Biochemical evidence for relaxed substrate specificity of Nα-acetyltransferase (Rv3420c/rimI) of Mycobacterium tuberculosis

Deepika Pathak1,*, Aadil Hussain Bhat1,*, Vandana Sapehia1, Jagdish Rai2 & Alka Rao1

Nα-acetylation is a naturally occurring irreversible modification of N-termini of proteins catalyzed by Nα-acetyltransferases (NATs). Although present in all three domains of life, it is little understood in bacteria. The functional grouping of NATs into six types NatA - NatF, in eukaryotes is based on subunit requirements and stringent substrate specificities. Bacterial orthologs are phylogenetically divergent from eukaryotic NATs, and only a couple of them are characterized biochemically. Accordingly, not much is known about their substrate specificities. Rv3420c of Mycobacterium tuberculosis is a NAT ortholog coding for RimIMtb. Using in vitro peptide-based enzyme assays and mass-spectrometry methods, we provide evidence that RimIMtb is a protein Nα-acetyltransferase of relaxed substrate specificity mimicking substrate specificities of eukaryotic NatA, NatC and most competently that of NatE. Also, hitherto unknown acetylation of residues namely, Asp, Glu, Tyr and Leu by a bacterial NAT (RimIMtb) is elucidated, in vitro. Based on in vivo acetylation status, in vitro assay results and genetic context, a plausible cellular substrate for RimIMtb is proposed.

1CSIR-Institute of Microbial Technology, Sector 39-A, Chandigarh, 160036, India. 2Institute of Forensic Science & Criminology, Panjab University, Sector 14, Chandigarh, 160014, India. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to A.R. (email: raoalka@imtech.res.in)
mode of Nα-acetylation is indicated. However, unequivocal in vitro proof for a posttranslational Nα-acetylation has been established for RimL15 alone, that acetylates N-terminal Ser residue of purified full-length ribosomal protein L12. Mycobacterium tuberculosis (Mtb) is the etiological agent of Tuberculosis. Posttranslational modifications of proteins are implicated in the virulence of this pathogen16. Lysine side chain acetylation of proteins is known for quite some time in Mtb, but acetylation of N-termini of proteins is taken note of, only recently. The first reported Nα-acetylated protein of Mtb, EsxA (ESAT-6, for early secreted antigenic target, 6kDa), is part of Mtb secretome in a human host cell. Its interaction with EsxB (CFP-10, for culture filtrate protein, 10kDa) is an important determinant of virulence in Mtb17,18. The mechanism behind the Nα-acetylation of EsxA and how this modification affects Esx-1-mediated secretion and virulence are unknown. In a more recent study, using transposon insertion mutagenesis of a non-cytotoxic Mycobacterium marinum strain a putative NAT (MMAR_0039 gene) is identified as protein responsible for homeostasis of EsxA Nα-acetylation. It also proposes an inverse correlation between EsxA acetylation and virulence19. However, our sequence searches did not reveal any corresponding homolog of MMAR_0039 in Mtb.

In two isolated studies that were aimed at annotating translational start sites (TSS), as many as 253 out of a total 874 peptides were identified to be Nα-acetylated in Mtb20. This indicates that at least 28% proteins constitute Mtb Nα-acetylome. Further, many of these proteins are important for virulence and form Mtb secretome in human macrophage cell21. Sequence analysis of mycobacterial Nα-acetylated peptides suggests the presence of at least one NAT with eukaryotic NatA-like substrate specificity that is involved in acetylating approximately 84% of the protein substrates (Supplementary Figure S1). The closest homolog of NatA catalytic subunit Ard1 is RimI, in bacteria22, sharing 31% similarity. As described already, bacterial Rim enzymes are considered ribosomal protein (substrate) specific. Therefore, it was intriguing to see if protein RimIMtb could acetylate ribosomal proteins or additional substrates in Mtb.

Interestingly, gene Rv3420c encoding RimIMtb is present in a genomic context that is conserved in 20% of bacterial genomes23. Most of the genes clustered along with rimI, namely groEL1 and groES (chaperone encoding genes) and tsaD, tsaB, and tsaE genes (encoding homologs of the tRNA-A37-tA transferase enzymes)23 are essential for Mtb survival and/or pathogenesis as indicated in Tuberculist database24. Therefore, we were interested in investigating the functional role of RimIMtb.

**Results**

**RimIMtb is a monomer in solution.** Rv3420c from Mtb was cloned, expressed and purified as a C-terminal His-tagged protein (RimIMtb) (Materials and Methods, Supplementary Tables S1 and S2). Gel filtration chromatography and intact mass analysis of purified protein by LC-ESI-MS suggested that RimIMtb is a monomer of 19.1 kDa (Fig. 1a,b). The identity of the protein was confirmed by peptide mass fingerprinting.

**RimIMtb acetylates N-terminus of rpS18 (ribosomal protein).** Ribosomal protein S18 (rpS18) is the only known substrate of RimI in bacteria namely, E. coli12 and S. typhimurium13. Mtb genome contains two putative ORFs for ribosomal protein S18 namely, Rv0055 (rpS1R) and Rv2055 (rpS2R). Therefore, acetyltransferase activity of RimIMtb was identified and assessed using substrate peptides representing first six residues (except N-terminal-Met) of rpS18 as well as rpsR1 and rpsR2 (Table 1). A clear mass increase of 42.0105 Da was identified in rpS18 that was concomitant with the addition of an acetyl group (Fig. 1c,e). Further, MS/MS analysis confirmed the site of acetylation as the N-terminal amino acid (Fig. 1d,f). However, N-term of rpsR1 (DP1) and rpsR2 (DP2) were not acetylated efficiently (Supplementary Figure S2). Therefore, acetylation of Mtb ribosomal proteins at N-terminal Ala could not be ascertained. Our result reinforces previous observations that Nα-acetylation signatures are not absolute and downstream residue differences might bring variable outcomes9.

**Relaxed acceptor substrate specificity of RimIMtb.** Protein sequence analysis of the Nα-acetylome of Mtb reveals that at least 54% of proteins have Thr at their N-term followed by Ser in 19% and Met in 12% of the proteins. In other words, Mtb Nα-acetylome primarily represents substrates preferred by eukaryotic NatA (Thr, Ser and Ala termini) and NatE (Met termini) (Supplementary Figure S1). Therefore, to explore the substrate specificity of RimIMtb, a small-scale tryptic peptide library (STPL) assay was developed (Fig. 2a, described in detail in Materials and Methods). Several peptides were detected as acetylated at their neo N-term (Fig. 2b). While the same remained un-acetylated in corresponding control (without RimIMtb) assays (Fig. 2b). N-terminal residues identified acetylated include Ala, Ser, Val, Leu, Tyr, Asp and Glu. While Ala and Ser are well known NatA substrates, Leu and Tyr are hitherto unidentified. Asp and Glu residues are modified by Naa10p catalytic subunit alone when not in complex with its auxiliary unit Naa15p8. From the results of STPL assays, we conclude that RimIMtb has relaxed specificity in vitro and it is promiscuous to substrates varying in length as well as sequence (Fig. 2b).

To decipher the precise substrate preference of RimIMtb, substrate peptide DPC (NatA substrate) was custom synthesized with single residue modifications at its N-term to represent substrate specificities of NatE (DP9), NatB (DP10), NatC (DP11), and novel substrate Leu (DP8) and tested, accordingly (Table 1). All the peptides were modified by RimIMtb (Fig. 3). A relative comparison of the ratios of intensities of modified vs. unmodified substrate peptides (having similar ionization efficiencies and reactions carried out under identical conditions) revealed that RimIMtb, in fact, could acetylate DP9 (NatE substrate) 18 fold better than DPC (NatA substrate) (Supplementary Table S3). Further, to ascertain the specificity of the enzyme, quantitative DTNB assays were conducted (as described in Materials and Methods) and specific activity of RimIMtb against typical NAT substrates was determined. As observed in Fig. 4, DTNB assay clearly confirmed the relaxed specificity of RimIMtb, encompassing that of at least three major eukaryotic NATs namely, NatA, NatC and NatE. However, RimIMtb acetylated...
NatE substrate more efficiently than that of other NATs. Further, from DTNB assay results a positive correlation between hydrophobicity of N-terminal residue of the substrate and enzyme activity was observed (Fig. 4).

Kinetic assessment of RimIMtb catalyzed acetylation of peptide substrate was carried out by selected ion monitoring (SIM) of the acetylated product through LC-ESI-MS/MS, where DP8 (ARYFRR) acted as an acceptor (Supplementary Methods and Supplementary Figure S4). Acetylated-DPC (Ac-ARYFRR) was chosen as the internal standard for its similar molecular weight and ionization efficiency as that of the product. Obtained $K_m$
(for DP8) 2.9 ± 0.9 mM indicated that acetylation efficiency of RimIMtb towards DP8 is similar to the efficiency observed previously in the case of RimI of S. typhimurium for Ala residue (2.2 ± 0.2 mM)13.

Stringent donor substrate specificity of RimIMtb. In addition to NAT activity, human and yeast NATs (Naa10p and Naa50p) are known to possess in vitro and in vivo, N-terminal-propionyltransferase activity9. Using acetyl coA or propionyl coA or succinyl coA as donor substrate while keeping the acceptor substrate identical (DP8 having sequence LRYFRR), RimIMtb was found to accept only acetyl coA as a donor (data not shown). Thus, RimIMtb has stringent donor substrate specificity in comparison to eukaryotic NATs.

RimIMtb acetylates N-terminus of neighboring non-ribosomal proteins. RimIMtb is present in an interesting conserved genomic context composed of homologs of the tRNA- A₃₇-t₆₄ transferase enzymes and essential chaperone proteins GroES and GroEL1 (as discussed in the introduction). GroES is known to be Nα-acetylated in the mycobacterial proteome, previously6 while both GroES and GroEL are observed to be Nα-acetylated in Pseudomonas aeruginosa8. Using STPL assay (as described already), we had successfully identified the Nα-acetylation of TsaD N-terminus by RimIMtb (Fig. 2b highlighted in red). Keeping in view the relaxed specificity of RimIMtb, we next explored if these neighboring proteins could be the substrate (s) for RimIMtb. Therefore, peptides representing N-terminus of other neighboring proteins namely, TsaB (DP4), TsaE (DP5), GroEL1 (DP6) and GroES (DP7) were custom synthesized and tested as substrates. RimIMtb could acetylate peptides representing N-terminus of GroES and GroEL1 (Table 1 and Fig. 5 and Supplementary Figure S3). However, highest specific activity of RimIMtb (231.69 μmoles/mg/min) was observed against GroES peptide (DP7) alone (Supplementary Figure S3).

| Substrate Protein | Gene Name (Organism) | Abbreviation | Peptide Sequence | Specificity |
|-------------------|----------------------|--------------|-----------------|------------|
| Ribosomal proteins | rpsR18 (S. typhimurium) | DPC | ARYFRR | NatA like |
| | rpsR1/Rv0055 (Mtb) | DP1 | AKSSKR | – |
| | rpsR2/Rv2055 (Mtb) | DP2 | AAKSAR | – |
| Neighboring proteins | tsaB/Rv3421c (Mtb) | DP4 | SRYQIS | NatA like |
| | tsaE/Rv3422c (Mtb) | DP5 | SREGIR | – |
| | groEL1/Rv3417c (Mtb) | DP6 | SLEKSY | NatA like |
| | groES/Rv3418c (Mtb) | DP7 | AKVNIK | NatA like |
| Custom synthesized peptides | DP8 | LRYFRR | Novel |
| | DP9 | MARYFRR | NatE like |
| | DP10 | MERYFRR | NatB like |
| | DP11 | MLRYFRR | NatC like |

Table 1. List of acceptor peptides used as substrates for in vitro NAT assays.

Figure 2. Small-Scale Tryptic Peptide Library (STPL) assay. (a) Schematic representation of STPL assay (b) Sequences of peptides (belonging to tryptic digests of CarD, RpoA and TsaD) Nα-acetylated at neo-terminal amino acid residues by RimIMtb, in vitro. Nα-acetylated N-terminus of TsaD is highlighted in red.
Figure 3. Relaxed substrate specificity of Rim\textsubscript{Mtb}. Results of NAT assays with substrate peptides DP8, DP9, DP10 and DP11 analyzed using MALDI-TOF/TOF. (a) MS/MS of 910.5 Da precursor ion (S/N = 4201) of unmodified and 952.5 Da precursor ion (S/N = 5894) of modified substrate peptide representing novel substrate specificity of Rim\textsubscript{Mtb} (b) MS/MS of 999.5 Da precursor ion (S/N = 7037) of unmodified and 1041.5 Da precursor ion (S/N = 4256) of modified substrate peptide representing N-terminus of conventional NatE substrate (c) MS/MS of 1057.8 Da precursor ion (S/N = 15916) of unmodified and 1099.5 Da precursor ion (S/N = 1011) of modified substrate peptide representing N-terminus of conventional NatB substrate (d) MS/MS of 1041.5 Da precursor ion (S/N = 17767) of unmodified and 1083.5 Da precursor ion (S/N = 14629) of modified substrate peptide representing N-terminus of conventional NatC substrate.

Figure 4. Determination of specific activity of Rim\textsubscript{Mtb} towards peptide substrates using DTNB assay. The amount of product (CoASH) generated as a result of acetyl transfer was monitored at 412 nm. The concentration of the product was quantified using CoASH calibration curve (inset) and specific activity of Rim\textsubscript{Mtb} against each substrate plotted in terms of μmoles/mg/min. The results shown here represent mean ± SD of experiments performed in triplicate.
Having understood that RimIMtb could acetylate N-terminus of at least three of its neighboring proteins with varying efficiencies, GroES, and TsAB-TsAD (complex) were expressed and purified successfully from *E. coli*. Enzyme assays utilizing C14 labeled acetyl-coa (Perkin Elmer) were set up and analyzed by SDS-PAGE. Interestingly, none of these folded proteins could be acetylated by RimIMtb. In the case of GroES, first 7–20 amino acid residues are known to be involved in monomerization23, therefore, might not be accessible to RimIMtb in folded conformation. However, nothing is known about the folded state of TsAB and TsAB-TsAD complex. This indicates that secondary structure of folded protein substrates could be an important determinant of Nα-acetylation activity of the enzyme as noted previously10.

**Plausible cellular substrate for RimIMtb.** In our experiments, we find that a) RimIMtb does acetylate peptides representing N-terminus of GroES, GroEL1, and TsAD proteins, in *vitro*; b) rimI, groES, groEL1, tsAB, tsaE and tsaD are co-regulated as part of a single operon (results are part of a separate study and communicated elsewhere); c) significant (P value < 0.05) specific activity of RimIMtb was observed against peptide representing N-terminus of GroES. Further, GroES and GroEL are identified Nα-acetylated in bacterial proteomes, previously6,8. Therefore, we propose GroES as one of the plausible cellular substrates for RimIMtb in *Mtb*. In eukaryotes, Nα-acetylation of GroES imparts increased resistance towards proteolytic degradation and longer half–life as compared to the non-acetylated (recombinant) counterpart38. However, nothing is yet understood about the role of such acetylation in bacteria.

**In silico modeling of RimIMtb, a comparison viz a viz other known NATs.** In bacteria, Rim enzymes possess stringent substrate specificities. They are known to acetylate Ala and Ser residues only that are conventional NatA substrate. It was interesting to observe that RimIMtb could acetylate a conventional NatA substrate (Met N-termini), more efficiently than the expected NatA substrates (Fig. 4). Therefore, we attempted in *silico* modeling of RimIMtb using ITSSARC27. The best model was selected based on lowest RMSD value that shared maximum similarity with TvArD of Thermoplasma volcanium (PDB ID: 4pV6)24. TvArD1 like its ortholog SsArD1 from *Sulfolobus solfataricus* (*S. solfataricus*) belongs to NatA family and acetylates Alba1 protein at N-terminal Ser residue. The selected model of RimIMtb was then docked with DP9 peptide (MARYFRR) (Supplementary Figure S5) using flexpepdock server29. Further, LPC/CSU server30 was employed to identify the key residues involved in stabilizing interactions, and contact analysis of substrate peptide docked on to the RimIMtb model. Sequence analysis and structural alignment of docked RimIMtb, Naa50p (3TFY/NatE), and TvArD1 (4pV6/NatA) show that the key catalytic residues that mediate recognition and acetylation of Met residue in Naa50p35 are conserved in both RimIMtb and TvArD1 (Supplementary Figure S5). V29 of Naa50p aligned completely with E34 of TvArD1 (Fig. 6a). However, according to LPC/CSU analysis, corresponding residue E25 of RimIMtb neither participated in substrate interactions nor aligned structurally with V29 and E34 of Naa50p and TvArD1, respectively. Rather, hydrogen bonding between N-terminal Met of substrate and Y140 of enzyme RimIMtb was found to be stabilizing (Fig. 6a). Residue Y140 of RimIMtb is conserved in Naa50p (Y139) and TvArD1 (Y141) and is previously known to be crucial for catalysis by Naa50p. Further, based on docking results and contact analysis, residues L28, F29, I82, L112 and A121 of RimIMtb provide a putative hydrophobic pocket for NatE like substrates (Fig. 6b). This pocket explains the higher specific activity of RimIMtb as observed towards substrates with a hydrophobic residue at N-termini (Fig. 4) and poor activity against substrates containing polar residues like Ser, Glu (Supplementary Figure S3). For efficient catalysis of NatA substrates, Glu35 is identified as an important residue in archaeal ortholog SsArD1. The same residue is attributed with lowering the acetylation efficiency of SsArD1 for Met N-termini due to steric hindrance32. Interestingly, this residue is conserved amongst prokaryotic NATs, TvArD1, SsArD1, and RimIMtb (Supplementary Figure S5, residues highlighted in red), yet RimIMtb is capable of efficient N-terminal Met acetylation of the substrate peptide.

Further, the shapes of the binding pocket of RimIMtb, Naa50p and TvArD1 were analyzed by calculating surface area and volume of the binding site using CASTp server31. The calculated ratios of surface area to volume of the binding pockets of RimIMtb, TvArD1 and Naa50p were 0.8, 0.83 and 0.57, respectively indicating deeper narrower substrate binding site of RimIMtb in comparison to that of Naa50p (Fig. 6c–e). The narrow tunnel shape of the substrate binding pocket suggests that the N-terminus of the substrate proteins needs to be accessible for Nα-acetylation, which could be the reason for the inability of RimIMtb to acetylate purified, folded proteins in our experiments.

**Curious genetic context of RimIMtb in the light of its protein-protein interactions.** As discussed under introduction section, rimI is positioned upstream of essential chaperone genes, groES and groEL1 and packed in-between homologs of essential tRNA-AAA- t4A transferase enzymes: tsaD, tsaB, tsaE. While chaperones are important for survival during various stress conditions, tRNA-AAA- t4A modification of tRNA is universally conserved and considered critical for maintaining translational fidelity25. Though the context is quite conserved, the significance of rimI here is not understood. Interestingly, in eukaryotes, NatA is known to interact with a chaperone-like protein HYPK (Huntington Interacting Protein)34. HYPK assists co-translational Nα-acetylation activity of NatA and in its absence, Nα-acetylation decreases. Also, rimI is seen fused with tsAB in *Bordetella* and *Mycobacterium leprae* genomes and a weak in *vivo* interaction between RimI and TsAB is reported in *E. coli*, previously35 (Fig. 7a).

Therefore, employing a previously described Mycobacterial Protein Fragment Complementation (MPFC) assay36, we explored the *in vivo* interactions of RimIMtb with proteins of its conserved neighborhood. For the assay, rimI, two upstream genes: tsaB, tsaE and three downstream genes: tsaD, groEL1 and groES were cloned independently and separately in vectors namely, pMD101 and pMD102 that harbor complementary domains of murine dihydrofolate reductase (mDHFR) (Supplementary Tables S1 and S2). These recombinant vectors were then co-transformed in *Mycobacterium smegmatis* (*M. smegmatis*) in all combinations. A successful *in vivo* protein-protein interaction reconstitutes mDHFR in co-transformants and aids survival of cells on the otherwise
Figure 5. Acetylation of N-termini of neighboring proteins by RimIMtb. (a) MS/MS of 672.4 Da precursor ion (S/N = 1999) of unmodified and 714.4 Da precursor ion (S/N = 557) of modified substrate peptide representing N-terminus of GroES. (b) MS/MS of 752.3 Da precursor ion (S/N = 11634) of unmodified and 794.4 Da precursor ion (S/N = 1984) of modified substrate peptide representing N-terminus of GroEL1. (c) MS/MS of 689.4 Da precursor ion (S/N = 409) of unmodified and 731.3 Da precursor ion (S/N = 138.5) of modified substrate peptide representing N-terminus of TsaB.
inhibitory concentration of Trimethoprim. Accordingly, in our experiments, we observed reliable in vivo interactions between RimIMtb and GroES (the chaperone) as well as between RimIMtb and TsaD (the main actor protein of tRNA-A37-t6A transferase enzymes) (Fig. 7b). GroEL1 was observed to interact with all the neighboring proteins including the mDHFR of vector backbone (data not shown). The promiscuous interaction of GroEL1 with several proteins including mDHFR for its chaperone nature is well documented. Interestingly, we could not detect anticipated interaction between RimIMtb and TsaB and between TsaB and TsaD that is previously well characterized, in our MPFC experiments. For the reason that we could successfully purify TsaD-TsaB as co-eluting protein complex from *E. coli*, it is possible that the fusion product of TsaB-DHFR adopts a conformation that is not accessible for protein-protein interaction in MPFC system. On the other hand, TsaD is not known to interact with TsaE directly but through the formation of a ternary complex in *E. coli* and *S. typhimurium*. Again, we found this to be contrary to our system and observed a direct in vivo interaction between TsaD and TsaE. Similar to our observation, recently TsaD-TsaE interaction were observed in vitro in calorimetric experiments with purified proteins in *S. typhimurium* in the presence of ATP. Nonetheless, our results indicate a plausible, biologically relevant involvement of RimIMtb with tRNA-A37-t6A transferase enzymes in mycobacteria and explain conservation of rimI genetic context, in part. Here observed cellular protein-protein interactions of RimIMtb may have far reaching implications or additional physiological relevance that needs to be explored further.

**Discussion**

Protein Nα-acetylation is a co/post-translational modification that is implicated in a variety of cellular processes and regulation thereof, in eukaryotes, but it is little understood in bacteria. A host of Nα-acetylated proteins have been identified in the pathogen, *Mtb* and *S. typhimurium*. Enzymes belonging to GNAT family of proteins that acetylate amino group of aminoglycosides are well known in mycobacteria. For example, Eis is a lysine acetyltransferase that
acetylates DUSP16/MKP-7 in *Mtb* while its ortholog acetylates antibiotic kanamycin in *M. smegmatis*. In the present work, we have characterized first protein/peptide Nα-acetyltransferase (RimIMtb) of *Mtb*. In our study, we identify that RimIMtb is a monomer in solution. We have developed a quick tryptic peptide-based method (STPL) to identify novel substrate specificities of such enzymes. RimIMtb acetylates a wide variety of N-terminal residues including hitherto unknown residues like Leu, Tyr, Glu, and Asp, in vitro. However, significant enzyme activity is observed against hydrophobic N-termini only. Further, the N-terminus of peptides representing important/essential proteins of *Mtb* namely, chaperones: GroES and GroEL1 were acetylated by RimIMtb. To conclude, in vitro enzyme assays showed that RimIMtb exhibits promiscuous acceptor substrate specificity which overrides the previous notion of specific or stringent substrate requirements of prokaryotic NATs. The only other known NAT of relaxed specificity is archaeal SsArd1. While acceptor specificity of SsArd1 is closer to eukaryotic NatA, RimIMtb acetylates substrate akin eukaryotic NatE substrates, more efficiently. These experimental observations were substantiated well by in silico modeling and peptide docking studies. In *Mtb* at least 12% of reported Nα-acetylome bears Nα-acetylated Met at N-terminus. Therefore, RimIMtb by its NatE like specificity could be the primary NAT responsible for acetylating such substrates in *Mtb*.

**Figure 7. In vivo protein-protein interactions between RimIMtb and neighboring proteins.** (a) Schematic representation of genetic context of rimI in *Mtb* (b) Mycobacterial Protein Fragment Complementation (MPFC) assay elucidating in vivo protein-protein interactions between 1) RimIMtb and GroES 2) RimIMtb and TsaD and 3) TsaD and TsaE, where corresponding co-transformants got selected on Trimethoprim supplemented at a concentration of 50 μg/ml.
Additionally, we have attempted to investigate the significance of conserved genomic context of rimI using in vivo MPFC assays. Our results accordingly suggest hitherto unnoticed involvement of RimIMtb with essential chaperones and tRNA-Asp-t6A transferase enzymes in Mtb. This study we believe is an important advance in the understanding of the process and proteins involved in No-acetylation in Mtb, in specific and bacteria, in general.

Materials and Methods
Chemicals and Reagents. All chemical reagents were purchased from Sigma-Aldrich unless otherwise noted. Materials used for protein production and purification were purchased from GE healthcare. Restriction enzymes, molecular ladders were purchased from NEB and Fermentas (Thermo Fisher Scientific). Synthetic peptides were purchased from GLS China.

Cloning and expression of Rv3420c/rimI from Mtb. Gene encoding Riml homolog in Mtb, Rv3420c (accession no.: CCF46242.3) was PCR amplified from Mtb genomic DNA (BEI Resources) using primers 20FP and 20RP (Supplementary Table S1) and cloned in NcoI and HindIII restriction sites of PET28a (Novagen) and confirmed by sequencing. The C-terminal His-tagged construct, RimIMtbPET28a was transformed in Lemo21 (DE3) (NEB #C2528H) cells for expression. The cells were grown at 37 °C till 0.6 O.D. was reached, induced with 0.1 mM Isopropyl-β-D-1-thiogalactopyranoside (IPTG) and incubated overnight, in shaking condition at 200 rpm, at 16 °C. The cells were subsequently harvested and pellets stored at −80 °C.

Purification of RimIMtb protein. The stored pellets were thawed and re-dissolved in Lysis buffer: (20 mM Tris pH 8.0, 1 M NaCl, 10% glycerol, 10 mM imidazole and 4 mM β-mercaptoethanol). The cells were sonicated and lysed so obtained centrifuged at 12000 rpm to remove debris. The supernatant was applied to Ni-NTA resin equilibrated with lysis buffer and protein was allowed to bind for one hour at 4 °C with gentle shaking. The beads were then washed with washing buffer (20 mM Tris pH 8.0, gradient of 1 M NaCl-300 mM NaCl, 10% glycerol and 30 mM imidazole). Bound protein was eluted in 10 ml of Elution buffer (20 mM Tris pH 8.0, 300 mM NaCl, 10% glycerol and 4 mM β-mercaptoethanol). A fraction of eluted protein was subjected to gel filtration chromatography using Superdex 75 10/300 GL column (GE healthcare, calibrated with GE LMW standards) with 20 mM Tris pH8.0, 300 mM NaCl, 10% glycerol and 4 mM β-mercaptoethanol. Fractions were collected, pooled and concentrated using Amicon 10 kDa molecular weight cut off filters. For enzyme assay, protein was dialyzed extensively at 4 °C in dialysis buffer (20 mM HEPES pH8.0, 150 mM NaCl, 10% glycerol, 2 mM DTT). The purified protein was cleaned up using ziptip C18 (Merck Millipore) to remove salts according to manufacturer’s protocol, and subjected to LC-ESI-MS (using Agilent 6550 iFunnel Q-TOF LC/MS) for intact mass analysis. RimIMtb was resolved using SDS-PAGE and in-gel trypsin digestion (Sigma # T6567) carried out according to manufacturer’s protocol. The MS spectra so obtained were searched against Mtb protein database using online MASCOT server (http://www.matrixscience.com/).

Qualitative acetyltransferase activity assay (NAT assay). For determining acetyltransferase activity of the enzyme, peptides consisting of first six amino acid residues (without the initiator Met) of putative protein substrates were synthesized (listed in Table 1). 100 μM of each peptide was added to a reaction mixture consisting of 20 mM HEPES pH 8.0, 100 μM acetyl coA and 4 μM RimIMtb protein. Reactions controls for the assays included, a) Substrate control A (Peptide alone), b) Substrate control B (Peptide + acetyl coA) and c) Enzyme control (Peptide + acetyl coA + dummy protein BSA). The reactions were incubated at 25 °C in a water bath, overnight. The samples were cleaned up using ziptip C18 and subjected to MALDI-TOF MS analysis (using AB SCIEX TOF/TOF™ 5800). The data was acquired using following settings: 1 μl of analyte was combined with 1 μL of the matrix (α-cyano-4-hydroxy-cinnamic acid matrix in 50% ACN/50% water with 0.1% TFA) and spotted onto an MALDI target and dried under ambient conditions before analysis. MS data was acquired in positive ion mode using fixed laser intensity of 3400, keeping the mass range 600–4000 Da, the total number of shots 2000, bin size 0.5 ns and pulse rate 400 Hz. MS/MS data was acquired using fixed laser intensity of 4300 keeping the mass range 400–700 Da and adduct tolerance 0.03. Precursors having S/N below 20 and resolution within 200 were excluded. The data obtained (2d files) was analyzed by Data Explorer software (version 4.6) and mass peaks plotted in Origin (OriginPro 2015 b.9.2.214).

Quantitative acetyltransferase assay (DTNB based). DTNB assay or Ellman’s assay is a colorimetric assay that detects generation of free sulphydryl group. In an acetyltransferase reaction, the products formed are acetylated substrate and CoASH. This free CoASH reacts with DTNB to form mixed disulfide and TNB which gives a yellow color at 412 nm. Thus, enzyme assays were set up according to published protocol45,46 along with control reactions a) Substrate control (Peptide + acetyl coA) b) Enzyme control (acetyl coA + RimIMtb). Briefly, reaction components were added in 50 mM Tris Buffer pH 8.0, 4 mM peptide substrate, 0.5 mM acetyl coA in a total reaction volume of 0.1 ml. Reactions were initiated by addition of 5 μM of RimIMtb and samples incubated at 25 °C for one hour. Enzyme activity was quenched by addition of 60 μl of 3.2 M Gn-Hcl and 10 μl of 2 mM DTNB was added. The reactions were vortexed and allowed to stand for 5 min before absorbance was measured at 412 nm in a plate reader. Absorbance of enzyme control was subtracted from enzyme assay readouts. For each peptide, the amount of product formed was calculated from CoASH calibration curve (0.1 mM–1 mM) and specific activity of the enzyme calculated. The mean average value and standard deviation for triplicate experiments were plotted.

Donor specificity assay. For determining donor substrate specificity, RimIMtb was incubated with 100 μM of peptide substrate (DP8) and 500 μM each of acetyl coA, succinyl coA and propionyl coA at 25 °C overnight. Data was acquired as mentioned above.
Small-scale Tryptic Peptide Library assay (STPL assay). Trypsin digestion. 10 µg of each purified, mass-verified protein was reconstituted in 100 µl of 6 M urea and trypsin digested according to manufacturers’ protocol. Briefly, the samples were reduced by addition of DTT and kept at 37 °C for one hour. The samples were cooled down and alkylated using iodoacetamide and again kept at 37 °C for one hour at RT in dark. 20 mM of ammonium bicarbonate (pH 8.0) was added to dilute concentration of urea to 1 M. Trypsin (Sigma: T6567) was added in a ratio of 1:50 and incubated at 37 °C overnight. The samples were speed vacuumed to 100 µl and ziptip purified.

Enzyme assay (NAT assay). For each set of proteins, one control reaction consisting of peptides generated by digest and acetyl coA and another enzyme reaction consisting of the tryptic digest, acetyl coA and RimL protein was set up in 20 mM HEPES buffer pH 7.5. The reactions were incubated at 25 °C overnight, ziptip purified and subjected to PMF analysis by MALDI.

Database search. Each control and enzyme MS spectra obtained was searched using online MASCOT Server against Mycobacterium complex in Swiss-Prot database with single missed cleavage of trypsin allowed. Carbamidomethylation of Cys (+57) was set as fixed modification while Nα-acetylation (+42) was set as variable modification. Mass tolerance for precursor ion was set as 0.2 Da and MASCOT report was set at AUTO to obtain only the protein/peptide hits with significant scores (threshold score for identified peptide was >30).

In vivo protein-protein interaction studies: MPFC. rimI and its neighboring genes namely, Rs3417c, Rs3418c, Rs3419c, Rs3421c, Rs3422c were PCR amplified from Mtb genomic DNA using respective primers (Supplementary Table S1) and cloned in EcoRI – HindIII and BamHI – HindIII sites of pMD101 (Kan+) and pMD102 (Hyg+)

Hygromycin (LB-Kan+pMD102 (HygR), respectively. The constructs (Supplementary Table S2) were co-transformed in all possible combinations in M. smegmatis mc2155 electrocompeent cells and grown on LB medium containing Kanamycin and Hygromycin (LB-Kan+Hyg) at concentration 50 µg/ml. The co-transformants were selected on LB-Kan+Hyg plates containing Trimethoprim (Trim) at conc. of 50 µg/ml to decipher positive interaction between corresponding proteins. Co-transformants bearing (pMD101 + pMD102), (pMD101 + pMD102-test gene fusion) and (pMD102-test gene fusion + pMD101) served as negative controls.

Statistical analysis. All statistical analysis was performed using one-way ANOVA with Dunnett’s multiple comparison test.

References
1. Stkarheim, K. K., Gevaert, K. & Arnesen, T. Protein N-terminal acetyltransferases: when the start matters. Trends Biochem. Sci. 37, 152–61 (2012).
2. Myklebust, L. M. et al. Biochemical and cellular analysis of Ogden syndrome reveals downstream N-acytlation defects. Hum. Mol. Genet. 24, 1956–76 (2015).
3. Kalvik, T. V. & Arnesen, T. Protein N-terminal acetyltransferases in cancer. Oncogene 32, 269–276 (2012).
4. Bienvenut, W. V., Giglione, C. & Meinnel, T. Proteome-wide analysis of the amino terminal status of Escherichia coli proteins at the steady-state and upon deacetylation inhibition. Proteomics 15, 2503–2518 (2015).
5. Falb, M. et al. Archaeal N-terminal protein maturation commonly involves N-terminal acetylation: a large-scale proteomics survey. J. Mol. Biol. 362, 915–24 (2006).
6. Kelkar, D. S. et al. Proteogenomic Analysis of Mycobacterium tuberculosis By High Resolution Mass Spectrometry. Mol. Cell. Proteomics 10, M111.01627–M111.01627 (2011).
7. Kirkland, P. A., Humbard, M. A. & Daniels, C. J. & Maupin-Ferret, J. Posttranslational Modifications of Ribosomal Proteins in Escherichia coli. Protein Sci. 17, 780–795 (2008).
8. Ouidir, T., Jarnier, F., Cosette, P. & Hardouin, J. Characterization of N-terminal protein modifications in Pseudomonas aeruginosa PA14. J. Proteomics 114, 214–25 (2015).
9. Varland, S., Osberg, C. & Arnesen, T. N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects. Proteomics n/a–n/a: doi: 10.1002/pmic.201400619 (2015).
10. Polevoda, B. & Sherman, F. N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. J. Mol. Biol. 325, 595–622 (2003).
11. Tanaka, S., Matsushita, Y., Yoshikawa, A. & Isoko, K. Cloning and molecular characterization of the gene riml which encodes an enzyme acetylating ribosomal protein L12 of Escherichia coli K12. Mol. Gen. Genet. 217, 289–93 (1989).
12. Yoshikawa, A., Isoko, S., Sheback, A. & Isoko, K. Cloning and nucleotide sequencing of the genes riml and rimj which encode enzymes acetylating ribosomal proteins S18 and S5 of Escherichia coli K12. Mol. Gen. Genet. 209, 481–8 (1987).
13. Vetting, W. M., Bareich, D. C., Yu, M. & Blanchard, J. S. Crystal structure of RimJ from Salmonella typhimurium LT2, the GNAT responsible for N(alpha)-acetylation of ribosomal protein S18. Protein Sci. 17, 1781–90 (2008).
14. Nesterchuk, M. V., Sergiev, P. V. & Donskoia, O. a. Posttranslational Modifications of Ribosomal Proteins in Escherichia coli. Acta Naturae 3, 22–33 (2011).
15. Vetting, M. W., de Carvalho, L. P. S., Roderick, S. L. & Blanchard, J. S. A novel dimeric structure of the RimL Nα-acetyltransferase from Salmonella typhimurium. J. Biol. Chem. 280, 22108–14 (2005).
16. Calder, B., Soares, N. C., Kock, E. & Blackburn, J. M. Mycobacterial proteomics: analysis of expressed proteomes and post-translational modifications to identify candidate virulence factors. Expert Rev. Proteomics 12, 21–35 (2015).
17. Brodin, P. et al. Functional analysis of early secreted antigenic target-6, the dominant T-cell antigen of Mycobacterium tuberculosis, reveals key residues involved in secretion, complex formation, virulence, and immunogenicity. J. Biol. Chem. 280, 33953–9 (2005).
18. Okkels, L. M. et al. CFP10 discriminates between nonacetylated and acetylated ESAT-6 of Mycobacterium tuberculosis by differential expression. Proteomics 4, 2954–2960 (2004).
19. Mbah Medie, F., Champion, M. M., Williams, E. A. & Champion, P. A. A. Homeostasis of N-α-terminal acetylation of EsxA correlates with virulence in Mycobacterium marinum. Infect. Immun. 82, 4572–86 (2014).
20. Rison, S. C. G., Mattow, J., Jungblut, P. R. & Stoker, N. G. Experimental determination of translational starts using peptide mass fingerprinting and tandem mass spectrometry within the proteome of Mycobacterium tuberculosis. Microbiology 153, 521–528 (2007).
21. Chande, A. G. et al. Selective enrichment of mycobacterial proteins from infected macrophages. Sci. Rep. 5, 13430 (2015).
22. El-Yacoubi, B. et al. A role for the universal Kae1/Qr7/YgdE family in rRNA modification. EMBO J. 30, 882–893 (2011).
23. Thiaville, P. C. et al. Cross Kingdom Functional Conservation of the Core Universally Conserved Threonine/Carbamoyl/prolyladenosine tRNA Synthesis Enzymes. Eukaryot. Cell 13, 1222–1231 (2014).
24. Lew, J. M., Kapopoulou, A., Jones, L. M. & Cole, S. T. Tuberculist–10 years after. Tuberculosis (Edinb). 91, 1–7 (2011).
How to cite this article: Pathak, D. et al. Biochemical evidence for relaxed substrate specificity of N-acetyltransferase (Rv3420c/rim1) of Mycobacterium tuberculosis. Sci. Rep. 6, 28892; doi: 10.1038/srep28892 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Acknowledgements

D.P. and A.H.B. received fellowships from Council of Scientific and Industrial Research (CSIR). The genomic DNA from Mtb, Strain H37Rv, NR-14865 was obtained from BEI Resources, NIAID. We thank Dr. Ashwani Kumar for helpful discussions and generous gift of mycobacterial protein fragment complementation (MPFC) system. Purified proteins namely, CarD and RpoA were gifted by Gundeep Kaur and Dr. Krishan Gopal. We also acknowledge Uday Agarwal for helpful discussions towards kinetic characterization. This work was supported by DBT (RGYI Grant No. BT/PR15106/GBD/27/317/2011 to Alka Rao) and CSIR (Grant No. BSC0210C to Alka Rao). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author Contributions

A.R. and D.P. conceptualized the project. D.P. and A.H.B. performed the experiments. D.P. and V.S. performed MPFC assays. D.P., J.R. and A.R. carried out in silico modeling of Rim126th. D.P. and A.R. wrote the manuscript. A.R. and J.R. reviewed and edited the manuscript. A.R. arranged the funding.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

www.nature.com/scientificreports/