

# SCIENTIFIC REPORTS



OPEN

## *Mycobacterium tuberculosis* Rv0366c-Rv0367c encodes a non-canonical PezAT-like toxin-antitoxin pair

Himani Tandon<sup>1</sup>, Arun Sharma<sup>2</sup>, Sankaran Sandhya<sup>1</sup>, Narayanaswamy Srinivasan<sup>1</sup> & Ramandeep Singh<sup>2</sup>

Toxin-antitoxin (TA) systems are ubiquitously existing addiction modules with essential roles in bacterial persistence and virulence. The genome of *Mycobacterium tuberculosis* encodes approximately 79 TA systems. Through computational and experimental investigations, we report for the first time that Rv0366c-Rv0367c is a non-canonical PezAT-like toxin-antitoxin system in *M. tuberculosis*. Homology searches with known PezT homologues revealed that residues implicated in nucleotide, antitoxin-binding and catalysis are conserved in Rv0366c. Unlike canonical PezA antitoxins, the N-terminal of Rv0367c is predicted to adopt the ribbon-helix-helix (RHH) motif for deoxyribonucleic acid (DNA) recognition. Further, the modelled complex predicts that the interactions between PezT and PezA involve conserved residues. We performed a large-scale search in sequences encoded in 101 mycobacterial and 4500 prokaryotic genomes and show that such an atypical PezAT organization is conserved in 20 other mycobacterial organisms and in families of class Actinobacteria. We also demonstrate that overexpression of Rv0366c induces bacteriostasis and this growth defect could be restored upon co-expression of cognate antitoxin, Rv0367c. Further, we also observed that inducible expression of Rv0366c in *Mycobacterium smegmatis* results in decreased cell-length and enhanced tolerance against a front-line tuberculosis (TB) drug, ethambutol. Taken together, we have identified and functionally characterized a novel non-canonical TA system from *M. tuberculosis*.

Bacterial toxin-antitoxin (TA) systems are plasmid or chromosome-encoded, mobile genetic elements expressed as part of the same operon<sup>1-3</sup>. Together with other mechanisms of stress tolerance like the SOS response or the second messenger guanosine pentaphosphate, TA systems have major contribution in bacterial persistence: a phenomenon where bacterial cells can transform into persister cells when under stress<sup>4</sup>. Under normal environmental conditions, antitoxins neutralize toxin activity by forming a tight TA complex. Under stress, antitoxins are proteolytically cleaved and the released toxin functions as an intracellular poison that acts on its cellular targets, which range from specific messenger ribonucleic acid (mRNA), transfer ribonucleic acid (tRNA), ribosomal ribonucleic acid (rRNA) to proteins like deoxyribonucleic acid (DNA) gyrase or elongation factor thermo-unstable (EF-Tu)<sup>5-9</sup>. In general, the toxins are always proteins whereas antitoxins can either be proteins or ribonucleic acids (RNAs)<sup>8</sup>.

Depending upon the nature of antitoxins, these systems can be categorised into six types from I to VI. Type I TA systems have antisense RNAs as antitoxins, which bind to the toxin mRNA whereas in type II TA systems, antitoxin is a protein moiety which directly interacts with the toxin<sup>10,11</sup>. Type III TA systems are composed of an RNA antitoxin which binds to its cognate toxin and suppresses its activity<sup>12</sup>. Type IV TA systems also possess a protein antitoxin but unlike type II, toxin and antitoxin bind to a common target<sup>13,14</sup>. Antitoxins in type V TA systems are endoribonucleases, which specifically cleave toxin mRNA<sup>15</sup>. In type VI TA systems, the toxin and antitoxin are both proteins. Here, the antitoxin does not neutralize the toxin directly on binding but promotes its degradation

<sup>1</sup>Molecular Biophysics Unit, Indian Institute of Science, Bangalore, 560012, India. <sup>2</sup>Tuberculosis Research Laboratory, Vaccine and Infectious Disease Research Centre, Translational Health Science and Technology Institute, NCR Biotech Science Cluster, 3rd Milestone, PO Box #4, Faridabad, Haryana, 121001, India. Himani Tandon and Arun Sharma contributed equally. Correspondence and requests for materials should be addressed to N.S. (email: [ns@iisc.ac.in](mailto:ns@iisc.ac.in)) or R.S. (email: [ramandeep@thsti.res.in](mailto:ramandeep@thsti.res.in))

by specific proteases<sup>9,16</sup>. Among these, type II TA systems are the most well characterized families of TA systems. Here usually, the antitoxins have two domains: an N-terminal DNA-binding domain which functions to regulate the TA operon expression and a C-terminal toxin-binding domain which inhibits the toxin from binding to its cellular targets<sup>17</sup>. In some systems, the DNA-binding function is carried out by a third protein in the operon<sup>18,19</sup>.

The TA systems are ubiquitously present in Gram-positive and Gram-negative bacteria, as well as in archaeobacteria. Several studies have been carried out previously, to explore the distribution of these systems across bacterial and archaeobacterial genomes<sup>11,20,21</sup>. Well-known examples of type II systems include RelBE, ParDE, HigBA, VapBC, MazEF, CcdBA, Phd-Doc and  $\epsilon/\zeta/\omega$ <sup>1</sup>. Out of these, the VapBC, RelBE, HigBA, MazEF and ParDE systems are the most abundant. The genus *Mycobacterium* is interesting because it includes not only the pathogenic *Mycobacterium tuberculosis* complex (MTBC) members but also non-pathogenic bacteria such as *Mycobacterium smegmatis*. The reported number of TA loci varies tremendously among these species e.g. *M. tuberculosis* harbours 79 systems, *M. smegmatis* has 3 such systems and *Mycobacterium leprae* has no known report of any TA system<sup>21,22</sup>. Within mycobacteria, *M. tuberculosis* possesses representatives from all major classes of type II TAs in single or multiple copies viz. 50 VapBC, 10 MazEF, 3 HigBA, 2 ParDE, 3 RelBE and few other unclassified novel systems<sup>1,23</sup>. But thus far, there have been no reports on any homologues of either CcdBA,  $\epsilon/\zeta/\omega$ , PezAT and Phd/Doc systems in *M. tuberculosis*.

The  $\epsilon/\zeta/\omega$  system was first reported on the plasmid pSM19035 of *Streptococcus pyogenes*<sup>24</sup>. Since then there have been numerous reports showing the presence of this system in various Gram-positive and Gram-negative bacteria<sup>11</sup>. The chromosomal homologue of  $\epsilon/\zeta/\omega$ , known as PezAT, was first reported in *Streptococcus pneumoniae*<sup>25</sup>. These two homologous TA systems show differences in their transcription regulation. An accessory protein called  $\omega$  is required for the transcriptional regulation of  $\epsilon/\zeta/\omega$  system whereas this function is carried out by the N-terminal region of PezA antitoxin, characteristic of type II TA systems<sup>26</sup>. Further, whereas the  $\omega$  protein adopts the RHH motif<sup>27</sup>, the N-terminal of PezA adopts the helix-turn-helix (HTH) motif<sup>25</sup>. The C-terminal of PezA antitoxin shows a sequence identity of 21% with the  $\epsilon$  antitoxin<sup>25,28</sup>. The toxin of both plasmid and chromosomal systems, namely the  $\zeta$  and PezT toxins, share a common phosphotransferase fold and possess 42% sequence identity<sup>25</sup>. The cytotoxic effect of  $\zeta$ /PezT toxin is conventionally exerted by phosphorylating the peptidoglycan precursor uridine diphosphate-N-acetylglucosamine (UNAG), utilising adenosine triphosphate (ATP), but more recently nicotinamide adenine dinucleotide (NAD) and uridine diphosphate-N-acetylmuramic acid (UNAM) have also been identified as substrates for these toxins<sup>29,30</sup>. In both PezAT and  $\epsilon/\zeta/\omega$  systems, the antitoxin functions by binding to the P-loop motif present in the toxin, thus preventing binding of ATP<sup>25,28</sup>. The plasmid  $\epsilon/\zeta/\omega$  system is responsible for post-segregational killing of the bacterial cell. In addition, PezAT has also been associated with virulence of *S. pneumoniae*<sup>24,31</sup>.

In the present study, we have performed rigorous sequence searches, using the predicted and characterised  $\epsilon/\zeta/\omega$  and PezAT TA systems known to-date, as queries, to probe the *M. tuberculosis* genome for homologues. Our searches identified Rv0366c-Rv0367c as a homologue for the PezAT using both  $\epsilon/\zeta/\omega$  and PezAT sequences from other microorganisms as queries. We demonstrate that inducible expression of PezT-like toxin of *M. tuberculosis*, Rv0366c inhibits growth of *Escherichia coli* and various mycobacterial species in a bacteriostatic manner. We were able to restore this growth defect by expressing Rv0366c along with the corresponding PezA-like antitoxin, Rv0367c. We also show that inducible expression of PezT-like toxin increases tolerance of *M. smegmatis* to ethambutol, a cell wall inhibitor. As per our knowledge, this is the first report on characterization of PezAT-like TA system in *M. tuberculosis* H<sub>37</sub>Rv. Future experiments would involve characterization of the functional target of PezT in *M. tuberculosis* and understanding its role in *M. tuberculosis* physiology and pathogenesis.

## Results

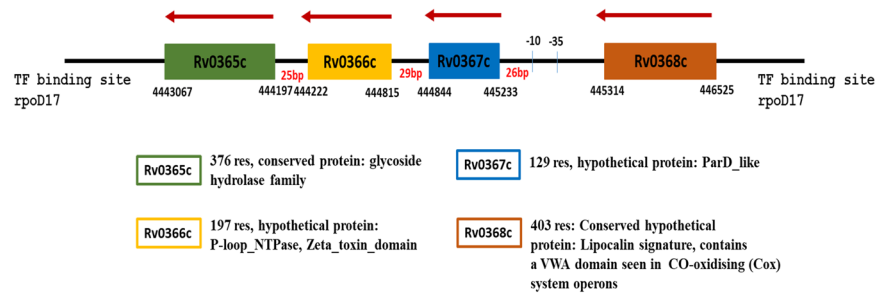
### Identification of a putative $\zeta$ /PezT-like toxin and PezA-like antitoxin in *M. tuberculosis* H<sub>37</sub>Rv.

Homologues of the PezT and  $\zeta$  toxin sequences from Pfam were identified in the *M. tuberculosis* genome using TBLASTN<sup>32,33</sup>. Of the 1054 query sequences, 337 identified the gene Rv0366c (O53701) as a homologue in *M. tuberculosis* H<sub>37</sub>Rv. The sequence identity of Rv0366c with its queries ranged from 30% to 80% with e-values between  $1e^{-10}$  and  $4e^{-96}$ . Pfam domain assignments for Rv0366c also predict, with very high confidence (e-value of  $2.1e^{-13}$ ), that this protein possesses the  $\zeta$  toxin domain (PF06414). A detailed sequence comparison of Rv0366c with known  $\zeta$  and PezT revealed that it lacks the N-terminal region, which in  $\zeta$  and PezT adopts a helical structure (Figure S1). This region in  $\zeta$  and PezT consists of a few antitoxin-binding residues. The remaining region of Rv0366c aligns well except for a short insertion in  $\zeta$  and PezT at its C-terminal region.

After the assignment of potential  $\zeta$ /PezT function to Rv0366c, we probed its neighbouring genes for a potential antitoxin. As shown in Fig. 1, Rv0365c, to the downstream of Rv0366c, is a 376-residue protein, which is annotated as a conserved glycoside hydrolase protein. To its upstream lies a 129 residue, hypothetical protein (Rv0367c), whose Pfam domain assignment suggested a ParD-like antitoxin (PF11903) domain for 65 out of 129 residues with an e-value of  $4.7e^{-22}$ . However, alignments with ParD-like antitoxin sequences showed poor conservation of toxin-binding residues (Figure S2). Secondary structure and disorder predictions of Rv0367c suggest that it is an ordered protein (data not shown), unlike the observed C-terminal of unstructured ParD antitoxin sequences<sup>34</sup>. Therefore, domain assignments of the Rv0367c to ParD antitoxin sequences were not considered for further evaluation.

### Rv0366c encodes a PezT/ $\zeta$ toxin homologue lacking a UNAG-binding site.

Rv0366c sequence was aligned with the sequences of known  $\zeta$  toxin of *S. pyogenes* (1gvn) and PezT toxin of *S. pneumoniae* (2p5t) along with other members of the  $\zeta$  toxin family from Pfam using MAFFT (Fig. 2)<sup>25,28,35</sup>.  $\zeta$ /PezT toxins are UDP-N-acetylglucosamine kinases having a well-defined nucleotide-binding motif (GXXGXXKT)<sup>25,28</sup>. Apart from nucleotide-binding, the substrate-binding as well as the antitoxin-binding site are also conserved among the  $\zeta$ /PezT toxins. The nucleotide-binding site in the  $\zeta$  toxin structure (1gvn) includes residues harbouring the 40-GXXGXXKT-47 motif, Arg171 and Glu116. For PezT (2p5t), equivalent nucleotide-binding residues include 39-GXXGXXKT-46 motif, Arg170, Glu115<sup>36</sup>. As shown in Fig. 2, the nucleotide-binding residues are



**Figure 1.** Representation of genomic region, ORF and promoters. The genomic locus for Rv0368c, Rv0367c, Rv0366c and Rv0365c are shown. Rv0367c, Rv0366c and Rv0365c were predicted to lie in the same operon. The promoter region was predicted to lie 26 bp upstream of Rv0367c<sup>58</sup>. The annotation for these gene neighbours, obtained from Tuberculist<sup>59</sup> is also given below the figure. Rv0367c was considered as a putative antitoxin for further analysis.

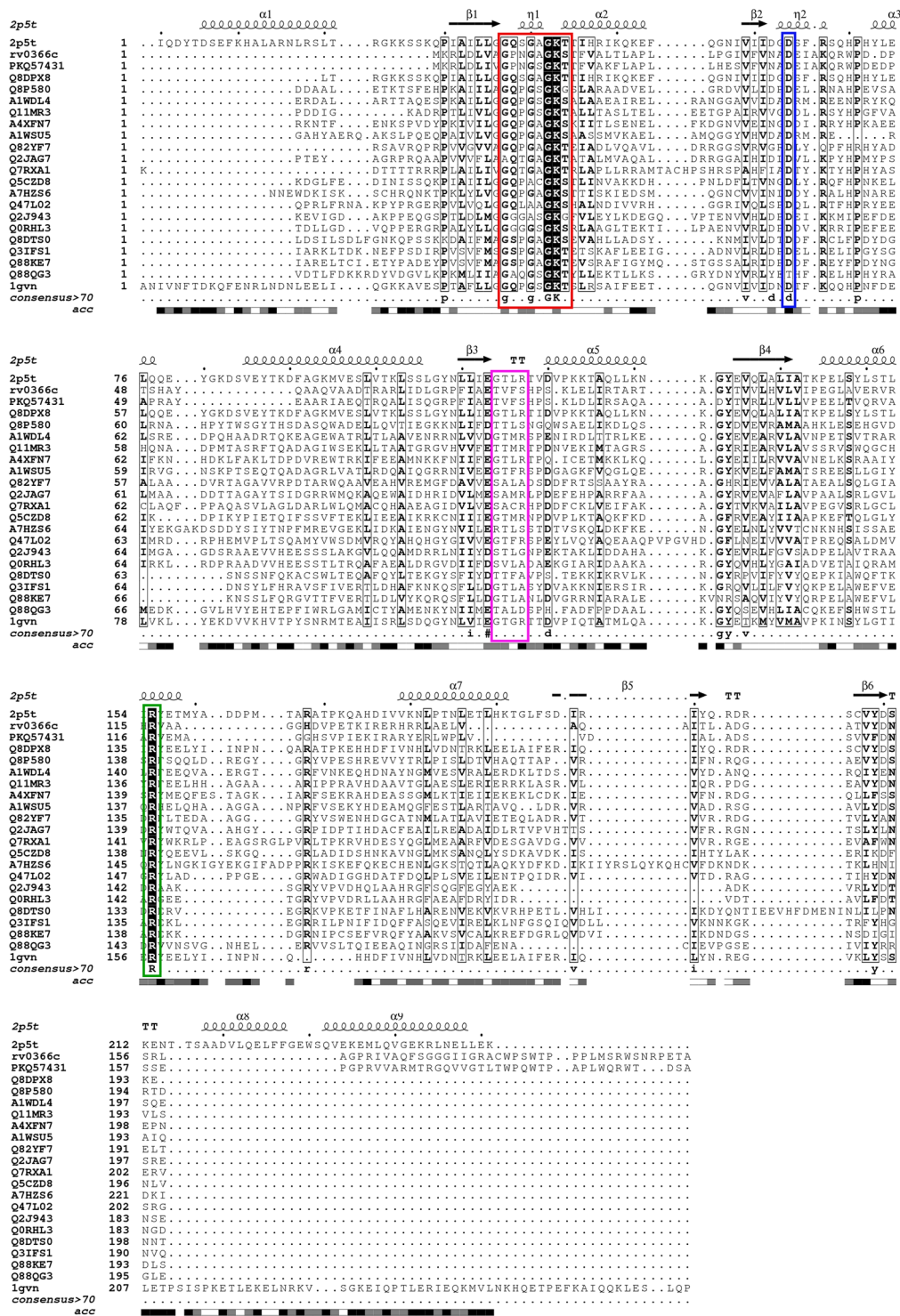
well-conserved in Rv0366c. These correspond to 9-GPNGAGKS-16, His122 and Glu75 in Rv0366c. An important antitoxin-binding residue is Arg158 in  $\zeta$  toxin while in PezT, it corresponds to Arg157. This residue, in addition to its role in antitoxin binding, has also been shown to be important for enzyme catalysis. It was observed that this antitoxin-binding and substrate-binding residue is also well conserved in Rv0366c (Arg116, Fig. 2).

The GTXR motif lying between Gly117 to Thr121 in  $\zeta$  and Gly116 to Thr120 in PezT toxins is required to bind UNAG<sup>36</sup>. In addition, Asp67, Glu100 in  $\zeta$  and Asp66, Glu99 in PezT are part of the substrate-binding site. The residues Asp67 in  $\zeta$  and Asp66 in PezT that lie in the nucleotide-binding site are important to deprotonate the substrate prior to phosphorylation and hence its activity<sup>25,28,37</sup>. The alignments in Fig. 2 show that the GTXR motif is not conserved in Rv0366c, but the Asp36 (equivalent to Asp67 and Asp66 in  $\zeta$  and PezT respectively), is present in Rv0366c. The conservation of the nucleotide-binding residues and the residue required to prime the bound substrate for subsequent phosphorylation suggests that Rv0366c is also likely to function as a nucleotide kinase. The lack of conservation of the UNAG-binding residues suggests that it may target other cellular substrates. Since, Rv0366c is the chromosomal homolog of  $\zeta$  and shares similarities with PezT, it makes a strong case for a non-canonical PezT toxin and would be referred as PezT<sup>Mtb</sup> in subsequent experiments.

In order to predict the three-dimensional organisation of the conserved residues and other surface exposed residues that may point to protein-protein interaction surfaces, PezT<sup>Mtb</sup> was modelled using Modeller<sup>38</sup>. For this, Igvn and 2p5t, which share 18% and 20% sequence identity respectively with Rv0366c, were used as templates<sup>25,28</sup>. Figure 3A shows the putative nucleotide-binding P-loop motif, Asp36 residue and the antitoxin-binding Arg116 residue. Other surface exposed and well conserved residues in accordance with the  $\zeta$ /PezT toxin homologues (Lys40, Trp43, Pro44, His122, Val77 and Asn34) are also shown (Fig. 3A). These residues lie in the vicinity of the predicted ATP-binding motif and have been predicted to be solvent exposed in the modelled structure. Such conserved, solvent-exposed patches on protein structures can interact with substrate or antitoxin and hence are likely important for toxin function. This analysis raises the possibility that PezT<sup>Mtb</sup> toxin can bind to a nucleotide and antitoxin despite being shorter than the canonical  $\zeta$ /PezT toxins.

**Comparison of PezT<sup>Mtb</sup> with  $\zeta$  toxins that bind other substrates.** Recently, it has been reported that  $\zeta$  toxin in *Neisseria gonorrhoeae* ( $\zeta_{ng}$ ) is unable to phosphorylate UNAG at C3'-OH but phosphorylates UNAG, UNAM and other UDP-activated sugars at C4'-OH, hence showing broader and altered substrate specificity than the  $\zeta$  and PezT from *S. pyogenes* and *S. pneumoniae* respectively<sup>30</sup>. Similarly, AvrRxo1 from *Xanthomonas Oryzae pv. oryzicola* is a  $\zeta$  toxin family member that phosphorylates NAD in plants<sup>29</sup>. Hence, we compared these two  $\zeta$  toxins with PezT<sup>Mtb</sup>. A structure-guided sequence alignment of PezT<sup>Mtb</sup> with  $\zeta_{ng}$  toxin is shown in Figure S3a and with AvrRxo1 in Figure S3b. Figure S3a shows that PezT<sup>Mtb</sup> which is shorter than  $\zeta_{ng}$  (197, Rv0366c; 401,  $\zeta_{ng}$ ), lacks the substrate-binding (UNAM and UNAM-4P) residues seen in  $\zeta_{ng}$ . Among the 10 substrate-binding residues identified in  $\zeta_{ng}$ , only 3 (Asp36, Arg116 and His122) are conserved in PezT<sup>Mtb</sup> (Table S1). For AvrRxo1, while precise information on the substrate-binding residues is not available, mutational experiments have identified two catalytic residues, Thr167 and Asp193 as important for phosphorylation of NAD<sup>39</sup>. Sequence alignment revealed that these two residues are conserved in PezT<sup>Mtb</sup> (Ser16 and Asp36). Additionally, the predicted active-site residues for AvrRxo1, Lys15 and Arg116 are also conserved in PezT<sup>Mtb</sup> (Table S1).

We further compared the structures of the canonical PezT (2p5t),  $\zeta_{ng}$  (6epg) and AvrRxo1 (4z8v) and the modelled PezT<sup>Mtb</sup> to analyse their ATP and substrate-binding pockets. Substrate-binding residues for the known structures along with predicted residues for AvrRxo1 were mapped on the available structures (Fig. 3B). We observed that while the ATP-binding P-loop motif is well conserved in all the structures, the substrate-binding cavity of PezT from *S. pneumoniae* differs from the other toxins. Between  $\zeta_{ng}$  and AvrRxo1, PezT<sup>Mtb</sup> showed better similarity to the substrate-binding region of AvrRxo1, both in terms of residue conservation and spatial location of the residues in the modelled structure (Figs 3B and S3). Hence, we speculate that the substrates for PezT<sup>Mtb</sup> might be closer to AvrRxo1 substrates. Altogether, we predict that although the UNAG-binding residues are not conserved, the nucleotide-binding, the antitoxin-binding and the aspartate residue (Asp36), important for the activity of  $\zeta$ /PezT toxins, are conserved, suggesting that Rv0366c might function as a PezT toxin but may target substrates other than UNAG.

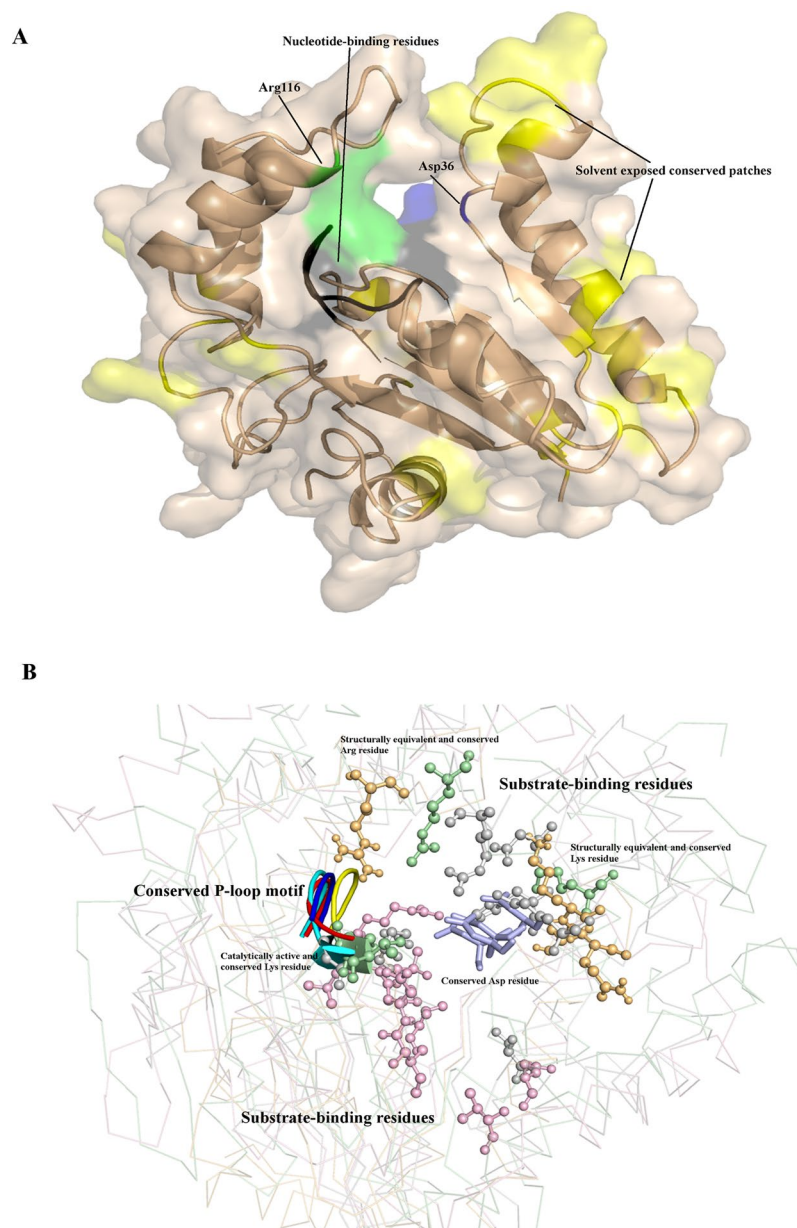


**Figure 2.** Structure-guided alignment of Rv0366c with C toxin from *Streptococcus pyogenes*, PezT from *Streptococcus pneumoniae* and other homologues. First sequence is the sequence of PezT of known structure.

The nucleotide-binding motif is completely conserved and marked in red. The aspartate residue, important for deprotonation of substrate, is also conserved and marked in blue. The conserved arginine residue, marked in green, is important for both antitoxin-binding and catalysis. The other UNAG-binding residues are marked in pink. Other fully conserved residues are shown in black background, while semi-conserved residues in black boxes. Alignment was generated using ESript 3.0<sup>60</sup>.

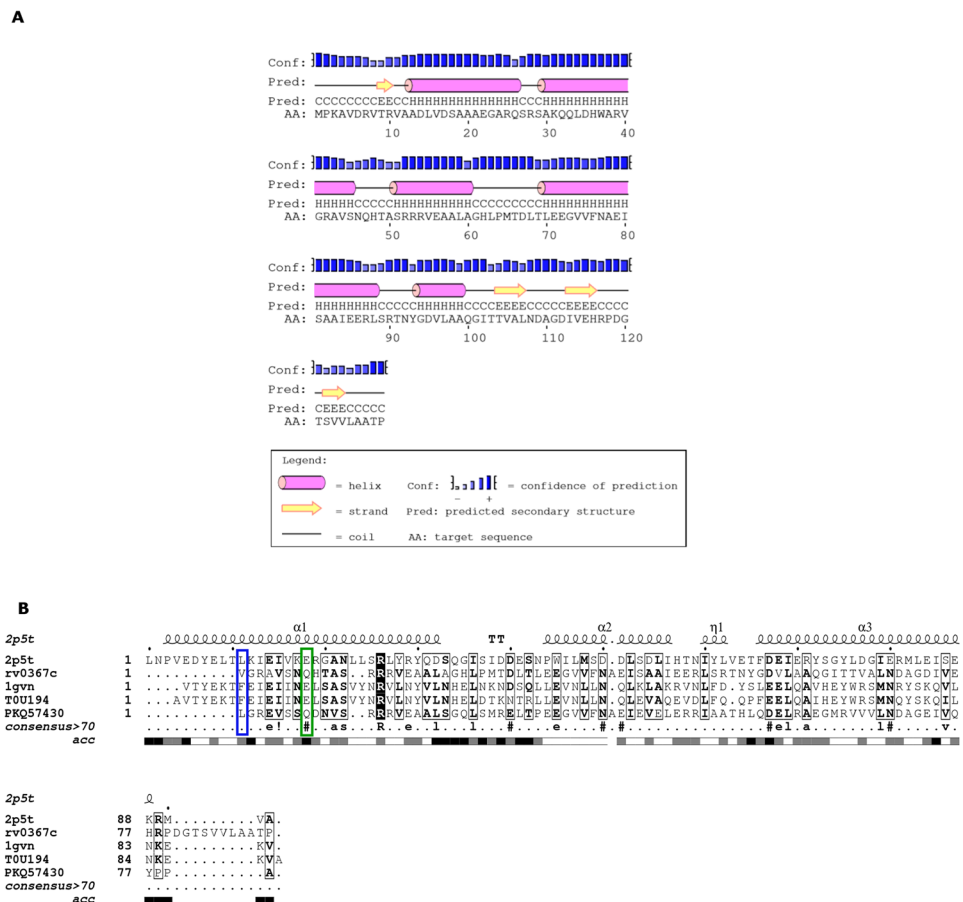
**Rv0367c shows signatures of PezA/ε antitoxin revealing a non-canonical PezAT-like TA system.** Since we were unable to establish a clear association of the putative antitoxin, Rv0367c, we predicted its fold using Phyre2 and HHpred<sup>40,41</sup>. Here, both methods assigned the RHH domain to the N-terminal region of





**Figure 3.** Modelled structure for PezT<sup>Mtb</sup> and comparison of substrate-binding pockets of known  $\zeta$  toxins with PezT<sup>Mtb</sup>. **(A)** Putative nucleotide-binding residues (black), conserved residue that may deprotonate the substrate (blue), residue with a potential role in both antitoxin-binding and catalysis (green) as well as solvent exposed, conserved residues, which include Lys40, Trp43, Pro44, His122, Val77 and Asn34 (yellow) are highlighted in the modelled PezT<sup>Mtb</sup> structure. **(B)** Ribbon representation of superposed crystal structures of PezT (2p5t),  $\zeta_{ng}$  (6epg), AvrRxo1 (4z8v) and the modelled structure of PezT<sup>Mtb</sup>. The nucleotide-binding P-loop motif is shown in cartoon representation, cyan- PezT, blue -  $\zeta_{ng}$ , yellow- AvrRxo1 and red - PezT<sup>Mtb</sup>. The substrate-binding residues are shown in ball and stick, magenta- PezT, grey -  $\zeta_{ng}$ , green - AvrRxo1 and orange- PezT<sup>Mtb</sup>. The conserved Asp residue, important for substrate-binding/deprotonating the substrate, is shown as sticks in light-blue. Other two residues (Lys40 and Arg116) known to bind substrate and are conserved in PezT<sup>Mtb</sup> are also labelled.

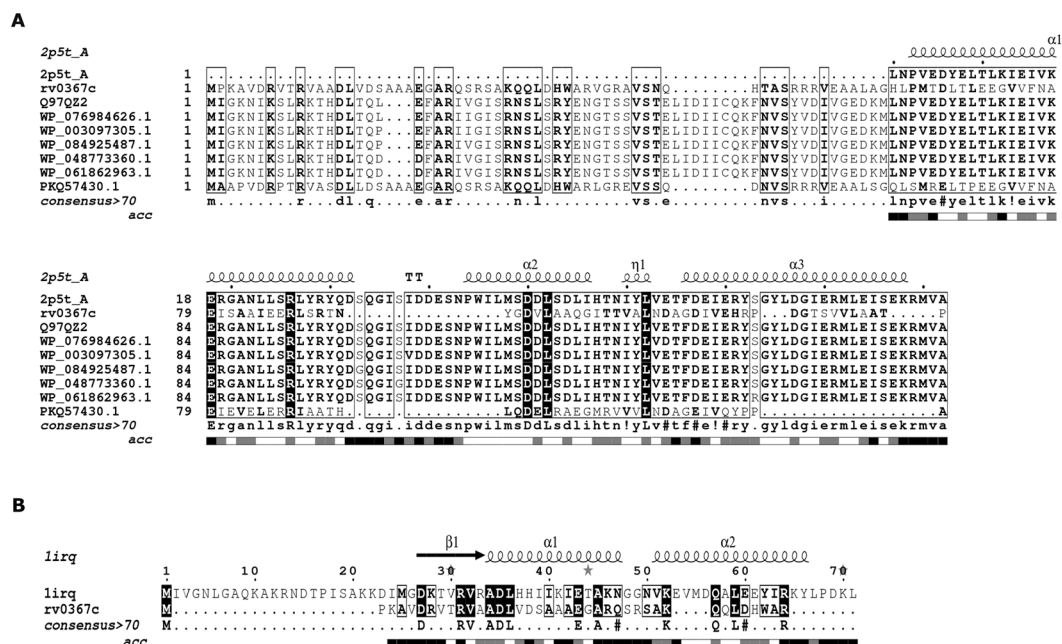
Rv0367c (Figure S4). PSIPRED, a secondary structure prediction algorithm, also predicted a small ribbon and two helices for the N-terminal region of Rv0367c (Fig. 4A)<sup>42</sup>. RHH domain is a structural motif known to bind to DNA and functions in auto-regulation of many TA modules<sup>43</sup>. In canonical PezA, this regulatory DNA-binding unit is predicted to adopt the HTH motif and is fused to the antitoxin sequence while in  $\epsilon/\zeta/\omega$  TA system, the regulatory protein  $\omega$  is a separate protein adopting the RHH motif. We additionally performed multiple sequence alignment of the C-terminal region of Rv0367c (residues 40–129) with the C-terminal region of PezA and  $\epsilon$  from *S. pneumoniae* and *S. pyogenes*, respectively (Fig. 4B). We observed that the equivalent residues for Leu77/Phe9, Glu84/Glu16, Tyr94/Leu26, Asp98/Leu30, and Asp116/Asn44, which in PezA/ $\epsilon$  are important for complex formation,



**Figure 4.** Alignment of PezA<sup>Mtb</sup> with  $\epsilon$  and PezA from other organisms. **(A)** The secondary structures predicted by PSIPRED show a ribbon-helix-helix (RHH) for the N-terminal region of Rv0367c<sup>42</sup>. C-terminal region is also predicted to be ordered with  $\alpha$ -helices and strands. **(B)** shows the multiple sequence alignment of C-terminal region of Rv0367c with  $\epsilon$  and its homologues, and the C-terminal of PezA. Structural alignment of both  $\epsilon$  and PezA were used to guide the alignment. Putative toxin-binding residue is marked in green and the residue predicted to occlude the nucleotide binding site is shown in blue. Other fully conserved residues are shown in black background, while semi-conserved residues in black boxes. Alignment was generated using ESPrnt 3.0<sup>60</sup>.

correspond to Val40, Gln47, Val55, Leu59 and Asn77, respectively in Rv0367c. The other solvent exposed and conserved residues, which can function as protein interaction surface in Rv0367c, include Ala50, Ser51, Arg53, Leu63, Asp67, Asp95 and Arg117. In order to determine the extent of sequence similarity in the DNA-binding regions, Rv0367c, full length PezA antitoxin and its homologues were aligned (Fig. 5A). This alignment showed better sequence conservation in the toxin-binding C-terminal than the N-terminal residues. The N-terminal of Rv0367c (residues 1–39), that is predicted to adopt the RHH motif exhibited 30% sequence identity with  $\omega$  protein from *S. pyogenes* (Iirq) in comparison to 10% sequence identity with N-terminal of PezA antitoxin<sup>25,27</sup>. Since the  $\zeta$  toxins are plasmid-borne and the PezT toxins are their established chromosomal homologues, we predict that Rv0366c-Rv0367c is an unusual PezAT<sup>Mtb</sup> TA system with N-terminal of PezA<sup>Mtb</sup> closer to  $\omega$ .

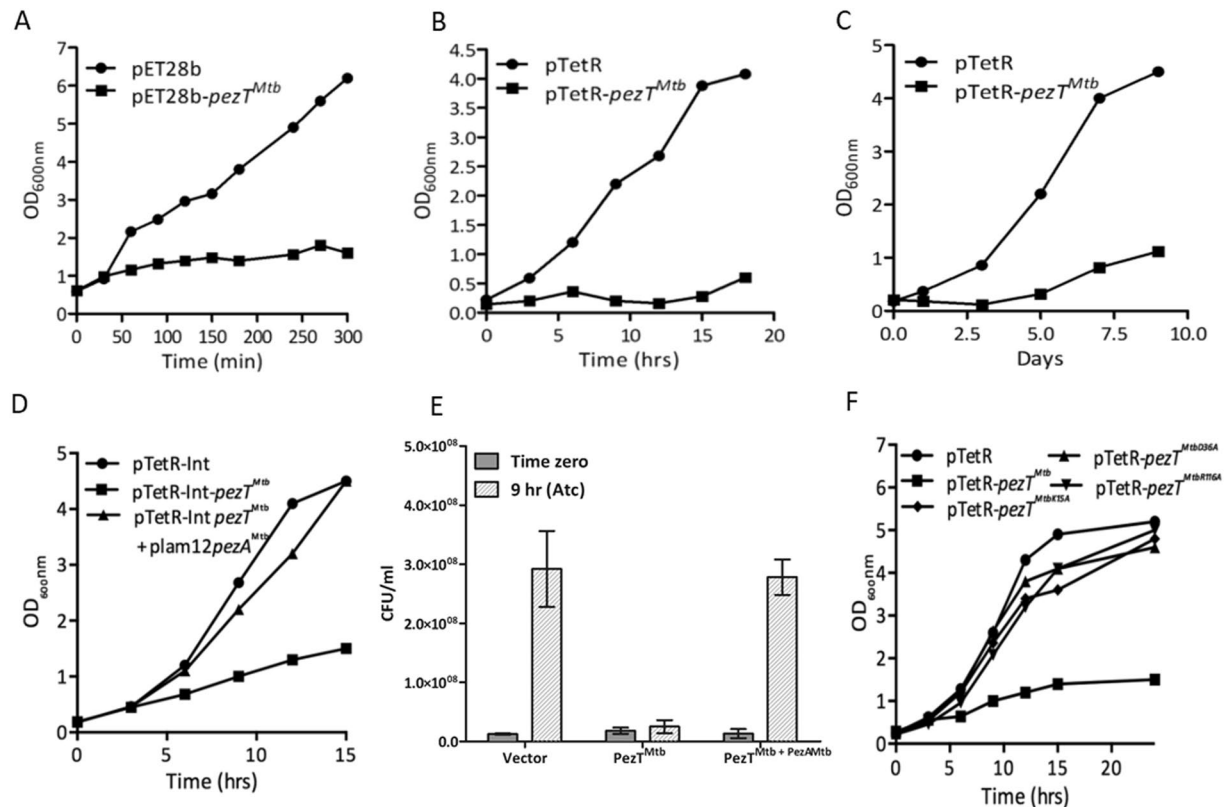
**Modelled complex structure predicts occlusion of PezT<sup>Mtb</sup> nucleotide-binding site by PezA<sup>Mtb</sup> residues.** Taking cues from the sequence alignments, we attempted to predict the binding mode of the putative PezAT-like TA system (PezAT<sup>Mtb</sup>). A full-length model was generated for PezA<sup>Mtb</sup> using Phyre2 intensive mode, using multiple templates (1u9p<sup>44</sup>, 3qoq<sup>45</sup>, 1mnt<sup>40,46</sup>). Though PezA<sup>Mtb</sup> could be modelled only with 70% confidence, certain secondary structures that can potentially form the interface between the toxin and antitoxin could be predicted with very high confidence. A model of the docked PezT<sup>Mtb</sup> and PezA<sup>Mtb</sup> (C-terminal region) was obtained using Z-dock and HADDOCK<sup>47,48</sup>. Since the templates for the TA complex were available (2p5t, 1gvn), restraints were derived based on known interfacial residues in the template that are conserved in the alignments (Fig. 2). Both the programs predicted at least one binding mode with hydrogen bond between Arg116 in PezT<sup>Mtb</sup> and Gln47 in PezA<sup>Mtb</sup>. This is equivalent to the interaction between Arg157 and Glu84 of PezT and PezA from *S. pneumoniae* and this interaction is important to inhibit toxic activity associated with PezT<sup>25</sup>. The HADDOCK-derived modelled complex structure predicts a binding mode in which, PezA<sup>Mtb</sup> occludes the nucleotide-binding site along with Asp36 of PezT<sup>Mtb</sup> (Figure S5).



**Figure 5.** Alignment of PezA<sup>Mtb</sup> with PezA homologues (A) and  $\omega$  from *S. pyogenes* (B). (A) Shows alignment of Rv0367c (O53702) with full-length PezA and its homologues. Though the N-terminal region shows sequence conservation, many residues that are a part of the helix–turn–helix motif in the N-terminal of PezA are not conserved in PezA<sup>Mtb</sup>. (B) Shows alignment of PezA<sup>Mtb</sup> (N-terminal) with  $\omega$  repressor (Irq) of  $\epsilon/\zeta$  TA system. Crystal structure of  $\omega$  was used to guide the alignment. The fully conserved residues between these sequences are highlighted in black. Alignment was generated using ESript 3.0<sup>60</sup>.

**Overexpression of PezT<sup>Mtb</sup> induces bacteriostasis.** In order to functionally characterize PezT<sup>Mtb</sup>, growth inhibition upon overexpression of PezT<sup>Mtb</sup> in both *E. coli* and mycobacterial strains was studied. The expression of Rv0366c in BL-21 ( $\lambda$ DE3, plysS) was induced by the addition of 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). We observed that overexpression of PezT<sup>Mtb</sup> resulted in severe growth inhibition of *E. coli* (Fig. 6A). As expected, no growth inhibition was seen in *E. coli* harbouring vector control (Fig. 6A). To investigate the effect of PezT<sup>Mtb</sup> overexpression in fast growing (*M. smegmatis*) and slow growing (*M. tuberculosis*) bacteria, it was cloned in an anhydrotetracycline-based inducible vector, pTetR<sup>49</sup>. Consistent with *E. coli* results, inducible expression of PezT<sup>Mtb</sup> inhibited growth of both *M. smegmatis* and *M. tuberculosis* H<sub>37</sub>Rv (Fig. 6B,C). We also noticed that this growth inhibition was restored by co-expressing PezT<sup>Mtb</sup> along with the predicted antitoxin partner, PezA<sup>Mtb</sup> (Fig. 6D). As shown in Fig. 6E, we observed that PezT<sup>Mtb</sup> overexpression resulted in bacteriostatic effect as reported earlier in the case of other type II toxins such as RelE, MazF and VapC<sup>23,49–52</sup>. To further characterize PezT<sup>Mtb</sup>, we performed growth inhibition experiments with *M. smegmatis* overexpressing Rv0366c harbouring mutations in the predicted catalytically important residues such as Asp36, Lys15 and Arg116 (Fig. 6F). As expected, mutations of these residues to either alanine or valine abrogated the growth inhibition activity associated with PezT<sup>Mtb</sup>.

**Overexpression of PezT<sup>Mtb</sup> reduces cell length and increases tolerance of *M. smegmatis* to ethambutol *in vitro*.** There are accumulating evidences that morphological changes occur in bacteria upon exposure to different stress conditions. We have shown earlier that overexpression of VapC toxins resulted in bacterial population that possessed a bulge<sup>49</sup>. Previously, it has been shown that overexpression of PezT results in cell wall lysis and formation of mid-cell positioned bulges in *E. coli*<sup>53</sup>. Hence, we next compared the cell morphology and length of vector harbouring and PezT<sup>Mtb</sup> overexpressing *M. smegmatis* strains. As shown in Fig. 7A, we noticed that 4',6-diamidino-2-phenylindole (DAPI) uniformly stained the nucleoid in both strains. Interestingly we observed that length of bacilli in PezT<sup>Mtb</sup> overexpression strain was reduced in comparison to the parental strain (Fig. 7A). The bacilli in both parental and overexpression strains at 3 hrs and 6 hrs post-induction appeared as rod shaped (Fig. 7A). We observed that the average length of bacteria in PezT<sup>Mtb</sup> overexpression strain was 4  $\mu$ m in comparison to 6  $\mu$ m in strain harbouring vector control (Fig. 7B,C). As expected these differences were not observed in *M. smegmatis* upon overexpression of PezT<sup>Mtb</sup> toxin along with its cognate PezA<sup>Mtb</sup> antitoxin (Fig. 7A–C). The observed reduction in bacilli length upon overexpression of PezT<sup>Mtb</sup> in *M. smegmatis* was observed to be statistically significant (\*\*p < 0.001). We also observed that the cell length of *M. smegmatis* overexpressing mutant proteins was similar to that observed for the parental strain (Figure S6). Further, we determined the contribution of PezT<sup>Mtb</sup> to the formation of drug-tolerant persisters. We quantified the fraction of persisters in early-log phase *M. smegmatis* overexpression or vector control cultures upon exposure to either ethambutol or levofloxacin. As shown in Fig. 7D, ethambutol exposure increased the number of drug-tolerant persisters by 10.0-fold compared to empty vector control (\*p < 0.05). However, this increase in tolerance was not

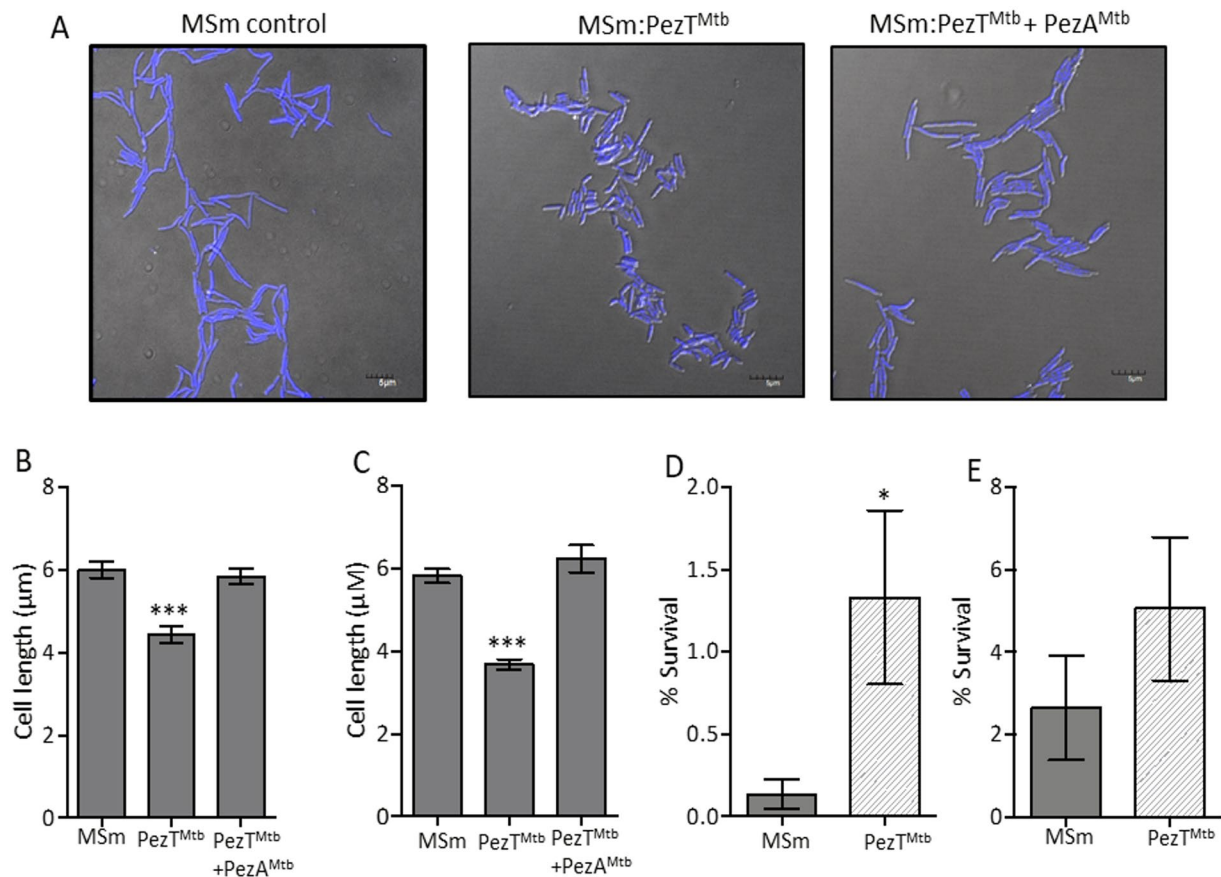


**Figure 6.** Effect of overexpression of PezT<sup>Mtb</sup> on bacterial growth. (A) *E. coli* BL-21, plyS, λDE3 was transformed with either pET28b or pET28b-pezT<sup>Mtb</sup>. The expression of PezT<sup>Mtb</sup> was induced by addition of 1 mM IPTG and growth was monitored by measuring OD<sub>600nm</sub> at indicated time points. (B,C) Expression of PezT<sup>Mtb</sup> in both *M. smegmatis* mc<sup>2</sup>155 and *M. tuberculosis* H<sub>37</sub>Rv was induced by addition of 50 ng/ml anhydrotetracycline in liquid cultures. The growth of induced cultures was monitored by measuring OD<sub>600nm</sub>. (D,E) For co-expression studies, the expression of PezT<sup>Mtb</sup> and PezA<sup>Mtb</sup> was induced by addition of 50 ng/ml anhydrotetracycline and 0.2% acetamide, respectively. The growth of induced cultures was monitored by measuring OD<sub>600nm</sub> (D) and viable counts (E). For bacterial enumeration 10.0-fold serial dilutions were prepared and plated on MB7H11 plates. The plates were incubated at 37 °C for 2–3 days. The data shown in this panel is representative of two independent experiments. (F) *M. smegmatis* strains harbouring vector or PezT<sup>Mtb</sup> or PezT<sup>Mtb</sup>K15A or PezT<sup>Mtb</sup>D36A or PezT<sup>Mtb</sup>R116A were grown in MB7H9 medium till OD<sub>600nm</sub> of 0.2. The effect of expression of wild type and mutant PezT<sup>Mtb</sup> proteins on *M. smegmatis* growth was monitored by measuring OD<sub>600nm</sub>. The data shown in these panels is representative of three independent experiments.

observed in cultures upon exposure to levofloxacin (Fig. 7E). The number of levo-tolerant persisters were comparable in *M. smegmatis* PezT<sup>Mtb</sup> overexpressing or vector harbouring strains. The percentage of surviving bacteria varied between 2.0–4.0% after 24 hrs post-drug exposure (Fig. 7E). These results are in concordance with previous studies demonstrating that overexpression of RelE contributes to increase in drug-specific persisters *in vitro*<sup>50</sup>.

**Conservation of PezAT-like system of H<sub>37</sub>Rv in other mycobacterial species.** The presence of a non-canonical PezAT<sup>Mtb</sup> prompted us to probe for the prevalence of such a non-canonical system in 101 other mycobacterial genomes using TBLASTN (Fig. 8). We observed that homologs of either PezT<sup>Mtb</sup> or PezA<sup>Mtb</sup> were present in 57 mycobacterial species while no hits were identified in the remaining 44 mycobacterial genomes. Out of the 57, these proteins existed as a TA pair in 20 bacterial genomes. Interestingly, a majority of these TA pairs are present in pathogenic bacteria such as the MTBC (marked in green) and shared a sequence identity of greater than 80% amongst themselves (Fig. 8). Others such as *Mycobacterium avium*, *Mycobacterium chimaera*, *Mycobacterium gordonae*, *Mycobacterium mucogenicum*, and *Mycobacterium peregrinum* also possess homolog for PezAT<sup>Mtb</sup> and are non-tuberculosis bacteria capable of causing various human diseases. We also noticed that homolog of PezT<sup>Mtb</sup> is present in *Mycobacterium gastri* and *Mycobacterium kansasii* with a sequence identity lower than 40%. Similarly, in other non-pathogenic bacteria like *Mycobacterium austroafricanum*, *Mycobacterium vanbaalenii* etc. the homologs of either the PezT<sup>Mtb</sup> toxin or PezA<sup>Mtb</sup> antitoxin exist but with poor sequence identity (Fig. 8). A separate search of the canonical UNAG-binding ζ toxins in mycobacterial species does not provide any noteworthy results, further strengthening the observation that the genus *Mycobacterium* may harbour only such non-canonical systems.



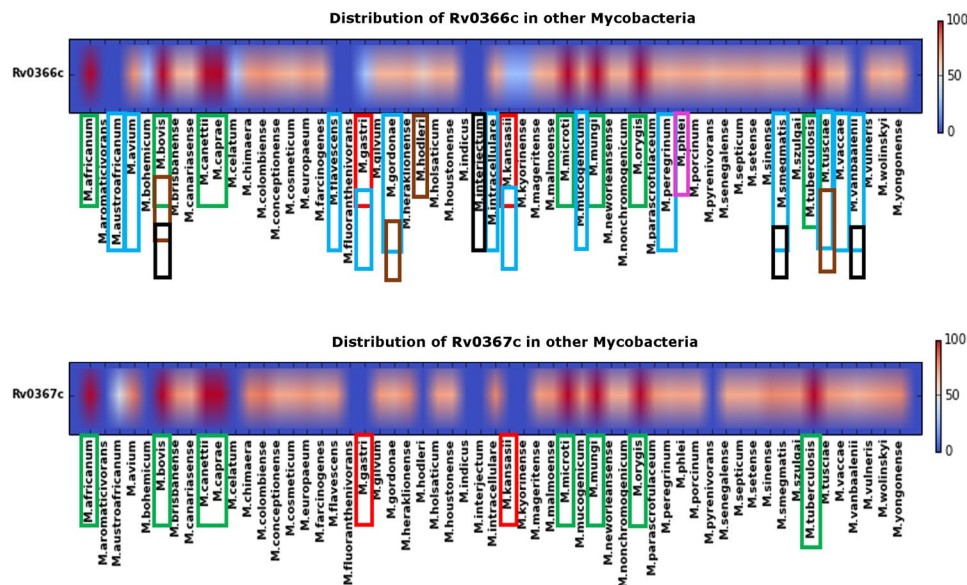


**Figure 7.** Effect of PezT<sup>Mtb</sup> overexpression on cell length and drug tolerance of *M. smegmatis*. (A) DAPI stained fixed images of *M. smegmatis* harbouring vector or expressing PezT<sup>Mtb</sup> in the absence or presence of PezA<sup>Mtb</sup> after 3 hrs of induction. (B,C) The panel depicts cell length of *M. smegmatis* harbouring either pTetR-Int or pTetR-Int-PezT<sup>Mtb</sup> or pTetR-Int-PezT<sup>Mtb</sup> and pLam12-PezA<sup>Mtb</sup>. The cultures were induced for either 3 hrs (B) or 6 hrs (C) and bacilli length was measured using Fluoview FV1000 software. The y-axis represents mean  $\pm$  S.E. of cell length for bacilli per strain. The data shown in this panel is representative of three independent experiments. (D,E) Induced *M. smegmatis* cultures harbouring either pTetR-Int or pTetR-Int-PezT<sup>Mtb</sup> were exposed to either 15.6  $\mu$ M ethambutol or 2.0  $\mu$ M levofloxacin. For bacterial enumeration, 12 hrs post-exposure, cultures were harvested, washed and 100  $\mu$ l of 10.0-fold serial dilutions were plated on MB7H11 plates at 37  $^{\circ}$ C for 2–3 days. Significant differences were observed for the indicated groups (paired, two-tailed, t-test, \* $P$  < 0.05, \*\*\* $P$  < 0.001).

**Probing the presence of PezAT<sup>Mtb</sup> system outside genus *Mycobacterium*.** The classical  $\epsilon/\zeta/\omega$  or PezAT systems are ubiquitously present in Gram-positive as well as Gram-negative bacteria<sup>36</sup>. Hence, we investigated, whether this non-canonical PezAT<sup>Mtb</sup> system also occurs in other prokaryotes. To address this question, we searched for its homologues in nearly 4500 bacterial and archaeobacterial representative genomes using TBLASTN (Table 1). Interestingly, only 26 organisms outside *Mycobacterium* possess homologues for the PezAT<sup>Mtb</sup> pair. Majority of these organisms belong to the family Gordoniaceae and Nocardiaceae which are members of the order Corynebacteriales as is *Mycobacteriaceae*. *Kineosphaera limosa*, which belongs to the same class as *Mycobacterium*, also possesses a homolog of this non-canonical system. In all these organisms, the putative toxin is annotated as ATPase while the upstream putative antitoxin is a hypothetical protein. A separate search for canonical  $\epsilon/\zeta/\omega$  or PezAT systems (1gvn and 2p5t) found homologs mainly in families Enterococcaceae and Streptococcaceae, belonging to class Bacilli, but no hits were observed in the organisms where the homologs of PezAT<sup>Mtb</sup> system were found. This analysis shows that non-canonical PezAT<sup>Mtb</sup> system are largely confined to the order Corynebacteriales with no homologues in organisms outside this order.

## Discussion

The reasons for the expanded repertoire of TA systems in organisms like *M. tuberculosis* remain a mystery. While it is speculated that many have been acquired through lateral gene transfer and gene duplication, the functional implications of such an expanded repertoire have been suggested as beneficial to bacterial pathogenesis, virulence and survival mechanisms<sup>54</sup>. Although various type II TA systems have been identified in *M. tuberculosis*, this study reports for the first time, the identification and characterization of a PezAT-like TA system in this organism. PezT and its plasmid homologues  $\zeta$  toxins that have been demonstrated to phosphorylate UNAG



**Figure 8.** Distribution of Pez<sup>T</sup><sup>Mtb</sup> (upper panel) and PezA<sup>Mtb</sup> (lower panel) in the genus Mycobacterium. The hits were obtained at an e-value better than 1e<sup>-04</sup> and query coverage better than 70% from TBLASTN. The organisms belonging to MTBC are marked in green in both panels. Red marking indicates the organisms in which maximum number of other H<sub>37</sub>Rv toxin-antitoxin systems have been found to be conserved (unpublished). The other colours represent the habitat that these organisms belong to: cyan – water, brown – soil, pink – plants, black – air.

S. no	Name of the organism	Order
1	<i>Amorphus coralli</i>	Rhizobiales
2	<i>Brevibacterium linens</i>	Corynebacteriales
3	<i>Conexibacter woesei</i>	Solirubrobacterales
4	<i>Corynebacterium-like bacterium</i>	Corynebacteriales
5	<i>Gordonia aichiensis</i>	Corynebacteriales
6	<i>Gordonia araii</i>	Corynebacteriales
7	<i>Gordonia desulfuricans</i>	Corynebacteriales
8	<i>Gordonia hirsute</i>	Corynebacteriales
9	<i>Gordonia jacobaea</i>	Corynebacteriales
10	<i>Gordonia malaquae</i>	Corynebacteriales
11	<i>Gordonia polyisoprenivorans</i>	Corynebacteriales
12	<i>Gordonia rhizosphere</i>	Corynebacteriales
13	<i>Gordonia terrae</i>	Corynebacteriales
14	<i>Hydrocarboniphaga effuse</i>	Nevskiales
15	<i>Kineosphaera limosa</i>	Micrococcales
16	<i>Novosphingobium barchaimii</i>	Shingomonadales
17	<i>Oceanicaulis alexandrii</i>	Rhodobacteriales
18	<i>Parvibaculum lavamentivorans</i>	Rhizobiales
19	<i>Rhodococcus fascians</i>	Corynebacteriales
20	<i>Rhodococcus jostii</i>	Corynebacteriales
21	<i>Rhodococcus pyridinivorans</i>	Corynebacteriales
22	<i>Rhodococcus triatomae</i>	Corynebacteriales
23	<i>Roseivivax atlanticus</i>	Rhodobacteriales
24	<i>Salinisphaera shabanensis</i>	Salinisphaerales
25	<i>Smaragdicooccus niigatensis</i>	Corynebacteriales
26	<i>Tomitella biformata</i>	Corynebacteriales

**Table 1.** Distribution of homologues of Rv0366c and Rv0367c, as pair, in organisms outside genus *Mycobacterium*.

function as nucleotide kinases. The phosphorylated form of UNAG inhibits MurA and other enzymes involved in peptidoglycan biosynthesis and maintenance of cell wall integrity<sup>25,28,30</sup>. The homologue in *M. tuberculosis* was identified through computational searches involving many canonical and non-canonical  $\zeta$  and PezT toxins. Since the homologue was not plasmid encoded, we annotated Rv0366c as PezT<sup>Mtb</sup>. This *M. tuberculosis* homologue harboured antitoxin-binding arginine residue (at position 116), an aspartate residue (at position 36) that prepares the substrate for phosphorylation through deprotonation, and the P-loop motif for NTP binding. However, the residues of  $\zeta$  and PezT toxins that bind to known substrates such as UNAG were not conserved in PezT<sup>Mtb</sup>.

Our searches for homologues of the canonical antitoxin using  $\varepsilon$ /PezA were less successful in identifying the cognate antitoxin due to two reasons: a) extensive sequence dispersion of the antitoxin and b) differences in the DNA-binding regulatory region of the antitoxin. Most antitoxins are composed of an N-terminal DNA binding region and a C-terminal toxin-binding region: two functions encoded in a single protein. However, in  $\varepsilon/\zeta/\omega$  system,  $\varepsilon$  possess only the toxin-binding region and  $\omega$  encodes a separate protein that participates in DNA-binding and regulation. In the PezAT<sup>Mtb</sup> TA system that we have identified in *M. tuberculosis*, the N-terminal of PezA<sup>Mtb</sup> shares 30% sequence identity with  $\omega$  protein that binds DNA in the  $\varepsilon/\zeta/\omega$  TA pair. Its C-terminal shares 17% sequence identity to the canonical, chromosomally encoded, PezA- antitoxins. Therefore, the non-canonical PezAT<sup>Mtb</sup> TA that we report here shows signatures that are characteristic of both the  $\varepsilon/\zeta/\omega$  and PezAT TA families.

We further show that overexpression of PezT<sup>Mtb</sup> toxin induces bacteriostasis in various bacterial systems and co-expression of PezA<sup>Mtb</sup> antitoxin could rescue this growth defect. The inhibition of toxic activity of PezT<sup>Mtb</sup> by co-expression of the PezA<sup>Mtb</sup> antitoxin in *M. smegmatis* and *E. coli* corroborates that the Rv0366c-Rv0367c locus encodes a TA pair. Indeed, their binding modes, predicted through docking studies, proposes a complex that is similar to known PezAT and  $\varepsilon/\zeta/\omega$  TA systems. As reported in the abundant  $\zeta$  toxins of Gram-positive bacteria, here also the antitoxin binds in a manner that occludes the nucleotide-binding pocket of the toxin. Contrary to published reports in *E. coli*, we did not observe any bulge formation in PezT<sup>Mtb</sup> overexpression *M. smegmatis* strains<sup>53</sup>. The findings that the presence of free PezT<sup>Mtb</sup> results in decreased bacterial length is suggestive of the fact that phosphorylation of UNAG might not be the mechanism by which PezT<sup>Mtb</sup> inhibits bacterial growth. A homologue from *X. oryzae* (AvrRxo1), is known to phosphorylate coenzyme NAD and its precursors<sup>39</sup>. More recently, a hitherto unknown homologue from *N. gonorrhoeae* that targets UDP-activated sugars, resulting in cell lysis, has also been characterized<sup>30</sup>. However, an alignment of PezT<sup>Mtb</sup> with the *N. gonorrhoeae*  $\zeta$  toxin and *X. oryzae* (AvrRxo1) shows conservation of only few of the substrate-binding residues, mainly with AvrRxo1. We reckon therefore, that PezT<sup>Mtb</sup>, which does not show high conservation of UNAG-binding residues, might target NAD or an altogether different cellular substrate. However, more detailed experiments will be required to further characterize the enzymatic activity associated with PezT<sup>Mtb</sup>. We also show that ectopic expression of PezT<sup>Mtb</sup> results in a metabolic state that results in drug-tolerance specific to ethambutol. These observations are in concordance with previous studies where the effect of overexpression of toxins on mycobacterial tolerance has been shown to be drug-specific<sup>50</sup>.

Recently, Jaén-Luchoro and co-workers reported the presence of a putative  $\zeta$  toxin in the sequenced genome of an isolate of *Mycobacterium* MSDH3, a close homologue of *Mycobacterium chelonae*<sup>55</sup>. While the toxin sequence in their study bears considerable similarity with the known  $\zeta$  toxins, no conclusive functional annotation was made for the antitoxin sequence. They reported remote similarities with the ParD antitoxin and shared similarities with the Arc repressor Pfam domain family whose members possess the RHH DNA-binding motif. We aligned the proteins reported in their study to PezAT<sup>Mtb</sup> (Figure S7) and found that PezT<sup>Mtb</sup> and PezA<sup>Mtb</sup> share high sequence identity with the MHS3 systems. Our study has attempted a more detailed annotation of the antitoxin sequence and we find that it shows high similarity with the  $\omega$  of the  $\varepsilon/\zeta/\omega$  toxin antitoxin system. While they limited their analysis to assessments of similarities with close hits detected in BLAST searches, we have performed a more exhaustive survey for the presence of such a TA system in all mycobacteria and extended the searches to nearly 4500 prokaryotic genomes. Indeed, our comparative study shows that homologues of the PezAT<sup>Mtb</sup> were restricted to the order Corynebacteriales.

A separate search for homologues of classical  $\varepsilon/\zeta/\omega$  systems finds hits in the genus *Enterococcus*, *Streptococcus* and *Lactococcus* but no significant hits in *Mycobacterium*. Nearly 20 pathogenic mycobacterial species show conservation of the non-canonical PezT<sup>Mtb</sup> as pairs. Although further experimental evidence will be required to demonstrate the cellular targets of this non-canonical TA pair and its role in conferring virulence to H<sub>37</sub>Rv, its homology to proteins from pathogenic bacteria raises the possibility that PezAT<sup>Mtb</sup> might also be implicated in bacterial pathogenesis and virulence. A search for homologues of this system in species outside the genus *Mycobacterium* revealed that it is conserved in a few micro-organisms outside Mycobacteriaceae family. Interestingly these families belong to the order Actinomycetales that includes the genus *Mycobacterium*. This strengthens the idea that this non-canonical PezAT-like system is largely confined to the class Actinobacteria and in fact mainly to Mycobacteriaceae. The results presented in the study indicate that this system has likely undergone selective evolutionary pressure to remain confined within this class of bacteria. Taken together, this study reports the identification, characterisation and conservation analysis of a non-canonical PezAT-like TA system in *M. tuberculosis*. Future experiments would aim i) to understand the plausible mechanism of growth inhibition by PezT<sup>Mtb</sup> and (ii) construction of knock out strains to determine the essentiality of this TA pair *in vivo*.

## Experimental Procedures

**Search for  $\zeta$ /PezT toxin homologue in *M. tuberculosis* H<sub>37</sub>Rv.** Putative  $\zeta$  toxin was searched in *M. tuberculosis* using two parallel approaches. In the first approach, 1036  $\zeta$  toxin sequences from the Pfam database were queried in the *M. tuberculosis* genome using TBLASTN<sup>32,33</sup>. These queries included both experimentally validated as well as predicted  $\zeta$  and PezT toxin sequences from various organisms. In addition, atypical  $\zeta$  toxin-like sequences that do not have a conserved UNAG-binding motif (GTXR) were also employed as queries. To increase the search sensitivity, other bacterial genome sequences were also included in the search database. Hits

were filtered using an e-value cut-off of  $1e^{-06}$  and a query coverage of 60%. Separately, Pfam domains were also assigned to the *M. tuberculosis* proteome, using the PfamScan program in Pfam, to identify proteins with the  $\zeta$  toxin domain.

**Identification of  $\epsilon$ /PezA antitoxin homologue in *M. tuberculosis* H<sub>37</sub>Rv.** After the identification of putative  $\zeta$  toxin loci in *M. tuberculosis*, the neighbouring genes of each hit were analysed to search for a cognate antitoxin. The neighbouring gene was considered as the putative antitoxin only if (1) it precedes the putative toxin and (2) its sequence distance from the putative toxin gene is  $\leq 150$  bp<sup>23</sup>. Pfam domains were assigned to each such putative antitoxin to scan for the  $\epsilon$ /PezA antitoxin domain. As in the search for the toxins, an independent sequence-based search for  $\epsilon$ /PezA sequences from the Pfam database was also performed in the *M. tuberculosis* proteome, using TBLASTN. Rv0367c was identified as the putative antitoxin sequence and aligned with  $\epsilon$  and PezA antitoxin sequences from the Pfam database, using MAFFT, to predict toxin-binding residues<sup>35</sup>. As for the toxin, a homology model was derived for the putative antitoxin, using templates identified by Phyre2<sup>40</sup>.

**Prediction of putative toxin structure and antitoxin-binding residues.** Rv0366c was predicted as a homologue of the  $\zeta$  and PezT toxins from our query dataset. It was modelled with modeller using the templates identified in Phyre2 and HHpred searches<sup>40,41</sup>. Igvn is a  $\zeta$  toxin from *S. pyogenes* and 2p5t is a PezT toxin from *S. pneumoniae*. Both templates show a Phyre2 confidence score  $>99\%$  and share 18% and 20% sequence identity with Rv0366c. A structure-guided alignment, derived using Promals3D, was employed to build the model<sup>36</sup>. 100 models were generated and the model with the least discrete optimized protein energy (DOPE) score was selected. MAFFT was employed to align Rv0367c with its homologous sequences from the Pfam database. Conserved, surface-exposed residues were recognized from the multiple sequence alignment and mapped onto the modelled structure, to evaluate the compatibility of its interaction interfaces with the antitoxin or substrate.

**Bacterial strains, plasmids and growth conditions.** All bacterial strains and plasmids used in the study are listed in Table S2. Bacterial strains were grown with shaking at 200 rpm, 37°C in Luria Bertani medium (LB) for *E. coli* or Middlebrook MB7H9 medium for mycobacteria. The growth of cultures was monitored by either spotting or measuring optical densities at 600 nm. When required, antibiotics were used at the following concentrations; kanamycin (25  $\mu$ g/ml for *E. coli*), ampicillin (100  $\mu$ g/ml for *E. coli*), hygromycin (150  $\mu$ g/ml for *E. coli*, 50  $\mu$ g/ml for mycobacteria) and tetracycline (10  $\mu$ g/ml for *E. coli*). The sequences of oligonucleotides used in the study are shown in Table S3.

**DNA manipulations.** DNA manipulations and other molecular biology techniques were performed according to standard protocols. The ORF encoding Rv0366c was amplified from *M. tuberculosis* genomic DNA using primers Rv0366c-F and Rv0366c-R. The PCR amplicons were digested and cloned into either IPTG inducible pET28b (for *E. coli* studies) or pTetR or pTetR-Int (for mycobacterial studies<sup>49</sup>). The recombinant pET28b-PezT<sup>Mtb</sup> was confirmed by sequencing and transformed into BL-21 ( $\lambda$ DE3, plysS) and expression was induced by addition of 1 mM IPTG. Recombinant pTetR constructs were electroporated into either *M. smegmatis* mc<sup>2</sup>155 or *M. tuberculosis* H<sub>37</sub>Rv and transformants were selected on MB7H11 medium plates supplemented with hygromycin. The ORF encoding for antitoxin, Rv0367c was PCR amplified using primers Rv0367c-F and Rv0367c-R and cloned into acetamide inducible vector plam12<sup>37</sup>. Further, the hygromycin resistance gene was replaced with kanamycin resistance gene resulting into plam-Rv0367c-kan. The construct was sequenced prior to transformation in mc<sup>2</sup>155 harbouring an integrative copy of PezT<sup>Mtb</sup>. Positive transformants were selected on MB7H11 plates supplemented with kanamycin and hygromycin.

**Activity of PezT<sup>Mtb</sup> in *E. coli* and mycobacteria.** For growth inhibition assays in *E. coli*, expression of PezT<sup>Mtb</sup> was induced by the addition of 1 mM IPTG. For overexpression of PezT<sup>Mtb</sup> in *M. smegmatis* and *M. tuberculosis*, cultures were grown till OD<sub>600nm</sub> of 0.2 and toxin expression was induced by the addition of 50 ng/ml anhydrotetracycline. For co-expression studies, the expression of antitoxin was induced by the addition of 0.2% acetamide. The experiments involving *M. smegmatis* recombinant strains were performed in standard Bio-safety level 2 containment facilities. The growth curves using various *M. tuberculosis* strains were carried out in Bio-safety level 3 containment facilities.

**Drug tolerance assays in *M. smegmatis*.** For these experiments, expression of toxins in *M. smegmatis* was induced by the addition of 50 ng/ml anhydrotetracycline. *M. smegmatis* strains harbouring pTetR or pTetR-PezT<sup>Mtb</sup> were diluted to OD<sub>600nm</sub> of 0.05 in medium containing either ethambutol (cell wall inhibitor) or levofloxacin (DNA replication inhibitor) and incubated at 37°C for 24 hrs. For bacterial enumeration, 10.0-fold serial dilutions were prepared and 100  $\mu$ l was plated on MB7H11 plates in duplicates. Percentage survival was calculated from the mean colony-forming unit (CFU) obtained in the culture after drug incubation divided by CFU before the addition of drug.

**Cell length experiments.** In order to investigate the effect of PezT<sup>Mtb</sup> overexpression on cell length, toxin expression in *M. smegmatis* was induced by the addition of 50 ng/ml anhydrotetracycline for either 3 hrs or 6 hrs. At designated time points, induced cultures were harvested, washed, fixed with 4% paraformaldehyde and stained with DAPI as previously described<sup>49</sup>. DAPI stained bacilli were viewed using FV1000 confocal microscope (Olympus, Japan). The morphology and length of bacteria was visualised and analysed using fluoview software (Olympus, Japan).



**Distribution of putative *M. tuberculosis* PezAT-like TA system in Mycobacteria and other prokaryotes.** TBLASTN searches were performed, by querying for homologues of PezT<sup>Mtb</sup> and PezA<sup>Mtb</sup> against a database consisting of 101 mycobacterial and 4500 other bacterial and archaeobacterial genomes. A cut-off of  $1e^{-04}$  and a query coverage of 70% was applied to filter the hits. Care was taken to consider only those homologues of PezT<sup>Mtb</sup> and PezA<sup>Mtb</sup> that were not separated by >150 bp in various genomes.

### Data Availability

All data pertaining to the manuscript have been provided in the form of figures. Supporting information has been made available as Tables S1–S3 and supporting Figures S1–S7. Datasets pertaining to the sequence searches described here are available from the corresponding author on reasonable request.

### References

- Sala, A., Bordes, P. & Genevoux, P. Multiple toxin-antitoxin systems in *Mycobacterium tuberculosis*. *Toxins (Basel)* **6**, 1002–1020 (2014).
- Gerdes, K., Christensen, S. K. & Løbner-Olesen, A. Prokaryotic toxin-antitoxin stress response loci. *Nature Reviews Microbiology* **3**, 371–382 (2005).
- Blower, T. R., Salmond, G. P. C. & Luisi, B. F. Balancing at survival's edge: The structure and adaptive benefits of prokaryotic toxin-antitoxin partners. *Curr. Opin. Struct. Biol.* **21**, 109–118 (2011).
- Harms, A., Maisonneuve, E. & Gerdes, K. Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science (80-)*. **354** (2016).
- Melderer, L. V. *et al.* Molecular Basis of Gyrase Poisoning by the Addiction Toxin CcdB. 1091–1102, <https://doi.org/10.1016/j.jmb.2005.03.049> (2005).
- Sharp, J. D. *et al.* Growth and translation inhibition through sequence-specific RNA binding by *Mycobacterium tuberculosis* VapC toxin. *J. Biol. Chem.* **287**, 12835–12847 (2012).
- Cruz, J. W. *et al.* Doc toxin is a kinase that inactivates elongation factor Tu. *J. Biol. Chem.* **289**, 7788–7798 (2014).
- Goeders, N. & Van Melderer, L. Toxin-antitoxin systems as multilevel interaction systems. *Toxins* **6**, 304–324 (2013).
- Harms, A., Brodersen, D. E., Mitarai, N. & Gerdes, K. Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Molecular Cell*. <https://doi.org/10.1016/j.molcel.2018.01.003> (2018).
- Fozo, E. M., Hemm, M. R. & Storz, G. Small Toxic Proteins and the Antisense RNAs That Repress Them. *Microbiol. Mol. Biol. Rev.* **72**, 579–589 (2008).
- Leplae, R. *et al.* Diversity of bacterial type II toxin-antitoxin systems: A comprehensive search and functional analysis of novel families. *Nucleic Acids Res.* **39**, 5513–5525 (2011).
- Fineran, P. C. *et al.* The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc. Natl. Acad. Sci.* **106**, 894–899 (2009).
- Brown, J. M. & Shaw, K. J. A Novel Family of *Escherichia coli* Toxin-Antitoxin Gene Pairs. *J. Bacteriol.* **185**, 6600–6608 (2003).
- Masuda, H., Tan, Q., Awano, N., Wu, K. P. & Inouye, M. YeeU enhances the bundling of cytoskeletal polymers of MreB and FtsZ, antagonizing the CbtA (YeeV) toxicity in *Escherichia coli*. *Mol. Microbiol.* **84**, 979–989 (2012).
- Wang, X. *et al.* A new type V toxin-antitoxin system where mRNA for toxin GhoT is cleaved by antitoxin GhoS. *Nat. Chem. Biol.* **8**, 855–861 (2012).
- Aakre, C. D., Phung, T. N., Huang, D. & Laub, M. T. A bacterial toxin inhibits DNA replication elongation through a direct interaction with the  $\beta$  sliding clamp. *Mol. Cell* **52**, 617–628 (2013).
- Smith, J. A. & Magnuson, R. D. Modular Organization of the Phd Repressor/Antitoxin Protein. *J. Bacteriol.* <https://doi.org/10.1128/JB.186.9.2692-2698.2004> (2004).
- Zielenkiewicz, U. & Ceglowski, P. The toxin-antitoxin system of the streptococcal plasmid pSM19035. *J. Bacteriol.* **187**, 6094–6105 (2005).
- Hallez, R. *et al.* New toxins homologous to ParE belonging to three-component toxin-antitoxin systems in *Escherichia coli* O157:H7. *Mol. Microbiol.* **76**, 719–732 (2010).
- Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Comprehensive comparative-genomic analysis of Type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol. Direct* **4**, 19 (2009).
- Pandey, D. P. & Gerdes, K. Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res.* **33**, 966–976 (2005).
- Frampton, R., Aggio, R. B. M., Villas-Bóas, S. G., Arcus, V. L. & Cook, G. M. Toxin-antitoxin systems of *Mycobacterium smegmatis* are essential for cell survival. *J. Biol. Chem.* **287**, 5340–5356 (2012).
- Ramage, H. R., Connolly, L. E. & Cox, J. S. Comprehensive functional analysis of *Mycobacterium tuberculosis* toxin-antitoxin systems: Implications for pathogenesis, stress responses, and evolution. *PLoS Genet.* **5** (2009).
- Ceglowski, P., Boitsov, A., Karamyan, N., Chai, S. & Alonso, J. C. Characterization of the effectors required for stable inheritance of *Streptococcus pyogenes* pSM19035-derived plasmids in *Bacillus subtilis*. *MGG Mol. Gen. Genet.* **241**, 579–585 (1993).
- Seok, K. K. *et al.* Molecular and structural characterization of the PezAT chromosomal toxin-antitoxin system of the human pathogen *Streptococcus pneumoniae*. *J. Biol. Chem.* **282**, 19606–19618 (2007).
- de la Hoz, A. B. *et al.* Plasmid copy-number control and better-than-random segregation genes of pSM19035 share a common regulator. *Proc. Natl. Acad. Sci.* **97**, 728–733 (2000).
- Murayama, K., Orth, P., De La Hoz, A. B., Alonso, J. C. & Saenger, W. Crystal structure of  $\omega$  transcriptional repressor encoded by *Streptococcus pyogenes* plasmid pSM19035 at 1.5 Å resolution. *J. Mol. Biol.* **314**, 789–796 (2001).
- Meinhart, A., Alonso, J. C., Strater, N. & Saenger, W. Crystal structure of the plasmid maintenance system epsilon/zeta: functional mechanism of toxin zeta and inactivation by epsilon 2 zeta 2 complex formation. *Proc Natl Acad Sci USA* **100**, 1661–1666 (2003).
- Shidore, T. *et al.* The effector AvrRxo1 phosphorylates NAD in planta. *PLoS Pathog.* **13** (2017).
- Rocker, A. *et al.* The ng $\_c1$  toxin of the gonococcal epsilon/zeta toxin/antitoxin system drains precursors for cell wall synthesis. *Nat. Commun.* **9**, 1686 (2018).
- Brown, J. S., Gilliland, S. M., Spratt, B. G. & Holden, D. W. A Locus Contained within A Variable Region of Pneumococcal Pathogenicity Island 1 Contributes to Virulence in Mice. *Infect. Immun.* **72**, 1587–1593 (2004).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).
- Finn, R. D. *et al.* The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285 (2016).
- Oberer, M., Zangger, K., Gruber, K. & Keller, W. The solution structure of ParD, the antidote of the ParDE toxin antitoxin module, provides the structural basis for DNA and toxin binding. *Protein Sci.* **16**, 1676–1688 (2007).
- Katoh, K. & Standley, D. M. MAFFT: Iterative refinement and additional methods. *Methods Mol. Biol.* **1079**, 131–146 (2014).
- Mutschler, H. & Meinhart, A. In *Prokaryotic {Toxin}-{Antitoxins}* 205–223, <https://doi.org/10.1007/978-3-642-33253-1> (2013).
- Izard, T. & Ellis, J. The crystal structures of chloramphenicol phosphotransferase reveal a novel inactivation mechanism. *EMBO J.* **19**, 2690 (2000).

38. Sali, A. MODELLER A Program for Protein Structure Modeling Release 9v4, r6262. *Structure* 779–815 (2008).
39. Han, Q. *et al.* Crystal structure of Xanthomonas AvrRxo1-ORF1, a type III effector with a polynucleotide kinase domain, and its interactor AvrRxo1-ORF2. *Structure* 23, 1900–1909 (2015).
40. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858 (2015).
41. Hildebrand, A., Remmert, M. & Biegert, A. & S?ding, J. Fast and accurate automatic structure prediction with HHpred. *Proteins Struct. Funct. Bioinforma.* 77, 128–132 (2009).
42. McGuffin, L. J., Bryson, K. & Jones, D. T. *The PSIPRED protein structure prediction server.* 16, 404–405 (2000).
43. Schreiter, E. R. & Drennan, C. L. Ribbon-helix-helix transcription factors: Variations on a theme. *Nat. Rev. Microbiol.* 5, 710–720 (2007).
44. Tabtiang, R. K., Cezairliyan, B. O., Grant, R. A., Cochrane, J. C. & Sauer, R. T. Consolidating critical binding determinants by noncyclic rearrangement of protein secondary structure. *Proc. Natl. Acad. Sci.* 102, 2305–2309 (2005).
45. Pryor, E. E. *et al.* The transcription factor AmrZ utilizes multiple DNA binding modes to recognize activator and repressor sequences of *Pseudomonas aeruginosa* virulence genes. *PLoS Pathog.* 8 (2012).
46. Burgering, M. J. M. *et al.* Solution Structure of Dimeric Mnt Repressor (1–76). *Biochemistry* 33, 15036–15045 (1994).
47. De Vries, S. J., Van Dijk, M. & Bonvin, A. M. J. J. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* 5, 883–897 (2010).
48. Pierce, B. G. *et al.* ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics.* <https://doi.org/10.1093/bioinformatics/btu097> (2014).
49. Agarwal, S. *et al.* System wide analysis unravels differential regulation and *in vivo* essentiality of VapBC TA systems from *Mycobacterium tuberculosis*. *J. Infect. Dis.* <https://doi.org/10.1093/infdis/jiy109> (2018).
50. Singh, R., Barry, C. E. & Boshoff, H. I. M. The three RelE homologs of *Mycobacterium tuberculosis* have individual, drug-specific effects on bacterial antibiotic tolerance. *J. Bacteriol.* 192, 1279–1291 (2010).
51. Tiwari, P. *et al.* MazF ribonucleases promote *Mycobacterium tuberculosis* drug tolerance and virulence in guinea pigs. *Nat. Commun.* 6 (2015).
52. Winther, K., Tree, J. J., Tollervey, D. & Gerdes, K. VapCs of *Mycobacterium tuberculosis* cleave RNAs essential for translation. *Nucleic Acids Res.* 44, 9860–9871 (2016).
53. Mutschler, H., Gebhardt, M., Shoeman, R. L. & Meinhart, A. A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis. *PLoS Biol.* 9 (2011).
54. Van Melderen, L. Toxin-antitoxin systems: Why so many, what for? *Current Opinion in Microbiology* 13, 781–785 (2010).
55. Jaén-Luchoro, D. *et al.* First insights into a type II toxin-antitoxin system from the clinical isolate *Mycobacterium sp.* MHSD3, similar to epsilon/zeta systems. *PLoS One* 12 (2017).
56. Pei, J., Kim, B. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. 36, 2295–2300 (2008).
57. van Kessel, J. C. & Hatfull, G. F. Recombineering in *Mycobacterium tuberculosis*. *Nat. Methods* 4, 147–152 (2007).
58. Cortes, T. *et al.* Genome-wide Mapping of Transcriptional Start Sites Defines an Extensive Leaderless Transcriptome in *Mycobacterium tuberculosis*. *Cell Rep.* 5, 1121–1131 (2013).
59. Lew, J. M., Kapopoulou, A., Jones, L. M. & Cole, S. T. TubercuList - 10 years after. *Tuberculosis* 91, 1–7 (2011).
60. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* 42(W1), W320–W324 (2014).

## Acknowledgements

RS and NS acknowledge the financial support received from Department of Biotechnology (BT/COE/34/15219/2015). HT is a DST-INSPIRE fellow. AS is thankful to DBT for junior research fellowship in the project (BT/PR5510/MED/29/513/2012). This research is also supported by Indian Institute of Science-DBT partnership program and Mathematical Biology program of Department of Science and Technology to N.S., who is also a J.C. Bose National Fellow. Support for infrastructural facilities from Fund for Improvement of Science and Technology infrastructure (FIST), DST and Centre for Advanced Structures (CAS), University Grants Commission (UGC) is also acknowledged. The authors are thankful to University of Delhi, South Campus for access to Biosafety level 3 laboratories. The technical help provided by Mr. Rajesh and Mr. Sher Singh is highly acknowledged. The authors are thankful to RS lab members for reading of the manuscript.

## Author Contributions

H.T., A.S. and S.S. performed the analysis. R.S. and N.S. conceptualized and closely supervised the project. All authors read, wrote and approved of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-37473-y>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019