Molecular profiling of Mycobacterium tuberculosis identifies tuberculosinyl nucleoside products of the virulence-associated enzyme Rv3378c

Molecular profiling of *Mycobacterium tuberculosis* identifies tuberculosinyl nucleoside products of the virulence-associated enzyme Rv3378c

Emilie Layre[a,1], Ho Jun Lee[b,1], David C. Young[a], Amanda Jezek Martinot[c], Jeffrey Buter[d], Adriaan J. Minnaard[d], John W. Annand[a], Sarah M. Fortune[e], Barry B. Snider[f], Isamu Matsunaga[g,2], Eric J. Rubin[c], Tom Alber[b,3], and D. Branch Moody[a,3,4]

[a]Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115; [b]Department of Molecular and Cell Biology and California Institute for Quantitative Biosciences, University of California, Berkeley, CA 94720; [c]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115; [d]Stratingh Institute for Chemistry, University of Groningen, 9747 AG, Groningen, The Netherlands; [e]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115; [f]Department of Chemistry MS 015, Brandeis University, Waltham, MA 02453-2728; and [g]Laboratory of Cell Regulation, Department of Viral Oncology, Institute for Virus Research, Kyoto University, Sakyo-ku, Kyoto 606-8507, Japan

Edited by David W. Russell, University of Texas Southwestern Medical Center, Dallas, TX, and approved January 15, 2014 (received for review August 22, 2013)

To identify lipids with roles in tuberculosis disease, we systematically compared the lipid content of virulent *Mycobacterium tuberculosis* with the attenuated vaccine strain *Mycobacterium bovis* bacillus Calmette–Guérin. Comparative lipidomics analysis identified more than 1,000 molecular differences, including a previously unknown, *Mycobacterium tuberculosis*-specific lipid that is composed of a diterpene unit linked to adenosine. We established the complete structure of the natural product as 1-tuberculosinyladenosine (1-TbAd) using mass spectrometry and NMR spectroscopy. A screen for 1-TbAd mutants, complementation studies, and gene transfer identified *Rv3378c* as necessary for 1-TbAd biosynthesis. Whereas Rv3378c was previously thought to function as a phosphatase, these studies establish its role as a tuberculosinyl transferase and suggest a revised biosynthetic pathway for the sequential action of Rv3377c-Rv3378c. In agreement with this model, recombinant Rv3378c protein produced 1-TbAd, and its crystal structure revealed a *cis*-prenyl transferase fold with hydrophobic residues for isoprenoid binding and a second binding pocket suitable for the nucleoside substrate. The dual-substrate pocket distinguishes Rv3378c from classical *cis*-prenyl transferases, providing a unique model for the prenylation of diverse metabolites. Terpene nucleosides are rare in nature, and 1-TbAd is known only in *Mycobacterium tuberculosis*. Thus, this intersection of nucleoside and terpene pathways likely arose late in the evolution of the *Mycobacterium tuberculosis* complex; 1-TbAd serves as an abundant chemical marker of *Mycobacterium tuberculosis*, and the extracellular export of this amphipathic molecule likely accounts for the known virulence-promoting effects of the Rv3378c enzyme.

TbAd | terpenyl transferase

With a mortality rate exceeding 1.5 million deaths annually, *Mycobacterium tuberculosis* remains one of the world's most important pathogens (1). *M. tuberculosis* succeeds as a pathogen because of productive infection of the endosomal network of phagocytes. Its residence within the phagosome protects it from immune responses during its decades long infection cycle. However, intracellular survival depends on active inhibition of pH-dependent killing mechanisms, which occurs for *M. tuberculosis* but not species with low disease-causing potential (2). Intracellular survival is also enhanced by an unusually hydrophobic and multilayered protective cell envelope. Despite study of this pathogen for more than a century, the spectrum of natural lipids within *M. tuberculosis* membranes is not yet fully defined. For example, the products of many genes annotated as lipid synthases remain unknown (3), and mass spectrometry detects hundreds of ions that do not correspond to known lipids in the MycoMass and LipidDB databases (4, 5).

To broadly compare the lipid profiles of virulent and avirulent mycobacteria, we took advantage of a recently validated metabolomics platform (4). This high performance liquid chromatography–mass spectrometry (HPLC-MS) system uses methods of extraction, chromatography, and databases that are specialized for mycobacteria. After extraction of total bacterial lipids into organic solvents, HPLC-MS enables massively parallel detection of thousands of ions corresponding to diverse lipids that range from apolar polyketides to polar phosphoglycolipids. Software-based (XCMS) ion finding algorithms report reproducibly detected ions as molecular features. Each feature is a 3D data point with linked mass, retention time, and intensity values from one detected molecule or isotope. All features with equivalent mass and retention time from two bacterial lipid extracts are aligned, allowing pairwise comparisons of MS signal intensity to enumerate molecules that are overproduced in one strain with a false-positive rate below 1% (4).

Significance

Whereas most mycobacteria do not cause disease, *Mycobacterium tuberculosis* kills more than one million people each year. To better understand why *Mycobacterium tuberculosis* is virulent and to discover chemical markers of this pathogen, we compare its lipid profile with that of the attenuated but related mycobacterium, *Mycobacterium bovis* Bacillus Calmette–Guérin. This strategy identified a previously unknown *Mycobacterium tuberculosis*-specific lipid, 1-tuberculosinyladenosine, which is produced by the Rv3378c enzyme. The crystal structure of Rv3378c provides information supporting drug design to inhibit prenyl transfer. Discovery of 1-tuberculosinyladenosine provides insight into how *Mycobacterium tuberculosis* resists killing in macrophages and a new target for diagnosing tuberculosis disease.

Author contributions: E.L., H.J.L., I.M., E.J.R., T.A., and D.B.M. designed research; E.L., H.J.L., D.C.Y., A. Jezek Martinot, J.B., and J.W.A. performed research; A. Jezek Martinot, J.B., A. J. Minnaard, S.M.F., B.B.S., and E.J.R. contributed new reagents/analytic tools; E.L., H.J.L., D.C.Y., A. Jezek Martinot, A. J. Minnaard, J.W.A., B.B.S., E.J.R., T.A., and D.B.M. analyzed data; and E.L., H.J.L., B.B.S., T.A., and D.B.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4CMV, 4CMW, and 4CMX).

[1]E.L. and H.J.L. contributed equally to this work.

[2]Present address: Okatani Hospital, 1-25-1 Kyobate-cho, Nara 630-8141, Japan.

[3]T.A. and D.B.M. contributed equally to this work.

[4]To whom correspondence should be addressed. E-mail: bmoody@rics.bwh.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1315883111/-/DCSupplemental.

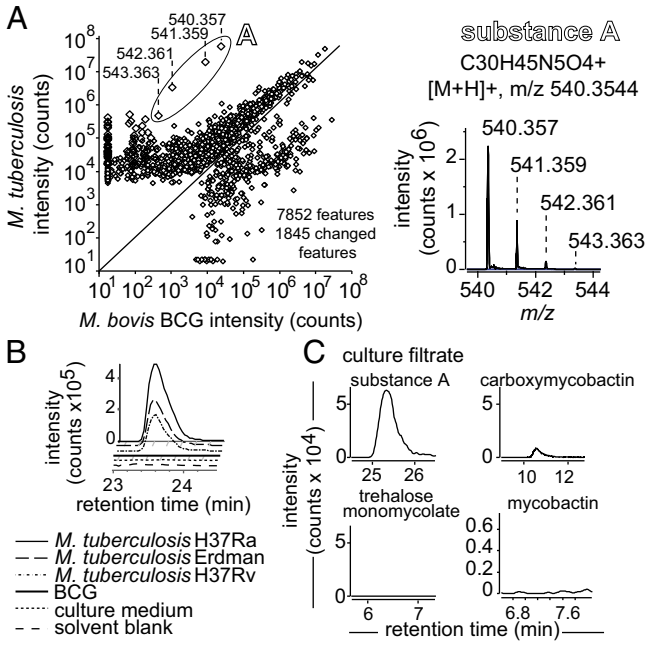

Fig. 1. Comparative lipidomic analysis of *M. tuberculosis* and BCG reveals a natural product constitutively produced and exported by *M. tuberculosis*. (*A*) Detected molecular features are shown as a scatterplot of intensity derived from *M. tuberculosis* H37Rv and BCG lipid extracts. Each feature corresponds to a detected ion and contains retention time and *m/z* values, which are detailed in *SI Appendix*, Dataset S1; 1,845 features out of 7,852 total features showed intensity ratios that deviate significantly from 1 (corrected *P* value, <0.05). The mass spectrum corresponds to the four *M. tuberculosis*-specific features of substance A. (*B*) Ion chromatograms extracted at *m/z* (540.3545) and retention time of substance A were used for the analysis of lipid extracts of reference strains. (*C*) Ion chromatograms from lipidomic analysis of filtered conditioned medium were extracted at the *m/z* of substance A or control compounds that are secreted (carboxymycobactin) and cell wall-associated lipids (trehalose monomycolate, mycobactin).

This comparative lipidomics system allowed an unbiased, organism-wide analysis of lipids from *M. tuberculosis* and the attenuated vaccine strain, *Mycobacterium bovis* Bacillus Calmette–Guérin (BCG). BCG was chosen because of its worldwide use as a vaccine and its genetic similarity to *M. tuberculosis* (6). We reasoned that any features that are specifically detected in *M. tuberculosis* might be clinically useful as markers to distinguish tuberculosis-causing bacteria from vaccines. Furthermore, given the differing potential for productive infection by the two strains, any *M. tuberculosis*-specific compounds would be candidate virulence factors. Comparative genomics of *M. tuberculosis* and BCG successfully identified "regions of deletion" (RD) that encode genes that were subsequently proven to promote productive *M. tuberculosis* infection (7), including the 6-kDa early secreted antigenic target (ESAT-6) secretion system-1 (ESX-1) (8, 9). We reasoned that a metabolite-based screen might identify new virulence factors because not all functions of RD genes are known. Also, biologically important metabolites could emerge from complex biosynthetic pathways that cannot be predicted from single-gene analysis.

Comparison of *M. tuberculosis* and BCG lipid profiles revealed more than 1,000 differences, among which we identified a previously unknown *M. tuberculosis*-specific diterpene-linked adenosine and showed that it is produced by the enzyme Rv3378c. Previously, Rv3378c was thought to generate free tuberculosinol and isotuberculosinol (10–12). This discovery revises the enzymatic function of Rv3378c, which acts as a virulence factor to inhibit phagolysosome fusion (13). Whereas current models of prenyl transferase function emphasize iterative lengthening of prenyl pyrophosphates using one binding pocket, the crystal structure of Rv3378c identifies two pockets in the catalytic site, establishing a mechanism for heterologous prenyl transfer to nonprenyl metabolites.

Results

Comparative Lipidomics of *M. tuberculosis* and *M. bovis* Bacillus Calmette–Guérin. Using HPLC-MS for comparative analysis of lipid extracts of *M. tuberculosis* H37Rv and BCG (Pasteur strain), we detected 7,852 molecular features (Fig. 1 and *SI Appendix*, Dataset S1). By aligning datasets and seeking features that significantly differed in intensity (corrected *P* value <0.05), we identified 1,845 features that were overdetected in one bacterium or the other (Fig. 1*A*). Among these features, we focused on molecules selectively detected in *M. tuberculosis* that showed the highest fold-change ratios and intensity. We identified four molecular features corresponding to a singly charged molecular ion at *m/z* 540.357 ($C_{30}H_{45}N_5O_4$) and its isotopes (Fig. 1*A*), but this chemical formula did not match entries in the MycoMass (4) or other public databases. We named the unknown molecule substance A.

Substance A Is an Abundant Natural Product of *M. tuberculosis*. The molecular ion of substance A was one of the most intense ions in the *M. tuberculosis* lipidome (Fig. 1*A*), suggesting that it was produced in abundance. Identification of an apparently abundant molecule in a widely studied pathogen was unexpected, leading to questions about whether substance A was truly a natural product. However, this compound was absent in media, solvent blanks, and BCG lipid extracts but was reproducibly detected in three reference strains of *M. tuberculosis* (Fig. 1*B*). As observed with cell-associated compounds (Fig. 1*A*), culture filtrate (Fig. 1*C*) yielded bright ion at m/z 540.357 whose intensity was higher than that of the abundantly secreted siderophore, carboxymycobactin. Its release into the extracellular space likely results from transmembrane transport, rather than budding of intact cell wall fragments, because cell wall-embedded lipids, trehalose monomycolate and mycobactin, were not detected in filtered supernatants (Fig. 1*C*). We detected substance A in *M. tuberculosis* during exponential or stationary phase and several types of media or when subject to acid stress (*SI Appendix*, Fig. S1 *A* and *B*). Thus, substance A is a natural product that is constitutively produced in many conditions and accumulates within and outside *M. tuberculosis*.

M. tuberculosis often compartmentalizes lipid biosynthesis so that lipids are assembled after transport across the plasma membrane. Sulfoglycolipids and phthiocerol dimycocerosates become undetectable when MmpL transporters are interrupted, even when biosynthetic genes are intact (14–16). Because ESX-1 is a transport system lacking in BCG, lack of export of an ESX-1–dependent lipid synthase might account for the loss of substance A. However, ESX-1–deficient *M. tuberculosis* lacking either the *espA* gene (*Rv3616c*) or the entire RD1 locus (17), which are both necessary for ESX-1 function, produces substance A at normal

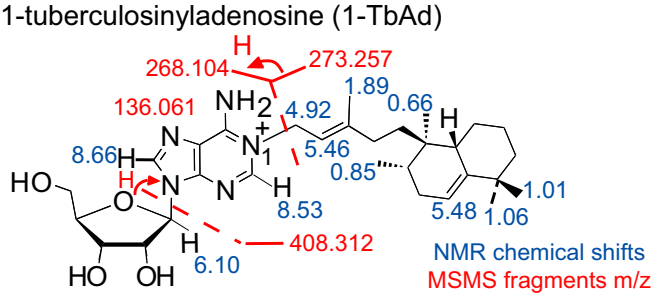

Fig. 2. Identification of 1-TbAd. The structure of substance A purified from *M. tuberculosis* lipid extract was characterized using CID-MS and NMR (800 MHz) analyses yielding key collision products and resonances detailed in *SI Appendix*, Figs. S2–S9.

BIOCHEMISTRY

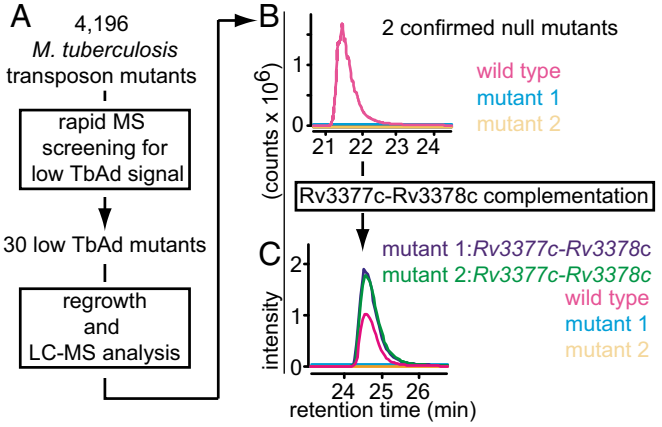

Fig. 3. *M. tuberculosis* biosynthesis of substance A requires Rv3378c. (*A*) The screening of 4,196 transposon mutants of *M. tuberculosis* H37Rv using a rapid 3-min HPLC-MS method yielded 30 strains with reduced 1-TbAd signal. (*B*) Rescreening with the 40-min HPLC-MS method confirmed absence of 1-TbAd signal in two mutants. (*C*) Both mutants were found to have spontaneous, non–transposon-induced mutations in *Rv3378c* and were subject to complementation of *Rv3377c-Rv3378c* and reanalysis for 1-TbAd production.

levels (*SI Appendix*, Fig. S1*C*). After ruling out a major known species-specific difference in transport, we devised a screen to detect genes responsible for substance A biosynthesis.

Substance A Is a 1-Tuberculosinyladenosine. Collision-induced mass spectrometry (CID-MS) identified the structural components of substance A as adenine ($[M+H]^+$, $C_5H_6N_5$, *m/z* 136.0618), adenosine ($[M+H]^+$, $C_{10}H_{14}N_5O_4$, *m/z* 268.1040), and a polyunsaturated C20 hydrocarbon ($[M+H]^+$, $C_{20}H_{33}$, *m/z* 273.2576) (Fig. 2 and *SI Appendix*, Figs. S2–S4). A common C20 diterpene is geranylgeraniol, and *M. tuberculosis* produces two C20 lipids containing bicyclic halimane skeletons: tuberculosinol and isotuberculosinol (18–20). Initially, CID-MS spectra could not distinguish among these three candidate diterpenes (*SI Appendix*, Figs. S3 and S4), but multistage CID-MS studies isolated the diterpene unit of substance A (*m/z* 273.3) and yielded collision patterns that matched tuberculosinol more closely than geranylgeraniol (*SI Appendix*, Fig. S4).

After purification of the natural product, we carried out NMR spectroscopy analyses using ^{1}H 1D, 2D COSY, HMQC, and NOESY spectra (*SI Appendix*, Figs. S5–S9), which unequivocally established the structure of substance A as 1-tuberculosinyladenosine (1-TbAd) (Fig. 2). The NMR signals of the diterpene moiety matched those of tuberculosinol (10, 19–21) except for the expected difference in the side-chain protons and carbons. The spectral data of the adenosine and adjacent atoms correspond closely to those of 1-prenyladenosine analogs (22–24). The allylic methylene group absorbs downfield as a doublet at δ 4.92 (J = 6.6 Hz). A NOESY cross peak between the adenine H-2 at δ 8.53 and the alkene hydrogen and allylic methylene and methyl groups at δ 5.46, 4.92, and 1.89, respectively, confirm that the tuberculosinyl group is attached to the adenine at position 1. Thus, *M. tuberculosis* produces a previously unknown type of diterpene nucleoside.

Rv3378c Produces 1-TbAd. To identify the genes necessary for 1-TbAd production, an existing library of random transposon insertional mutants (25) was screened in high throughput (4,196 mutants) for 1-TbAd production using a simplified 3-min HPLC-MS method (Fig. 3*A*). Thirty mutants showing low or absent signals were rescreened using the original, high-resolution lipidomic separation method (Fig. 1). Reporting only mutants with complete signal loss of 1-TbAd signal in both assays, we identified two 1-TbAd–null mutants carrying transposons in Rv1796 (mutant 1) and Rv2867c (mutant 2) (Fig. 3*B*). The concurrently performed biochemical studies described above identified the highly characteristic tuberculosinyl moiety as a component of 1-TbAd, and the *Rv3377c-Rv3378c* locus was known to encode enzymes needed for tuberculosinol and isotuberculosinol production (10, 11, 18–21). Sequencing identified spontaneous mutations in *Rv3378c* in both mutants (10, 18–21). Mutant 1 encoded a predicted Asp→Gly substitution at residue 34, and mutant 2 encoded a Pro→Ser substitution at residue 231. We generated complementation constructs to separately test whether the point mutations in *Rv3378c* or the transposon insertions were responsible for 1-TbAd loss. Transfer of Rv1796 and Rv2867c failed to restore 1-TbAd production (*SI Appendix*, Fig. S10), but transfer of *Rv3377c-Rv3378c* reconstituted 1-TbAd production in both mutants (Fig. 3*C*). Thus, *Rv3377c-Rv3378c* genes are necessary for 1-TbAd biosynthesis in *M. tuberculosis*.

The Biosynthetic Pathway of 1-TbAd. Furthermore, the known role of Rv3377c and Rv3378c enzymes in tuberculosinol production potentially provided a mechanism to connect *Rv3377c* and *Rv3378c* genes with the production of a nucleoside-modified tuberculosinol. Rv3377c is a terpene cyclase, which acts on geranylgeranyl pyrophosphate (GGPP) to generate tuberculosinyl pyrophosphate (TbPP). Rv3378c was thought to be a phosphatase, which converts TbPP to free tuberculosinol (10, 21). Extending current models (Fig. 4*A*), 1-TbAd might result from downstream action of an unknown enzyme on free tuberculosinol to transfer it to adenosine. Polyprenol synthase genes and the *Rv3377c-Rv3378c* locus are coordinately regulated and encoded at adjacent sites on the chromosome (26). Therefore, we searched *M. tuberculosis* databases for genes located near this locus that might plausibly function as adenosine transferases. We failed to find candidates and

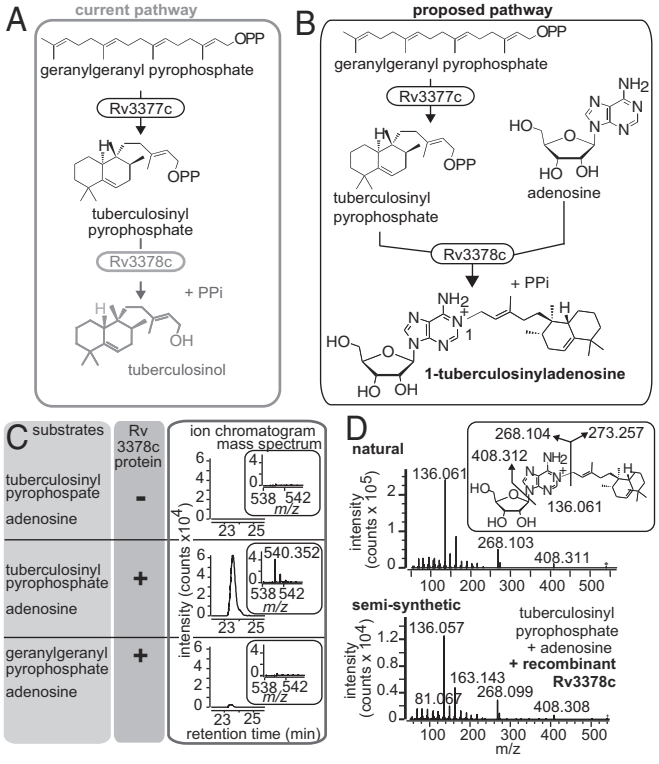

Fig. 4. Rv3378c acts as a tuberculosinyl transferase. (*A*) Rv3377c and Rv3378c are currently thought to produce tuberculosinol and isotuberculosinol. (*B*) The existence of 1-TbAd might be explained by a revised function of the Rv3378c enzyme, which acts as a tuberculosinyl transferase. (*C* and *D*) Ion chromatograms and mass spectra (*Insets*) (*C*) and CID-MS (*D*) of the 1-TbAd standard and reaction products of enzymatic assays performed using recombinant Rv3378c protein.

noted that no transposon insertion that blocked 1-TbAd production mapped to genes adjacent to this locus.

Therefore, we considered a revised biosynthetic model in which Rv3378c protein is not a simple phosphatase, as currently believed, but instead acts with combined phosphatase and tuberculosinyl transferase functions, using adenosine as the nucleophilic substrate (Fig. 4*B*). This model is mechanistically simple and might explain the lack of an apparent stand-alone transferase gene. Also, whereas current models predict that tuberculosinol is the end product of this pathway, we did not detect tuberculosinol in lipidomics experiments (Fig. 1*A* and *SI Appendix*, Dataset S1). The revised model posits that 1-TbAd is the end product of Rv3378c pathway, explaining why it accumulates to high levels as one of the brightest ions in the lipidome (Fig. 1*A*). After chemical synthesis of TbPP (*SI Appendix*, Fig. S11), we tested TbPP and GGPP as substrates for the recombinant Rv3378c protein (18). Rv3378c catalyzed the condensation of adenosine and TbPP to generate 1-TbAd but produced little or no product from GGPP. Free tuberculosinol was not detected in these assays (Fig. 4 *C* and *D*). Thus, Rv3378c shows tuberculosinyl transferase activity, which rules in the revised biosynthetic pathway (Fig. 4*B*).

***Rv3377c-Rv3378c* Is Sufficient for 1-TbAd Biosynthesis in Cells.** To test the sufficiency of this locus for 1-TbAd production in cells, we transferred the *Rv3377c-Rv3378c* locus to *Mycobacterium smegmatis*. In all three clones tested, expression of *Rv3377c-Rv3378c* transferred production of a molecule with the mass, retention time, and CID-MS spectrum of 1-TbAd (Fig. 5 and *SI Appendix*, Fig. S12). Thus, no other *M. tuberculosis*-specific cofactor or transporter is needed for 1-TbAd production. *Rv3377c-Rv3378c* is sufficient to synthesize 1-TbAd from ubiquitous cellular precursors present in most bacteria, likely GGPP and adenosine.

Crystal Structure of Rv3378c. To understand whether the active site of Rv3378c is compatible with the revised function as a tuberculosinyl transferase, we determined its crystal structure. Lacking structures with high sequence similarity, single-wavelength anomalous dispersion phasing was used to calculate the initial electron-density map. The model was refined against native data to 2.2-Å resolution (*SI Appendix*, Table S1). As expected from gel-filtration studies, Rv3378c formed a homodimer (Fig. 6*A*). Although structural similarity was not predicted by sequence comparisons, Rv3378c adopts the fold seen in (*Z*)-prenyl or *cis*-prenyl transferases (27), including *M. tuberculosis* (*Z*)-farnesyl diphosphate synthase (Rv1086) and decaprenyl pyrophosphate synthase (Rv2361c), as well as *Escherichia coli* undecaprenyl pyrophosphate synthase (UPP) (28, 29) (Fig. 6*B*). These enzymes condense an allyl pyrophosphate and the five-carbon isopentyl pyrophosphate building block to produce linear isoprenoids (28, 29).

Structural Insight into Prenyl Unit Binding. In considering competing models that Rv3378c might simply hydrolyze the TbPP or carry out the newly proposed role in adenosine transfer (Fig. 5 *A* and *B*), we superimposed Rv3378c with the pseudosubstrate and product complexes of Rv2361c (29) to model an enzyme–substrate (ES) complex. In contrast to other (*Z*)-prenyl transferases, Rv3378c has a unique C-terminal helical segment (from residue 251 to end), which contributes to domain swapping. An extra N-terminal helical segment (residues 6–24) packs via hydrophobic interactions with adjacent helices (Fig. 6*A* and *SI Appendix*, Fig. S13).

Rv3378c shares functional motifs with the (*Z*)-prenyl transferases, including residues for substrate binding and catalysis: Asp34, Arg37, and Arg38 (Fig. 6*B*). (*Z*)-prenyl transferases bind the allyl pyrophosphate substrate through a characteristic DGNG/RRW amino acid sequence motif starting two residues before the N terminus of an α-helix (α3). The aspartate chelates a magnesium ion, whereas the glycine, the helix terminus, and the arginine(s) engage the pyrophosphate (Fig. 6 *B* and *C*) (27, 28, 30). In Rv3378c, Asp34 sits in the expected position to carry out its essential catalytic function providing a specific mechanism that likely explains why mutant 1, which contains an Asp34→Gly alteration, does not

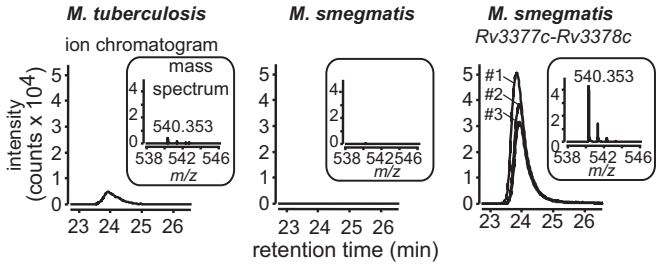

Fig. 5. The expression of *Rv3377c-Rv3378c* is sufficient for production of 1-TbAd in *M. smegmatis*. Extracted ion chromatograms and mass spectra (*Insets*) of 1-TbAd (*m/z* 540.3545) for the HPLC-MS analysis of lipid extracts from *M. tuberculosis* (*Left*), *M. smegmatis* parental (*Center*), or each of three *M. smegmatis Rv3377c-Rv3378c* knock-in (*Right*) strains.

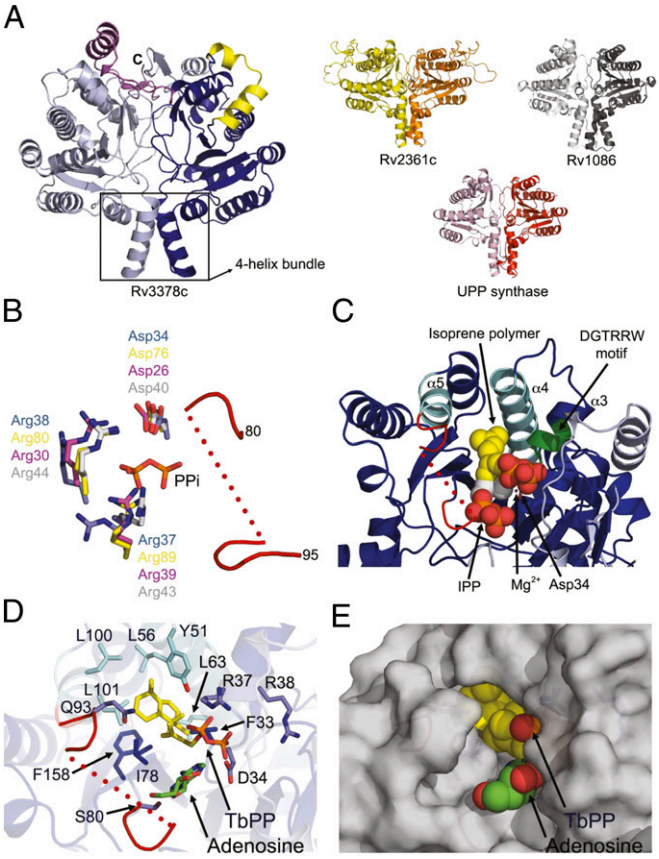

Fig. 6. Rv3378c adopts a (*Z*)-prenyl transferase fold. (*A*) Structure of the Rv3378c dimer is compared with conventional (*Z*)-prenyl transferases. (*B*) Superposition of the active site of Rv3378c and other (*Z*)-prenyl transferases with the pyrophosphate bound to Rv2361c (stick) shows conserved key residues for substrate binding and catalysis (Rv3378c, blue; Rv2361c, yellow; Rv1086, gray; *E. coli* UPP synthase, magenta for carbon atoms). (*C*) The monomeric subunits of Rv3378c and Rv2361c were superimposed and Rv2361c substrates (sphere) (carbon, yellow/gray; oxygen, red; phosphate, orange) are modeled in the active site of Rv3378c. The conserved residue, Asp34, is shown as a stick model, and the magnesium ion is shown as a magenta sphere. (*D*) Proposed model of Rv3378c shows two substrate pockets with hydrophobic residues lining the predicted prenyl binding pocket and D34 positioned adjacent to the predicted adenosine binding pocket. (*B–D*) The flexible P-loop of Rv3378c (residues 80–95) is colored in red, with the dotted line for disordered region (residues 84–90). (*E*) The translucent surface of Rv3378c was modeled with substrates (spheres) using the same view as *D*.

BIOCHEMISTRY

produce 1-TbAd. As predicted by prior studies showing the role of aspartate in prenyl transfer (27, 28, 30) and the conserved location of Asp34 vis-à-vis the prenyl binding site (Fig. 6 *A* and *B*), mutation to asparagine or alanine abolished the prenyl transferase function of Rv3378c (*SI Appendix*, Fig. S14). In Rv2361c, the isoprenoid binding site is a hydrophobic pocket located between the β-sheet and the α2 (residues 89–110) and α3 (residues 129–152) helices (29). Rv3378c contains all of these features (Fig. 6*C*), including the 34-DGTRRW-39 motif and a deep pocket adjacent to helices α4 (residues 51–68) and α5 (residues 96–103). Hydrophobic residues (L56, L63, L100, and L101) line the pocket created by helices α4 and α5, and other hydrophobic residues (F33, I78, F158) further contribute to the hydrophobic character of the pocket. This binding pocket is predicted to position the pyrophosphate group of TbPP, which can interact with Arg37 and Arg38 from the DGTRRW motif and Tyr51 from the N terminus of helix α4 (Fig. 6*D*).

A Second Pocket at the Catalytic Site. The binding mode of the nucleophilic adenosine substrate is harder to model, because the binding site is likely to be completed by the closure of the P-loop over the active site when native substrate is present. The P-loop is disordered in the unliganded structure, but it becomes ordered in a nonphysiological complex with mellitic acid (*SI Appendix*, Fig. S13). This structure suggests a specific mechanism by which substrate binding provides polar interactions with the P-loop to exclude water from the active site. Other considerations provide pertinent clues about the adenosine binding mode. As contrasted with Rv2361c, Rv1086, and UPP synthase, Rv3378c has a second pocket that can accommodate adenosine (Fig. 6 *D* and *E*). Superimposing N1 of the adenine on the isopentenyl pyrophosphate (IPP) nucleophile in complex with Rv2361c (29) guides the positioning of the adenosine substrate in the active site of Rv3378c. In Rv2361c, the pyrophosphate of IPP interacts with Arg244 and Arg250 (29). Corresponding to the fact that adenosine lacks the pyrophosphate, Rv3378c lacks this conserved pair of arginines, which are replaced with glycine and serine. These features distinguish Rv3378c from known (*Z*)-prenyl transferases and are consistent with adenosine binding and transfer.

Discussion

Overall, structural, genetic, and biochemical data strongly suggest a revised function of Rv3378c as a tuberculosinyl transferase that produces 1-TbAd, an abundant amphiphile that is exported outside *M. tuberculosis*. This result establishes the efficacy of unbiased lipidomic screens to identify previously unknown compounds. A C35 terpene cyclase activity is found in nonpathogenic mycobacteria (31, 32), but Rv3377c orthologs are only known within *M. tuberculosis* complex. Higher-order terpene-nucleosides are rare in nature, and we have not identified a precedent for 1-linked prenyl adenosines. Plant and marine sponges produce terpene-purine derivatives, such as cytokinins and agelasines, which regulate growth or show antimicrobial effects (33). However, these natural products contain adenine rather than adenosine, and the terpene moiety is carried at the N6 position of the adenine in the cytokinins and N7 or N9 in the agelasines. Furthermore, among microbes studied to date, we have only detected 1-TbAd in members of the *M. tuberculosis* complex, suggesting that 1-TbAd production is limited to pathogenic mycobacteria. Orthologs of Rv3377c or RV3378c are limited to the *M. tuberculosis* complex. Although *M. bovis* and BCG strains encode orthologous genes, strains examined to date encodes a frameshift mutation in *Rv3377c* (11), and the Pasteur strain used here encodes a second coding point mutation in *Rv3378c*. Thus, both genetic and biochemical evidence suggest that 1-TbAd is a specific marker of *M. tuberculosis*, supporting the development of 1-TbAd or 1-TbAd–specific immune responses as candidate targets for diagnostic tests for tuberculosis.

The lack of 1-TbAd in BCG might represent evidence that changes in *Rv3377c-Rv3378c* might contribute to the vaccine strain's attenuation. More direct evidence for a role of this locus in virulence comes from transposon studies showing that *Rv3377c* and *Rv3378c* play nonredundant roles in phagosome–lysosome fusion and survival in macrophages (13). This key finding initiated an intensive search for the actual functions of these virulence-associated genes. Rv3377c is a terpene cyclase (18–20). Rv3378c has few orthologs in nature, and its biochemical function was not apparent from predictive folding algorithms. Based on in vitro studies, Rv3378c is currently thought to function as a TbPP pyrophosphatase that yields free tuberculosinol (10). Tuberculosinol coupled to beads blocks phagosomal acidification (21). However, end products of biosynthetic pathways typically accumulate, and to our knowledge, the extent of accumulation of free tuberculosinol as a natural product in intact *M. tuberculosis* remains unknown. We did not detect free tuberculosinol or isotuberculosinol in lipidomics analysis of *M. tuberculosis* or among in vitro products of Rv3378c. This result does not rule out biosynthesis of free tuberculosinol, but it is notable that 1-TbAd is not only detected but substantially accumulates within and outside *M. tuberculosis*. Furthermore, we prove that the action of Rv3378c is a combined phosphatase and tuberculosinol transferase through in vitro study of purified proteins, gene transfer to *M. tuberculosis* and *M. smegmatis*, and a structural analysis of Rv3378c. Based on parallel lines of genetic, biochemical, and structural evidence, we propose that Rv3378c should be known as "tuberculosinyl adenosine transferase."

The structures of enzymes that transfer prenyl pyrophosphates to substrates other than linear isoprenoids have not been determined previously. Like other (*Z*)-prenyl transferases, Rv3378c contains a characteristic allyl pyrophosphate-binding site, catalytic aspartate, and flexible P-loop in position to close over the active site. The canonical TbPP binding pocket structure is sufficiently conserved that it may be sensitive to available drugs or analogs that target other (*Z*)-prenyl transferases. However, the nucleophile binding site lacks conserved features that mediate recognition of pyrophosphate moiety of isoprene building blocks seen in previously characterized (*Z*)-prenyl transferases. Instead, Rv3378c active site contains a second pocket in which the adenosine can be positioned for nucleophilic attack on C1 of TbPP. We observed this reaction in vitro and found that Rv3378c does not act on GGPP and specifically produces the 1-linked form of 1-TbAd, defining two aspects of the substrate specificity. Whereas most prenyl transferases have one identifiable pocket, this revised two-pocket model suggests a broader paradigm for the prenylation of metabolites catalyzed by members of the (*Z*)-prenyl transferase family. Whereas current models emphasize iterative elongation through the repeated use of one pocket, the dual-substrate pocket of Rv3378c provides a general model for prenylation of nonprenyl substrates. Product specificity is determined by a conventional allyl pyrophosphate binding site and a second pocket tailored to bind and activate each target nucleophile.

The larger 1-TbAd biosynthetic pathway starts with two evolutionarily conserved systems, which produce geranylgeranyl pyrophosphate and adenosine. These pathways operate separately in most organisms, but *M. tuberculosis* joins these two pathways to generate a terpene nucleoside. The appearance of TbAd after transfer of *Rv3377c* and *Rv3378c* genes to *M. smegmatis* proves that additional *M. tuberculosis*-specific genes, such as transporters, are not required for 1-TbAd biosynthesis. More generally, these data represent an experimental demonstration that transfer of two genes is sufficient to reconstitute a complex metabolite, which likely requires more than 20 genes for its complete biosynthesis. Combining this observation with data suggesting that the ancestral *Rv3377c* and *Rv3378c* genes were acquired by horizontal gene transfer (26), a scenario emerges by which evolutionarily ancient terpene and nucleotide biosynthetic pathways were joined together by transfer of two genes late in the evolution of the *M. tuberculosis* complex (26).

Rv3378c gene likely carries out its known effects in promoting *M. tuberculosis* infectivity via the production of 1-TbAd. Within minutes of phagocytosis, *M. tuberculosis* inhibits host defenses, including phagosome acidification and phagolysosome fusion

(34, 35). The *Rv3377c-Rv3378c* locus is required for optimal phagosome maturation arrest (13). The discovery of extracellular 1-TbAd provides specific insight into mechanisms by which an enzyme that is thought to be localized in the cytosol affects events outside the bacterium (13). To our knowledge, neither Rv3378c nor free tuberculosinol has been detected in culture filtrates (18). In contrast, 1-TbAd is an amphiphile that is released into the extracellular space using an export mechanism that is independent of ESX-1.

Future studies will be needed to understand the particular mechanisms by which 1-TbAd contributes to the effects of Rv3377c-Rv3378c on phagosome maturation. Adenosine is almost exclusively found inside cells, and terpene chains catalyze the transfer of pyrophosphate across the mycobacterial envelope for the biosynthesis of arabinogalactan (36). By analogy, prenylation might promote the transit of the nucleoside to the phagosomal space, where the adenosine could engage host receptors. Alternatively, tuberculosinol or isotuberculosinol might be the active moiety (12, 21), whose solubility or transport is influenced by adenosine. The cellular mechanism leading to altered mycobacterial survival might include changed integrity of the phagosomal membrane, intraphagosomal proton capture, or escape of 1-TbAd across the phagosomal membrane and into the host, where it might signal global changes in macrophage function.

Methods

Bacteria were cultured and extracted by chloroform/methanol mixtures or ethyl acetate, respectively, as described (4, 37). Lipid extracts were analyzed using an Agilent 6520 Accurate-Mass Q-Tof and a 1200 series HPLC system with a Varian MonoChrom Diol column (4, 37) with data output from XCMS and MultiplotPreprocess and Multiplot modules of GenePattern (Broad Institute) (38). Rv3378c and GroES/GroEL chaperones were coexpressed in BL21-CodonPlus (Stratagene) cells and purified on a precharged Ni Sepharose Fast Flow column (Ni-NTA HisTrap FF, GE Healthcare). Purified Rv3378c (10 mg/mL) was crystallized by vapor diffusion, and 2.20-Å-resolution data were collected on the Advanced Light Source. The structure of Rv3378c was solved by single-wavelength anomalous diffraction phasing of a mercury derivative using Phenix AutoSol. Enzymatic assays were performed by incubating 56 μg of diterpene in presence of 33 μg of adenosine (Sigma) and 80 μg of purified Rv3378c in 1 mL of pH 7.4 Tris·HCl buffer 4 h at 37 °C under magnetic agitation. *M. tuberculosis* transposon mutants from a random library (25) were grown in 96-well format and heat-killed, followed by lipid extraction by 70:30 methanol:isopropanol. Lipids were analyzed by HPLC-MS to monitor 1-TbAd production; 1-TbAd–null strains were confirmed regrowing the bacteria and using a full lipidomic analysis method. 1-TbAd was purified from mycobacterial cell-associated lipid extract using normal and reversed-phase chromatography. Structures were solved using CID-MS and NMR spectroscopy using a Bruker Avance 800.

ACKNOWLEDGMENTS. We thank Lisa Prach for advice. This work was supported by Mark and Lisa Schwartz and the Burroughs Wellcome Fund Program in Translational Medicine, National Institutes of Health (NIH) National Institute of Allergy and Infectious Disease Grants R01 AI049363, R01 AI049313, and P01 AI095208, the Dutch Science Foundation (Netherlands Organization for Scientific Research), and the Canadian Institutes of Health Research (H.J.L.). The 800-MHz spectrometer in the Landsman Research Facility, Brandeis University, was purchased under the NIH National Center for Research Resources High-End Instrumentation Program (Grant 1S10RR017269-01).

1. Dye C, Glaziou P, Floyd K, Raviglione M (2013) Prospects for tuberculosis elimination. *Annu Rev Public Health* 34:271–286.
2. Sturgill-Koszycki S, et al. (1994) Lack of acidification in Mycobacterium phagosomes produced by exclusion of the vesicular proton-ATPase. *Science* 263(5147):678–681.
3. Camus JC, Pryor MJ, Médigue C, Cole ST (2002) Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv. *Microbiology* 148(Pt 10):2967–2973.
4. Layre E, et al. (2011) A comparative lipidomics platform for chemotaxonomic analysis of Mycobacterium tuberculosis. *Chem Biol* 18(12):1537–1549.
5. Sartain MJ, Dick DL, Rithner CD, Crick DC, Belisle JT (2011) Lipidomic analyses of Mycobacterium tuberculosis based on accurate mass measurements and the novel "Mtb LipidDB" *J Lipid Res* 52(5):861–872.
6. Behr MA, et al. (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 284(5419):1520–1523.
7. Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK (1996) Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. *J Bacteriol* 178(5):1274–1282.
8. Brodin P, Rosenkrands I, Andersen P, Cole ST, Brosch R (2004) ESAT-6 proteins: Protective antigens and virulence factors? *Trends Microbiol* 12(11):500–508.
9. Fortune SM, et al. (2005) Mutually dependent secretion of proteins required for mycobacterial virulence. *Proc Natl Acad Sci USA* 102(30):10676–10681.
10. Nakano C, et al. (2011) Characterization of the Rv3378c gene product, a new diterpene synthase for producing tuberculosinol and (13R, S)-isotuberculosinol (nosyberkol), from the Mycobacterium tuberculosis H37Rv genome. *Biosci Biotechnol Biochem* 75(1):75–81.
11. Mann FM, et al. (2009) Characterization and inhibition of a class II diterpene cyclase from Mycobacterium tuberculosis: Implications for tuberculosis. *J Biol Chem* 284(35): 23574–23579.
12. Mann FM, et al. (2009) Edaxadiene: A new bioactive diterpene from Mycobacterium tuberculosis. *J Am Chem Soc* 131(48):17526–17527.
13. Pethe K, et al. (2004) Isolation of Mycobacterium tuberculosis mutants defective in the arrest of phagosome maturation. *Proc Natl Acad Sci USA* 101(37):13642–13647.
14. Domenech P, et al. (2004) The role of MmpL8 in sulfatide biogenesis and virulence of Mycobacterium tuberculosis. *J Biol Chem* 279(20):21257–21265.
15. Jain M, Cox JS (2005) Interaction between polyketide synthase and transporter suggests coupled synthesis and export of virulence lipid in M. tuberculosis. *PLoS Pathog* 1(1):e2.
16. Converse SE, et al. (2003) MmpL8 is required for sulfolipid-1 biosynthesis and Mycobacterium tuberculosis virulence. *Proc Natl Acad Sci USA* 100(10):6121–6126.
17. Garces A, et al. (2010) EspA acts as a critical mediator of ESX1-dependent virulence in Mycobacterium tuberculosis by affecting bacterial cell wall integrity. *PLoS Pathog* 6(6):e1000957.
18. Prach L, Kirby J, Keasling JD, Alber T (2010) Diterpene production in Mycobacterium tuberculosis. *FEBS J* 277(17):3588–3595.
19. Nakano C, Okamura T, Sato T, Dairi T, Hoshino T (2005) Mycobacterium tuberculosis H37Rv3377c encodes the diterpene cyclase for producing the halimane skeleton. *Chem Commun (Camb)* (8):1016–1018.
20. Nakano C, Hoshino T (2009) Characterization of the Rv3377c gene product, a type-B diterpene cyclase, from the Mycobacterium tuberculosis H37 genome. *ChemBioChem* 10(12):2060–2071.
21. Hoshino T, Nakano C, Ootsuka T, Shinohara Y, Hara T (2011) Substrate specificity of Rv3378c, an enzyme from Mycobacterium tuberculosis, and the inhibitory activity of the bicyclic diterpenoids against macrophage phagocytosis. *Org Biomol Chem* 9(7): 2156–2165.
22. Ottria R, Casati S, Baldoli E, Maier JA, Ciuffreda P (2010) N^6-Alkyladenosines: Synthesis and evaluation of in vitro anticancer activity. *Bioorg Med Chem* 18(23):8396–8402.
23. Casati S, Manzocchi A, Ottria R, Ciuffreda P (2010) 1H, 13C and 15N NMR assignments for N6-isopentenyladenosine/inosine analogues. *Magn Reson Chem* 48(9):745–748.
24. Casati S, Manzocchi A, Ottria R, Ciuffreda P (2011) 1H, 13C and 15N NMR spectral assignments of adenosine derivatives with different amino substituents at C6-position. *Magn Reson Chem* 49(5):279–283.
25. Sassetti CM, Boyd DH, Rubin EJ (2001) Comprehensive identification of conditionally essential genes in mycobacteria. *Proc Natl Acad Sci USA* 98(22):12712–12717.
26. Mann FM, Xu M, Davenport EK, Peters RJ (2012) Functional characterization and evolution of the isotuberculosinol operon in Mycobacterium tuberculosis and related Mycobacteria. *Front Microbiol* 3:368.
27. Kurokawa H, Koyama T (2010) (2010) Prenyltransferase. *Comprehensive Natural Products II: Chemistry and Biology*, eds Mander L, Liu H-W (Oxford, Oxford, UK), 557–583.
28. Chang SY, Ko TP, Chen AP, Wang AH, Liang PH (2004) Substrate binding mode and reaction mechanism of undecaprenyl pyrophosphate synthase deduced from crystallographic studies. *Protein Sci* 13(4):971–978.
29. Wang W, et al. (2008) The structural basis of chain length control in Rv1086. *J Mol Biol* 381(1):129–140.
30. Guo RT, et al. (2005) Crystal structures of undecaprenyl pyrophosphate synthase in complex with magnesium, isopentenyl pyrophosphate, and farnesyl thiopyrophosphate: Roles of the metal ion and conserved residues in catalysis. *J Biol Chem* 280(21): 20762–20774.
31. Sato T, Kigawa A, Takagi R, Adachi T, Hoshino T (2008) Biosynthesis of a novel cyclic C35-terpene via the cyclisation of a Z-type C35-polyprenyl diphosphate obtained from a nonpathogenic Mycobacterium species. *Org Biomol Chem* 6(20):3788–3794.
32. Sato T, Takagi R, Orito Y, Ono E, Hoshino T (2010) Novel compounds of octahydroheptaprenyl mycolic acyl ester and monocyclic C35-terpene, heptaprenylcycline B, from non-pathogenic mycobacterium species. *Biosci Biotechnol Biochem* 74(1): 147–151.
33. Vik A, et al. (2007) Antimicrobial and cytotoxic activity of agelasine and agelasimine analogs. *Bioorg Med Chem* 15(12):4016–4037.
34. Yates RM, Hermetter A, Russell DG (2005) The kinetics of phagosome maturation as a function of phagosome/lysosome fusion and acquisition of hydrolytic activity. *Traffic* 6(5):413–420.
35. Rohde K, Yates RM, Purdy GE, Russell DG (2007) Mycobacterium tuberculosis and the environment within the phagosome. *Immunol Rev* 219:37–54.
36. Alderwick LJ, et al. (2011) Biochemical characterization of the Mycobacterium tuberculosis phosphoribosyl-1-pyrophosphate synthetase. *Glycobiology* 21(4):410–425.
37. Madigan CA, et al. (2012) Lipidomic discovery of deoxysiderophores reveals a revised mycobactin biosynthesis pathway in Mycobacterium tuberculosis. *Proc Natl Acad Sci USA* 109(4):1257–1262.
38. Reich M, et al. (2006) GenePattern 2.0. *Nat Genet* 38(5):500–501.