the binding of protein–metal–promoter complex, the structural significance

**Table 1.** Physicochemical characteristics of CadR from *Pseudomonas aeruginosa* BC15 by ProtParam.

| Amino acid composition | CadR of *P. aeruginosa* | | |
|---|---|---|---|
| | Amino acids | Residues | % |
| | Ala (A) | 14 | 9.0 |
| | Arg (R) and Glu (E) | 19 | 12.2 |
| | Asn (N) and Met (M) | 3 | 1.9 |
| | Asp (D) | 5 | 3.2 |
| | Cys (C), Gln (Q) and Ser (S) | 6 | 3.8 |
| | Gly (G) | 11 | 7.1 |
| | His (H) | 9 | 5.8 |
| | Ile (I) and Pro (P) | 8 | 5.1 |
| | Leu (L) | 18 | 11.5 |
| | Lys (K) | 2 | 1.3 |
| | Sec (U), Pyl (O) and Trp (W) | 0 | 0.0 |
| | Phe (F) | 1 | 0.6 |
| | Thr (T) and Val (V) | 7 | 4.5 |
| | Tyr (Y) | 4 | 6 |
| Molecular weight | 17,694.1 | | |
| Theoretical Pi | 8.70 | | |
| Instability index (II) | 65.48 | | |
| Aliphatic index | 86.99 | | |
| Grand average of hydropathicity | −0.559 | | |
| Total number of negatively charged residues | (Asp + Glu) 24 | | |
| Total number of positively charged residues | (Arg + Lys) 21 | | |
| Atomic composition: | Carbon (C) 756<br>Hydrogen (H) 1230<br>Nitrogen (N) 242<br>Oxygen (O) 231<br>Sulphur (S) 9 | | |
| Formula | C756H1230N242O231S9 | | |
| Estimated half-life: | The estimated half-life is:<br>• 30 h (mammalian reticulocytes, *in vitro*)<br>• >20 h (yeast, *in vivo*)<br>• >10 h (*Escherichia coli*, *in vivo*) | | |

*Phyre2 Results*

*Dimer interface*: F49, H52, L56, D57, L66, C77, G78, N81, I84, L88, R94, L98, L101, R102, G109, V114, A115, C121, I123
*DNA-binding residues*: K2, I3, G4, R18, N35, Y36, R37
*Putative metal binding site*: C77, C112, C121

in Cd sensitivity using computational approach. Furthermore, we have performed molecular cloning of $P_{cadR}$ in order to validate our computational results.

Herein, we used the transcriptional regulatory protein (CadR) sequences from *P. aeruginosa* BC15 for *in silico* studies. Table 1 shows the physicochemical properties of CadR, which has a molecular weight of 17,694.1, and computed pI of 6.31 indicating its precipitation in acidic buffer. The GRAVY value − 0.559 shows the hydrophilic and soluble nature of the target protein. The CadR has 156 amino acids, which are similar to MerR transcription regulators (Table 1). In contrast, the CadR from *Pseudomonas putida* contains 147 amino acids (Lee et al., 2001). The amino acid composition shows an abundance of Arg (R) and Glu (E) residues (12.2%) in the CadR. The predicted subcellular localization of CadR was highest in cytoplasmic localization (reliability – 4.453). Kobae et al. (2004) explored the metal tolerance protein (MTP) MTP1 expression in cytoplasm. In addition, Li et al. (2016) described that cytoplasmic CadR expression significantly reduced the Cd content in chloroplasts. The Pfam database result exhibited CadR coming under the MerR helix-turn-helix (HTH) regulatory

**Table 2.** Peptide masses and motif analysis of CadR from *Pseudomonas aeruginosa* BC15.

| Mass | Position | Peptide sequence |
|---|---|---|
| 3004.4403 | 145–172 | ELEQPAPLSPIAEECAEAGHMHVPGVHR |
| 1942.7423 | 7–21 | OTEINSEQUENCEMK |
| 1496.7968 | 102–113 | LIDEHLHHVEVR |
| 1392.6722 | 131–144 | CTVAGNSQACGILR |
| 1304.5648 | 90–101 | DRPEADCGTANR |
| 1221.5780 | 74–83 | SLDMTQEEIR |
| 1116.5796 | 57–65 | QYTLAHVER |
| 981.5363 | 42–50 | EGLLPEPAR |
| 975.4928 | 29–37 | TGCPVETIR |
| 913.5465 | 114–121 | IAELQALR |
| 725.3213 | 51–56 | SEGNYR |
| 717.3460 | 1–6 | CADRPR |
| 686.4559 | 84–89 | TLLALR |
| 635.3875 | 66–70 | LSFIR |
| 630.3821 | 22–27 | IGELAK |
| 630.2882 | 38–41 | YYER |
| 547.2834 | 126–130 | DLGSR |
| 545.3042 | 122–125 | EQLR |

| Pfam | Position (independent *E*-value) | Description |
|---|---|---|
| MerR_1 | 1.68 (5.5e22) | PF13411, MerR HTH family regulatory protein |
| MerR DNA bind | 44.108 (3.1e20) | PF09278, MerR, DNA binding |
| MerR | 2.39 (4.4e18) | PF00376, MerR family regulatory protein |
| HTH_20 | 120.140 (0.44) | PF12840, HTH domain |
| Pkip-1 | 42.72 (0.2) | PF06878, Pkip-1 protein |
| GnsAB_toxin | 2.24 (0.21) | PF08178, GnsA/GnsB toxin of bacterial toxin antitoxin system |
| HTH_3 | 4.23 (0.53) | PF01381, HTH |

HTH: helix-turn-helix.

protein, which is concordant with previously reported MerR transcriptional activators (Brown et al., 2003). The motif analysis showed that the amino acid residues from the second HTH motif of CadR was responsible for DNA binding and interactions (position = 44–108 region (independent *E*-value =3.1e20); Table 2). The secondary structure analysis revealed that CadR had 58.91% $\alpha$-helix (76 residues), 3.88% extended strand (five residues), 8.53% $\beta$-turn (11 residues), 28.68% random coil (37 residues), 26% disorder regions and 1% $\beta$-strand, which is very low as predicted by SOPMA. The Conserved Domain Database (CDD), evolutionary tree of CadR and protein–protein interaction results demonstrate that CadR has high homology with other transcriptional regulatory proteins (see Online Supplementary Figure S1A). The STRING tool predicted that PA3689 (CadR) homologs are uncharacterized HTH-type transcriptional regulator PA4778 (similarity bit score is 69.7) and probable transcriptional regulator PA2718 (similarity bit score is 53.1). The evolutionary relationship of this CadR is shown in Online Supplementary Figure S1B. The peptide mass is shown in Table 2, and the longest peptide was present in the 145–172 position. UniProt BLAST of CadR with 1000 protein sequences showed the highest homology (94.9%) with the Cd(II)/Pb(II)-responsive transcriptional regulator from *P. aeruginosa* (strain PA7). CadR from *P. putida* contains HSHVGRSHGH/histidine-rich C-terminal extension, which is not present in the MerR family transcriptional regulatory proteins (Lee et al., 2001). The transcriptional regulatory protein CadR from *P. aeruginosa* also has C-terminal histidine-rich extension/HMHVPGVHRRHG. These histidine-rich C-terminal extensions may conserve or facilitate the sensitivity against Cd and chelation of heavy metals in *Pseudomonas* sp.

## Homology modelling

In order to identify the *P. aeruginosa* BC15 CadR 3D structure, the homology modelling for structure-based functional annotation was carried out. Phyre results

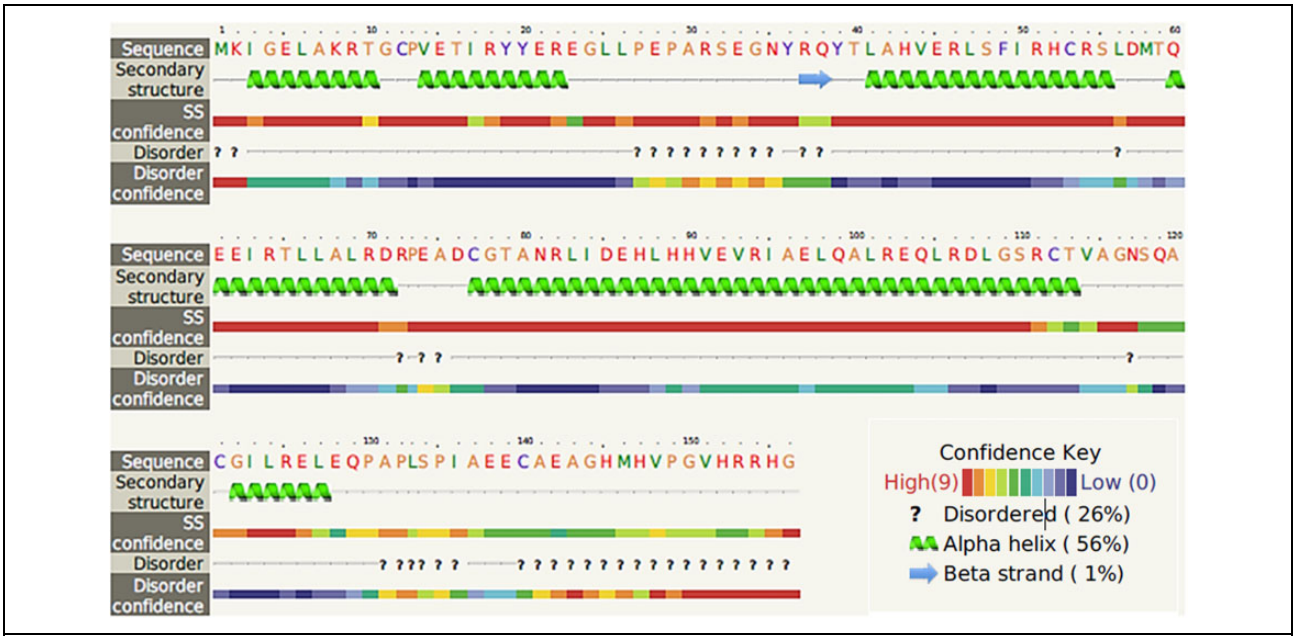



**Figure 1.** Phyre analysis and secondary structure elements of CadR.

revealed CadR has the highest (37%) identity with CueR protein of *E. coli*. The CadR was modelled with 100% confidence and 78% coverage with the single highest scoring template. Figure 1 exhibited the Phyre analysis for secondary structure, disorder prediction and CadR consensus. In addition, the COFACTOR is a structure-based method exhibited 1q06 as a best functional homology template for CadR (Template modeling (Tm) score: 0.83, root-mean-square deviation (RMSD): 2.05, identity: 0.341 and coverage are 0.961; Figure 2(a)). Finally, the structure of CadR was constructed using 1q06 as the template and viewed with PyMOL (Figure 2(b)). The Ramachandran plot and QMEAN score were exhibited; the obtained model was reliable (see Online Supplementary Figures S2 and S3). Additionally, SCOP info shows that CadR is under the super family of putative DNA-binding domain (DBD) and family of DNA-binding N-terminal domain of transcription activators. The results of the present study have good agreement with Heldwein and Brennan (2001) identified MerR transcription activator, CueR of *E. coli*.

## *Structure of CadR from* P. aeruginosa

The DNA-binding region of the CadR was present at the N-terminus and comprised of two HTH motifs (amino acid residues K2, I3, G4, R18, N35, Y36 and R37). Similarly, Wang et al. (2016) reported that the mercury regulatory protein from *P. aeruginosa* contains N-terminal DBD. A long 10-turn $\alpha$-helix followed the N-terminal region of the protein joined by loop regions. This long helix was basically responsible for dimerization of the CadR serving as the dimerization interface between the two CadR monomers (Figure 1). The dimer interface was constituted in the antiparallel coiled-coil domain between C77, C112 and Cys121. The sequence alignment revealed the presence of the metal ion-binding region in the CadR (Figure 1). The C77, C112 and C121 amino acids are responsible for metal ions binding to the CadR. Similarly, Lee et al. (2001) reported that the three conserved amino acids (C77, C112 and C121) from *P. putida* were responsible for the cation binding to the CadR. The metal ion-binding region was present at the C-terminal end of the CadR which is concordance with the result of Hobman et al. (2012).

## *Analysis of the Cd-sensing* $P_{cadR}$ *promoter region*

Metal-sensing promoters from microorganisms are useful for the heavy metal remediation in polluted environments (Newberry and Brennan, 2004). Lee et al. (2001) reported that the Cd-recognizing promoter region was lying between ATG codons for the *cadA* and *cadR* genes from *P. putida.* In this study, the 213-bp length of the Cd-sensing promoter region from *P. aeruginosa* BC15 was identified using a *cadR* gene upstream region, which was retrieved from NCBI database. The identified $P_{cadR}$ sequences by *in silico* were amplified from *P. aeruginosa* BC15 and sequenced (Figure 3). Based on the relative

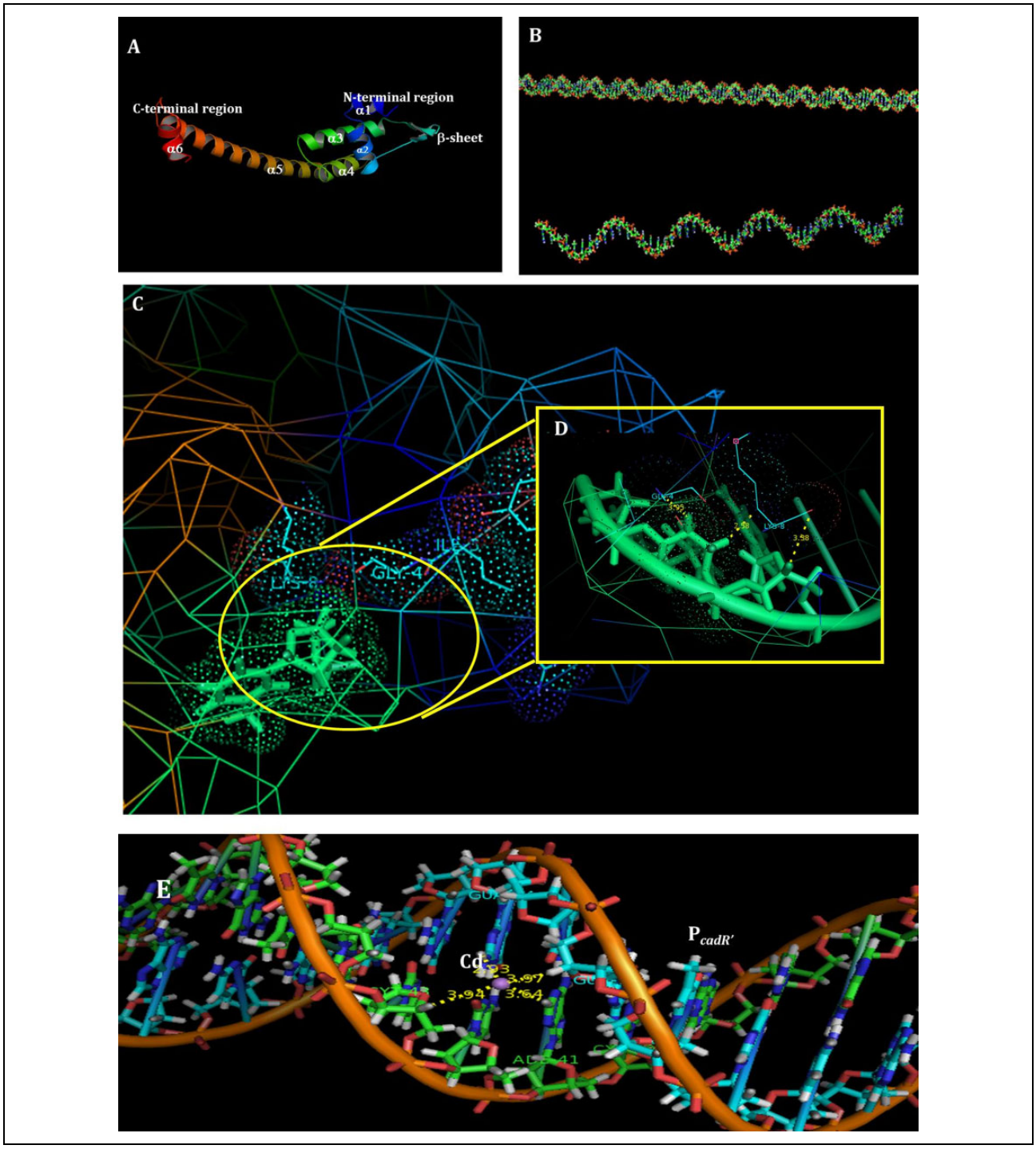



**Figure 2.** CadR protein 3D model by COFACTOR. (a) Cartoon representation CadR protein 3D model. (b) Cartoon representation of promoter DNA–CadR interactions. (c) The important DNA-binding amino acid residues (K2, I3, G4, R18, N35, Y36 and R37) are presented as sticks and dotted lines. (d) The H-bonding is presented as a dotted line. (e) Cartoon representation of $P_{cadR}$-Cd complex.

analysis of the sequences of strain *P. aeruginosa*, BC15 with previously available database showed that the 213-bp length sequence was closely related to the members of the transcriptional regulator from genus *P. aeruginosa*. The Basic Local Alignment Analysis of the $P_{cadR}$ sequence showed a relationship (100%) to each of the transcriptional regulators (Online Supplementary Table S3). The evolutionary relationship of this $P_{cadR}$ shown in Figure 3 and the identified sequences was named as $P_{cadR}$. The multiple sequence alignment

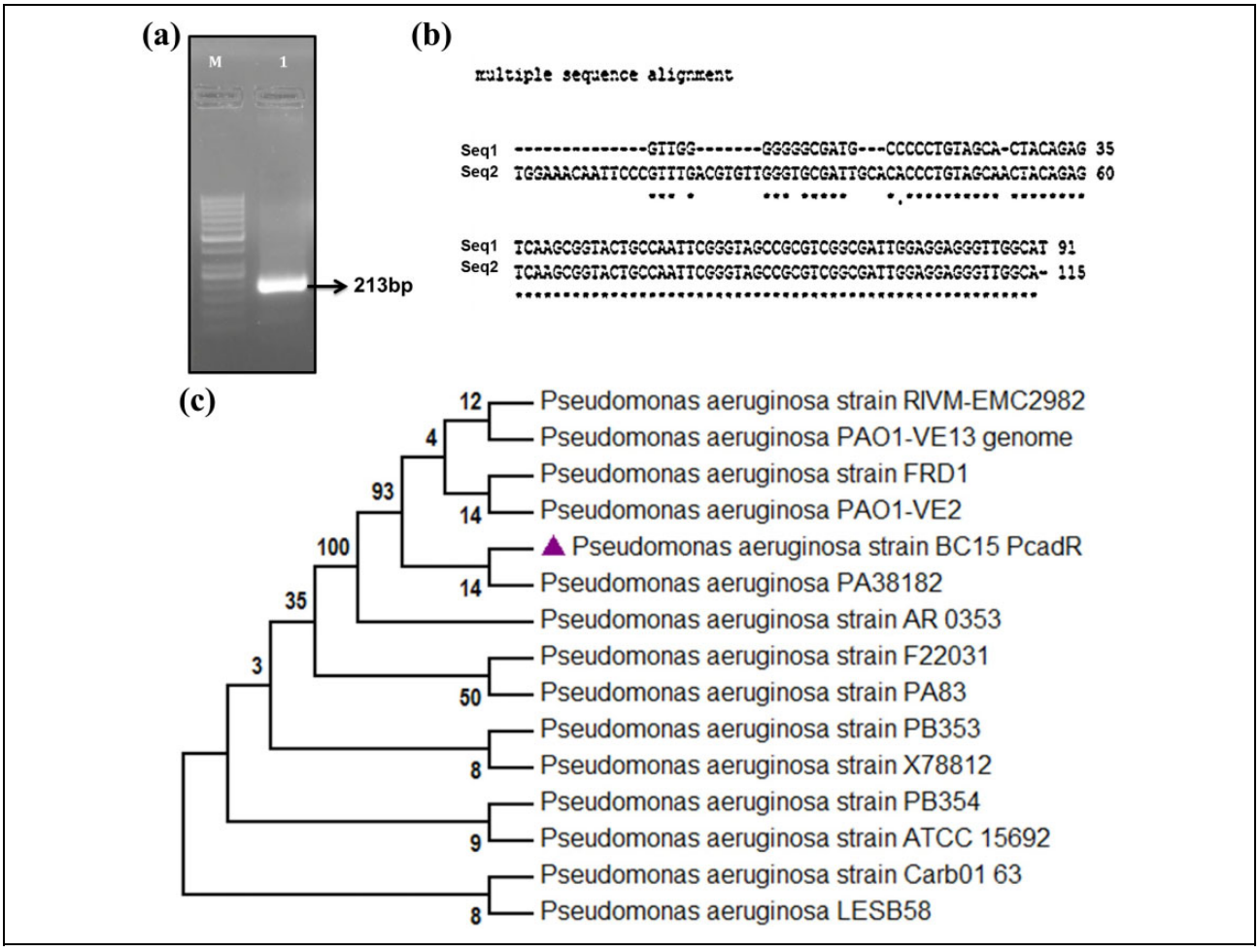


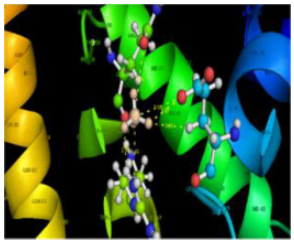
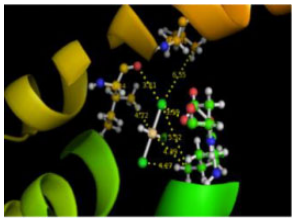
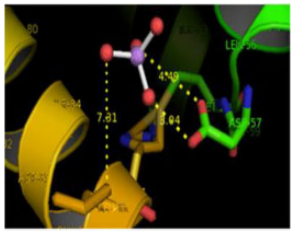
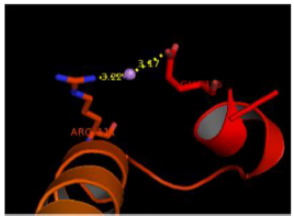
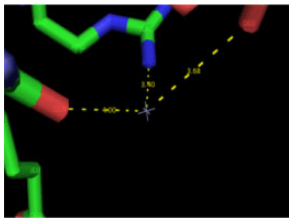
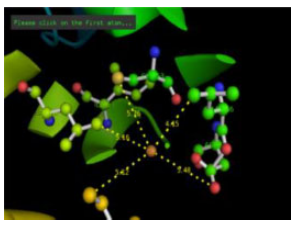
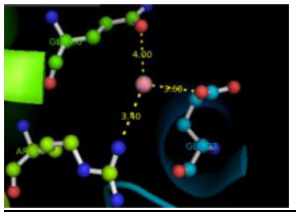

**Figure 3.** Identification and PCR amplification of $P_{cadR}$. (a) PCR amplification of $P_{cadR}$ from *Pseudomonas aeruginosa* BC15. M – 50 bp ladder, Lane 1 $P_{cadR}$ PCR-amplified product. (b) Pairwise alignment of $P_{cadR}$ promoter sequence. Alignment of the promoter region of CadR from *P. aeruginosa* BC15 (Seq1) and *P. aeruginosa* PA01 (Seq2). Identical bases are shown as asterisks. (c) Phylogenetic tree of the relatedness of $P_{cadR}$ from pUCP-$P_{cadR}$-$GFP^{mut2}$ with other *Pseudomonas* sp. strains based on the promoter sequences. PCR: polymerase chain reaction.

result for the $P_{cadR}$ promoter sequence with MerR family promoter sequences were shown in Online Supplementary Figure S5. The promoter sequence was used for the development of the 3D model. Simultaneously, the $P_{cadA}$ sequence was retrieved from NCBI database. The $P_{cadR}$ and $P_{cadA}$ promoters contain a 19-bp spacer region, the 6-bp −35 element and 1 bp of the −10 element. The results of the present study agreed well with Newberry and Brennan (2004).

## *Understanding the interaction between $P_{cadR}$ and CadR*

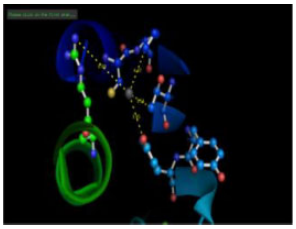
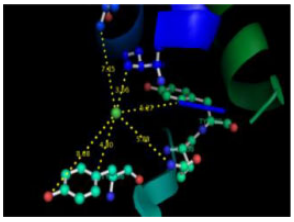
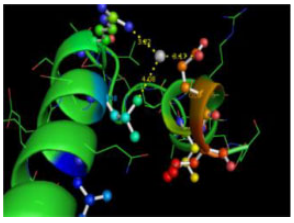
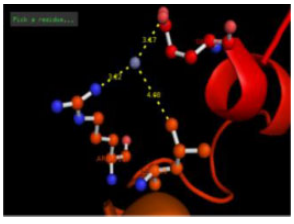
In order to analyse the Cd-sensing ability, remediation and transcription regulation process of CadR and promoters, a 3D structure of the $P_{cadR}$ and $P_{cadA}$ promoter DNA regions was constructed. These models were energy minimized and docked onto the model of the corresponding homodimeric CadR and heavy metals. In addition, the analysis of the $P_{cadR}$ promoter curvature reveals that the DNA curvature angle is 2.3728 and curvature is 0.0234 by bendit server (http://www.lfd.uci.edu/~gohlke/dnacurve/; Online Supplementary Figure S6). Newberry and Brennan's (2004) result confirmed that the MerR family promoters bend showed shortening of the spacers between −35 and −10 promoter elements and threefold increase in the MtaN-DNA binding affinity. In the present study, docking of CadR with Cd(II) and $CdCl_2$ demonstrated significant binding affinity towards CadR (ACE value of −15.04 kcal/Mol; Table 3). Similarly, Liu et al. (2015) reported that CadR from *P. putida* was

**Table 3.** Binding efficiency of As, Cr, Zn, Co, Pb, Ag, Ni, Cu and Cd (ligands) with CadR and $P_{cadR}$ in *Pseudomonas aeruginosa* BC15 by PatchDock.[a]

| CadR protein binding efficiency with heavy metals | | | | |
|---|---|---|---|---|
| PatchDock result | Heavy metals | Score | Area | Atomic contact energy |
|  | Cadmium | 852 | 101.3 | −15.04 |
|  | Cadmium chloride | 1736 | 186.40 | −46.93 |
|  | Arsenate | 386 | 146.90 | −11.84 |
|  | Arsenic | 358 | 49.50 | 0.00 |
|  | Chromium | 852 | 101.30 | − 15.04 |
|  | Copper | 852 | 101.30 | − 15.04 |
|  | Cobalt | 852 | 101.30 | − 15.04 |

*(continued)*

**Table 3.** (continued)

| CadR protein binding efficiency with heavy metals | | | | |
|---|---|---|---|---|
| PatchDock result | Heavy metals | Score | Area | Atomic contact energy |
|  | Lead | 544 | 61.10 | −27.74 |
|  | Nickel | 534 | 67.60 | −18.21 |
|  | Silver | 358 | 49.50 | 0.00 |
|  | Zinc | 358 | 49.50 | 0.00 |

| $P_{cadR}$ promoter binding efficiency with heavy metals | | | | |
|---|---|---|---|---|
| Promoter | Heavy metals | Score | Area | Atomic contact energy |
| $P_{cadR}$ | Cadmium | 2528 | 285.50 | −40.18 |
| $P_{cadR}$ | Arsenic | 404 | 54.70 | 0.00 |
| $P_{cadR}$ | Chromium | 404 | 54.70 | 0.00 |
| $P_{cadR}$ | Copper | 2086 | 237.00 | −126.51 |
| $P_{cadR}$ | Cobalt | 240 | 24.40 | 0.79 |
| $P_{cadR}$ | Lead | 676 | 81.30 | 37.42 |
| $P_{cadR}$ | Nickel | 670 | 80.60 | −27.79 |
| $P_{cadR}$ | Silver | 404 | 54.70 | 0.00 |
| $P_{cadR}$ | Zinc | 404 | 54.70 | 0.00 |

As: arsenic; Cr: chromium; Zn: zinc; Co: cobalt; Pb: lead; Ag: silver; Ni: nickel; Cu: copper; Cd: cadmium.
[a]The increasing binding efficiency is shown by high score. The high binding area also indicates that Cd binds with CadR and $P_{cadR}$ strongly.

specifically recognizing the Cd with good binding affinity. In order to analyse the interactions between CadR and its promoter DNA region, a model of $P_{cadR}$ and dimeric model of CadR were generated and docked. PatchDock result of CadR with promoters showed highest binding affinity

with $P_{cadR}$ promoter (score: 11,664, area: 1454.70 and ACE value: −250.40) compared with $P_{cadA}$.

MerR family regulators act on an elongated spacer region between −10 and −35 elements of the metal responsive promoter (Kidd and Brown, 2003). The ATPase gene promoter $P_{cadA}$ is overlapping and divergent to the *cadR* gene promoter. On docking of the metals As(II), Cr(II), Co(II), Ni(II), Pb(II), Zn(II), Co(II), Cu(I), Ag(I) and Hg(II), the affinity of CadR for the DNA decreased, as it occurs with $P_{cadR}$ in the presence of Cd(II) (Figure 3). This docking result shows that a regulator protein encoded by *P. aeruginosa* genome uniquely responds to Cd(II). The CadR is a strong activator of transcription from the $P_{cadR}$ (Table 3) and $P_{cadA}$ in the presence of Cd (see Online Supplementary Table S4); it also acts as a weak repressor in the absence of Cd. Lee et al. (2001) and Wu et al. (2009) studies revealed that $P_{cadR}$ from *P. putida* responds to Cd alone. In both states, it binds a dyad symmetrical sequence between the −35 and −10 promoter elements of $P_{cadR}$. The spacer sequence between the −10 and −35 elements is 19 bp long. CadR contacts with the DNA are altered upon Cd binding to CadR. This result was similar to O'Halloran and Walsh (1987) study results. The sigma 70 factor was docked with metal ions, and the resulting complex was docked with $P_{cadR}$ and $P_{cadA}$ promoters. The result indicates that sigma factor–Cd complex strongly binds with $P_{cadR}$ as compared with $P_{cadA}$ and other heavy metals (Table 4). This result has high concordance with the report of Li et al. (2016) who identified; the CadR expression has increased Cd detoxification activity and specificity.

**Table 4.** Binding efficiency of promoters with sigma factor in the presence and absence of heavy metals such as As, Cr, Zn, Co, Pb, Ag, Ni, Cu and Cd (ligands).[a]

| PatchDock result | Heavy metals | Score | Area | Atomic contact energy |
|---|---|---|---|---|
| | Sigma factor + $P_{cadA}$ | 9662 | 1500.40 | 344.62 |
| | Sigma factor + $P_{cadR}$ | 11,870 | 1556.00 | 142.98 |
| Sigma Cd + $P_{cadA}$ | | 9568 | 1499.10 | 346.20 |
| Sigma Cd + $P_{cadR}$ | | 11,938 | 1471.50 | 439.24 |
| Sigma Cu + $P_{cadR}$ | | 11,938 | 1471.50 | 439.24 |
| Sigma As + $P_{cadR}$ | | 11,870 | 1556.00 | 142.98 |
| Sigma Cr + $P_{cadR}$ | | 11,870 | 1556.00 | 142.98 |
| Sigma Co + $P_{cadR}$ | | 11,870 | 1556.00 | 142.98 |
| Sigma Ni + $P_{cadR}$ | | 11,870 | 1556.00 | 142.98 |
| Sigma Ag + $P_{cadR}$ | | 11,870 | 1556.00 | 142.98 |
| Sigma Zn + $P_{cadR}$ | | 11,870 | 1556.00 | 142.98 |

As: arsenic; Cr: chromium; Zn: zinc; Co: cobalt; Pb: lead; Ag: silver; Ni: nickel; Cu: copper; Cd: cadmium.

[a]The increasing binding efficiency is shown by high score. The high binding area also indicates that sigma factor–Cd complex binds with $P_{cadR}$ strongly.

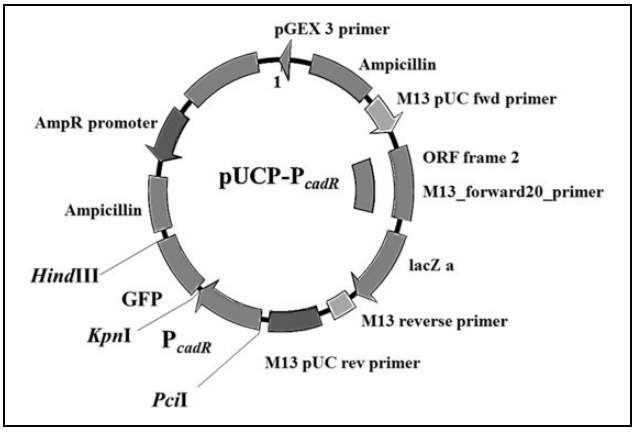



**Figure 4.** Schematic diagram of vector construct pUCP-$P_{cadR}$-$gfp^{mut2}$.

## *Promoter activity assay*

The identified 213 bp length of $P_{cadR}$ was introduced into the fluorescent-based reporter system (pUCP-$P_{lac}$-$gfp^{mut2}$; Online Supplementary Figure S6B). The positive recombinant *E. coli* (pUCP-$P_{cadR}$-$gfp^{mut2}$) strains, designated GSTC3 and GSTC8, were analysed for their responses to a range of divalent metal salts (Figure 4). The $P_{cadR}$ harbouring cells show *gfp* expression in the presence of Cd, and no *gfp* expression was observed when the cells cultured in $MnCl_2$, $NiSO_4$, $CuSO_4$, $HgCl_2$, $Pb(C_2H_3O_2)_2$, $ZnCl_2$ and without Cd (Figure 5). Similarly, *in silico* analysis results also demonstrate that Cd is the powerful inducer for the expression of CadR and $P_{cadR}$ transcription. In concordance with the present study, Brocklehurst et al. (2003) and Bereza-Malcolm et al. (2017) constructed $P_{cadR}$ harbouring biosensors that were only induced by Cd. These results may be used to know exactly how the $P_{cadR}$ works in *P. aeruginosa*, in which Cd ions may have different accessibility to the CadR. Further detailed analysis of the metal-responsive promoter is required in order to understand the discriminatory metal recognition.

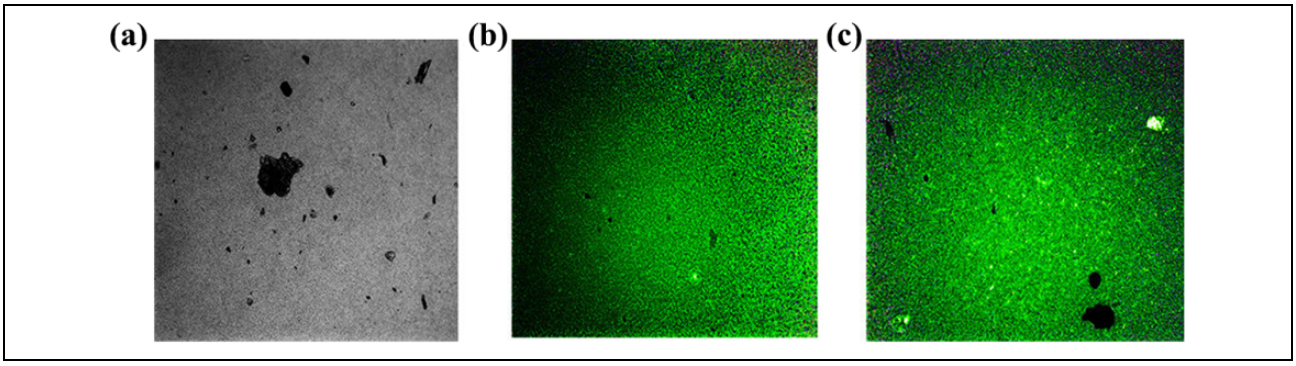



**Figure 5.** Green fluorescence images of the *Escherichia coli* carrying pUCP-$P_{cadR}$-$gfp^{mut2}$ exposed to $CdCl_2$ by high content screening. (a) Wild type *E. coli* (negative control), (b) *E. coli* DH5$\alpha$ harbouring pUCP-$P_{lac}$-$gfp^{mut2}$ exposed to 1 mM of $CdCl_2$ and (c) *E. coli* DH5$\alpha$ harbouring pUCP-$P_{cadR}$-$gfp^{mut2}$ exposed to 0.5 mM of $CdCl_2$. $CdCl_2$: cadmium chloride.

## Conclusion

First, we successfully predicted and characterized the CadR 3D structure by *in silico,* identified the $P_{cadR}$ promoter from *P. aeruginosa* BC15 and cloned the $P_{cadR}$ promoter in *E. coli*. By computational analysis, we understood that the DBDs may be possibly present in the N-terminal region. In addition, the dimeric interface occurred in the HTH and metal ion binding in the C-terminal region. Furthermore, we found that the cysteine residues in the metal-binding region initiate the expression of CadR from their promoter ($P_{cadR}$).

Second, we analysed the Cd-sensing ability and promoter interaction of the gene *cadR* (PA3689) coding for the transcriptional regulatory protein from *P. aeruginosa*. The 3D structure predictions of CadR and promoter might facilitate to add the functional annotation of protein and promoter sequence. Through molecular docking simulations, we identified the amino acid residues from the second HTH motif of CadR, which is responsible for the DNA-binding activity. Bendit analysis showed a curvature in $P_{cadR}$, which implies that the promoter bends allow the RNA polymerase accessing the full promoter site for transcription. So far, this is the first report that describes the mode of DNA binding and regulation by CadR from *P. aeruginosa*. This study would pave the pathway for future genetic and mutational approach to find the role of CadR amino acids used in gene PA3689 regulation. Finally, the 3D structure prediction of CadR and its metal binding as well as identification and molecular cloning of $P_{cadR}$ results revealed that the *cadR* gene promoter was only induced by Cd.

## Acknowledgements

The authors would like to thank Prof. Pradeep Singht, University of Washington for providing the vector, and also the central instrumentation facility at SBS, MKU, through CEGS, NRCBS, DST-FIST, DBT-IPLS and DST-PURSE programme.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was financially supported by the Department of Biotechnology (project number – BT/PR3857/MED/32/222/2011), Government of India.

## ORCID iD

Govindan Sadasivam Selvam http://orcid.org/0000-0002-7786-2034

## Supplemental material

Supplemental material for this article is available online.

## References

Albano LJ and Macfie SM (2016) Investigating the ability of *Pseudomonas fluorescens* UW4 to reduce cadmium stress in *Lactuca sativa* via an intervention in the ethylene biosynthetic pathway. *Canadian Journal of Microbiology* 62: 1057–1062.

Altschul SF, Gish W, Miller W, et al. (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Attwood TK, Blythe MJ, Flower DR, et al. (2002) PRINTS and PRINTS-S shed light on protein ancestry. *Nucleic Acids Research* 30: 239–241.

Ausubel FM, Brent R, Kingston RE, et al. (eds.) (1999) *Introduction of DNA into Mammalian Cells. Current*

*Protocols in Molecular Biology*. Chapter 9. New York: John Wiley.

Bereza-Malcolm L, Aracic S, Kannan R, et al. (2017) Functional characterization of Gram-negative bacteria from different genera as multiplex cadmium biosensors. *Biosensors and Bioelectronics* 94: 380–387.

Bereza-Malcolm L, Mann G and Franks AE (2015) Environmental sensing of heavy metals through whole cell microbial biosensors: a synthetic biology approach. *ACS Synthetic Biology* 4: 535–546.

Brocklehurst KR, Megit SJ and Morby AP (2003) Characterisation of CadR from *Pseudomonas aeruginosa*: a Cd(II)-responsive MerR homologue. *Biochemical and Biophysical Research Communications* 308: 234–239.

Brown NL, Stoyanov JV, Kidd SP, et al. (2003) The MerR family of transcriptional regulators. *FEMS Microbiology Reviews* 27: 145–163.

Chen WP and Kuo TT (1993) A simple and rapid method for the preparation of gram-negative bacterial genomic DNA. *Nucleic Acids Research* 21: 2260.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39(4): 783–791.

Furnholm TR and Tisa LS (2014) The ins and outs of metal homeostasis by the root nodule actinobacterium Frankia. *BMC Genomics* 15: 1092.

Heldwein EE and Brennan RG (2001) Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature* 409: 378–382.

Hobman JL, Julian DJ and Brown NL (2012) Cysteine coordination of Pb(II) is involved in the PbrR-dependent activation of the lead-resistance promoter, P*pbrA*, from *Cupriavidus metallidurans* CH34. *BMC Microbiology* 12: 109.

Kallberg M, Wang H, Wang S, et al. (2012) Template-based protein structure modeling using the RaptorX web server. *Nature Protocols* 7: 1511–1522.

Kang YS, Brame K, Jetter J, et al. (2016) Regulatory activities of four ArsR proteins in *Agrobacterium tumefaciens* 5A. *Applied and Environmental Microbiology* 82: 3471–3480.

Kidd SP and Brown NL (2003) ZccR—a MerR-like regulator from *Bordetella pertussis* which responds to zinc, cadmium, and cobalt. *Biochemical and Biophysical Research Communications* 302: 697–702.

Kim HJ, Lim JW, Jeong H, et al. (2016) Development of a highly specific and sensitive cadmium and lead microbial biosensor using synthetic CadC-T7 genetic circuitry. *Biosensors and Bioelectronics* 79: 701–708.

Kobae Y, Uemura T, Sato MH, et al. (2004) Zinc transporter of *Arabidopsis thaliana* AtMTP1 is localized to vacuolar membranes and implicated in zinc homeostasis. *Plant and Cell Physiology* 45: 1749–1758.

Lee SW, Glickmann E and Cooksey DA (2001) Chromosomal locus for cadmium resistance in *Pseudomonas putida* consisting of a cadmium-transporting ATPase and a MerR family response regulator. *Applied and Environmental Microbiology* 67: 1437–1444.

Li J, Wei X, Yu P, et al. (2016) Expression of *cadR* enhances its specific activity for Cd detoxification and accumulation in Arabidopsis. *Plant and Cell Physiology* 57: 1720–1731.

Li Y, Duan Z, Liu G, et al. (2015) Evaluation of the possible sources and controlling factors of toxic metals/metalloids in the Florida Everglades and their potential risk of exposure. *Environmental Science & Technology* 49: 9714–9723.

Liao VH and Ou KL (2005) Development and testing of a green fluorescent protein-based bacterial biosensor for measuring bioavailable arsenic in contaminated groundwater samples. *Environmental Toxicology and Chemistry* 24: 1624–1631.

Liu Q, Yuan F, Liang Y, et al. (2015) Cadmium adsorption by *E. coli* with surface displayed CadR. *RSC Advances* 5: 16089–16092.

Liu X, Tian G, Jiang D, et al. (2016) Cadmium (Cd) distribution and contamination in Chinese paddy soils on national scale. *Environmental Science and Pollution Research* 23: 17941–17952.

Lovell SC, Davis IW, Arendall WB3rd, et al. (2003) Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins* 50: 437–450.

Newberry KJ and Brennan RG (2004) The structural mechanism for transcription activation by MerR family member multidrug transporter activation, N terminus. *Journal of Biological Chemistry* 279: 20356–20362.

O'Halloran T and Walsh C (1987) Metalloregulatory DNA-binding protein encoded by the merR gene: isolation and characterization. *Science* 235: 211–214.

Philips SJ, Canalizo-Hernandez M, Yildirim I, et al. (2015). Allosteric transcriptional regulation via changes in the overall topology of the core promoter. *Science* 349: 877–881.

Raja CE, Anbazhagan K and Selvam GS (2006) Isolation and characterization of a metal-resistant Pseudomonas aeruginosa strain. *World Journal of Microbiology and Biotechnology* 22(6): 577–585.

Raja CE and Selvam GS (2012) Characterization of chromosomal mediated cadmium resistance in *Pseudomonas aeruginosa* strain BC15. *Journal of Basic Microbiology* 52: 175–183.

Rizwan M, Ali S, Adrees M, et al. (2017) A critical review on effects, tolerance mechanisms and management of cadmium in vegetables. *Chemosphere* 182: 90–105.

Sambrook J and Russel D (2001) *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press.

Schirawski J, Hagens W, Fitzgerald GF, et al. (2002) Molecular characterization of cadmium resistance in *Streptococcus thermophilus* strain 4134: an example of lateral gene transfer. *Applied and Environmental Microbiology* 68: 5508–5516.

Schneidman-Duhovny D, Inbar Y, Nussinov R, et al. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Research* 33: W363–W367.

Schrodinger LLC (2010) The PyMOL molecular graphics system, version 1.3 r1, pp. 00119–9. (accessed 10 April 2016).

Sigrist CJ, de Castro E, Cerutti L, et al. (2013) New and continuing developments at PROSITE. *Nucleic Acids Research* 41: D344–D347.

Snel B, Lehmann G, Bork P, et al. (2000) STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Research* 28: 3442–3444.

Tamura K, Stecher G, Peterson D, et al. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30: 2725–2729.

Waldron KJ, Rutherford JC, Ford D, et al. (2009) Metalloproteins and metal sensing. *Nature* 460: 823–830.

Wang D, Huang S, Liu P, et al. (2016) Structural analysis of the Hg(II)-regulatory protein *Tn501* MerR from *Pseudomonas aeruginosa*. *Scientific Reports* 6: 33391.

Wu CH, Le D, Mulchandani A, et al. (2009) Optimization of a whole-cell cadmium sensor with a toggle gene circuit. *Biotechnology Progress* 25: 898–903.

Yu CS, Lin CJ and Hwang JK (2004) Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on *n*-peptide compositions. *Protein Science* 13: 1402–1406.

Zdobnov EM and Apweiler R (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847–848.