Bioinformatic and mutational studies of related toxin–antitoxin pairs in *M. tuberculosis* predict and identify key functional residues

Running title: Toxin-Antitoxin relationships in *M.tuberculosis*

Himani Tandon[1], Arun Sharma[2], Saruchi Wadhwa[2], Raghavan Varadarajan[1], Ramandeep Singh[2], Narayanaswamy Srinivasan[1*], and Sankaran Sandhya[1*]

[1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore- 560012, India.

[2]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.

* To whom correspondence should be addressed. Sankaran Sandhya: Molecular Biophysics Unit, Indian Institute of Science, Bangalore- 560012; sandhyas@iisc.ac.in; Tel: +918022932837; Fax: +918023600535.

Correspondence may also be addressed to Narayanaswamy Srinivasan. Molecular Biophysics Unit, Indian Institute of Science, Bangalore- 560012; ns@iisc.ac.in; Tel: +918022932837; Fax: +918023600535.

Keywords: Toxin-antitoxin, homology, paralogues, structure-function, *Mycobacterium tuberculosis*, genome analysis, bacteriostasis, VapBC, PIN domain, phylogeny, molecular modelling, protein evolution

S1. Trends observed in the distribution of homologues of *M.tuberculosis* TA systems within MTBC and conservation pattern of Rv0909-Rv0910 and Rv1546 in MTBC and other organisms.

Ramage *et. al*, have earlier probed the spread of *M.tuberculosis* type II TA in 5 of the 10 genomes in the MTBC complex (1). In addition to these genomes, we have included the genomes of *M.orygis, M.caprae and M.mungi* that are now available since their study, for our analysis. A search of *M.tuberculosis* TA in MTBC revealed that not all TAs were found as a pair with the same confidence in *M.mungi, M.orygis* and *M.canetti*. For example, Rv2653c-Rv2654c, that is absent in *M.bovis* and *M.africanum*, is only conserved in *M.microti* and *M.tuberculosis*. Little is known about the role of this predicted TA system which is closest in sequence to the PhiRv2 phage protein and implicated in colony formation (1). All genes in its neighbourhood are also encoded on a prophage implying that these genes may have been transferred from other genomes. As has been shown earlier, we also find that the Rv0059-Rv0060 TA pair is present in all MTBC but *M. canetti*, which is believed to be a member of a progenitor species from which all MTBC members might have emerged (2). Searches in mycobacterial genomes other than MTBC could identify homologues for some TA pairs (Figures 4, 5 and S6). For the 10 MazEF TAs (Figure S8), we find that while no homologues are obtained for MazEF1, MazEF3, MazEF4 and MazEF8 in the mycobacterial genomes, homologues for one or more of the other MazEF TA are found in 19 of the 93 genomes. Interestingly, the MTBC harbour homologues for each of the MazEF type II TA (Tables S5, S6, and Figure S6). For the 51 VapBC TA pairs (Figure 5), we find that VapBC47, VapBC5, VapBC4 etc. are conserved as a pair in many mycobacteria while VapBC49, VapBC48 and others find homologues in only one other mycobacterial species. It is interesting to note that VapBC41, VapBC25, VapBC23, VapBC21, VapBC18, VapBC12 and VapBC8 are conserved only within the MTBC (Figures 5) and do not find conserved TA pairs in any of the other mycobacterial genomes. Similar distribution profiles are plotted for the 20 other TAs in Figures S6 and S8.

Figure S8 shows the distribution profile of conserved TA pairs of other types in all the mycobacterial species, excluding the MTBC. Here we find that while Rv0910 and Rv0909 individually find homologues in *M.abscessus*, they fail to meet the search criteria to be considered as a conserved TA). Indeed, *M.abscessus* does not appear to possess a homologue for any of the TA pairs involved in the study (Figure S8). Homologues for Rv0909-Rv0910 were not found in *M. crocinum, M.hodleri, M.pallens* and *M.xenopi*, all of which are environmental species with the exception of *M.xenopi*, which has been implicated in pulmonary infections and is transmitted to humans through the aerosol route (3). Further,

although sequence identities for the toxin were >60% across most homologues, the hits in *M.setense* and *M.aromaticivorans* have diverged considerably (showing <45% sequence identity). Rv1546 (Y1546) is a conserved hypothetical protein that is uniformly conserved in most mycobacteria (Figure 4 and Table S6). Interestingly, the eight species in which they are not conserved, are all rapidly growing species that are usually water-borne, but pathogenic in immunocompetent or immunocompromised individuals. Two of the species, *M.senegalense* and *M.farcinogenes* are closely related bovine pathogenic species.

Outside of the *MTB* complex, the organisms which possess a large number of the TA pairs (Figure 4, Table S6) include 21 for *M.brumae*, 22 for *M.celatum*, 25 for *M.gastri*, 16 for *M.kansasii*, 12 for *M.tusciae*, 14 for *M kyorinense* and 10 for *M.heckeshornense*, some of which such as *M.celatum* and *M.lentiflavum* are nontuberculous mycobacteria that causes pulmonary infections in immunocompromised patients (4, 5). Of these *M. gastrii,* that is closest to *M. kansasii,* conserves the most TA modules (~25) amongst the mycobacterial species studied here. Both organisms, show the presence of far more toxin homologues (24 and 21) in comparison to antitoxins (~14). Strikingly, although mycobacteria include species that are pathogenic and non-pathogenic, we find that the distribution of TAs does not seem to make a distinction. While *M. kansasii* has been implicated in pulmonary infections, *M.gastrii* is a harmless environmental saprophyte that is occasionally a casual resident of the human gut (6). Further, *M.gastrii* with 25 TA pairs, has homologues of each of the *M.tuberculosis* TA query types in our dataset (Figures 5 and S8), while *M.kansasii* possesses homologues of all TA types, except for the MazEF type TA system which functions as a cell growth modulator through its function as an mRNA interferase (Figure S8). Our searches reveal, in general, that the conservation of *M.tuberculosis* TA types is seen far more in slow growing species than in rapidly growing species such as *M. chelonae, and M. abscessus* (Figures 4, 5 and Table S6). The toxins Rv2231A (VapC16), Rv2829c (VapC22), Rv2274c (MazE8), Rv1955 (HigB1), Rv2142c (ParE2), Rv2826c (P71626) and Rv3189 (O53335) occur sparsely in Mycobacteria (<10) (Table S6). Rv2231A is conserved only in *MTB*C. Rv2829c conserved in *MTB*C at high sequence identity, is also observed in *M.celatum* at 35% sequence identity. Rv3189, is observed in *M.lentiflavum* at a sequence identity of 78%, outside of the *MTB*C.

S2. Statistics of TA pairs present in various orders and families of class Actinobacteria, Gamma, Delta and Alpha-proteobacteria.

Hits were obtained for toxins and antitoxins either individually or as a pair in ~2220 genomes (Table S8). Figures S9-S10 show the result of this exhaustive and comprehensive survey of hits obtained in various genomes, which are grouped based on their taxonomy using phyloT (https://phylot.biobyte.de). The circular, unrooted tree, coloured based on class, shows that TA homologues could be detected majorly in Actinobacteria, of which *M.tuberculosis* is a member, and also in Bacilli, Clostridia, Gamma, Alpha and Delta-proteobacteria (Table S10). For more detailed data visualization, a link to the interactive versions of the trees at the iTOL server is provided at the following URLs: https://itol.embl.de/tree/14139128221380415181740 18 and https://itol.embl.de/tree/1413912822265121518597108

Mycobacteria are acid-fast Gram-positive bacteria with high GC content that belong to the class *Actinobacteria* and are members of the order *Corynebacteriales*. The order *Corynibacteriales* comprises of seven families *Nocardiaceae, Gordoniaceae, Tsukamurellaceae, Dietziaceae, Corynebacteriaceae, Mycobacteriaceae and Segniliparaceae*. While TA pairs show a patchy distribution within the order *Corynebacteriales* with families such as *Nocardiaceae* and *Gordoniaceae* conserving as many as 63 and 31 TA pairs, families such as *Tsukamurellaceae* conserve only two TA pairs (Table S10). The likely absence of homologues of the TA system in the *Hoyosella,* another genus within the *Mycobacteriaceae* family, that are typically residents of oil reservoirs and moderate halophiles, suggests that TAs within the Mycobacterial genus have evolved independently of the other genus in this family (Figures S9-S10). Outside of the *Corynebacteriales*, homologues of the most TA pairs (between 9 to 11) were observed in different orders of the class *Actinobacteria* such as in the *Jiangella* genus (12 pairs) that are members of order *Jiangellales* (high GC, Gram positive), *Kineosphaera limosa, Serinicoccus marinus, Tetrasphaera jenkinsii, Knoellia sinensis* (order *Micrococcales*) and *Pseudonocardia asaccharolytica, Pseudonocardia acaciae (Pseudonocardiales)* (Figure S9, Tables S9, S10). *Actinobacteria* constitute one of the largest phyla among bacteria and represent Gram-positive bacteria with a high G+C content in their DNA. In addition to various morphologies, members of this class have adopted a number of lifestyles ranging from pathogenic (*Mycobacteria, Nocardia, Corynebacterium*), soil inhabiting (*Streptomyces*), plant commensals (*Leifsonia*), or gastrointestinal commensals (*Bifidobacterium*) (7).

Outside of the class *Actinobacteria*, 11 pairs of TA (in addition to 11 lone toxin homologues and 2 antitoxin homologues) were identified in *Acidithrix ferrooxidans and Candidatus microthrix* (7 TA pairs, 12 lone toxin homologues) which are both members of class *Acidimicrobiia* (Table S10). It has been reported that many of the proteins that appeared to be specific for *Actinobacteria* were transferred to an *Alphaproteobacterium- Magnetospirillum magnetotacticum,* through horizontal gene transfer (7). However, our results show that two other members of the *Alphaproteobacteria- Sinorhizobium meliloti* and *Mesorhizobium alhagi* have 7 and 5 TA pairs conserved (Table S10). Strikingly, although *Delta proteobacterium* (Gram-negative) harbours only 2 TA pairs that are homologous to the *M.tuberculosis* TA, we could find homologues for several toxins (of different types) in this bacteria.

S3. Rv3642c-Rv3641c complex

We examined the multiple sequence alignment of Rv3641c with members of the Fic toxin family (PF02661) from PFAM (Figure S14A). We find that the typical features of the Fic toxin family (based on the analysis of a number of Fic toxin proteins such as the conserved central core of four helices (α2 to α5) that is flanked by three helices (α1, α4 and α6) and a loop formed by helices α4 and α5 and the N-terminal cap of helix α5 are also predicted in Rv3641c using PSIPRED. Further, the alignment shows that the active centre represented by the signature motif HxFx(D/E)GNGRxxR, seen in all FIC domain containing proteins, is present in Rv3641c as well (8). The alignment presented in Figure S14 shows that an inhibitory signature (S/T)xxxE(G/N) present in FicA (8), although absent in Rv3641c, is present on its gene neighbour and the putative FicA antitoxin, Rv3642c (NXXXEG/N), suggesting that Rv3641c-Rv3642c might function as a class I FicTA system in *M.tuberculosis* (Figures S14 and S15). Since the templates for both toxin and antitoxin were identified with 95% confidence, we performed a template-based modelling of the Rv3642c-Rv3641c complex using Modeller v9.14 with *B. schoenbuchensis* VbhTA (PDB id: 3shg) as template.

References

1. Ramage, H. R., Connolly, L. E., and Cox, J. S. (2009) Comprehensive functional analysis of Mycobacterium tuberculosis toxin-antitoxin systems: Implications for pathogenesis, stress responses, and evolution. *PLoS Genet.* 10.1371/journal.pgen.1000767

2. M Cristina, G., Brisse, S., Brosch, R., Fabre, M., Omaïs, B., Marmiesse, M., Supply, P., and Vincent, V. (2005) Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. *PLoS Pathog.* **1**, 0055–0061

3. Van Ingen, J., Boeree, M. J., De Lange, W. C. M., Hoefsloot, W., Bendien, S. A., Magis-Escurra, C., Dekhuijzen, R., and Van Soolingen, D. (2008) Mycobacterium xenopi clinical relevance and determinants, the Netherlands. *Emerg. Infect. Dis.* **14**, 385–389

4. Piersimoni, C., Zitti, P. G., Nista, D., and Bornigia, S. (2003) Mycobacterium celatum pulmonary infection in the immunocompetent: Case report and review. *Emerg. Infect. Dis.* **9**, 399–402

5. Molteni, C., Gazzola, L., Cesari, M., Lombardi, A., Salerno, F., Tortoli, E., Codecasa, L., Penati, V., Franzetti, F., and Gori, A. (2005) Mycobacterium lentiflavum infection in immunocompetent patient. *Emerg. Infect. Dis.* **11**, 119–122

6. Velayati, A. A., Boloorsaze, M. R., Farnia, P., Mohammadi, F., Karam, M. B., Soheyla-Zahirifard, and Masjedi, M. R. (2005) Mycobacterium gastri causing disseminated infection in children of same family. *Pediatr. Pulmonol.* **39**, 284–287

7. Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F., and van Sinderen, D. (2007) Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum. *Microbiol. Mol. Biol. Rev.* **71**, 495–548

8. Engel, P., Goepfert, A., Stanger, F. V., Harms, A., Schmidt, A., Schirmer, T., and Dehio, C. (2012) Adenylylation control by intra-or intermolecular active-site obstruction in Fic proteins. *Nature*. **482**, 107–110