Mycobacterial STAND adenylyl cyclases: the HTH domain binds DNA to form biocrystallized nucleoids

A. Zaveri[#], A. Bose[#], S. Sharma, A. Rajendran, P. Biswas, A.R. Shenoy and S. S.Visweswariah*

#: authors contributed equally to this study

Keywords: STAND proteins; mycobacteria; pseudoenzyme; DNA binding protein; RNA binding protein; atomic force microscopy

Abstract

Mycobacteria harbour a unique class of adenylyl cyclases with a complex domain organization, consisting of an N-terminal putative adenylyl cyclase domain fused to an NB-ARC (nucleotide-binding adaptor shared by apoptotic protease-activating factor-1, plant resistance proteins and CED-4) domain, a tetratricopeptide repeat (TPR) domain and a C-terminal helix-turn-helix (HTH) domain. The products of the *rv0891c-rv0890c* genes represent a split gene pair, where Rv0891c has sequence similarity to adenylyl cyclases, and Rv0890c harbours the NB-ARC-TPR-HTH domains. Rv0891c had very low adenylyl cyclase activity so it could represent a pseudoenzyme. By analysing the genomic locus, we could express and purify Rv0890c and find that the NB-ARC domain binds ATP and ADP, but does not hydrolyze these nucleotides. Using systematic evolution of ligands by exponential enrichment (SELEX) we identified DNA sequences that bound to the HTH domain of Rv0890c. Uniquely, the HTH domain could also bind RNA. Atomic force microscopy revealed that binding of Rv0890c to DNA was sequence-independent, and binding of adenine nucleotides to the protein induced the formation of higher order structures which may represent biocrystalline nucleoids. This represents the first characterization of this group of proteins and their unusual biochemical properties warrant further studies into their physiological roles in future.

Statement of Significance

- Unique domain fusions are seen in adenylyl cyclases from Mycobacteria
- Characterization of a split gene pair reveals that the adenylyl cyclase has very low activity, representing a pseudoenzyme
- The second member of the gene pair harbours an NB-ARC domain that binds adenine nucleotides and a novel HTH domain that binds DNA and RNA in a sequence independent manner
- In the presence of adenine nucleotides, protein binding to DNA results in the formation of biocrystallized nucleoids
- This family of enzymes, found only in slow growing mycobacteria, have unusual properties warranting a closer study of their role in this slow growing pathogen.

INTRODUCTION

The prolonged co-existence of *Mycobacterium tuberculosis* (Mtb)[1] with humans has resulted in the evolution of several distinctive features in the pathogen that presumably play a role in its complex interaction with the host. These include the increased frequency of genes devoted to lipid metabolism and the expansion of signalling pathways similar to those seen in eukaryotes, such as serine/threonine kinases (1) and adenylyl cyclases. Mtb encodes for a rich cAMP machinery, consisting of sixteen putative adenylyl cyclases (ACs), ten proteins which harbour cyclic nucleotide-binding domains, and a single characterised phosphodiesterase (2, 3). The presence of multiple genes encoding for adenylyl cyclases is correlated with the ability of this organism to synthesize and secrete elevated amounts of cAMP, providing a mechanism for the pathogen to hijack signalling within the macrophage (3, 4).

The diversity in domain organization and the biochemical properties of adenylyl cyclases argues against redundancy (5) in terms of function. Indeed, deletion of some individual adenylyl cyclase genes in the bacterium did not alter the virulence of the bacterium in macrophages (6-8). However, deletion of only *rv0386* was reported to impair the virulence of Mtb in mice (8). Rv0386 has a complex, multidomain organisation consisting of an adenylyl cyclase, an NB-ARC (nucleotide-binding adaptor shared by apoptotic protease-activating factor-1, certain R gene products and cell death protein-4), a TPR (tetratricopeptide repeat) and a tetrahelical LuxR-type HTH (helix-turn-helix) domain (C-A-T-H). This domain organisation is shared by two other Mtb adenylyl cyclases, Rv1358 and Rv2488c, which are as yet uncharacterised (9). The C-A-T-H domain organisation also places these proteins in a larger group of Signal Transduction ATPases with Numerous Domains (STAND) proteins, an emerging subgroup of AAA+ ATPases that exhibit complex signalling paradigms by formation of oligomeric hubs (10, 11). While most STAND proteins have a single effector domain towards the N-terminus of the NB-ARC domain, Rv0386 and its paralogues present an interesting case of two potential effector domains at each terminus, an adenylyl cyclase and an HTH domain. Furthermore, this domain organization, where a putative adenylyl cyclase domain is present along with an HTH domain in a single protein, is exclusive to this set of mycobacterial proteins and not found in any other organism (9).

The STAND-adenylyl cyclases are large proteins and difficult to express and purify (12). Indeed, only the adenylyl cyclase domain of Rv0386 has been characterized to date. In order to characterize the properties of this group of unusual proteins from mycobacteria, we chose to work with the split gene pair, *rv0891c-rv0890c* where the gene encoding the adenylyl

cyclase is separated from the adjacent A-T-H domains. We describe here the biochemical properties of this split gene pair (*rv0891c-rv0890c*), and find that while adenylyl cyclase activity is severely compromised, rendering Rv0891c a 'pseudoenzyme', the NB-ARC domain in Rv0890c binds ATP. Moreover, the HTH domain is able to bind DNA and forms biocrystallized nucleoids in the presence of adenine nucleotides. Our study indicates that the roles of these STAND ACs in mycobacteria must be viewed in the context of additional biochemical properties they possess, and not solely on their ability to generate cAMP.

MATERIALS AND METHODS

Sequence alignments Domain boundaries were identified using Pfam 29.0 (13), CDD (14) or Interpro Scan 5 (15). Sequence alignments were generated using Clustal Omega (16). Conserved motifs were identified using existing literature (5, 10). All primers used in this study are listed in Table 1.

RNA isolation and reverse transcription PCR (RTPCR) *M. tuberculosis* H37Rv cells were grown in Middlebrook's 7H9 broth containing 0.05% Tween-80 and supplemented with 0.2% glycerol, 0.2% glucose or 0.2% propionate as indicated and cultured with shaking at 37 ˚C till the exponential phase of growth was reached for each culture (4-8 days; OD 0.3-1.0, depending on the culture medium). Pelleted cells were resuspended in appropriate amounts of TRI reagent (Sigma-Aldrich, USA) or RNAiso Plus (Takara Bio, Japan). Approximately 800 µL were used per 5 x 10^9 cells. Cells were lysed by bead beating (0.5 mm glass beads, BioSpec products) in Mini-Beadbeater-16 (BioSpec products) followed by centrifugation at 16000 rcf and 4 ˚C for 10 min. The supernatant was mixed with 0.2 vol of chloroform and centrifuged at 16000 rcf and 4 ˚C for 10 min. The aqueous layer was mixed with an equal volume of isopropanol and centrifuged at 16000 rcf and 4 ˚C for 10 min. The obtained pellet was washed with 70% ethanol and dried and resuspended in water. 5 µg RNA was treated with RNase-free DNaseI (2 units). 2 µg of DNase-free RNA was mixed with 200 ng of random hexamers followed by incubation at 70 ˚C for 5 min and snap chilled on ice for 5 minutes. To this 500 µM of dNTPs, and an appropriate amount of RT buffer (Thermo Scientific, USA) were added to a final volume of 18 µL and the reaction was split into two equal halves. To one, 200 units of RevertAid reverse transcriptase was added. The tubes were incubated at 25 ˚C for 10 min followed by incubation at 42 ˚C for 1 h. Enzyme inactivation was brought about by incubation at 70 ˚C for 10 min. The cDNA was subjected to RT-PCR using appropriate primers (Rv0890c RT rvs and Rv0890c

RT fwd; Rv0891c RT fwd and Rv0891c RT rvs; Rv0891c RT fwd and Rv0891c-Rv0890c junction rvs) for testing the operon (Table 1).

Expression of Rv0891c-Rv0890c in M. smegmatis To generate a clone encompassing the *rv0891c-rv0890c* operon under the control of the *sigA* promoter, PCR was performed on *M. tuberculosis* H37Rv genomic DNA as the template using primers Rv0891c BamHI fwd and Rv0891c XhoI rvs (Table 1). The resulting 893 bp amplicon was digested with BamHI and XhoI and ligated with similarly digested pPROEX-HTb to generate pPRO-Rv0891c_start. The 899 bp long BamHI-HindIII fragment obtained from pPRO-Rv0891c_start and a 340 bp long KpnI-BamHI fragment from a plasmid containing the *sigA* promoter (amplified from genomic DNA) was ligated with the ~ 3.8 kb KpnI-HindIII fragment from pMV 10-25 to generate pMV-sigA-Rv0891c. Next, the 3.1 kb NheI-HindIII fragment from pPRO-Rv0891c-Rv0890c (see below) was ligated with the ~4.5 kb NheI-HindIII fragment from pMV-sigA-Rv0891c to generate pMV-sigA-Rv0891c-Rv0890c.

Table 1 List of oligonucleotide primers used in the study

Name	**Sequence (5' to 3')**
Rv0891c BamHI fwd	CAGGATCCGTGCTCTTCAACGCAGTTC
Rv0891c T128STOP XbaI fwd	GTGTCGGAAGCTATCTAGAATCACGGTGGGG
Rv0890c TPR EcoR1 fwd	CCGGAATTCGCCACCGACATCATCGC
Rv0890c TPR XbaI rvs	CCGGTCTAGATTATTCACCGCGACCGCGTTGTGC
Rv0891c RT fwd	TCTATGTCGGCCCCACTATCAA
Rv0891c RT rvs	CGGCCAGTCACTAAATCACCTG
Rv0890c RT fwd	TGTCCGGAGCTGACGATCTT
Rv0890c RT rvs	GCAAACAATTCGACTGCTTCG
Rv0891c-Rv0890c junction rvs	GCCAGACGGGTCTTGCCGACACC
Rv0891c T128STOP XbaI rvs	CCCCACCGTGATTCTAGATAGCTTCCGACAC
Rv0891c EcoR1 fwd	CGGAATTCGCTCTTCAACGCAGTTCATAACAGC
Rv0891c90-285 BamHI fwd	TGCGGATCCCGGTGACGGTGACATTGCTCTTAGC
Rv0891c XhoI rvs	CGGTCTCGAGCCAACAGTGCCCGCACCTCAG
Rv0891c d90 EcoRI fwd	TGCGAATTCCGGTGACGGTGACATTGCTCTTAGC
Rv0891c d81 BamHI fwd	TGCGGATCCCGGTGGACGTGAGCAGATTGCCGCC
Rv0890c EcoRI fwd	CTGGAATTCGGGCACTGTTGGCGCAGAACC

Rv0890c XbaI rvs	TGCCTCTAGAATCGTCACGAGGGTGAGCC
Rv0891c-Rv0890c ΔSTOP	CGTGCGCAAATAAGCGAGGTGCGGGCACTGTTGG
Rv0890c -60 StuI fwd	TCTAGGCCTATGTCGATTCTCCCGGCGCAGTTC
T7 rvs	AATACGACTCACTATAGGGCG
Rv0890c R844A StuI fwd	CAAGGACATTGCAAAGGCCTTATTCGTCTCGCCGCGC
Rv0890c R850A NaeI fwd	CTTTTCGTCTCGCCGGCTACTGTGCAAACCCAC
SELEX linker 1(phosphorylated)	GATCCGGAATTCGATTACATTAATTAAG
SELEX linker 2	CTTAATTAATGTAATCGAATTCCG
Rv3133c BamH1fwd (DosR)	GTGGATCCATGGTAAAGGTCTTCTTGG
Rv3133c HindIII rev (DosR)	GGCAAGCTTTTGTCATGGTCCATCACC

To generate a clone to investigate translational coupling, PCR was performed using primers Rv0891cBamHI fwd and Rv0891cT128STOP XbaI rvs. The resulting 400 bp amplicon was digested with BamHI and XbaI. PCR was also performed using primers Rv0891c T128STOP XbaI fwd and Rv0891c XhoI rvs and the 480 bp amplicon was digested with XbaI and XhoI. Digested 400 and 480 bp amplicons were ligated with pBKSII digested with BamHI and XhoI to generate pBKS-Rv0891c T128STOP. The 3.1 kb NheI-HindIII fragment from pPRO-Rv0891c-Rv0890c (see below) were ligated with the 4.3 kb BamHI-HindIII fragment from pMV-sigA-Rv0891c-Rv0890c to generate pMV-sigA-Rv0891c-Rv0890c T128STOP.

Plasmids pMV-sigA-Rv0891c-Rv0890c and pMV-sigA-Rv0891c-Rv0890c T128STOP were electroporated into electrocompetent *M. smegmatis* cells and transformed cells were selected using 50 µg/mL hygromycin on Middlebrook's 7H10 agar plates supplemented with 0.5% glycerol. Individual colonies were grown in 5 mL of Middlebrook's 7H9 broth supplemented with 0.2% glycerol and 0.05% Tween-80 at 37 ˚C till an OD_{600} of 1. Cells were lysed by bead beating (0.5 mm glass beads, BioSpec products) in Mini-Beadbeater-16 (BioSpec products) in 50 mM Tris-Cl pH 8.0 at 4 ˚C, 100 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol, 1 mM benzamidine and 1 mM PMSF and the protein concentration in the lysate was quantitated by Bradford's method. 50 µg of lysate were analysed by western Blotting using anti-Rv0891c and anti-Rv0890c specific antibodies (see below).

Cloning of Rv0891c and Rv0890c The Rv0891c gene was cloned using primers Rv0891c EcoRI fwd and Rv0891c XhoI rvs. PCR was carried out on the genomic DNA of *M. tuberculosis*

H37Rv, and the product was digested with EcoRI and XhoI and cloned into similarly digested pPROEX-HTc to generate pPRO-Rv0891c.

Rv0891c$_{CHD}$ was cloned using primers Rv0891c90-285 BamHI fwd and Rv0891c XhoI rvs. PCR was carried out on pPRO-Rv0891c and the product was digested with BamHI and XhoI and cloned into similarly digested pPROEX-HTc vector to generate pPRO-Rv0891c$_{CHD}$.

Rv0891c$_{\Delta 81}$ (representing a protein with an additional 9 amino acids at the N-terminus) was cloned using primers Rv0891c d81 BamHI fwd and Rv0891c XhoI rvs. PCR was carried out using plasmid pPRO-Rv0891c as template, the product obtained digested with BamHI and XhoI, and cloned into similarly digested pPROEX-HTc vector to generate pPRO-Rv0891c$_{\Delta 81}$.

Rv0890c was cloned using primers Rv0890c EcoRI fwd and Rv0890c XbaI rvs. PCR was carried out on the genomic DNA of *M. tuberculosis* H37Rv, and the product was digested with EcoRI and XbaI and cloned into similarly digested pPROEX-HTb to generate pPRO-Rv0890c. Note that this and all clones containing Rv0890c harbor an Ala at position 866 instead of a Pro residue, as reported in the published H37Rv genome sequence . Inspection of a number of *M. tuberculosis* strains sequenced to date all show an Ala at position 866 (17). It is therefore possible that the strain used for obtaining the published genome sequence harbored a single nucleotide polymorphism at this position that converted the amino acid to a Pro residue at this position, or it represented a sequencing artefact.

Rv0891c-Rv0890c were cloned using Rv0891c EcoRI fwd and Rv0890c XbaI rvs. PCR was carried out on the genomic DNA of *M. tuberculosis* H37Rv, and the product was digested with EcoRI and XbaI and cloned into similarly digested pPROEX-HTc to generate pPRO-Rv0891c-Rv0890c.

Rv0890c$_{+20}$ was generated by first performing deletion mutagenesis (18) using a PstI-BamHI fragment of Rv0891c-Rv0890c cloned in pBKSII as the template and Rv0891c-Rv0890c ΔSTOP as the primer to generate pBKS-Rv0891c-90c junction delta stop. The deletion of T856 would result in incorporation of a Glu residue (Figure. 7A). This plasmid was used as template for PCR amplification with Rv0890c -60 StuI fwd and T7 rvs. The product was digested with BamHI and ligated with SmaI-BamHI digested pBKSII vector to generate pBKS-Rv0890c$_{+20}$ junction. The EcoRI-BamHI fragment from pBKS-Rv0890c$_{+20}$ junction and the BamHI-HindIII fragment from pPRO-Rv0891c-Rv0890c were ligated with EcoRI-HindIII

digested pPROEX-HTa to generate pPRO-Rv0890c$_{+20}$. This would generate a protein which encompasses residues S268-S285 of Rv0891c, followed by a Glu and then entire Rv0890c.

For cloning Rv0891c-Rv0890c$_{FUS}$, Rv0891c was first re-cloned to obtain compatible sites. PCR was performed using pPRO-Rv0891c$_{CHD}$ as template and primers Rv0891c d90 EcoRI fwd and Rv0891c XhoI rvs. The product was digested with EcoRI and XhoI and ligated into similarly digested pBKSII to generate pBKS-Rv0891c$_{CHD}$. Next, the EcoRI-PstI fragment from pBKS-Rv0891c$_{CHD}$ was ligated with similarly digested pBKS-Rv0891c-90c junction delta stop. The EcoRI-BamHI fragment from the resulting plasmid and the BamHI-HindIII fragment from pPRO-Rv0891c-Rv0890c were ligated with pPROEX-HTb vector digested with EcoRI and HindIII, thus generating pPRO-Rv0891c-Rv0890c$_{FUS}$.

The TPR-HTH domain of Rv0890c was cloned using primers Rv0890c TPR EcoRI fwd and Rv0890c XbaI rvs. PCR was performed on genomic DNA of *M. tuberculosis* H37Rv, and the product was digested with EcoRI and XbaI and cloned into similarly digested pBKSII to generate pBKS-Rv0890c$_{TPR-HTH}$. The EcoRI-XbaI fragment from pBKS-Rv0890c$_{TPR-HTH}$ was ligated with similarly digested pPROEX-HTa to obtain pPRO-Rv0890c$_{TPR-HTH}$.

The TPR domain of Rv0890c was PCR amplified using primers Rv0890c TPR EcoRI fwd and Rv0890c TPR XbaI rvs. PCR was performed on pBKS-Rv0890c$_{TPR-HTH,}$ and the product was digested with EcoRI and XbaI and ligated with similarly digested pPROEX-HTa to obtain pPRO-Rv0890c$_{TPR}$.

For cloning Rv0891c-Rv0890c$_{FUS}$ ΔHTH, the EcoRI-AgeI fragment from pPRO-Rv0891c-Rv0890c$_{FUS}$ and the AgeI-HindIII fragment from pPRO-Rv0890c$_{TPR}$ were ligated with pPROEX-HTb vector digested with EcoRI-HindIII to generate pPRO-Rv0891c-Rv0890c$_{FUS}$ ΔHTH.

The region coding for DosR was PCR amplified using primers Rv3133c BamHI fwd and Rv3133c HindIII rvs on *M. bovis* BCG gDNA as template. The amplicon was digested with BamHI and HindIII and ligated with similarly digested pPROEX-HTb to obtain pPRO-DosR.

Site directed mutagenesis (18) was performed on pPRO-Rv0890c$_{TPR-HTH}$ as template using primers Rv0890c R844A StuI fwd and Rv0890c R850A NaeI fwd to generate the respective single mutants.

All clones and mutants were confirmed by sequencing (Macrogen, South Korea).

Generation of antibodies to Rv0891c and Rv0890c and Western blot analysis Recombinant His-tagged Rv0891c$_{CHD}$ and a fragment of Rv0890c (Rv0890c gene was digested with Nco1 followed by relegation; this protein contains regions of the NB-ARC and the TPR-HTH domains) were purified by Ni-NTA affinity chromatography followed by size-exclusion chromatography and used as an immunogen. The primary dose (400 µg protein) was administered to rabbits in Freund's complete adjuvant subcutaneously and booster doses were administered after 14, 28 and 42 days. Sera were collected on day 56.

Protein samples were electrophoresed on an SDS-polyacrylamide gel of appropriate percentage and transferred to PVDF membranes in transfer buffer (25 mM Tris, 192 mM glycine, 20 % methanol, pH 8.3) for 2 hours at constant current of 200 mA. Prior to probing with appropriate antibody, membranes were blocked with 2.5 % blocking agent (RP2109, GE Healthcare) in TBS-T buffer (10 mM Tris-Cl pH 7.5, 0.9% NaCl and 0.1% Tween-20) for 1 hour at room temperature. Blot was probed with primary antibody in TBS-T containing 0.2% BSA (TBT) for 8-12 h at 4 ˚C with gentle agitation. Membranes were washed thrice with TBS-T and incubated with anti-rabbit IgG conjugated to horseradish peroxidase (GE Healthcare; 1:50,000) for 1 hour at room temperature. Membrane was washed thrice with TBS-T and bound antibody was detected using enhanced chemiluminescence with Luminata Crescendo (Millipore).

Expression and purification of proteins Rv0891c$_{CHD}$, Rv0891c$_{\Delta 81}$, Rv0891c-Rv0890c$_{FUS,}$ Rv0891c-Rv0890c$_{FUS}$ΔHTH and Rv0890c$_{+20}$, proteins were expressed in *E. coli* BL21(DE3) pLysS endo^{-} strain on induction with 500 µM IPTG at 16 ˚C for 16 h. Cells were lysed by sonication in lysis buffer (50 mM Tris-Cl pH 8.0, 5 mM 2-mercaptoethanol (2-ME), 100 mM NaCl, 10% glycerol). 1 mM benzamidine, and 2 mM phenylmethylsulfonyl fluoride were added prior to sonication, followed by centrifugation at 30,000 g. The supernatant was interacted with nickel-nitrilotriacetic acid beads (GE healthcare, UK) and bound protein was washed with wash buffer (50 mM Tris-Cl pH 8.0, 5 mM 2-mercaptoethanol (2-ME), 500 mM NaCl, 10% glycerol, 20 mM imidazole). Protein was eluted in lysis buffer supplemented with

300 mM imidazole and dialysed into lysis buffer. For kinetic analysis of Rv0891c$_{CHD}$ the protein was dialysed into buffer with 10 mM NaCl, since this was found to potentiate activity. Aliquots of protein were stored at -70 ˚C.

For purification of Rv0890c$_{TPR-HTH}$, the process was identical except 50 mM NaCl was maintained in all buffers during purification. Under these conditions, the protein had pre-bound RNA. Hence, the purification was altered to include 500 mM NaCl in the wash buffer, while maintaining 50 mM NaCl elsewhere. This abrogated co-purification with RNA. Hence, all other constructs of TPR-HTH domains and the DosR protein were purified in this manner.

Size exclusion chromatography analyses were performed at a flow rate of 0.25 mL/min using a Superose 12 10/300 prepacked column (GE Healthcare, UK) on an AKTA FPLC system (GE Healthcare, UK). The column was equilibrated in buffer (50 mM Tris-Cl pH 8.0 at 4 ˚C, 50 mM NaCl, 5 mM β-mercaptoethanol, containing 10% glycerol) for 3 column volumes prior to loading the protein. The column was calibrated using commercially available gel filtration standards (Bio-Rad Laboratories, USA) containing bovine thyroglobulin (670 kDa), bovine γ-globulin (158 kDa), chicken ovalbumin (44 kDa), horse myoglobin (17 kDa), and Vitamin B_{12} (1.35 kDa).

Adenylyl cyclase assays Adenylyl cyclase assays were performed in final volume of 50 µL. For comparisons between cyclase domains and longer proteins, Rv0891c$_{CHD}$ or Rv0891c$_{\Delta 81}$ (2 µg each) or Rv0891c-Rv0890c$_{FUS}$ (4 µg) were incubated in 50 mM Tris-Cl pH 8.0, 100 mM NaCl, 5 mM 2-ME, 10% glycerol, 1 mM MnATP and 10 mM free Mn^{2+} at 37 ˚C for 30 min. For kinetic analysis of Rv0891c$_{CHD}$ or Rv0891c$_{\Delta 81}$ (5 µg each), increasing concentrations of MnATP were used with the free Mn^{2+} concentration constant at 10 mM, in 50 mM Tris-Cl pH 8.0, 10 mM NaCl, 5 mM 2-ME, 10% glycerol at 37 ˚C for 30 min. All reactions were terminated by the addition of 450 µl of cold sodium acetate buffer (pH 4.75) and boiling of the samples for 10 min. Suitable volumes were used for cAMP estimation after acetylation by radioimmunoassay or ELISA (Cayman Chemical, USA).

ATPase assays For ATPase assays, Rv0890c$_{+20}$ were incubated in 50 mM Tris-Cl pH 8.0, 100 mM NaCl, 5 mM 2-ME, 10% glycerol, 1 mM ATP, 10 mM $MgCl_2$, and ~ 0.2 µCi of [γ-^{32}P]-ATP at 37 ˚C for 30 min in a final volume of 20 µL. Approximately 3-4 µL of the reaction was spotted on PEI-cellulose plates (Merck, Germany) and run for 15 cm in a pre-saturated TLC chamber containing a solvent system of 0.5 M LiCl and 0.5 M formic acid. Plates were allowed to dry and then scanned using a phosphoimager (Typhoon FLA 9500).

Genomic SELEX M. bovis BCG genomic DNA was digested with Sau3AI and ligated to SELEX linkers 1 and 2. For the first attempt of genomic SELEX, 4 µg of purified DNA library was pre-cleared with Ni-NTA resin for 1 h at 4 ˚C in SELEX buffer (50 mM Tris-Cl pH 8.0 at 4 ˚C, 50 mM NaCl, 5 mM 2-ME, 10 % glycerol, 0.05% NP-40, 5 mM $MgCl_2$, 100 µM EDTA, 20 µg/mL poly dI-dC). Pre-cleared DNA library was interacted with 2 µg of His-tagged $Rv0890c_{TPR-HTH}$ in a final volume of 50 µL for 1 h at 4 ˚C and protein-DNA complexes were pulled-down by interaction with Ni-NTA resin. Beads were washed and bound DNA was eluted by boiling in TE (10 mM Tris-Cl pH 8.0, 1 mM EDTA). The eluate was used as template for PCR with SELEX linker 2 and the resulting products were used as input for the next panning. In the second attempt at SELEX, 300 ng of protein was used at each panning, poly dI-dC was eliminated from the buffer, and the library was pre-cleared using $Rv0890c_{TPR}$ before interaction with $Rv0890c_{TPR-HTH}$. Three rounds of panning were performed in each attempt and the final products were cloned into the TA vector (pGEMT) and sequenced. A clone pGEMT-S41 (which contained a fragment from 281342-281835 in *M. bovis* BCG genome) was used in further experiments since it exhibited a single shift with complete depletion of free DNA in EMSAs (see below).

Sequences obtained from SELEX (Supplemental Table 1) were analysed in MEME (19) using all combinations of parameters specifying the frequency of occurrence of motifs in each sequence and the requirement for the motif to be palindromic. As a control, three to five sets of random sequences with identical size distributions were extracted from the *M. tuberculosis* H37Rv genome.

Electrophoretic mobility shift assays (EMSA) 500 ng of the indicated SELEX DNA fragment was prepared using PCR with SELEX linker 2 as primer and the corresponding SELEX fragment cloned into pGEMT vector as template. For radioactive EMSAs, dsDNA was end-labelled using [γ-^{32}P]-ATP with T_4 PNK according to manufacturer's instructions (Thermo Scientific, USA). The labelled DNA was purified from the unincorporated nucleotide by passing it through a Sephadex G-25 column equilibrated with sterile MilliQ water.

For EMSA, 40 pmol of proteins were used except where indicated otherwise. Protein was incubated with 500 ng of DNA (for non-radioactive EMSAs) or 10^5 cpm of [γ-^{32}P]-labelled DNA (for radioactive EMSAs) in EMSA buffer (50 mM Tris-Cl pH 8.0, 50 mM NaCl, 5 mM 2-ME, 10% glycerol, 5 mM $MgCl_2$, 0.05% NP-40, 100 µM EDTA, and 20 µg/mL salmon sperm DNA) for 2 h at 4 ˚C in a total volume of 20 µL. The complexes were resolved on 5%

polyacrylamide gels (29:1 acrylamide:bisacrylamide) containing TAE (40 mM Tris-acetate, 1 mM EDTA, pH 8.2) and 5% glycerol and run in TAE at 4 ˚C for 6-8 h at 50 V. Post-run, non-radioactive gels were stained with EtBr and visualised under 302 nm. Radioactive gels were dried and scanned using a phosphoimager (Typhoon FLA 9500).

The affinity of Rv0890c$_{TPR-HTH}$ was estimated by densitometric scanning of the unbound DNA and subtracting that from the total DNA used in the reaction, to calculate the fraction bound.

For competition with unlabelled nucleic acids S41 was used as dsDNA and total RNA from *M. bovis* BCG as RNA (isolated as described previously (20)). 13.2 µg of each (or indicated amounts of salmon sperm DNA) were pre-incubated with protein in EMSA buffer before addition of [γ-^{32}P]-labelled DNA.

RNA pull-down For RNA pull down experiments, 30 µg of total *M. smegmatis* RNA was pre-cleared on Ni-NTA resin at 4 ˚C for 1 h in 50 mM Tris-Cl pH 8.0, 50 mM NaCl, 5 mM 2-ME, 10 % glycerol. Pre-cleared RNA was incubated with 660 pmol of protein at 4 ˚C for 2 h (final volume 100 µL) followed by interaction with Ni-NTA resin at 4 ˚C for 1 h. Beads and supernatant were separated and beads were washed thrice with the interaction buffer. RNA was extracted by treatment with 2% SDS and 1 unit of proteinase K (Thermo Scientific, USA) at 37 ˚C for 1 h. Extracted RNA was treated with phenol: chloroform: isoamyl alcohol (25:24:1) and precipitated using 0.1 vol of 3 M sodium acetate (pH 5.2) and 2.5 vol of 100% ethanol. 1/20th of the load and breakthrough samples and half of the pull-down samples were analysed by denaturing agarose gel electrophoresis on a 1 % agarose/formaldehyde gel.

Thermal shift assay Proteins were purified in buffer containing 50 mM HEPES pH 8.0 at 4^{o}C, 150 mM NaCl, 10 mM β-mercaptoethanol and 20% glycerol. The thermal shift assays were performed in HEPES buffer (50 mM, pH = 8.0 at 4^{o}C) containing 150 mM NaCl, 20% glycerol and 10 mM $MgCl_2$ in StepOnePlus real-time PCR machine (Applied Biosystems, USA) using 1X SYPRO® Orange dye (Sigma-Aldrich, USA). The samples were heated from 19°C to 95°C with an increase of 1°C per step and a hold of 2 min at initial and final temperature. The data from 37°C to 55°C was taken for analysis. The initial temperature of the block of the PCR machine varied from 25°C to 30°C. This resulted in fluctuations in intensities at initial temperature points. Hence, the data was always acquired from lower temperatures than that used for analysis. Protein concentration used in the assay was 4 µM. ATP or ADP was added

with 10 mM $MgCl_2$ to achieve 5 mM MgATP or MgADP in the specified reaction conditions (calculated using MaxChelator).

Atomic Force Microscopy Fresh DNA (SELEX sequences cloned into pGEMT vector) was isolated from *E. coli* DH10B cells using Unipro plasmid miniprep kit (Dr. KPC Life Sciences Pvt. Ltd.) and eluted in water. The DNA quality was checked by agarose gel electrophoresis and the band representing supercoiled DNA was eluted from the gel and quantitated. The reaction (total volume of 20 µL) was set up in ice with the required concentration of DNA and protein in AFM buffer containing 50 mM Tris-Cl (pH 8 at 4°C), 50 mM NaCl, 10% glycerol, 5 mM $MgCl_2$ and 5 mM β-mercaptoethanol, and incubated at 4°C for 2 h. ATP, ADP or AMP-PNP (5 mM) were added along with 10 mM $MgCl_2$.

Muscovite mica (V1 quality from Electron Microscopy Sciences) was freshly cleaved using Scotch Tape. Post incubation, the total reaction mix was deposited on the freshly cleaved mica and allowed to dry for 30 mins. The mica surface was washed thoroughly with water and left to dry in a desiccator. Samples were then taken for imaging. All reagents used were autoclaved and filtered with a 0.22 µm filter to avoid debris. Imaging of samples on mica substrate was done via the non-contact mode in Park Systems XEI Atomic Force Microscope using 40 N/m ACTA cantilevers from Park Systems. From each sample a minimum of 6 images were taken representing different parts of the mica surface. Images were acquired at 500 nm x 500 nm, 2µm x 2µm or 5µm x 5µm according to the size of the DNA or DNA-protein complex being visualised. Image analysis was done using XEI software from Park Systems. Sizes of protein particles were calculated based on (21).

Statistical analysis Statistical analysis was performed using Student's t-test in GraphPad Prism 5.

RESULTS

The *rv0891c-rv0890c* split-gene pair transcript is expressed as an operon in *M. tuberculosis*

We began by confirming that *rv0891c* and *rv0890c* are expressed in *M. tuberculosis*. Scrutiny of transcription start sites (TSS) (Figure 1A) as identified in a high-throughput analysis (22) indicated the presence of a site that would direct production of an mRNA encoding Rv0891c (marked as 'a' in Figure 1A). Internal TSS's in *rv0891c* ('b' and 'c') were also identified, and

an RNA transcript beginning at 'c' could encode the entire coding sequence of Rv0890c. Since no TSS was identified close to the start site of *rv0890c*, and a number of mRNAs in *M. tuberculosis* are transcribed as leaderless mRNAs (22), it is possible that these two genes could represent an operon and be transcribed as a single RNA. RT-PCR using primers specific for the two genes indicated that these genes were expressed in *M. tuberculosis* grown in different media (Figure 1B), with glucose and propionate representing carbon sources which the bacterium is thought to experience within the host macrophage and granuloma (23, 24). Moreover, a transcript was also detected using primers that would amplify an mRNA produced from an operon encoding *rv0891c-rv0890c* (Figure 1C).

Examination of the genomic locus encoding for Rv0891c revealed that *rv0891c* is separated by a single base pair from the downstream gene, *rv0890c* (Figure. 1D; see also Figure. 8A). Interestingly, *rv0890c* encodes for a protein containing NB-ARC, TPR and HTH domains, making this a 'split gene' pair (*rv0891c-rv0890c*) in comparison with the other STAND ACs, namely Rv0386, Rv1358 and Rv2488c (Figure 1D). Since *rv0890c* and *rv0891c* are expressed as a single mRNA, we asked if they could also be translationally coupled, whereby translation termination and reinitiation occurs at the junction of the two genes (25). We cloned the *rv0891c-rv0890c* operon under the strong *sigA* promoter (Figure 1E, left panel) and monitored expression of the two proteins in *M. smegmatis* which does not harbour homologs of these two genes. As shown in Figure 1E right panel, a band of ~ 25 kDa was detected specifically using antibodies raised to Rv0891c. This is a size smaller than the molecular weight of 31 kDa predicted from the annotated ORF. Ribosome profiling in *M. tuberculosis* (https://doi.org/10.1101/665208 and https://mtb.wadsworth.org/) does suggest that translation could initiate from transcriptss originating from the TSS 'b' (Figure 1A) found downstream of the annotated start codon. Either of the two translated proteins from this transcript could would encode shorter proteins of molecular weight < 31 kDa.

A band of ~ 90 kDa was seen in cell lysates using antibodies to Rv0890c, indicating expression of this protein from the transcript expressed under the SigA promoter The molecular weight predicted of the annotated Rv0890c is 94.5 kDa. Ribosome profiling (26) has identified ribosome binding sites at sequences near the start of the Rv0890c ORF (Figure 1A). Therefore, to monitor the extent of translational coupling in the expression of the two proteins encoded in the operon, we introduced a stop codon following T128 in Rv0891c, ~ 700 bp upstream of the start codon of rv0890c, assuming that this would reduce translational coupling and/or ribosome skipping at the junction of the two genes. While western blot analysis revealed a complete loss of expression of Rv0891c, a band corresponding to Rv0890c could

still be detected. We therefore conclude that translation initiation of Rv0890c within a single operonic mRNA can occur independently of translation of Rv0891c, which is in agreement with ribosome profiling data (Figure 1A).

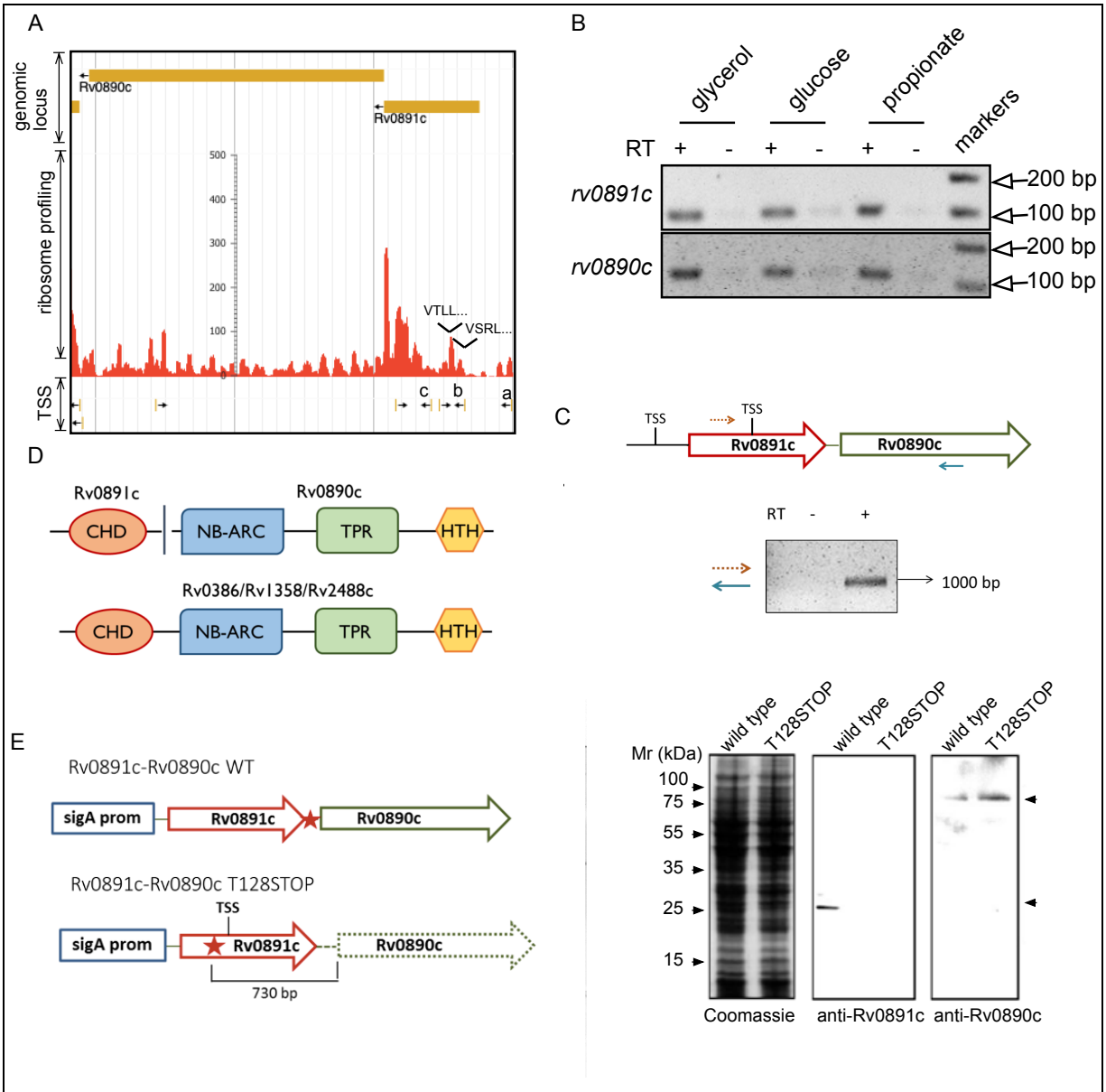

FIGURE 1. A. A snapshot of the genomic region of *rv0891c-rv0890c* (shown as mustard bars) in *M. tuberculosis* indicating the ribosome binding sites detected by ribosome profiling (enrichment shown as red peaks) and transcription start sites (black arrows, direction indicating the direction of transcription). Marked are predicted amino acid sequences that could represent start sites for the protein (VSRL…(Δ81) and/or VTLL..(Δ90) in the vicinity of a ribosome-

enriched sequence of the mRNA. Also annotated are three transcription start sites that are found upstream (a) and within the *rv0891c* coding sequence (b,c), and discussed in the text. The image is obtained from the web server https://mtb.wadsworth.org/. B. PCR products obtained from cDNA prepared from RNA isolated from cultures of *M. tuberculosis* grown in the presence of different carbon sources. Primers used for PCR are contained within the coding sequence of the two individual genes. + and - represent samples of RNA taken for PCR with or without reverse transcription being performed, to monitor the extent of genomic DNA contamination that could generate a PCR product. Data shown is representative of experiments repeated in two independent experiments. C. Primers spanning the two genes, as indicated by arrows in the schematic, were used for PCR with cDNA prepared from RNA obtained from glycerol grown cultures. + and - represent samples of RNA taken for PCR with or without reverse transcription being performed, respectively . A PCR product of ~ 1 kb would be produced from a single mRNA produced by the putative operon. D. Schematic showing the domain organisation of the STAND (signal transduction ATPases with numerous domains) adenylyl cyclases. Vertical line indicates the split gene organisation of Rv0891c and Rv0890c. Domain representations are not to scale. E. Lysates were prepared from *M. smegmatis* cells harbouring plasmids where the *rv0891c-rv0890c* operon (either wild type, or incorporating a stop codon after T128 in Rv0891c (indicated as a red star)) was transcribed under the control of the *sigA* promoter. Lysates were subjected to western blot analysis and probed with antibodies raised to Rv0891c or Rv0890c as indicated. A Coomassie stained gel of lysates taken for western blotting is shown to normalize protein loading across samples. Arrow heads on the right indicate expression of Rv0891c (~ 25 kDa; smaller than the size predicted from the annotated gene) and Rv0890c (~ 90 kDa). Data shown is a representative blot from experiments repeated at least thrice.

The adenylyl cyclase domain of Rv0891c is a 'pseudoenzyme'

A multiple sequence alignment of Rv0891c with the adenylyl cyclase domains (CHD) of known mycobacterial adenylyl cyclases is shown in Figure 2A. The protein has a 90-amino acid long N-terminal extension preceding the adenylyl cyclase domain (residues colored in blue, Figure 2A). This extension did not bear significant similarity to known protein sequences, as inferred from a pBLAST analysis (27). Recombinant expression of Rv0891c as annotated in the genome resulted in its localisation to inclusion bodies in *E. coli* (data not shown).

As mentioned above, the nucleotide sequence encoding this N-terminal extension contains an internal transcription start site (TSS; 'b' in Figure 1A) upstream of two putative start codons (GTG; represented by Valine 82 and Valine 91; red circles in Figure 2A) close to the start of residues that show similarity to the adenylyl cyclase domain of Rv0891c. We expressed truncated proteins lacking either 81 residues from the N-terminus, (Rv0891c$_{\Delta 81}$) or 90 residues (Rv0891c$_{CHD}$, which represents the shortest protein encompassing the cyclase homology domain) in *E. coli* and were able to obtain substantial amounts of both proteins (Figure 2B).

While Rv0891c had conserved residues (indicated by asterisks in Figure 2A) for metal-binding (D97 and D142) and transition-state stabilization (open arrow heads, N193 and R197) seen in nucleotidyl cyclases (28), the protein lacked canonical residues for ATP recognition (depicted as vertical arrows), substituting an R138 and L186 in place of K and D seen usually in adenylyl cyclases. As revealed in the structure of the mycobacterial adenylyl cyclase, Rv1900c (29), non-canonical substrate specifying residues (N342 and D395; Figure 2A) are placed distant from the adenine ring and no base-specific interactions are seen at the active site. This results in the formation of an asymmetric active site, reminiscent of the mammalian adenylyl cyclases (29). We have previously reported that mutations in these ATP-binding residues abolishes dimerization of Rv1625c, another mycobacterial adenylyl cyclase (30, 31), even though Rv1900c exists as a dimer in the absence of canonical substrate-specifying residues (29). Both Rv0891c$_{\Delta 81}$ and Rv0891c$_{CHD}$ exhibited low V_{max} (104 ± 13 μmol cAMP/min/mol protein and 156 μmol ± 40 cAMP/min/mol protein respectively; Figure 2C), compared to ~ 5 mol cAMP/min/mol of Rv1625c (30). Indeed, activity was detectable only because of the very sensitive assays used to measure cAMP (12, 29, 30, 32). There was no significant difference in activity between Rv0891c$_{\Delta 81}$ and Rv0891c$_{CHD}$, which is shorter by 9 amino acids, using different preparations of protein (inset Figure 2C). Adenylyl cyclase activity displayed allosteric behaviour with K' for MnATP of 300 -500 μM) across different protein preparations, and Hill coefficients of ~ 2 (Figure 2B). Since the presence of an additional 9 amino acids at the N-terminus did not increase the Vmax significantly, we proceeded to work with Rv0891c$_{CHD}$, since this would represent the smallest protein that would harbour the catalytic domain.

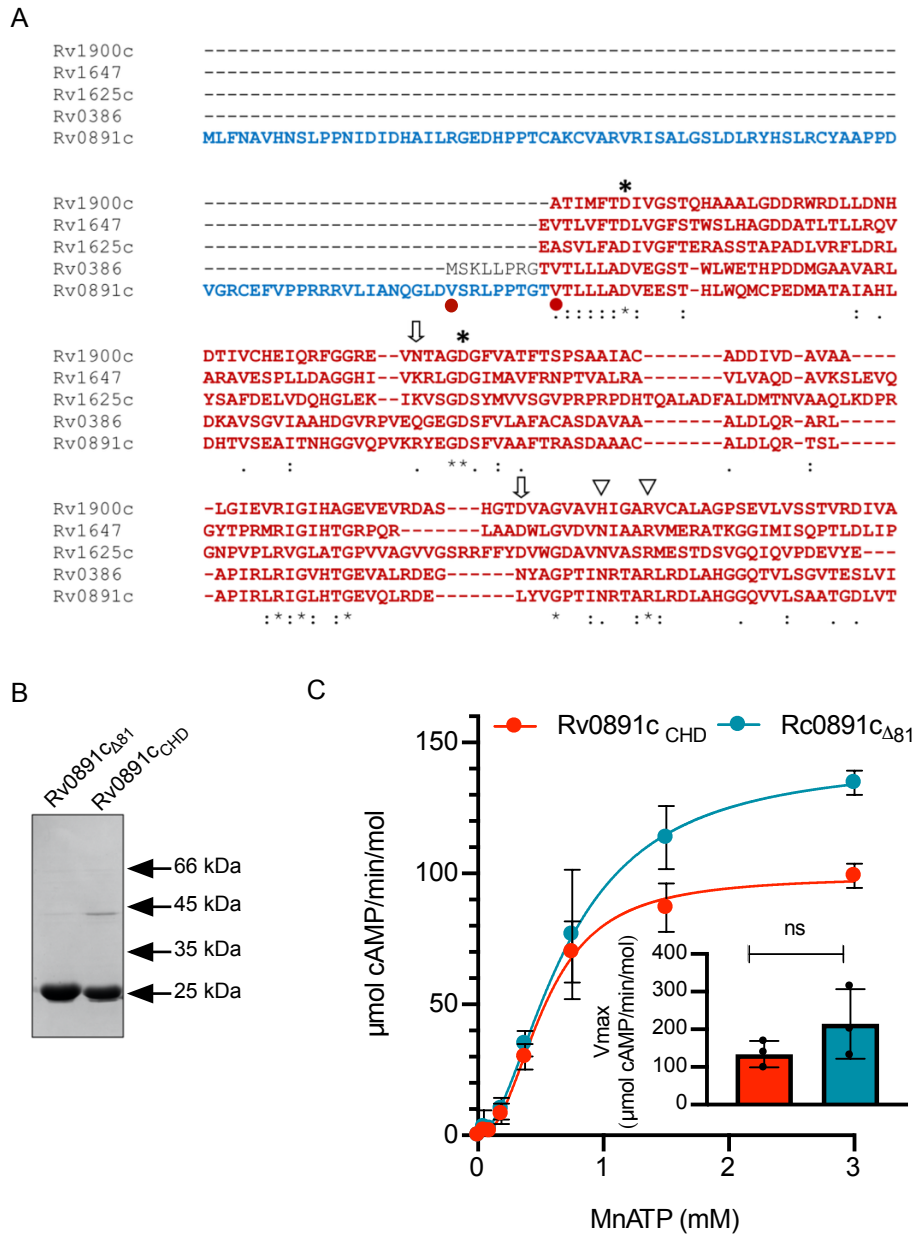

FIGURE 2. A. Multiple sequence alignment of the adenylyl cyclase domains (residues colored in red) of Rv1900c, Rv1647 and Rv1625c with Rv0891c and Rv0386. * denote metal-binding residues, arrows denote substrate specifying residues and inverted triangles denote transition state stabilising residues. Residues colored in blue constitute the N-terminal extension in Rv0891c. Also indicated are the putative start codons represented by Valine 81 and Valine 91 (red circles) B. Purified proteins used for adenylyl cyclase assays. C. Cyclic AMP production by Rv0891c$_{\Delta 81}$ and Rv0891c$_{CHD}$ in the presence of 10 mM Mn^{2+} and increasing concentrations of MnATP. Data shown is the mean ± SD. Inset: Vmax values determined from three different assays using two different protein preparation.

Rv0891c$_{CHD}$ was largely monomeric as determined by gel filtration and unable to convert GTP to cGMP, nor was activity increased in the presence of Mg^{2+} as the metal ion, or in the presence of bicarbonate or detergents (data not shown). Thus, Rv0891c is a poorly active adenylyl cyclase under the conditions tested, possibly as a consequence of its reduced capacity to form stable and catalytically-competent dimers and poorly conserved substrate binding residues.

ATP/ADP binding to the NB-ARC domain of Rv0890c

We next examined the properties of the second member of the split gene pair, Rv0890c, beginning with its N-terminal NB-ARC domain (Figure 1D). The NB-ARC domain of Rv0890c (Figure 3A) lacked a catalytic glutamate in the Walker B (the hhhhDD/E motif), and a conserved histidine in the MHD motif. These residues are essential for ATPase activity, suggesting that the protein may not hydrolyse ATP (33-35). However, the Walker A motif, required for ATP binding was conserved (34)(cite Hanson). Rv0890c was cloned and expressed in *E. coli*, but the protein localised to inclusion bodies (Figure 3B).

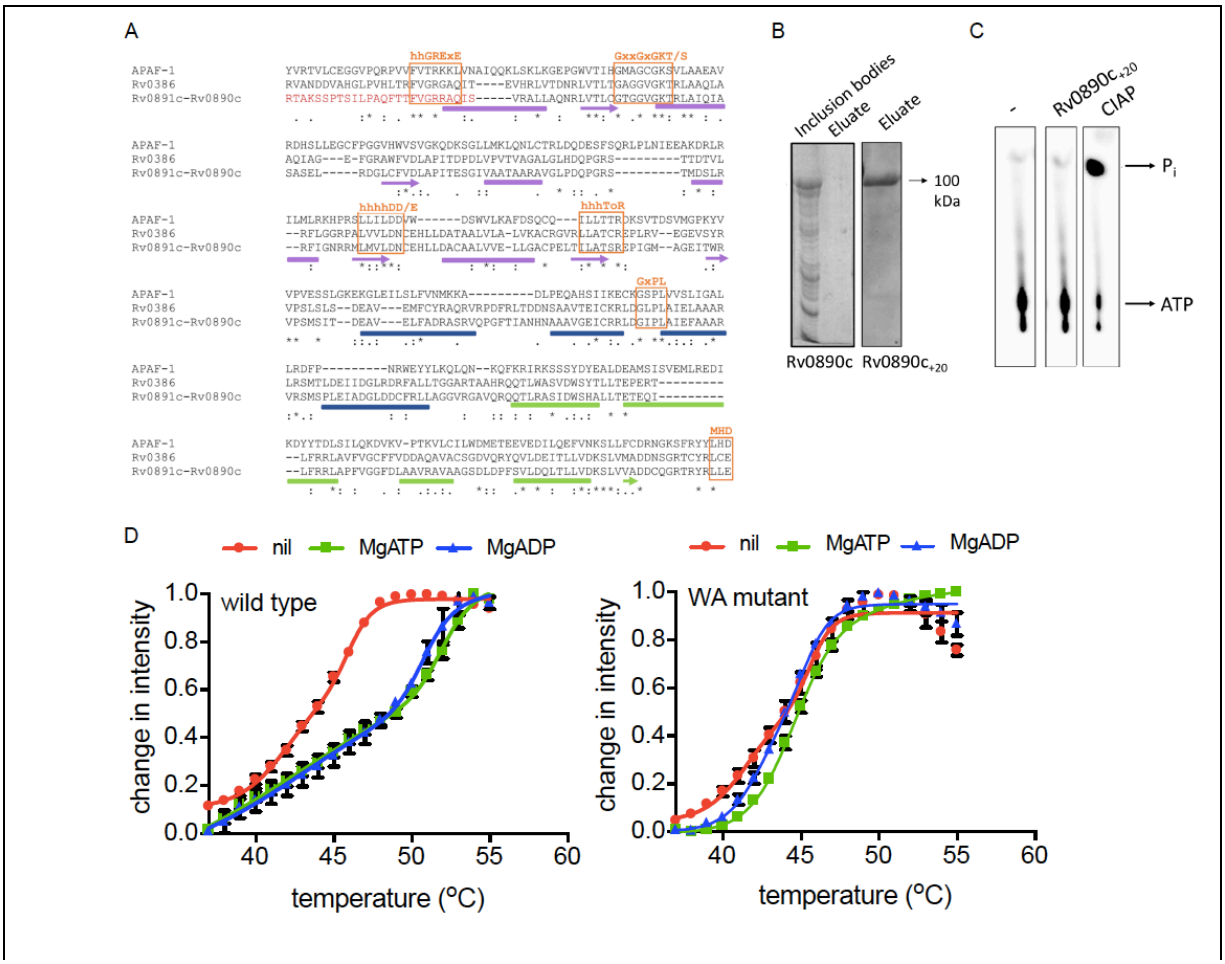

FIGURE 3. A. Multiple sequence alignment of the NB-ARC domains of the indicated proteins with predicted secondary structure for Rv0386. Blocks depict α helices and arrows, β strands. Purple, blue and green colors in the secondary structure elements denote the nucleotide-binding, helical domain 1 and winged-helix subdomains respectively. Colored in red are residues from Rv0891c. Highlighted with orange boxes and text are the conserved motifs (x, any amino acid; h, hydrophobic amino acid; o, amino acid with alcoholic side chain). B. Coomassie stained SDS-polyacrylamide gel showing protein obtained from the purification of Rv0890c *(left)* and Rv0890c$_{+20}$ (*right)*. C, [γ-^{32}P]-ATP was incubated alone (-) or in presence of Rv0890c$_{+20}$ and the resulting products were separated by thin-layer chromatography. Calf intestinal alkaline phosphatase (CIAP) was used to identify the mobility of ATP and $P_{i.}$. D. Thermal shift assay with 4μM Rv0890c$_{+20}$ or Rv0890c$_{+20}$ WA in the absence or presence of 5 mM MgATP/MgADP (n=3). Data are representative of experiments performed using three independent protein preparations.

Based on the interrupted nature of the *rv0891c-rv0890c* gene pair, we hypothesized that a region of Rv0891c may be required to assist folding of Rv0890c to generate a soluble protein that would represent the domain seen in full length homologs like Rv0386. The split between Rv0891c and Rv0890c occurs within a predicted α-helix (predicted using PSIPRED (36); Figure 3A), and the final twenty residues of Rv0891c include the motif hhGRExE, thought to be important for signal propagation across STAND proteins (10). Thus, we cloned Rv0890c with these additional 20 residues of Rv0891c fused at the N-terminus. This protein, Rv0890c$_{+20}$, was expressed to high levels and could be purified efficiently (Figure 3B).

As predicted from the absence of important residues in the Walker B motif, the protein was unable to hydrolyze ATP (Figure 3C). We then asked if this domain could bind ATP. We used a fluorescence-based thermal shift assay, where binding of a hydrophobic fluorescent dye increases as a protein unfolds, and the temperature at which this unfolding occurs could be altered in the presence of small molecule ligands. We incubated Rv0890c$_{+20}$ with MgATP and performed thermal shift assays. The melting profile of Rv0890c$_{+20}$ in the absence of MgATP was broad, with initial higher intensity of SYPRO Orange fluorescence indicating poorly folded protein. In the presence of MgATP the profile was biphasic and a new cooperatively melting species, with higher melting temperature was present. Similar shifts were observed in the presence of MgADP but not in the presence of AMP (Figure 3D; data not shown).

The Walker A motif (GxxGxGKT/S) has a conserved lysine that is involved in binding to the phosphates in nucleotides, and mutation of this lysine residue abrogates nucleotide binding. The melting profile of the Rv0890c$_{+20K80A}$ (WA) mutant protein was not altered in the presence of the MgATP (Figure 3D), indicating that the shift observed in the melting profile of Rv0890c$_{+20}$ in the presence of MgATP resulted from ATP binding to the Walker A motif of the NB-ARC domain.

The HTH domain of Rv0890c binds DNA

STAND proteins are known to engage in intra-molecular interactions involving the NB-ARC domain with downstream regions (10, 11), which in the case of Rv0890c is a C-terminal tetrahelical LuxR type HTH domain (Figure 1D). In order to identify DNA sequences with which the HTH domain could interact with, we purified the Rv0890c$_{TPR-HTH}$ protein (Figure 4A, inset). The protein was found to elute on gel filtration as monomeric and dimeric species, and a construct without the HTH domain ((Rv0890c$_{TPR}$; Figure 4A) showed similar oligomeric states. Using the Rv0890c$_{TPR-HTH}$ protein, we performed a genomic SELEX (systemic evolution of ligands by exponential enrichment) using a library of *Sau*3AI digested fragments from *M. bovis* BCG genomic DNA. We identified 46 independent sequences that bound to the protein which lay in both intergenic as well as intragenic regions of the genome (Supplemental Table 1). These sequences did not share a common sequence motif as analysed by MEME (19) (http://meme-suite.org).

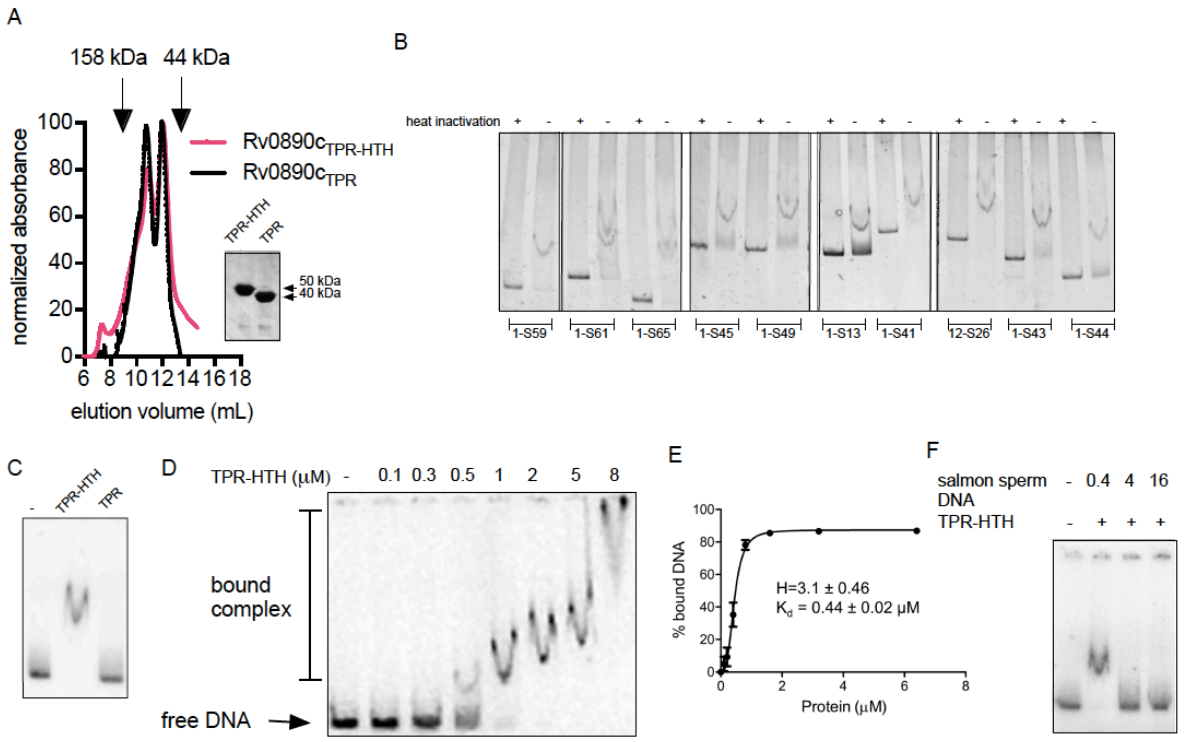

FIGURE 4. A. Size exclusion chromatography of Rv0890c$_{TPR-HTH}$ and Rv0890c$_{TPR}$. Inset shows a Coomassie stained SDS-polyacrylamide gel of the purified proteins. B. EMSAs were performed by incubating dsDNA sequences (500 ng) generated by PCR amplification of clones obtained from genomic SELEX, with native or heat-inactivated Rv0890c$_{TPR-HTH}$ (2 μM total protein) followed by native PAGE and EtBr staining. C. EMSAs with ^{32}P-labelled 1-S41 dsDNA with 2 μM of Rv0890c$_{TPR-HTH}$ or Rv0890c$_{TPR}$. D. EMSA with radiolabelled 1-S41 dsDNA with increasing concentration of Rv0890c$_{TPR-HTH}$ as indicated. E. Affinity of DNA-binding was estimated by densitometric quantitation of free and bound probe. 'H' indicates the Hill coefficient. Data shown is mean ± SD of duplicate determinations from assays performed twice. F. EMSAs with ^{32}P-labelled 1-S41 dsDNA with 2 μM of Rv0890c$_{TPR-HTH}$ and indicated amounts of unlabelled salmon sperm DNA in μg.

We confirmed that Rv0890c$_{TPR-HTH}$ bound to several independent sequences using electrophoretic mobility shift assays (EMSA; Figure 4B). The DNA fragment 1-S41 was found to bind efficiently to Rv0890c$_{TPR-HTH}$ with a clear single shift and complete depletion of free DNA. Hence, 1-S41 was used as the representative SELEX DNA fragment in further experiments. Heat-inactivation of the protein abolished DNA binding, and Rv0890c$_{TPR}$ showed no detectable binding to 1-S41 (Figure 4C). Increasing protein concentrations resulted in the formation of complexes with decreasing mobility (Figure 4D), indicating that multiple

protein molecules could bind to the DNA fragment. Co-operative binding of Rv0890c$_{TPR-HTH}$ for the ~ 550 bp DNA sequence 1-S41, obtained from the genomic SELEX, was observed (Figure 4E). The DNA sequences used in these experiments do not possess any internal repeats and analysis of the sequences obtained in the genomic SELEX showed a lack of sequence similarity. This suggested that the binding of the HTH domain was sequence independent. In agreement with this, incorporation of an excess amount of unlabelled non-specific competitor, salmon sperm DNA, during the EMSA, abolished DNA-binding to Rv0890c$_{TPR-HTH}$ (Figure 4F). Therefore, Rv0890c$_{TPR-HTH}$ may bind to multiple sites on a single DNA molecule in a sequence-independent manner.

The HTH domain present in Rv0890c is similar to DosR, the response regulator in a two-component system involved in the regulation of genes that are induced under conditions of hypoxia and nitric oxide stress, with 46% sequence identity (37). The crystal structure of DosR along with bound DNA (38) has identified residues involved in DNA interaction (Figure 5A). Alignment of the HTH domain of Rv0890c and DosR (Figure 5B) reveals that a Lys residue in DosR (Lys179) is conservatively substituted with an Arg residue (Arg 850), while the other two residues in DosR that interact with DNA (Lys 182 and Asn 183) are replaced by Gln 853 and Thr 854 in Rv0890c. We mutated Arg 850 to an Ala residue in the Rv0890c$_{TPR-HTH}$ construct of Rv0890c, purified the protein and tested its DNA binding ability. As shown in Figure 5C, the Arg850Ala mutant did not bind DNA, while mutation of Arg 844 to Ala (corresponding to conserved Arg 173 in DosR that does not interact with DNA; Figure. 5B, 5A) , continued to interact with DNA. Purified DosR, which shows sequence-specific binding (37), did not bind to the DNA used for EMSA (Figure 5C), since this particular DNA did not contain sequences required for DosR binding.

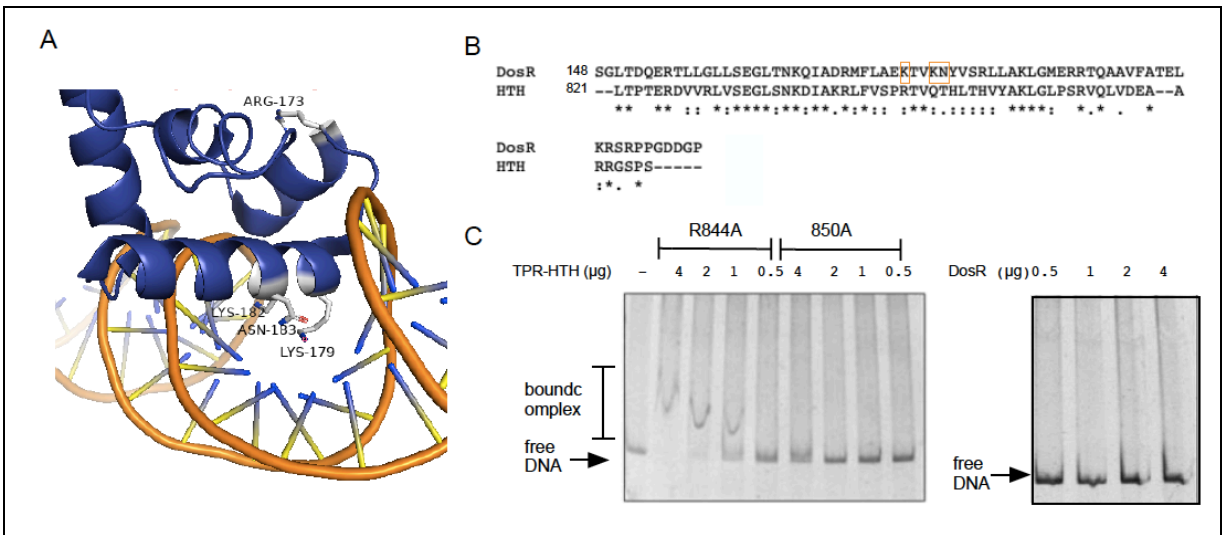

FIGURE 5. A. Structure of DosR (blue) bound to DNA (pdb:1zlK) indicating amino acids discussed in the text that are important for DNA binding (K179, K182 and R183) as well as a residue exposed to solvent (R173). B. Sequence alignment of DosR and the HTH domain of Rv0890c. Residues boxed in orange indicate those important for DNA binding in DosR. C. EMSA performed by incubating 1-S41 dsDNA as a PCR product (500 ng) with varying concentrations of Rv0890c$_{TPR\text{-}HTHR844A}$, Rv0890c$_{TPR\text{-}HTHR850A}$ or DosR. Samples were analysed by native PAGE followed by EtBr staining.

The HTH domain of Rv0890c binds RNA

We hypothesized that the lack of sequence specificity in DNA binding ability of Rv0890c$_{TPR\text{-}HTH}$ may allow binding of the protein to RNA. We competed binding of the Rv0890c$_{TPR\text{-}HTH}$ to DNA using double or single stranded DNA and RNA, and noted that RNA was effectively able to compete for binding to the protein (Figure 6A).

In an alternative approach, we interacted the purified protein with mycobacterial total RNA, followed by immobilization of the protein-RNA complex on Ni-NTA beads. Beads were washed and bound nucleic acid detected by denaturing agarose gel electrophoresis, followed by EtBr staining. As shown in Figure 6B, a significant amount of RNA could bind to the protein, and this binding was again dependent on the HTH domain. Interestingly, DosR failed to bind RNA (Figure 6B). The Arg850Ala mutant TPR-HTH protein did not bind RNA, while the Arg844Ala protein continued to bind RNA (Figure 6B). Therefore, residues critical for interacting with DNA are also essential to allow RNA binding. In summary, the HTH domain of Rv0890c is a nucleic acid binding domain, that, despite its similarity to DosR and other LuxR-like DNA binding motifs, can accommodate RNA. The non-sequence specificity of DNA binding is reminiscent of HU, the nucleoid-associated protein from bacteria, including mycobacteria (39, 40) and Dps from mycobacteria which has been shown to organize the DNA into nucleoid-like structures (41).

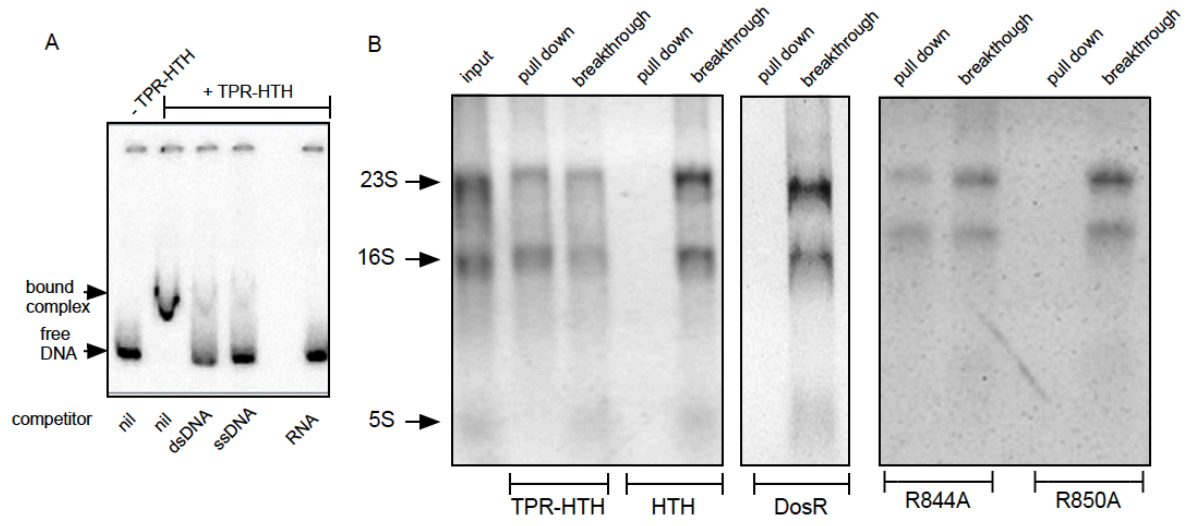

FIGURE 6. A. EMSA with ^{32}P-labelled 1-S41 dsDNA with 2 µM of Rv0890c$_{TPR-HTH}$ in the presence of 13.2 µg of indicated unlabelled nucleic acids. B. 660 pmol of indicated proteins were interacted with 30 µg of total *M. smegmatis* RNA. Bound complexes were pulled down by Ni-NTA agarose resin and bound (pull down) and unbound (breakthrough) samples were subjected to proteinase-K and phenol-chloroform-isoamyl alcohol to remove residual protein. Nucleic acids were precipitated and analyzed by agarose/formaldehyde gel electrophoresis followed by EtBr staining. Data are representative of experiments performed using two independent protein preparations.

AFM identifies DNA-protein complexes that form biocrystalline nucleoids in the presence of adenine nucleotides

We then studied the role of the HTH domain in association with the NB-ARC domain, but found that the complexes formed with Rv0890c$_{+20}$ and sequences obtained by SELEX did not enter acrylamide gels (data not shown) either in the presence or absence of ATP/ADP. We therefore asked if these proteins were oligomeric. We utilized atomic force microscopy (AFM) to investigate oligomeric states of Rv0890c$_{+20}$ in the presence and absence of adenine nucleotides. Rv0890c$_{+20}$ protein molecules alone had a globular shape of several sizes, with the volume spread across a wide range (Figure 7A), suggesting that the protein could exist as oligomers that may facilitate DNA/RNA binding. However, upon addition of ATP and AMP-PNP, a non-hydrolyzable analog of ATP, the volume of the protein molecules became more uniform and smaller. The mean volume of the protein molecules represented hexameric and dimeric forms in the presence of ATP and AMP-PNP, respectively. In the presence of ADP,

particle volumes increased significantly (Figure 7A). These results demonstrate that the oligomeric status of the protein can be regulated by the presence of adenine nucleotides.

The apparent lack of sequence specificity of the HTH domain indicated a role for this class of proteins as being nucleoid-associated proteins (NAP). DNA was prepared from the clone used for EMSA and imaged alone or following addition of protein. Rv0890c$_{+20}$ protein molecules with globular shape of various sizes were found bound to the DNA (Figure 7B). No such structures were seen when only the TPR domain was used, in agreement with the EMSA (data not shown; Figure 4C).

Addition of ATP to the DNA-protein complex brought about dramatic changes and organized, assembled network-like structure of DNA could be seen (Figure 7C). To ensure that the formation of this higher order structure was due to binding of ATP or ADP by the NB-ARC domain, the Rv0890c$_{+20}$ WA mutant was used in the presence of 5 mM MgATP or MgADP. Interestingly, the formation of the higher order organized DNA structure was lost and the DNA-protein complexes showed similar conformation as those in the absence of nucleotides (Figure 7D). Thus, ATP or ADP binding by the NB-ARC domain brings about a conformational change across the protein that causes it to form such higher order structures with DNA.

Since both ATP and ADP binding resulted in the formation of organized higher order structures, we determined whether ATP hydrolysis post DNA-binding is required. AMP-PNP, a non-hydrolyzable analog of ATP**,** was used with Rv0890c$_{+20}$. In the presence of AMP-PNP, the formation of ordered structures can be seen, but the efficiency of packing the DNA into such structures was not as efficient as with ATP (Figure 7D), perhaps as a consequence of different conformational changes induced in the protein in the presence of ATP/ADP or AMP-PNP. Nevertheless, since organized structures were formed in the presence of AMP-PNP, ATP hydrolysis was not necessary to form complex DNA structures.

Finally, to confirm that DNA binding and nucleoid formation was not dependent on the sequence of DNA, we used pGEMT vector alone and observed that crystalloid formation was again seen in the presence of ATP (Figure 7E).

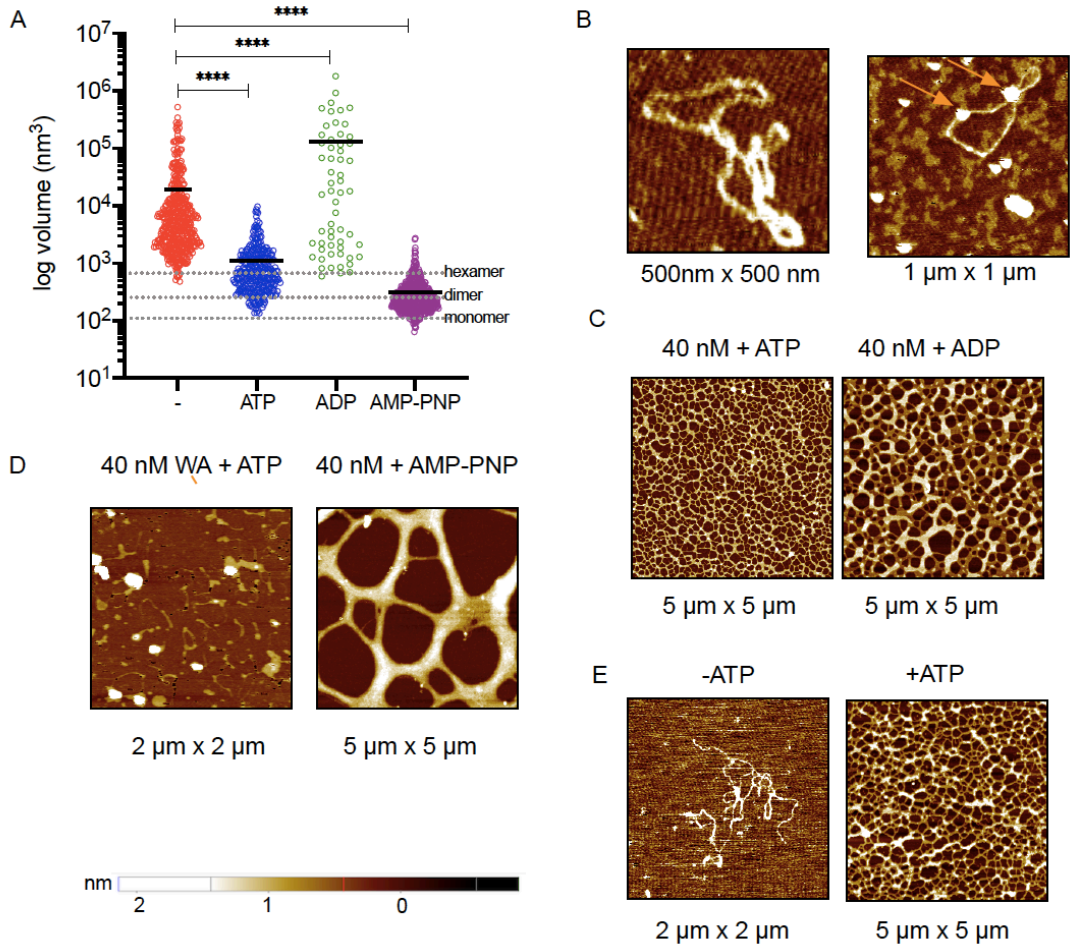

FIGURE 7. A. Volume of Rv0890c$_{+20}$ alone (40 nM in a total volume of 20 µl) or in the presence of 5 mM MgATP, MgADP and MgAMP-PNP as calculated from AFM images. Data points represent individual protein molecules seen with 3 independent protein preparations. Shown are sizes predicted for the monomer, dimer and hexamer, as estimated from (21). B. Supercoiled DNA on mica surface *(left)* and DNA bound to Rv0890c$_{+20}$ *(right)* (n>3). Arrows indicate protein bound to DNA. C. Conformation of DNA-protein complex in presence of 40 nM Rv0890c$_{+20}$ with 40 nM DNA in presence of 5 mM MgATP *(left)* and 5 mM MgADP *(right)* (n>3). D. Conformation of DNA-protein complex in presence of 40 nM Rv0890c$_{+20}$ WA with 40 nM DNA in presence of 5 mM MgATP *(left)* and 5 mM MgAMP-PNP *(right)* (n=3). E. Conformation of DNA-protein complex using pGEMT vector alone (40 nM) and Rv0890c$_{+20}$ (40 nM) in presence and absence of 5 mM MgATP. All images were processed using Park XEI software and data analysed using GraphPad Prism 8. The brightness palette at the bottom of the Figure indicates elevation of the sample from the surface for panels B-E. Data are representative of experiments performed using three independent protein preparations.

Characterization of the Rv0891c-Rv0890c fusion protein

In *M. canetti*, the *rv0891c-rv0890c* genes are fused, and encode a protein with all four domains present in a single polypeptide (Figure 1A, 8A). Upon closer examination of the sequence, we noted that this fusion is a result of the absence of the T residue in the TGA codon of *rv0891c* thereby removing a stop codon, and allowing the remaining base pairs to code for a Glu residue (Figure 8A). Other than this change, the amino acid sequences of the predicted *M. canetti* protein and Rv0891c-Rv0890c are identical. We therefore generated the corresponding deletion in *rv0891c-rv0890c* construct and were able to purify the recombinant fused protein (Rv0891c-Rv0890c$_{FUS}$) (Figure 8B, inset). Note that this fusion protein does not contain the first 90 amino acids of the predicted product from *M. canetti*, where an additional two variations are seen at (V37G and P86L) when compared to Rv0891c-Rv0890c. This fusion protein would serve as a model to understand the features of other STAND-ACs in *M. tuberculosis*, as well as the properties of the protein in *M. canetti*, if it too is produced from an internal TSS.

Rv0891c-Rv0890c$_{FUS}$ showed a marginal increase in adenylyl cyclase activity compared to the isolated cyclase domain (~ 40 μmol/min/mmol protein). The fusion protein showed no ATPase activity (data not shown), but thermal shift assays showed that it bound ATP (Figure. 8B). However, strikingly, the curve showed a sharp denaturation, in contrast to the clearly biphasic curves seen with Rv0890c$_{+20}$ (Figure 3D), indicating that the presence of the cyclase domain allowed the protein to form a single species which denatured at a lower temperature. The Rv0891c-Rv0890c$_{FUS}$ protein bound to DNA, but complexes did not enter gels in an EMSA analysis (data not shown). As seen in AFM analysis, the fusion protein in the presence of ATP compacted the DNA more efficiently, in comparison to structures formed by Rv0890c$_{+20}$ (Figure. 8C). Importantly, RNA binding to the fusion protein was also observed, and this binding was dependent on the presence of the HTH domain (Figure 8D).

DISCUSSION

In this study, we characterize a novel adenylyl cyclase pseudoenzyme and its adjacent gene, representing a 'split gene' that belongs to the family of STAND adenylyl cyclases found in mycobacteria. The data we have presented indicate that the two genes are present in an operon in *M. tuberculosis*. Proteomic evidence indicated that both Rv0891c and Rv0890c are present in *M. tuberculosis* lysates (42), in agreement with observations made from ribosome profiling

(Figure 1A) (26). However, the antibodies we have raised were unable to detect endogenous expression of the proteins reproducibly (data not shown), but only when expressed in *M. smegmatis* under the strong *sigA* promoter (Figure 1D).

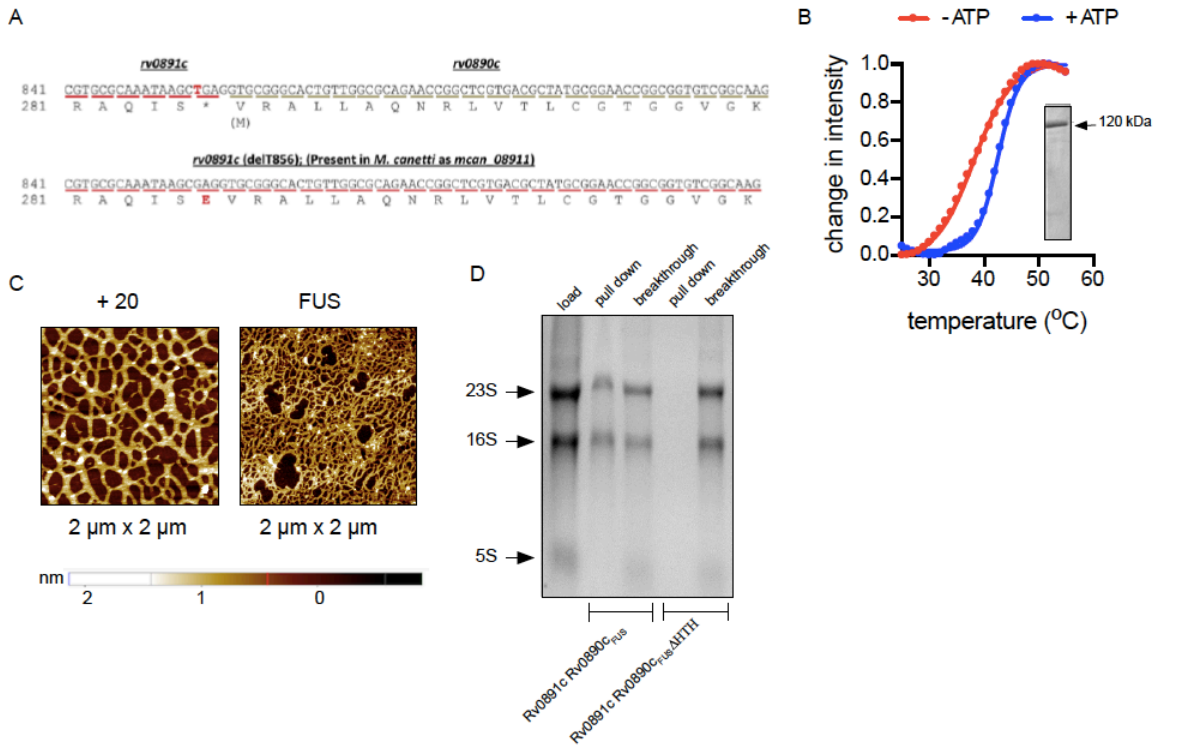

FIGURE 8. A. Sequence alignment of genomic regions in *M. tuberculosis* and *M. canetti* demonstrating the rationale for generation of Rv0891c-Rv0890c$_{FUS}$. B. Thermal shift assay with Rv0891c- Rv0890c$_{FUS}$ (4 µM)in the absence or presence of 5 mM MgATP. Data was analysed using GraphPad Prism 5, and is representative of assays repeated thrice. Inset shows the purified protein used for assays. C. Comparison of DNA-protein complexes formed by Rv0890c$_{+20}$ *(left)* and Rv0891c-Rv0890c$_{FUS}$ *(right)* in presence of 5 mM MgATP with 40 nM of DNA and protein (n=2). All images were processed using Park XEI software. The brightness palette indicates elevation of the sample from the surface. D. Indicated proteins (660 pmol) were interacted with 30 µg of total *M. smegmatis* RNA (load). Reaction mixtures were interacted with Ni-NTA agarose and RNA was extracted from the bound (pull down) and unbound (breakthrough) samples. 1/20th of the input and breakthrough samples and half of the pull down samples were analysed by denaturing agarose gel electrophoresis on a 1 % agarose/formaldehyde gel and stained with EtBr. Data are representative of experiments performed using two independent protein preparations.

The cyclase domain of Rv0891c could be a vestigial remnant of a once active nucleotide cyclase following a gene splitting event. For example, the cyclase domain of Rv0386 has robust catalytic activity (12). Rv0891c has mutations at the catalytic centre, suggesting it is a pseudoenzyme and indicating positive selection for loss-of-function (Figure 2A). As is seen in other pseudoenzymes, we speculate that the cyclase domain present in the *M. cannetti* ortholog for example, may allosterically regulate its associated domains after binding MgATP, as is seen in many pseudokinases (43). Indeed, fusion of these domains in a single protein could allow regions of the two domains to interact with each other robustly. Thermal inactivation assays showed no significant change in stability of $Rv0891c_{CHD}$ domain in the presence of ATP (data not shown). However, presence of the cyclase domain imparted thermal stability when fused to the NB-ARC-TPR-HTH domain in the fusion protein (Figure. 8B). We have been unable to detect robust interactions between Rv0891c and $Rv0890c_{+20}$ *in vitro*, but the products of the 'split gene' may interact with each other in the cell, allowing cyclase domain regulation of nucleic acid binding by Rv0890c.

Rv0890c is an inactive ATPase (Figure 3C), but the Walker A motif present in the protein can facilitate ATP binding (Figure 3D). A catalytically important glutamate residue in ATPases is absent in Rv0890c (Figure 3A) (33), and could account for the absence of ATP-hydrolysis activity. For several STAND proteins, ATP-binding and hydrolysis is contingent on interaction with an inducer, which may be the case in Rv0890c (11). Alternatively, the STAND ACs as a group may function without ATPase activity, as exemplified by CED-4 and Dark (44, 45). In fact, recent reports find that Dark can form apoptosomes even in the absence of ATP-binding, highlighting the functional flexibility in nucleotide-binding and hydrolysis by the NB-ARC module across STAND proteins (46).

Importantly, we show here that the HTH domain of mycobacterial STAND proteins act like NAPs and can organize DNA topology in the presence of adenine nucleotides. There are reports of histones forming such assembled DNA structures with lambda DNA (47) and DosR binds to DNA as a dimer (48). Gel filtration data demonstrated that the TPR domain in Rv0890c could determine oligomerization of the protein to facilitate DNA binding (Figure 4A). Inside a cell, ATP levels can change in response to external and internal stimuli such as starvation, hypoxia etc. The [ATP]/[ADP] ratio can be sensed by Rv0890c, which may then bind the adenine nucleotides and promote the formation of a biocrystallized nucleoid, protecting the organism from imminent stress. Several nucleoid associated proteins in prokaryotes including mycobacteria bring about diverse conformational changes in the nucleoid in response to external and internal stimuli. *M. smegmatis* adopts nucleoid 'toroidal' and 'coral-reef'

structures in the stationary phase of growth. (41). Such 'toroidal' structures are similar to the structures formed by Rv0890c$_{+20}$ and Rv0891c-Rv0890c$_{FUS}$ with DNA as observed via AFM. These structures have been referred to as the "last resort for survival" of a bacterial species under stress and are formed due to a phenomenon called 'biocrystallization' (49). Thus, under conditions when the dynamic order of life cannot be maintained, an equilibrium order is attained by reversible formation of tightly packed and highly ordered structures, within which vital components are protected through physical sequestration.

The sizes of structures formed by adenine-nucleotide bound protein and DNA seen using AFM are larger than the size of a mycobacterial cell, which is ~ 3 microns in length with a diameter of ~ 0.5 - 1 micron (50). However, the sizes we see in the AFM could be a consequence of the concentrations of DNA and protein used for experiments. For example, it is estimated that the concentrations of protein, RNA and DNA of *M. bovis* BCG growing in a low carbon chemostat approximate 72, 10 and 6 fg/cell (51), at which concentrations AFM would be difficult to perform. Alternatively, such large DNA-protein structures could form following lysis of cells in biofilms as has been reported in other bacteria, where DNA-binding proteins are found in the biofilm matrix (52, 53).

Rv0890c binds to both DNA and RNA, a property hitherto unattributed to LuxR-type HTHs. We engineered an artificial construct to characterize the properties of this protein which would represent the domains seen in other STAND adenylyl cyclases in *M. tuberculosis*. Therefore, we posit that in a protein like Rv0386, where deletion of this gene has been shown to attenuate the infection seen in mice (8), the effects of this deletion may not rest solely on the reduction in cAMP levels in the cell, but could also lie in abrogation of critical DNA binding properties of this protein. This possibility in fact has not been addressed to date.

The lack of sequence specificity in nucleic acid recognition by the HTH domain of Rv0890c, and perhaps the HTH domain seen in other STAND proteins, may suggest that these proteins may function as NAPs and influence the global architecture of the DNA. While such proteins typically function by bending, wrapping or bridging DNA, recent reports have also implicated RNA-binding by NAPs as a means of maintaining the nucleoid morphology (54). Non-coding RNAs associate with the nucleoid via the protein HU leading to DNA condensation (55). It is plausible that RNA binding to Rv0890c enhances or modulates interaction with DNA in a similar way. While Rv0890c showed binding to rRNA, rRNA is typically coated with a large number of other proteins in the cell, thereby decreasing its accessibility. Therefore, to access rRNA, Rv0890c may need to co-localize with the ribosome as has been demonstrated for the HTH of an archaeal protein (56). Alternatively, Rv0890c

may be associated with leaderless RNAs or small RNAs produced by promiscuous transcription of the mycobacterial genome (22, 57, 58). Given the complex domain architecture of this group of STAND adenylyl cyclases, and their unique properties and presence almost exclusively in slow growing-mycobacterial species, it is anticipated that their roles in mycobacteria are complex and identifiable only under certain environmental conditions.

SUPPLEMENTARY DATA

Supplementary Data are available online.

Author Contributions: Anisha Zaveri: conceptualization, investigation, original draft preparation; Avipsa Bose: conceptualization, investigation, original draft preparation; Suruchi Sharma: investigation ; Abinaya Rajendran: investigation; Priyanka Biswas: investigation; Avinash R Shenoy: conceptualization, investigation; Sandhya S. Visweswariah: conceptualization, supervision, writing-reviewing and editing, funding acquisition.

ACKNOWLEDGMENTS

We thank Monisha M. for help with atomic force microscopy experiments in the Centre for Biosystems Science and Engineering in the Indian Institute of Science.

FUNDING

Support from the Department of Biotechnology, Government of India is acknowledged (BT/PR15216/COE/34/02/2017) and funding from DBT-IISc Partnership Program Phase-II (BT/PR27952/INF/22/212/2018/21.01.2019). SSV is a JC Bose National Fellow (SB/S2/JCB-18/2013) and a Margdarshi Fellow supported by the Wellcome Trust DBT India Alliance.

CONFLICT OF INTEREST

Authors declare no conflict of interest.

References

1. Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.

A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell. 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393(6685):537-544.
2. Shenoy, A. R., and S. S. Visweswariah. 2006. New messages from old messengers: cAMP and mycobacteria. Trends Microbiol 14(12):543-550.
3. Zaveri, A., and S. S. Visweswariah. 2013. Cyclic AMP in Mycobacteria: the second messenger comes first. Curr Sci 105(5):666-675.
4. McDonough, K. A., and A. Rodriguez. 2012. The myriad roles of cyclic AMP in microbial pathogens: from signal to sword. Nat Rev Microbiol 10(1):27-38.
5. Shenoy, A. R., and S. S. Visweswariah. 2006. Mycobacterial adenylyl cyclases: biochemical diversity and structural plasticity. FEBS Lett 580(14):3344-3352.
6. Dittrich, D., C. Keller, S. Ehlers, J. E. Schultz, and P. Sander. 2006. Characterization of a Mycobacterium tuberculosis mutant deficient in pH-sensing adenylate cyclase Rv1264. Int J Med Microbiol 296(8):563-566.
7. Guo, Y. L., U. Kurz, A. Schultz, J. U. Linder, D. Dittrich, C. Keller, S. Ehlers, P. Sander, and J. E. Schultz. 2005. Interaction of Rv1625c, a mycobacterial class IIIa adenylyl cyclase, with a mammalian congener. Mol Microbiol 57(3):667-677.
8. Agarwal, N., G. Lamichhane, R. Gupta, S. Nolan, and W. R. Bishai. 2009. Cyclic AMP intoxication of macrophages by a Mycobacterium tuberculosis adenylate cyclase. Nature 460(7251):98-102.
9. Shenoy, A. R., K. Sivakumar, A. Krupa, N. Srinivasan, and S. S. Visweswariah. 2004. A survey of nucleotide cyclases in actinobacteria: unique domain organization and expansion of the class III cyclase family in Mycobacterium tuberculosis. Comp Funct Genomics 5(1):17-38.
10. Leipe, D. D., E. V. Koonin, and L. Aravind. 2004. STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. J Mol Biol 343(1):1-28.
11. Danot, O., E. Marquenet, D. Vidal-Ingigliardi, and E. Richet. 2009. Wheel of Life, Wheel of Death: A Mechanistic Insight into Signaling by STAND Proteins. Structure 17(2):172-182.
12. Castro, L. I., C. Hermsen, J. E. Schultz, and J. U. Linder. 2005. Adenylyl cyclase Rv0386 from Mycobacterium tuberculosis H37Rv uses a novel mode for substrate selection. FEBS J 272(12):3085-3092.
13. Finn, R. D., P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. A. Salazar, J. Tate, and A. Bateman. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44(D1):D279-285.
14. Marchler-Bauer, A., M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng, and S. H. Bryant. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res 43(Database issue):D222-226.
15. Mitchell, A., H. Y. Chang, L. Daugherty, M. Fraser, S. Hunter, R. Lopez, C. McAnulla, C. McMenamin, G. Nuka, S. Pesseat, A. Sangrador-Vegas, M. Scheremetjew, C. Rato, S. Y. Yong, A. Bateman, M. Punta, T. K. Attwood, C. J. Sigrist, N. Redaschi, C. Rivoire, I. Xenarios, D. Kahn, D. Guyot, P. Bork, I. Letunic, J. Gough, M. Oates, D. Haft, H. Huang, D. A. Natale, C. H. Wu, C. Orengo, I. Sillitoe, H. Mi, P. D. Thomas, and R. D. Finn. 2015. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res 43(Database issue):D213-221.
16. Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539.
17. Ioerger, T. R., Y. Feng, K. Ganesula, X. Chen, K. M. Dobos, S. Fortune, W. R. Jacobs, Jr., V. Mizrahi, T. Parish, E. Rubin, C. Sassetti, and J. C. Sacchettini. 2010. Variation among genome sequences of H37Rv strains of Mycobacterium tuberculosis from multiple laboratories. J Bacteriol 192(14):3645-3653.

18. Shenoy, A. R., and S. S. Visweswariah. 2003. Site-directed mutagenesis using a single mutagenic oligonucleotide and DpnI digestion of template DNA. Anal Biochem 319(2):335-336.
19. Bailey, T. L., J. Johnson, C. E. Grant, and W. S. Noble. 2015. The MEME Suite. Nucleic Acids Res 43(W1):W39-49.
20. Nambi, S., K. Gupta, M. Bhattacharyya, P. Ramakrishnan, V. Ravikumar, N. Siddiqui, A. T. Thomas, and S. S. Visweswariah. 2013. Cyclic AMP-dependent protein lysine acylation in mycobacteria regulates fatty acid and propionate metabolism. J Biol Chem 288(20):14114-14124.
21. Li-Y-L., M., Y-F., Zhang, Z-M., Jiang, Y. 2014. Detecting the Oligomeric State of Escherichia coli MutS from Its Geometric Architecture Observed by an Atomic Force Microscope at a Single Molecular Level. J. Phys. Chem. B 118(31):9218-9224.
22. Cortes, T., O. T. Schubert, G. Rose, K. B. Arnvig, I. Comas, R. Aebersold, and D. B. Young. 2013. Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in Mycobacterium tuberculosis. Cell Rep 5(4):1121-1131.
23. Huang, L., E. V. Nazarova, and D. G. Russell. 2019. Mycobacterium tuberculosis: Bacterial Fitness within the Host Macrophage. Microbiol Spectr 7(2).
24. Ehlers, S., and U. E. Schaible. 2012. The granuloma in tuberculosis: dynamics of a host-pathogen collusion. Front Immunol 3:411.
25. Huber, M., G. Faure, S. Laass, E. Kolbe, K. Seitz, C. Wehrheim, Y. I. Wolf, E. V. Koonin, and J. Soppa. 2019. Translational coupling via termination-reinitiation in archaea and bacteria. Nat Commun 10(1):4006.
26. Smith, C., Canestrari, J.G., Wang, J, Derbyshire, K.M. Gray, T.A., Wade, J.T. 2019. Pervasive translation in *Mycobacterium tuberculosis*. bioRxiv.
27. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J Mol Biol 215(3):403-410.
28. Linder, J. U., and J. E. Schultz. 2003. The class III adenylyl cyclases: multi-purpose signalling modules. Cell Signal 15(12):1081-1089.
29. Sinha, S. C., M. Wetterer, S. R. Sprang, J. E. Schultz, and J. U. Linder. 2005. Origin of asymmetry in adenylyl cyclases: structures of Mycobacterium tuberculosis Rv1900c. EMBO J 24(4):663-673.
30. Shenoy, A. R., N. Srinivasan, M. Subramaniam, and S. S. Visweswariah. 2003. Mutational analysis of the Mycobacterium tuberculosis Rv1625c adenylyl cyclase: residues that confer nucleotide specificity contribute to dimerization. FEBS Lett 545(2-3):253-259.
31. Ketkar, A. D., A. R. Shenoy, U. A. Ramagopal, S. S. Visweswariah, and K. Suguna. 2006. A structural basis for the role of nucleotide specifying residues in regulating the oligomerization of the Rv1625c adenylyl cyclase from M. tuberculosis. J Mol Biol 356(4):904-916.
32. Shenoy, A. R., N. P. Sreenath, M. Mahalingam, and S. S. Visweswariah. 2005. Characterization of phylogenetically distant members of the adenylate cyclase family from mycobacteria: Rv1647 from Mycobacterium tuberculosis and its orthologue ML1399 from M. leprae. Biochem J 387(Pt 2):541-551.
33. Takken, F. L., M. Albrecht, and W. I. Tameling. 2006. Resistance proteins: molecular switches of plant defence. Curr Opin Plant Biol 9(4):383-390.
34. Hanson, P. I., and S. W. Whiteheart. 2005. AAA+ proteins: have engine, will work. Nat Rev Mol Cell Biol 6(7):519-529.
35. Riedl, S. J., W. Li, Y. Chao, R. Schwarzenbacher, and Y. Shi. 2005. Structure of the apoptotic protease-activating factor 1 bound to ADP. Nature 434(7035):926-933.
36. Buchan, D. W., F. Minneci, T. C. Nugent, K. Bryson, and D. T. Jones. 2013. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res 41(Web Server issue):W349-357.
37. Park, H. D., K. M. Guinn, M. I. Harrell, R. Liao, M. I. Voskuil, M. Tompa, G. K. Schoolnik, and D. R. Sherman. 2003. Rv3133c/dosR is a transcription factor that mediates the hypoxic response of Mycobacterium tuberculosis. Mol Microbiol 48(3):833-843.

38. Wisedchaisri, G., M. Wu, A. E. Rice, D. M. Roberts, D. R. Sherman, and W. G. Hol. 2005. Structures of Mycobacterium tuberculosis DosR and DosR-DNA complex involved in gene activation during adaptation to hypoxic latency. J Mol Biol 354(3):630-641.
39. Dey, D., V. Nagaraja, and S. Ramakumar. 2017. Structural and evolutionary analyses reveal determinants of DNA binding specificities of nucleoid-associated proteins HU and IHF. Mol Phylogenet Evol 107:356-366.
40. Stojkova, P., P. Spidlova, and J. Stulik. 2019. Nucleoid-Associated Protein HU: A Lilliputian in Gene Regulation of Bacterial Virulence. Front Cell Infect Microbiol 9:159.
41. Ghatak, P., K. Karmakar, S. Kasetty, and D. Chatterji. 2011. Unveiling the role of Dps in the organization of mycobacterial nucleoid. PLoS One 6(1):e16019.
42. Schubert, O. T., C. Ludwig, M. Kogadeeva, M. Zimmermann, G. Rosenberger, M. Gengenbacher, L. C. Gillet, B. C. Collins, H. L. Rost, S. H. Kaufmann, U. Sauer, and R. Aebersold. 2015. Absolute Proteome Composition and Dynamics during Dormancy and Resuscitation of Mycobacterium tuberculosis. Cell Host Microbe 18(1):96-108.
43. Ribeiro, A. J. M., S. Das, N. Dawson, R. Zaru, S. Orchard, J. M. Thornton, C. Orengo, E. Zeqiraj, J. M. Murphy, and P. A. Eyers. 2019. Emerging concepts in pseudoenzyme classification, evolution, and signaling. Sci Signal 12(594).
44. Rodriguez, A., H. Oliver, H. Zou, P. Chen, X. Wang, and J. M. Abrams. 1999. Dark is a Drosophila homologue of Apaf-1/CED-4 and functions in an evolutionarily conserved death pathway. Nat Cell Biol 1(5):272-279.
45. Yan, N., J. Chai, E. S. Lee, L. Gu, Q. Liu, J. He, J. W. Wu, D. Kokel, H. Li, Q. Hao, D. Xue, and Y. Shi. 2005. Structure of the CED-4-CED-9 complex provides insights into programmed cell death in Caenorhabditis elegans. Nature 437(7060):831-837.
46. Pang, Y., X. C. Bai, C. Yan, Q. Hao, Z. Chen, J. W. Wang, S. H. Scheres, and Y. Shi. 2015. Structure of the apoptosome: mechanistic insights into activation of an initiator caspase from Drosophila. Genes Dev 29(3):277-287.
47. Liu, Y., Guthold, M., Snyder, M., Lu, H. . 2015. AFM of self-assembled lambda DNA-histone networks. Colloids Surf B Biointerfaces 134:17-25.
48. Wisedchaisri, G., M. Wu, D. R. Sherman, and W. G. Hol. 2008. Crystal structures of the response regulator DosR from Mycobacterium tuberculosis suggest a helix rearrangement mechanism for phosphorylation activation. J Mol Biol 378(1):227-242.
49. Minsky, A., E. Shimoni, and D. Frenkiel-Krispin. 2002. Stress, order and survival. Nat Rev Mol Cell Biol 3(1):50-60.
50. Cook, G. M., M. Berney, S. Gebhard, M. Heinemann, R. A. Cox, O. Danilchanka, and M. Niederweis. 2009. Physiology of mycobacteria. Adv Microb Physiol 55:81-182, 318-189.
51. Beste, D. J., J. Peters, T. Hooper, C. Avignone-Rossa, M. E. Bushell, and J. McFadden. 2005. Compiling a molecular inventory for Mycobacterium bovis BCG at two growth rates: evidence for growth rate-mediated regulation of ribosome biosynthesis and lipid metabolism. J Bacteriol 187(5):1677-1684.
52. Kavanaugh, J. S., C. E. Flack, J. Lister, E. B. Ricker, C. B. Ibberson, C. Jenul, D. E. Moormeier, E. A. Delmain, K. W. Bayles, and A. R. Horswill. 2019. Identification of Extracellular DNA-Binding Proteins in the Biofilm Matrix. mBio 10(3).
53. Devaraj, A., J. R. Buzzo, L. Mashburn-Warren, E. S. Gloag, L. A. Novotny, P. Stoodley, L. O. Bakaletz, and S. D. Goodman. 2019. The extracellular DNA lattice of bacterial biofilms is structurally related to Holliday junction recombination intermediates. Proc Natl Acad Sci U S A 116(50):25068-25077.
54. Dorman, C. J., and P. Deighan. 2003. Regulation of gene expression by histone-like proteins in bacteria. Curr Opin Genet Dev 13(2):179-184.
55. Qian, Z., M. Macvanin, E. K. Dimitriadis, X. He, V. Zhurkin, and S. Adhya. 2015. A New Noncoding RNA Arranges Bacterial Chromosome Organization. mBio 6(4).
56. Blombach, F., H. Launay, A. P. Snijders, V. Zorraquino, H. Wu, B. de Koning, S. J. Brouns, T. J. Ettema, C. Camilloni, A. Cavalli, M. Vendruscolo, M. J. Dickman, L. D. Cabrita, A. La Teana, D. Benelli, P. Londei, J. Christodoulou, and J. van der Oost. 2014. Archaeal MBF1 binds to 30S and 70S ribosomes via its helix-turn-helix domain. Biochem J 462(2):373-384.

57. Shell, S. S., J. Wang, P. Lapierre, M. Mir, M. R. Chase, M. M. Pyle, R. Gawande, R. Ahmad, D. A. Sarracino, T. R. Ioerger, S. M. Fortune, K. M. Derbyshire, J. T. Wade, and T. A. Gray. 2015. Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape. PLoS Genet 11(11):e1005641.
58. Arnvig, K. B., and D. B. Young. 2009. Identification of small RNAs in Mycobacterium tuberculosis. Mol Microbiol 73(3):397-408.

[1]Abbreviations: AC, adenylyl cyclase; AFM, atomic force microscopy; EMSA, electrophoretic mobility shift assays; HTH, helix-turn-helix; Mtb, *Mycobacterium tuberculosis*; NB-ARC, nucleotide-binding adaptor shared by apoptotic protease-activating factor-1, certain R gene products and cell death protein 4; SELEX, systemic evolution of ligands by exponential enrichment ; STAND, signal transduction ATPases with numerous domains; TPR, tetratricopeptide repeat; TSS, transcription start site