Title

In Vivo Biosynthesis of Terpene Nucleosides Provides Unique Chemical Markers of Mycobacterium tuberculosis Infection

Journal

Cell Chemical Biology, 22(4)

ISSN

2451-9456

Authors

Young, David C
Layre, Emilie
Pan, Shih-Jung
et al.

Publication Date

2015-04-01

DOI

10.1016/j.chembiol.2015.03.015

Peer reviewed

HHS Public Access

Author manuscript

Chem Biol. Author manuscript; available in PMC 2016 April 23.

Published in final edited form as:

Chem Biol. 2015 April 23; 22(4): 516–526. doi:10.1016/j.chembiol.2015.03.015.

In vivo biosynthesis of terpene nucleosides provides unique chemical markers of *Mycobacterium tuberculosis* infection

David C. Young[1], **Emilie Layre**[1], **Shih-Jung Pan**[2], **Asa Tapley**[2,3], **John Adamson**[2,4], **Chetan Seshadri**[1], **Zhongtao Wu**[5], **Jeffrey Buter**[5], **Adriaan J. Minnaard**[5], **Mireia Coscolla**[6,7], **Sebastien Gagneux**[6,7], **Richard Copin**[8], **Joel D. Ernst**[8], **William Bishai**[2,4], **Barry B. Snider**[9], and **D. Branch Moody**[1,*]

[1]Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Smith Building Room 538, 1 Jimmy Fund Way, Boston, Massachusetts 02115, United States [2]K-RITH;KwuZulu-Natal Research Institute for Tuberculosis & HIV, Nelson R. Mandela School of Medicine-Univ. of Kwazulu-Natal, K-RITH Tower Building, 719 Umbilo Road, Durban, 4001 Private Bag X7, Congela-Durban, 4001, South Africa [3]University of California, San Francisco School of Medicine, California, USA [4]Center for Tuberculosis Research, Div. of Infections Diseases, Johns Hopkins University School of Medicine, 1550 Orleans St. Room108, Baltimore, Maryland,21231 USA [5]Stratingh Institute for Chemistry, University of Groningen, Nijenborgh 7, 9747 AG, Groningen, The Netherlands [6]Swiss Tropical Public Health Institute, Basel, Switzerland [7]University of Basel, Basel, Switzerland [8]Division of Infectious Diseases & Immunology, New York University School of Medicine, 522 First Avenue, SRB 901, New York, NY 10016 USA [9]Dept of Chemistry MS015, Brandeis University, Waltham, Massachusetts 02454-9110, United States

Summary

Although small molecules shed from pathogens are widely used to diagnose infection, such tests have not been widely implemented for tuberculosis. Here we show that the recently identified compound, 1-tuberculosinyladenosine (1-TbAd), accumulates to comprise > 1 percent of all *M. tuberculosis* lipids. In vitro and in vivo, two isomers of TbAd were detected that might serve as infection markers. Using mass spectrometry and NMR, we established the structure of the

*To whom correspondence should be addressed: D. Branch Moody, Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Smith Building Room 538, 1 Jimmy Fund Way, Boston, MA 02115, United States 617-525-1037; bmoody@partners.org.

Present Address-Chetan Seshadri, Division of Allergy and Infectious Diseases, University of Washington, Box 356423, Seattle, WA 98195

Emilie Layre, Department of Tuberculosis and Infections Biology, Institut de Pharmacologie et de Biologie Structurale, CNRS UPS UMR5089, 31077 Toulouse, France.

AUTHOR CONTRIBUTIONS

D.C.Y and E.L. performed normal phase LC-MS experiments; B.B.S. performed and interpreted NMR on N^6-TbAd; Z.W.,J.B., and A.J.M synthesized chemical analogs; S.P., A.T., and J.A. performed analysis of mouse samples; W.B, J.D.E., and D.B.M conceived the overall collaboration and experiments; and D.C.Y and D.B.M wrote the paper.

previously unknown molecule, N^6-tuberculosinyladenosine (N^6-TbAd). Its biosynthesis involves enzymatic production of 1-TbAd by Rv3378c followed by conversion to N^6-TbAd via the Dimroth rearrangement. Intact biosynthetic genes are observed only within *M. tuberculosis* complex bacteria, and TbAd was not detected among other medically important pathogens, environmental bacteria and vaccine strains. With no substantially similar known molecules in nature, the discovery and in vivo detection of two abundant terpene nucleosides support their development as specific diagnostic markers of tuberculosis.

Introduction

Tuberculosis (TB) remains a leading cause of death worldwide, resulting in 1.5 million deaths annually (World Health Organization, 2014), yet no rapid, sensitive and specific diagnostic test exists. Diagnosis based on detection of *M. tuberculosis* in patient samples mainly relies on sputum microscopy, which is insensitive, or on in vitro culture, which is slow, insensitive and infeasible in many clinics. T-cell antigen recall tests, such as intradermal injection of purified protein derivative (PPD) or interferon−γ release assays (IGRA) are in widespread use (Lalvani and Pareek, 2010), but give delayed results or are expensive and have suboptimal test characteristics related to sensitivity and specificity. Vaccination with live Bacille Calmette-Guerin (BCG) in most parts of the world leads to antigen-specific T cell responses, which create false positive results, rendering the PPD test unusable in many populations. Accordingly, there is now strong consensus that developing better diagnostic tests for *M. tuberculosis* infection is the key issue for improved disease control through rapid initiation of antibiotics and categorization of patients for vaccine trials (Hanekom et al., 2008).

Detection of pathogen-specific shed molecules or “antigens” provides rapid and specific diagnosis of many infectious diseases. Such antigen tests have long been a mainstay of diagnosis for infection by cryptococci, legionella and other pathogens (Shelhamer et al., 1996). The strengths of antigen test technology are high diagnostic specificity and rapid detection of molecules using a simple enzyme-linked immunosorbant assay (ELISA) of urine or serum (Couturier et al., 2014). Therefore, the key criterion for discovery of chemical targets for antigen tests is specific expression of the target by the disease-causing pathogen, combined with lack of expression among other microbes, especially those that are abundant in the environment or cause diseases that mimic the disease of interest. Other desirable criteria related to test sensitivity involve identifying targets with broad expression among most infecting strains in clinical settings, expression of the antigen at high concentrations, expression in vivo under conditions of infection and lack of host degradation or metabolism to unrecognizable chemical forms.

In addition to antigen capture, pathogen-specific molecules can be coated onto plastic and used to detect target-specific host antibodies, functioning as a ‘serological test'. Such tests have not yet moved into widespread clinical use for TB due in part of specificity concerns that may be related to immune responses to environmental mycobacteria, mildly pathogenic mycobacteria or live vaccine strains (Lawn et al., 2012) against tuberculosis (Baumann et al., 2014).

Despite the widespread use of antigen and serological tests in other infectious diseases, neither type of test is widely used for tuberculosis. Although the urine lipoarabinomannan (LAM) ELISA has usefulness for TB-human immunodeficiency virus (TB-HIV) coinfection (Lawn et al., 2012), no antigen or serological test has emerged as having widespread clinical usefulness for tuberculosis. The current chemical targets for testing were chosen based on their ready availability and represent only a small fraction of the candidate small molecules that could be developed. For example, among 169 subclasses of mycobacterial lipids in the *MycoMass* and Lipid DB databases, more than 90 percent are expressed only by mycobacteria (Layre, et al., 2011; Sartain et al., 2011). Thus, the potential range of specific mycobacterial targets for diagnostics development is vast and largely unexplored.

We initiated a comprehensive effort to discover specific chemical targets for diagnostic testing, using a newly developed HPLC-MS-based lipidomics platform (Layre, et al., 2011). A recent comparative lipidomics screen of mycobacteria sought to identify those molecules that are present in *M. tuberculosis* but are lacking in a BCG vaccine strain (Layre, et al., 2014). These two species are evolutionarily related and share more than 99 percent sequence identity, but only *M. tuberculosis* causes widespread disease. BCG has been administered to more than 3 billion people, and it induces immune responses that can cause false positive immunological tests for TB (World Health Organization, 2014). By identifying cell wall lipids absent in BCG, we reasoned that such molecules would also be lacking in BCG vaccines and in other, less related bacteria that confound the diagnosis of TB. Also, given the lower infectious potential of BCG, molecules selectively expressed in *M. tuberculosis* might be virulence factors. This screen identified a previously unknown diterpene nucleoside, 1-tuberculosinyladenosine (1-TbAd) (Layre, et al., 2014).

Having established the lack of 1-TbAd in BCG, here we sought to determine if 1-TbAd has the key test characteristics needed for use as a diagnostic target, including high abundance, specific expression and shedding from intact bacteria in ways that lead to detection in vivo. Unexpectedly, we identified an abundant but previously unknown terpene nucleoside derived from *M. tuberculosis* present in vivo. The discovery of two terpene nucleosides, which are specifically and abundantly expressed by *M. tuberculosis*, provides two highly attractive targets for development as targets for new diagnostic tests for tuberculosis.

Results

1-TbAd is a major lipid in *M. tuberculosis*

1-TbAd was identified in *M. tuberculosis* using electrospray ionization mass spectrometry (ESI-MS)-based lipidomics platform. Among 7,852 ions, 1-TbAd ($C_{30}H_{46}N_5O_4^+$; *m/z* 540.3545) was the second most intense ion in the lipidome (Layre et al., 2014). The high intensity might have resulted from 1-TbAd's accumulation to high concentration. Yet it was difficult to imagine that any abundant class of molecule would have escaped detection over decades of study of a pathogen of worldwide importance. Alternatively, 1-TbAd might have been present only in trace amounts, yet its intrinsically charged nature and amphipathic character might have promoted particularly efficient ionization in ESI-MS. To measure the mass of 1-TbAd as a percentage of all lipids, we analyzed extracts from *M. tuberculosis* strain H37Rv using the method of standard additions. We compared the area under the curve

of retention time versus intensity measured at the mass (*m/z* 540) of 1-TbAd (A_{540}) and phosphatidylethanolamine (A_{720}), which controls for the efficiency of ESI-MS detection. As with other polar lipids (Layre et al., 2011), we observed nearly linear relationship between mass input and 1-TbAd signal intensity (Figure 1A). In three experiments, we determined that 1-TbAd comprises 1.1, 1.4 and 1.5 percent of mycobacterial lipids. Thus, contrary to expectations that terpene nucleosides previously escaped detection due to scarcity, 1-TbAd is a highly abundant molecule that comprises a major class of lipid in *M. tuberculosis*. The basis for lack of prior detection remains unknown, but the near co-elution of 1-TbAd and abundant membrane phospholipids in normal phase TLC and HPLC methods might have obscured 1-TbAd detection in the past (Figure 1B).

Constitutive biosynthesis of 1-TbAd

1-TbAd was initially isolated from bacteria grown in a formulated rich media (7H9) in late logarithmic phase (Layre et al., 2014). Because Rv3377c and Rv3378c protect bacteria against low pH, we reasoned that TbAd production might be pH responsive. However, we detected equivalent production with high absolute intensity ($> 10^6$ counts) when growing at neutral (7.4) or lysosomal pH (4.5) (Figure 1C). Further, 1-TbAd production was not substantially inhibited during stationary phase, with laboratory strains of different origins (Layre et al., 2014) or when growing in minimal (Sauton) or rich (7H9) medium (Figure 1D). Thus, 1-TbAd is constitutively produced at high concentrations among diverse in vitro conditions, and we could not identify signals for its inhibition.

1-TbAd biosynthetic gene expression

To function as a sensitive marker of infection, molecules must be expressed in most strains of *M. tuberculosis* that infect patients and among the seven recognized *M. tuberculosis* lineages that exist worldwide (Comas et al.,2013). 1-TbAd derives from geranylgeranyl pyrophosphate (GGPP) and adenosine, which are present in nearly all organisms. Lipid cyclization and coupling to adenosine are performed by specialized enzymes, Rv3377c (prenyl cyclase) and Rv3378c (tuberculosinyl transferase), which together with polyprenyl synthases comprise a functional gene island (Mann and Peters, 2012) that is necessary and sufficient for 1-TbAd biosynthesis (Layre et al., 2014). Knowledge of the essential TbAd biosynthetic genes allowed a survey of the genomes of 432 phylogeographically diverse clinical isolates of the *M. tuberculosis* complex (MTB complex) for an intact biosynthetic locus. TbAd genes were detected among all clinical strains examined.

We identified 25 unique single nucleotide polymorphisms (SNPs) in Rv3377c and 22 unique SNPs in Rv3378c affecting a total of 150 *M tuberculosis* strains (Supplemental Table 1). While the proportion of non-synonymous SNPs (nSNP) (18/25 in rv3377c vs. 14/22 in rv3378c) in both genes was similar, more strains contained nSNPs in rv3377c (101 strains with nSNP in rv3377c vs. 52 strains with nSNP in rv3378c). To understand the selection forces driving Rv3377c and Rv3378c sequence diversity, we calculated the ratio of the rates of non-synonymous to synonymous single nucleotide changes (dN/dS) of both genes with respect to the inferred MTB complex ancestor. The whole-gene dN/dS ranged from 0.97 to 0.47 for Rv3377c and Rv3378c respectively. The dN/dS for Rv3378c was similar to the

global dN/dS of the *M. tuberculosis* genome (0.45–0.67 in MTB complex (Comas et al., 2010) whereas the dN/dS of Rv3377c was significantly higher.

Bioinformatic algorithms that seek to identify which mutations are likely to alter protein function suggested that 60% (15/25) of nSNPs identified in Rv3377c are not likely to affect protein structure (Supplemental Table S1). However, 10 SNPs are predicted to have notable effects on structure. The replacement of a glycine at position 31 with a valine is predicted to impact the protein structure in the entire lineage 6, also known as *M africanum*. In summary we found that most clinical isolates have an intact Rv3377c-3378c locus and no clear evidence for selection driving the mutations identified, with the exception that lineage 6 might have an inactivating mutation.

Pathogen-specific gene expression

To determine whether 1-TbAd biosynthesis exists in species other than *M. tuberculosis*, we sought orthologs of *Rv3377c* and *Rv3378c* across the biological kingdoms. Confirming a prior analysis (Mann and Peters, 2012), few candidate orthologs were found among non-mycobacterial species. Low stringency searches did return candidate Rv3378c orthologs in *Zea mays* and *Dictyostelium discoideum*. However, considering all non-mycobacterial species, we did not identify candidate orthologs organized into a locus with both essential genes present. Among mycobacteria, orthologs of R3377c and Rv3378c could not be identified in most species, including disease-causing members of the *M. avium* complex, *M. kansasii* and *M. marinum* (Figure 2A). Instead, orthologs were present only within the MTB complex: in *M. bovis*, *M. bovis BCG* (Pasteur strain), *M. cannetti* and *M. africanum*. However, coding mutations were found in the locus in all species other than *M. tuberculosis* as detailed in Figure 2A. Two mutations were found the Pasteur strain of BCG (Pasteur), including a frameshift mutation that is likely inactivating (Figure 2A). Thus, we could not identify an intact orthologous locus outside the MTB complex, and we found a wild type sequence only in *M. tuberculosis*.

However, gaps in the analysis remained. Some but not all of the Rv3378c mutations, including those present in *M. africanum*, have been proven to inactivate lipid biosynthesis (Chan et al., 2014; Layre et al., 2014). Second, BCG vaccine strains are of particular interest because they cause false positive immunological tests. Although the locus in BCG strain Pasteur was inactivated through frameshift, other BCG or *M. bovis* strains might never have acquired these mutations, or they might have done so and then reverted to wild-type sequences. The 12 widely used BCG vaccine strains show variations in efficacy (Ritz et al., 2008), so differential expression of TbAd might account for this (Pethe et al., 2004). However, further analysis of all common vaccine strains used worldwide (Pasteur, Copenhagen, Japan, Mexican, Australian, Russia, Glaxo, Prague, Phipps, Connaught, Denmark, Tice) confirmed the presence of the frameshift mutation in all cases, so 1-TbAd is likely absent from all major vaccines used worldwide.

1-TbAd expression in pathogens

A general limitation of genetic analysis is that genes might exist that are functionally equivalent to Rv3377c and Rv3378c, but lack the sequence identity needed for identification

as orthologs. Therefore, we undertook biochemical analysis for 1-TbAd production, focusing on non-tuberculous mycobacteria and microbes, whose infection can mimic tuberculosis disease. By adapting the existing HPLC-MS lipidomics protocol (Layre et al., 2011), we validated a rapid method to specifically detect 1-TbAd within complex lipid mixtures. Analyzing total lipids, molecular events (linked intensity, *m/z* and retention time values) corresponding to all lipids are recorded, and the events corresponding to the mass (*m*/*z* 540.35) and retention time (23 min) of authentic 1-TbAd are reported (Figure 2B). Monitoring of unfractionated lipids avoids time-consuming methods needed to purify 1-TbAd, as well as artifacts introduced through sample handling. Datasets can be subjected to further data mining for events corresponding to other lipid products that are upstream and downstream in biosynthetic pathways, isomers of 1-TbAd, as well as TbAd variants with defined changes in mass due to chain length variation or oxidation. Initial validation showed that no other mycobacterial lipid obscured the event corresponding to the mass and retention time (t_R) of 1-TbAd. Further, total lipid extracts of the human THP-1 macrophage-like cell line did not generate detectable signals at this mass and t_R coordinate, even when amplified 1000-fold (Figure 2B, inset). Thus, no host lipid interfered with TbAd detection in this system, which is a key consideration for diagnostic applications.

Lipid extracts from non-actinomycete bacteria (*Escherichia coli*, *Staphylococcus aureus*) and fungi (*Aspergillis fumigatus*, *Candida albicans*) did not produce a 1-TbAd signal. Turning to mycobacteria, we could not detect 1-TbAd among reference strains of environmental (*M. fallax*) or non-pathogenic laboratory strains of mycobacteria (*M. smegmatis*, *M. phlei*) that lack *Rv3378c* orthologs (Figure 2C). In agreement with genetic results, we did not detect 1-TbAd among disease causing bacteria that are related to *M. tuberculosis* but lack identifiable orthologs of *Rv3377c-3378c* (*M. avium*, *M. marinum)* or those with ortholgous loci containing a known frameshift mutation *(M. bovis).* Among all organisms tested to date, only *M. tuberculosis* produces 1-TbAd.

Detection of 1-TbAd *in vivo*

To determine if 1-TbAd is detectable *in vivo*, we infected balb-C mice via inhalation for 21 days followed by collection of whole lung homogenates. Target validation *in vivo* does not require scanning of the complete lipidome, but instead focuses on sensitive and specific detection of pre-determined targets, which are likely to be highly diluted within host tissues. We developed a reversed phase method to increase chromatographic resolution, while taking advantage of the specificity and accuracy of triple quadrupole mass detection. This method detected a 1-TbAd standard at 11.8 minutes based on its known transition for 1-TbAd (m/z 540->408->136). Direct analysis of unfractionated lipids from whole lung homogenates did not detect signal corresponding to 1-TbAd in uninfected lung, but did detect a signal corresponding to 1-TbAd that was well above background signals in 6 of 6 infected mice tested, with representative data shown in Figure 3A. Thus, 1-TbAd is produced *in vivo* and is readily detected *ex vivo* in a 1-step HPLC-MS method.

Unknown terpene nucleoside

Ion chromatograms of the ion 540->408 generated also showed a second, later-eluting peak (14.6 min) with transitions that were distinct from 1-TbAd in all six mice tested (Figure 3A).

Unlike 1-TbAd, the unknown lipid contained a 540->148 ion and a relatively bright 540->408 transition that was equivalent in intensity to the 540->136 ion. Returning to normal phase HPLC-Q-TOF-MS experiments using in vitro grown *M. tuberculosis*, ion chromatograms measured at *m/z* 540.4 also showed two peaks at ~25 and ~7 min, which corresponded to late eluting 1-TbAd and an 'early eluting' unknown lipid detected in mice (Figure 3B). CID-MS analysis of the unknown showed fragment ions of *m/z* 148.062, and 408.313, matching the pattern of the late eluting unknown from reversed phase chromatography, suggesting that they were the same molecule (Figure 3 A, C). Thus, the unknown later eluting compound detected in mice was likely derived from *M. tuberculosis* rather than the host. Also, separate tracking of the unknown and 1-TbAd in the bacterial pellet and conditioned supernatant demonstrated a much higher ratio of signal for the unknown in supernatants (Figure 3B). Thus, the early eluting unknown was likely generated during or after transit from the cytosol to the extracellular space.

1-TbAd isomers

The unknown (*m/z* 540.355) and 1-TbAd (m/z 540.3545, calculated) had identical molecular ion masses (+/− 2 ppm) and similar fragmentation patterns. The unknown has a much lower retention time in normal phase chromatography. Therefore, the unknown was likely a much less polar isomer of 1-TbAd. CID-MS ions were assigned unequivocally as $C_5H_6N_5^+$ (*m/z* 136.0618) and $C_{10}H_{14}N_5O_4^+$ (*m/z* 268.1040), matching the mass of adenine and adenosine, respectively. The neutral loss leading to *m/z* 268.1040 suggested the loss of $C_{20}H_{32}$, likely a diterpene moiety (Figure 3C). A fragment ion with the formula of $C_{25}H_{38}N_5$ (calculated *m/z* 408.3122) matches the loss of a pentose-derived fragment ($C_5H_8O_4$, calculated 132.0423 Da). Therefore, the unknown was a diterpene-substituted adenosine ($C_{30}H_{45}N_5O_4$).

Experiments were guided by two hypotheses. The unknown isomer and 1-TbAd might differ in the type of diterpene carried, or a tuberculosinyl group could have an alternate linkage to adenine. *M. tuberculosis* produces geranylgeranyl (GG) pyrophosphate, and tuberculosinyl pyrophosphate is hydrolyzed to isotuberculosinol in vitro (Nakano et al., 2011). Therefore, 1-GGAd and 1-isoTbAd were candidate structures (Figure 4A). However, they would be expected to have largely the same ionic properties as 1-TbAd. Alternatively, a tuberculosinyl linkage at the N^6 adenosine position would mimic common N^6-linked adenine compounds including zeatin (Heyl et al., 2012) and would be expected to alter the ionic properties observed in the unknown. Therefore, we considered N^6-tuberculosinyladenosine as a candidate structure (Figure 4A), and we synthesized N^6-(*E*,*E*,*E*)- and N^6-(*E*,*E*,*Z*)-geranylgeranyl-adenosine as a model N^6-linked compound to evaluate this hypothesis (Figure 4B and Supplemental Data 2).

Identification of N^6-TbAd

Detection of the isomer in cultured *M. tuberculosis* provided a route to purifying it in greater quantity. After growing *M. tuberculosis* cultures in roller bottles, ~50 mg of lipid was fractionated by normal phase open-column chromatography followed by reversed-phase HPLC. Adding 0.2% trifluoroacetic acid (TFA) to the mobile phase increased the retention of 1-TbAd in reversed phase chromatography, but did not affect the unknown isomer, confirming that the two compounds differed in their ionic properties. This purification

sequence produced ~ 300 μg of each isomer, so that the combined yield (~1.2%) confirmed ESI-MS studies (Figure 1A).

Analysis of the MS data and ^{1}H, COSY, NOESY, and HMQC NMR spectra (800 MHz) (Table 1) established the structure of the unknown as N^6-TbAd. A 1-linked geranylgeranyl side chain could not produce signals as low as the observed signals at 1 ppm. The single vinylic proton $H^{14"}$ would be replaced by the two $H^{15"}$ protons in the isotuberculosinyl group, ruling out this possibility. The spectral data for 1-TbAd and N^6-TbAd (Table 1) both correspond closely to those of tuberculosinol (Nakano et al., 2005, 2009, 2011; Hoshino et al., 2011) except for the expected difference in the side chain protons and carbons. The H-17" methyl doublets at δ 0.85, and H-20" methyl singlets at δ 0.65–0.66 are particularly characteristic of the tuberculosinyl ring system.

In the unknown, the signals of the adenosine moiety and the side chain protons (H-12" to H-16") correspond closely to those of N^6-(3-methyl-2-butenyl)adenosine (Casati et al., 2010, 2011; Ottria et al., 2010). The adenine protons at δ 8.24 (H-8) and 8.26 (H-8) are characteristic of an N^6-substituted adenosine and are different from those of 1-TbAd at δ 8.53 and 8.66. All COSY and NOESY correlations are consistent with the assignment of the unknown as N^6-TbAd. The absorption of the allylic methylene group is a broad peak at 4.22-4.16 ppm due to slow rotation at room temperature around the C-N^6 bond (Uzawa and Anzai, 1988). The diagnostic ion at *m/z* 148 in the N^6-TbAd is explained by an alternate protonation of the double bond at C-14" and subsequent charge-directed fragmentation, which would be suppressed in 1-TbAd, having its positive charge stabilized in the adenine ring (Figures 4D).

Dimroth rearrangement of 1-TbAd to N^6-TbAd

A plausible mechanism to account for the biological origin of N^6-TbAd is that 1-TbAd is converts to N^6-TbAD by a Dimroth rearrangement, which involves attack of a nucleophile (such as hydroxide or an amine) at C-2 to form a ring-opened intermediate followed by ring closure by the unsubstituted nitrogen to give the N^6-linked form (Figure 5A), (El Ashry et al., 2010; Fujii and Itaya, 1998; Grimm and Leonard, 1967; Leonard et al., 1966; Macon and Wolfenden, 1968; Ottria et al., 2010; Snyder and Adams, 2011). Products from recombinant Rv3378c enzyme showed a strong signal for 1-TbAd and no signal for N^6-TbAd, demonstrating that the latter is not a primary product of this enzyme (Figure 5B). Further, we observed that treating 1-TbAd under conditions known to favor the Dimroth reaction (Me_2NH in methanol) led to clean conversion into N^6-TbAd (Figure 5C).

The pK_a of the conjugate acid of N^6-(3-methylbut-2-enyl)adenosine and of the 1-(3-methyl-2-butenyl) adenosine cation are 3.76 and 8.47, respectively (Martin and Reese, 1968). Thus, the markedly differing chemical and physical properties of 1-TbAd and N^6-TbAd arise from their differing charge states, and they predict differing response to altered pH within the physiological range. At neutral pH, N^6-TbAd should exist as the neutral free base, explaining why it is uncharged and elutes more readily from normal-phase columns (Figure 5A–B). 1-TbAd is predominantly charged (protonated) at this pH and binds tightly to polar silica gel and diolmodified stationary phases.

The data suggests a two-step model of biosynthesis whereby Rv3378c mediates conjugation of tuberculosinyl pyrophosphate and adenosine to produce 1-TbAd in the cytosol, and 1-TbAd could later rearrange to N^6-TbAd in other compartments (Figure 3B). However, in vivo transformation of 1-TbAd to N^6-TbAd in mice could not be unequivocally established because mouse lung analysis required harvest, homogenization and HPLC-MS analysis at the bench. The *ex vivo* workup of lung might have inadvertently induced Dimroth rearrangement, but this possibility was ruled out by workup of lung spiked with 1-TbAd, which detected only 1-TbAd (Figure 5D). In summary, Rv3378c produces 1-TbAd which is transformed to N^6-TbAd by the Dimroth rearrangement, and both compounds occur in vivo and represent specific markers of *M. tuberculosis* infection (Figure 5E).

Discussion

These studies show that 1-TbAd and the newly identified in vivo biotransformation product, N^6-TbAd, possess key features of high quality targets for development in antigen capture diagnostic or serological tests. These features include abundant expression and efficient shedding, as well as absence of expression in common sources of false positive test results including vaccine strains, non-mycobacterial lung pathogens and environmental bacteria. Further, these studies provide evidence for intact biosynthetic genes among clinical strains and expression by reference strains under many different conditions surveyed, suggesting that it could be a sensitive biochemical marker of infection. Further, taking advantage of the broad lipidomics detection methods, the data demonstrate sensitive detection of both TbAd isomers without interference by any other host, bacterial or fungal lipids using a standardized and simple HPLC-MS test.

Another criterion for use of shed molecules as diagnostic tools is that they remain intact during in vivo infection or, if they are metabolized by the host, generate recognizable transformation products that are distinct from all host molecules. These data identify the complete structure of N^6-TbAd as a previously unknown *in vivo* biotransformation product of *M. tuberculosis*. Here we report a small proof-of-principle study in which both TbAd isomers are readily detected in the lung. Starting from a systematic screen of all *M. tuberculosis* specific compounds, these two terpene nucleosides meet all criteria needed for entry into human population studies.

Further, these data determined the origin and stepwise appearance of the TbAd isomers in ways that add to a basic understanding of tuberculosis pathogenesis. Prior studies by Russell identified that Rv3378c acts by some unknown mechanism to inhibit phagosomal acidification and promote mycobacterial survival in macrophages (Pethe et al., 2004). Rv3378c is localized in the bacterial cytosol, and yet might act in some way on events in the phagosomal space. Our results identify the product of Rv3378c as an amphipathic diterpene nucleoside (Layre et al., 2014) and show that the product of Rv3378c is efficiently exported to the extracellular space, so that this product does escape to the phagosomal compartment that is involved in pH regulation. Unexpectedly, 1-TbAd undergoes biotransformation to N^6-TbAd, a previously unknown in vivo metabolite within the infected host.

1-TbAd and N^6-TbAd show only one difference in structure, but the site of tuberculosinyl linkage determines their pKa and charge state. Quantitative consideration of the ionic properties of 1-TbAd (pKa = 8.5) and N^6-TbAd (pKa=3.8) (Kapinos et al., 2011) predict that N^6-TbAd is neutral at pH 7.4, whereas 1-TbAd exists as a mixture of [TbAd]- and $[TbAdH]^+$. During active infection, *M. tuberculosis* typically grows in moderately acidic phagosomal compartments or extracellular caseous material. Therefore, 1-TbAd would change in charge states as it leaves neutral pH environment of the mycobacterial cell wall, traverses the acidic phagosome and enters the neutral environment of infected tissues and finally the peripheral circulation. These effects would alter its water solubility and adherence to anionic membranes in the three compartments. Further, the Dimroth rearrangement is favored under basic conditions. At lower pH, the concentration of nucleophiles decreases, so 1-TbAd is less prone to rearrangement in the acidic environment of localized acute infection. Thus, 1-TbAd has an intrinsic chemical feature that would stabilize its structure while in the phagosome, and the N^6-form might be more efficiently generated in non-acid compartments that are distant from the site of infection. In this scenario, 1-TbAd is more important in control of the export to the phagosome, and N^6-TbAd might represent an altered, possibly inactivated, form of 1-TbAd whose importance derives its role as a recognizable infection marker.

These studies set the stage for human population studies to detect both TbAd isomers as targets for diagnosis in clinically assayable fluids like sputum, serum, and urine. Both isomers are readily detected *ex vivo* using a routine LCMS method. For serological applications, the difference in charge state strongly predicts differing host antibody response to 1-TbAD and N^6-TbAd, and knowledge of these differing ionic properties may guide the future, separate development of these terpene nucleosides as targets for ELISAs. Although LAM ELISAs have emerged as a clinical test with some utility, one limiting factor for this and other mycobacterial lipid-specific ELISAs is thought to be false positive results from subclinical exposures to non-tuberculous mycobacteria or BCG vaccination (Baumann et al., 2014; Lawn, 2012). Here we demonstrated the lack of both TbAd isomers and the Rv3377c-3378c genes in common sources of false positive immune response. These discovery-based chemical studies open two new paths to development of simple and specific antigen or serological tests, which are urgently needed for better care of patients and vaccine research.

Significance

The discovery and detection of 1-tuberculolsinyladenosine and its rearrangement product N^6-tuberculosinyladenosine, produced in-vivo during the course of mycobacterial infection in a murine model, provides unique chemical markers to diagnose tuberculosis disease.

Experimental Procedures

Bacteria cultures

M. tuberculosis was grown in the absence of detergent in 7H9 medium supplemented with 10% ADC or OADC or in Sauton's medium until mid-log phase with or without phosphate-citrate buffers adjusted at the given pH. Supernatant was separated from bacteria by

centrifugation at 2000g and filtration of supernatant cultures (0.2 μm). Cell-associated or secreted lipids were extracted using successive contact in chloroform and methanol mixtures or ethyl acetate, respectively, as previously described (Layre et al., 2011).

Purification of 1-TbAd and the unknown TbAd isomer

Lipids were prepared by serially extracting desiccated or wet pellets of *Mycobacterium tuberculosis* H37Rv and other named species with 1:2 (V:V), 1:1, and then 2:1 chloroform/methanol. The combined extracts were dried at room temperature and stored in 1:1 chloroform/methanol solution (~10 mg/mL). Fifty mg was concentrated under nitrogen resulting in a slurry that was loaded on an open silica gel column (2 cm × 1.6 cm). Lipids were eluted with 10 ml of solvent in the following sequence: chloroform, chlorofom/isopropanol (95:5, V/V), chloroform/isopropanol (90:10 V/V) and chloroform/methanol (50:50, V/V). Fractions were monitored for ions with a nominal *m/z* of 540 and the *m/z* 540 >*m/z* 408 >*m/z* 148 transitions in the nanospray MS^2 and MS^3, with the *m/z* 148 ions seen in the unknown (N^6-TbAd), not 1-TbAd. N^6-TbAd eluted using isopropanol/chloroform while 1-TbAd required methanol/chloroform for elution. After evaporation 1-TbAd and its isomer were further purified using reversed-phase HPLC using octadecylmodified silica (5 micron) semi-preparative column (250 × 10 mm) run under isocratic conditions at 3.0 mL/min. 1-TbAd was purified using 89.9:10:0.2 (v/v) methanol/water/trifluoroacetic acid monitoring by UV at 260 nm. Collected fractions were evaporated to dryness in 0.5-mL portions with 2.0 mL of acetonitrile added to help minimize exposure of the collected 1-TbAd to TFA. N^6-TbAd was purified using 5:90:5 chloroform:methanol:water (V:V:V) as the mobile phase, monitoring by UV at 260 nm, and eluted at.14.5 min.

Ion Trap MS

Multistage CID-MS experiments were completed on a linear ion trap (Thermoscientific LXQ) using a nanoelectrospray ionization source and borosilicate glass pipettes pulled to a 2 μm tip.

Basic Dimroth conversion

1-TbAd was converted to N^6-TbAd by dissolving a small portion of 1-TbAd in 2 M dimethylamine in methanol at room temperature for 3 days.

NMR analysis

A (Bruker Avance 800) NMR instrument was used to analyze samples dissolved in deuterated methanol (0.5 mg/mL) that were compared to a 1-TbAd standard (Layre et al., 2014).

HPLC-ESI-Q-TOF analyses

Extracted cell-associated or secreted lipids were concentrated under reduced pressured at room temperature. One hundred micrograms were resuspended at 0.5 mg/mL in 70:30 hexane:isopropyl alcohol and 20uL were injected for analysis by *HPLC-ESI-MS* (Agilent 6520 QToF) as described (Layre et al., 2014)

Reversed phase HPLC-MS

Mouse (Balb-c) lungs were harvested after three weeks and analyzed using an injection volume of 50 uL using a reversed phase, linear gradient elution method developed using a Zorbax Eclipse Plus-C18 column (2.1 mm × 50 mm, 3.5-μm particle size, Agilent Technologies, Santa Clara, CA, USA). The solvents were (A) 0.1% formic acid in milliQ water, and (B) 0.1% formic acid in acetonitrile. After 8 minute equilibration with 100% solvent (0.2 mL/min), a two-part linear solvent gradient was used as follows: 0–10 min, 0% to 100% solvent B; 10–22 min, 100% to 0% solvent B.

N^6-TbAd structural analysis

Synthesis of N^6-linked adenosine analogs and detailed descriptions of N^6 products are described in Supplemental Experimental Procedure.

Genetic analysis of Rv3377c and Rv3378c

Detection of SNPs in Rv3377c and Rv3378c was done from multiple alignments of the orthologous sequences of both genes retrieved from the genomes of 431 strains representative of the 7 main phylogenetic lineages of the MTB complex, comprising 271 published genomes (Comas et al., 2010), 161 whole genomes that are currently being analyzed and 12 published BCG strains (Copin et al., 2014). Nucleotides differing between the query strains and the reference strain (Comas et al., 2010) were recorded as SNPs. The genetic diversity of Rv3377c and Rv3378c in the MTB complex was done by multiple alignments of the orthologous sequences of both genes retrieved from the genome of 432 strains. CODEML from PAML 3.14b (Yang, Z., 1997) was used to estimate dN/dS rates. The codon frequencies were derived from the average nucleotide frequencies at the three-codon position (F3 × 4 model), and the model was chosen to compute one ratio of dN/dS fixed for all sites (model = 1 and NSsites = 0). The dN/dS was calculated using their orthologous sequences in the most recent common ancestor as previously described (Comas, et al., 2010). To predict the impact of amino acid changes on protein function, we used 3 algorithms in parallel: Sift (http://sift.jcvi.org/), Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/) and Provean (http://provean.jcvi.org/index.php).

Lipid extracts of microbial species

The chloroform/methanol extraction and sources of mycobacterial, non-mycobacterial, and fungal species used has been described (Ly et al., 2013). In addition, *M.marinum*, *C. albicans*, and *A. fumigatus* were obtained from the American Type Culture Collection.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors thank Lisa Prach and Thomas Alber for providing Rv3378c enzyme and Thomas Ennis for a BLAST search result. This work was funded by Mark and Lisa Schwartz, the Burroughs Wellcome Fund Program in Translational Medicine, the TB Vaccine Accelerator Award from the Bill and Melinda Gates Foundation and the NIAID (U19 AI076217 R01 049313) and the Dutch Science foundation NWO. The 800 MHz spectrometer in the

Landsman Research Facility, Brandeis University was purchased under NIH RR High-End Instrumentation program, 1S10RR017269-01.

REFERENCES

Baumann R, Kaempfer S, Chegou NN, Oehlmann W, Loxton AG, Kaufmann SH, van Helden PD, Black GF, Singh M, Walzl G. Serologic diagnosis of tuberculosis by combining Ig classes against selected mycobacterial targets. J Infect. 2014; 69:581–589. [PubMed: 24968240]

Casati S, Manzocchi A, Ottria R, Ciuffreda P. 1H-13C and 15N NMR assignments for N6-isopentenyladenosine/inosine analogues. Magnetic resonance in chemistry : MRC. 2010; 48:745–748. [PubMed: 20625985]

Casati S, Manzocchi A, Ottria R, Ciuffreda P. 1H, 13C and 15N NMR spectral assignments of adenosine derivatives with different amino substituents at C6-position. Magnetic resonance in chemistry : MRC. 2011; 49:279–283. [PubMed: 21387403]

Chan HC, Feng X, Ko TP, Huang CH, Hu Y, Zheng Y, Bogue S, Nakano C, Hoshino T, Zhang L, et al. Structure and inhibition of tuberculosinol synthase and decaprenyl diphosphate synthase from *Mycobacterium tuberculosis*. J Am Chem Soc. 2014; 136:2892–2896. [PubMed: 24475925]

Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, Ernst JD, Gagneux S. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. Nat Genet. 2010; 42:498–503. [PubMed: 20495566]

Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei L, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D, Gagneux S. Out-of-Africa migration and neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nat Genet. 2013; 45:1176–1182. [PubMed: 23995134]

Copin R, Coscollá M, Efstathiadis E, Gagneux S, Ernst JD. Impact of in vitro evolution on antigenic diversity of *Mycobacterium bovis* bacillus Calmette-Guerin (BCG). Vaccine. 2014; 32:5998–6004. [PubMed: 25211768]

Couturier MR, Graf EH, Griffin AT. Urine antigen tests for the diagnosis of respiratory infections: legionellosis, histoplasmosis, pneumococcal pneumonia. Clinics in laboratory medicine. 2014; 34:219–236. [PubMed: 24856525]

El Ashry ESH, Nadeem S, Shah MR, El Kilany Y. Recent advances in the Dimroth rearrangement: a valuable tool for the synthesis of heterocycles. Advances in Heterocyclic Chemistry. 2010; 101:161–228.

Fujii T, Itaya T. The Dimroth rearrangement in the adenine series: a review updated. Heterocycles. 1998; 48:359–390.

Grimm WA, Leonard NJ. Synthesis of the "minor nucleotide" N6-(gamma, gamma-dimethylallyl)adenosine 5'-phosphate and relative rates of rearrangement of 1-to N6-dimethylallyl compounds for base, nucleoside, and nucleotide. Biochemistry. 1967; 6:3625–3631. [PubMed: 6076616]

Hanekom WA, Dockrell HM, Ottenhoff TH, Doherty TM, Fletcher H, McShane H, Weichold FF, Hoft DF, Parida SK, Fruth UJ. Immunological outcomes of new tuberculosis vaccine trials: WHO panel recommendations. PLoS Med. 2008; 5:e145. [PubMed: 18597551]

Heyl A, Riefler M, Romanov GA, Schmulling T. Properties, functions and evolution of cytokinin receptors. European journal of cell biology. 2012; 91:246–256. [PubMed: 21561682]

Hoshino T, Nakano C, Ootsuka T, Shinohara Y, Hara T. Substrate specificity of Rv3378c, an enzyme from *Mycobacterium tuberculosis*, and the inhibitory activity of the bicyclic diterpenoids against macrophage phagocytosis. Organic & biomolecular chemistry. 2011; 9:2156–2165. [PubMed: 21290071]

Kapinos LE, Operschall DIBP, Larsen E, Sigel H. Understanding the acid–base properties of adenosine: the intrinsic basicities of N1, N3 and N7. Chem. Eur. J. 2011; 17:8156–8164. [PubMed: 21626581]

Lalvani A, Pareek M. A 100 year update on diagnosis of tuberculosis infection. Br Med Bull. 2010; 93:69–84. [PubMed: 19926636]

Lawn SD. Point-of-care detection of lipoarabinomannan (LAM) in urine for diagnosis of HIV-associated tuberculosis: a state of the art review. BMC Infect Dis. 2012; 12:103. [PubMed: 22536883]

Lawn SD, Kerkhoff AD, Vogt M, Wood R. Diagnostic accuracy of a low-cost, urine antigen, point-of-care screening assay for HIV-associated pulmonary tuberculosis before antiretroviral therapy: a descriptive study. Lancet Infect Dis. 2012; 12:201–209. [PubMed: 22015305]

Layre E, Lee HJ, Young DC, Martinot AJ, Buter J, Minnaard AJ, Annand JW, Fortune SM, Snider BB, Matsunaga I, et al. Molecular profiling of *Mycobacterium tuberculosis* identifies tuberculosinyl nucleoside products of the virulence-associated enzyme Rv3378c. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:2978–2983. [PubMed: 24516143]

Layre E, Sweet L, Hong S, Madigan CA, Desjardins D, Young DC, Cheng TY, Annand JW, Kim K, Shamputa IC, et al. A comparative lipidomics platform for chemotaxonomic analysis of *Mycobacterium tuberculosis*. Chemistry & biology. 2011; 18:1537–1549. [PubMed: 22195556]

Leonard NJ, Achmatowicz S, Loeppky RN, Carraway KL, Grimm WA, Szweykowska A, Hamzi QH, Skoog F. Development of cytokinin activity by rearrangement of 1-substituted adenines to 6-substituted aminopurines: inactivation by N6, 1-cyclization. Proceedings of the National Academy of Sciences of the United States of America. 1966; 56:709–716. [PubMed: 5229988]

Ly D, Kasmar AG, Cheng TY, de Jong A, Huang S, Roy S, Bhatt A, van Summeren RP, Altman JD, Jacobs WR, Adams EJ, Minnaard AJ, Porcelli SA, Moody DB. CD1c tetramers detect ex vivo T cell responses to processed phosphomycoketide antigens. J. Exp. Med. 2013; 210:729–741. [PubMed: 23530121]

Macon JB, Wolfenden R. 1-Methyladenosine. Dimroth rearrangement and reversible reduction. Biochemistry. 1968; 7:3453–3458. [PubMed: 5681457]

Mann FM, Peters RJ. Isotuberculosinol: the unusual case of an immunomodulatory diterpenoid from *Mycobacterium tuberculosis*. MedChemComm. 2012; 3:899–904. [PubMed: 23926455]

Martin MG, Reese CB. Some aspects of the chemistry of N(1)- and N(6)-dimethylallyl derivatives of adenosine and adenine. Journal of the Chemical Society Perkin transactions 1. 1968; 14:1731–1738. [PubMed: 5691105]

Nakano C, Hoshino T. Characterization of the Rv3377c gene product, a type-B diterpene cyclase, from the *Mycobacterium tuberculosis* H37 genome. Chembiochem : a European journal of chemical biology. 2009b; 10:2060–2071. [PubMed: 19618417]

Nakano C, Okamura T, Sato T, Dairi T, Hoshino T. *Mycobacterium tuberculosis* H37Rv3377c encodes the diterpene cyclase for producing the halimane skeleton. Chem Commun (Camb). 2005:1016–1018. [PubMed: 15719101]

Nakano C, Ootsuka T, Takayama K, Mitsui T, Sato T, Hoshino T. Characterization of the Rv3378c gene product, a new diterpene synthase for producing tuberculosinol and (13R, S)-isotuberculosinol (nosyberkol), from the *Mycobacterium tuberculosis* H37Rv genome. Bioscience, biotechnology, and biochemistry. 2011; 75:75–81.

Ottria R, Casati S, Manzocchi A, Baldoli E, Mariotti M, Maier JA, Ciuffreda P. Synthesis and evaluation of in vitro anticancer activity of some novel isopentenyladenosine derivatives. Bioorganic & medicinal chemistry. 2010; 18:4249–4254. [PubMed: 20494583]

Pethe K, Swenson DL, Alonso S, Anderson J, Wang C, Russell DG. Isolation of *Mycobacterium tuberculosis* mutants defective in the arrest of phagosome maturation. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101:13642–13647. [PubMed: 15340136]

Ritz N, Hanekom WA, Robins-Browne R, Britton WJ, Curtis N. Influence of BCG vaccine strain on the immune response and protection against tuberculosis. FEMS Microbiol Rev. 2008; 32:821–841. [PubMed: 18616602]

Sartain MJ, Dick DL, Rithner CD, Crick DC, Belisle JT. Lipidomic analyses of *Mycobacterium tuberculosis* based on accurate mass measurements and the novel "Mtb LipidDB". JLipid Res. 2011; 52:861–872. [PubMed: 21285232]

Shelhamer JH, Gill VJ, Quinn TC, Crawford SW, Kovacs JA, Masur H, Ognibene FP. The laboratory evaluation of opportunistic pulmonary infections. Ann Intern Med. 1996; 124:585–599. [PubMed: 8597323]

Snyder, NL.; Adams, TP. Name Reactions in Heterocylic Chemistry II. Li, JJ., editor. Wiley; 2011.

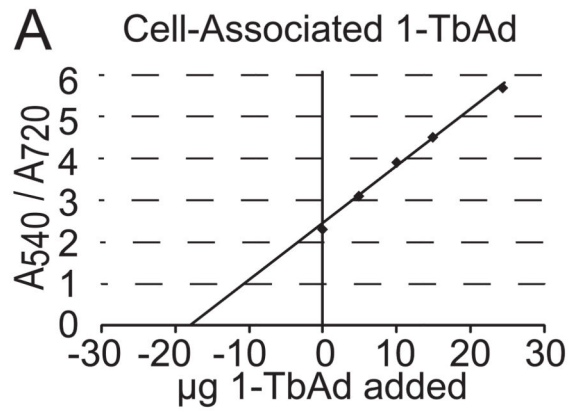
Uzawa J, Anzai K. Restriction of the C6-N^6 bond rotation of adenosine compounds investigated by ^{1}H- and ^{15}N-NMR spectra. Liebigs Ann Chem. 1988; 1988:1195–1196.

World Health Organization. Global Tuberculosis Report. 2014. http://www.who.int/tb/publications/global_report/en/.

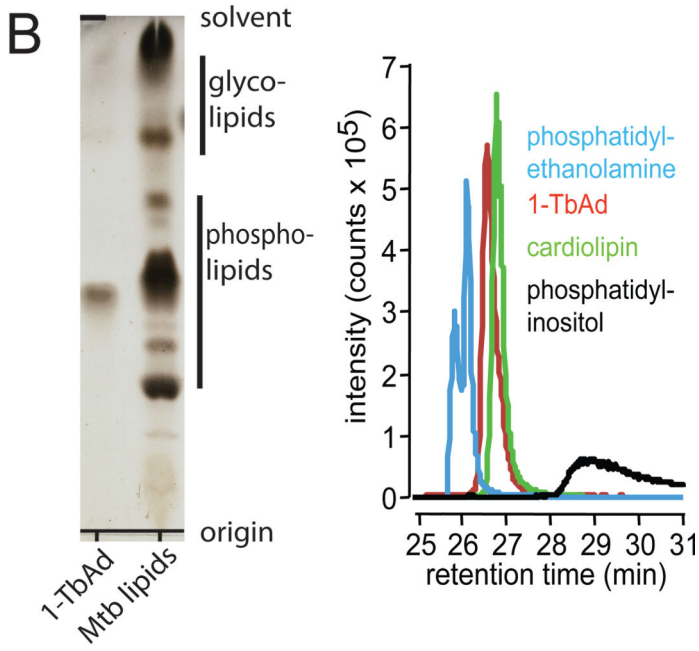
Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 1997; 13:555–556. [PubMed: 9367129]

Highlights

- N^6-tuberculosinyladenosine (rearrangement of 1-tuberculosinyladenosine) is described.
- These unique terpene nucleosides are detected in *M. tuberculosis* infected mouse lung.



Experiment	Lipid	1-TbAd	%TbAd
1	4.47 mg	64 µg	1.5
2	7.11 mg	100 µg	1.4
3	1.58 mg	18 µg	1.1



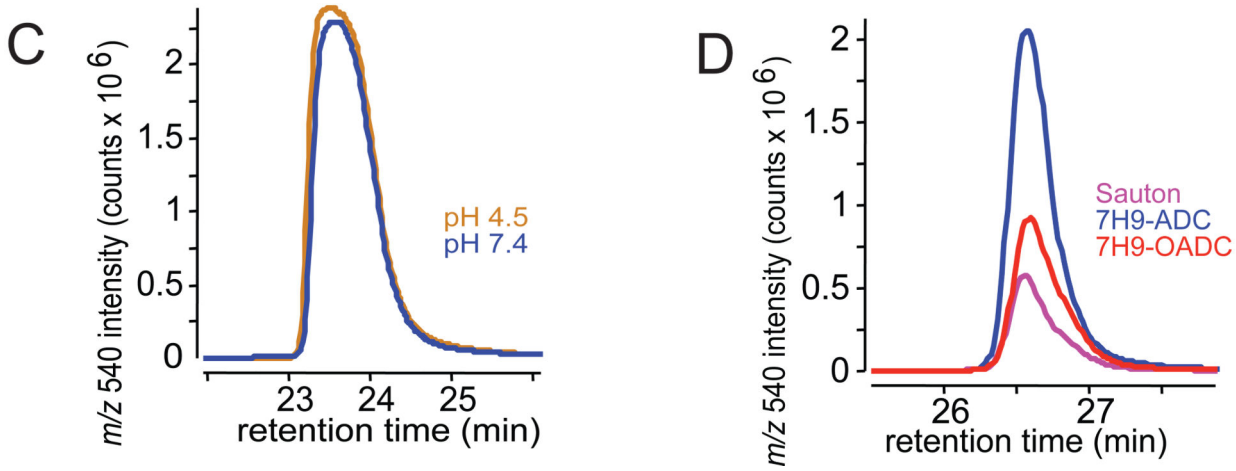


Figure 1.

1-TbAd is a major lipid in *M. tuberculosis*. (A) *M. tuberculosis* was grown in 7H9 medium in batches (experiments 1–3) and washed, extracted into organic solvents and weighed (Lipid). Lipids were spiked with 1-TbAD and subjected to positive mode HPLC-Q-TOF-MS analysis to estimate 1-TbAd concentration. (B) *M. tuberculosis* lipids were separated in normal phase silica TLC or HPLC-MS, demonstrating near co-elution of 1-TbAd with abundant membrane phospholipids. (C–D) *M. tuberculosis* grown in minimal (Sauton) or complete 7H9 with albumin (A), oleic acid (O), dextrose (D) and catalase (C) at the indicated pH was subjected to lipid extraction and positive mode normal phase HPLC-Q-TOF-MS.

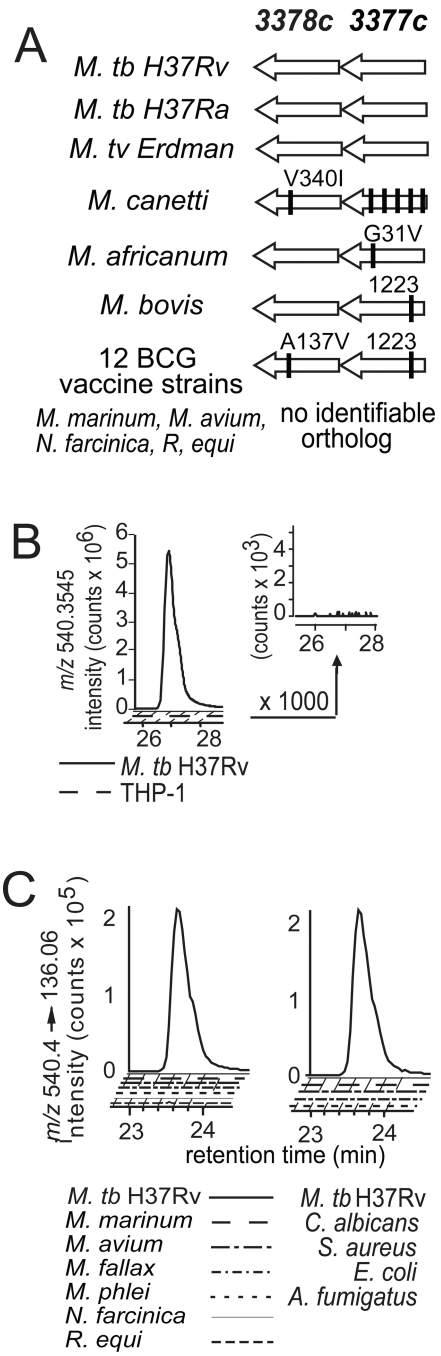


Figure 2.

Genetic and biochemical analysis of TbAd biosynthetic genes (Rv3378c, Rv3377c) and 1-TbAd in lung pathogens, environmental mycobacteria and vaccine strains. (A) The TbAd locus is comprised of a geranylgeranyl pyrophosphate cyclase (Rv3377c) and tuberculosinyl adenosine transferase (Rv3378c). Orthologous genes are shown by arrows, and SNPs are indicated by vertical lines. Supplemental table S1 lists polymorphisms for these genes in the *M. tuberculosis* complex. (B–C) Lipid extracts from human macrophage-like cells (THP-1), *M. tuberculosis* H37Rv or the indicated species were subjected to HPLC-MS.

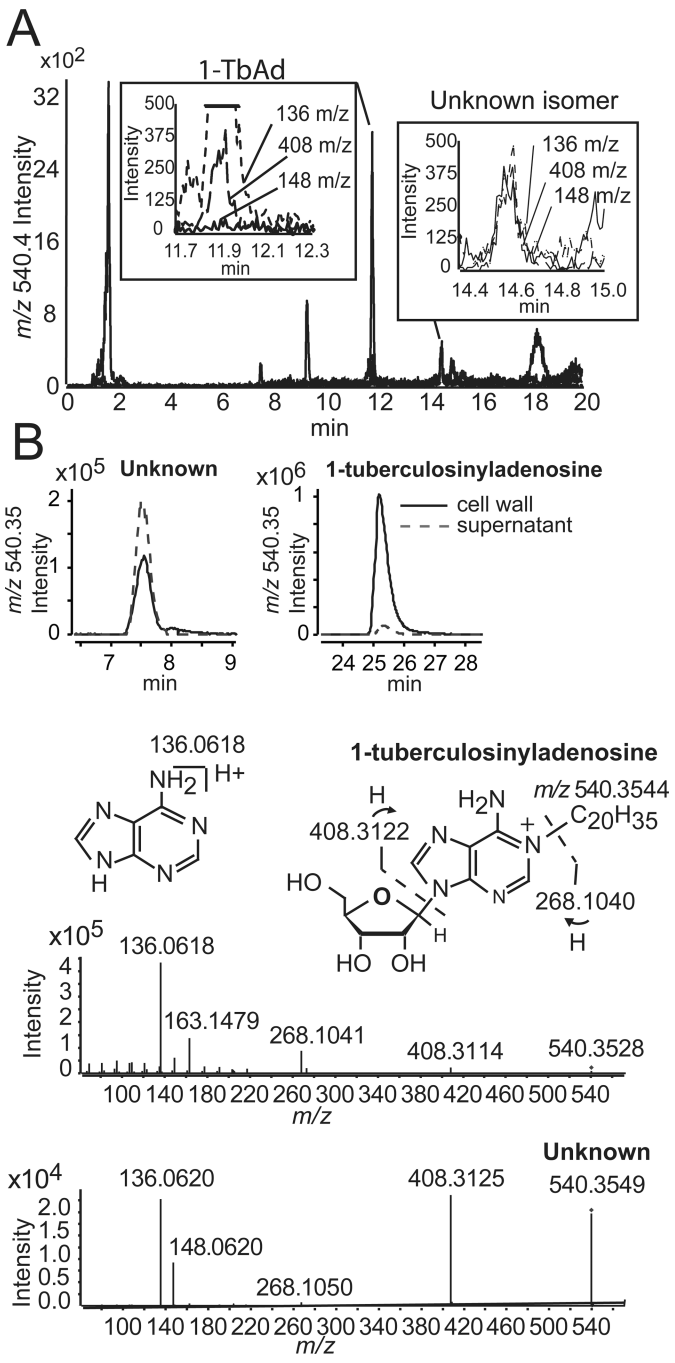


Figure 3.

Detection of 1-TbAd ex vivo in mice along with a previously unknown diterpene nucleoside. (A) Reversed phase, triple quadrupole HPLC-MS analysis of tissue extracts from *M.tuberculosis* infected mice with ion monitoring at *m/z* 540.4 detected a peak at 11.7 min that was identified as 1-TbAd based on near co-elution with an authentic 1-TbAd standard (not shown) and the diagnostic transitions (540->408->136) characteristic of 1-TbAd. Another peak with the same mass at 14.1 minutes demonstrates transitions that are not seen in 1-TbAd including a 540->148. Results are representative of six experiments in

which both ions were detected in all cases. (B) Separate experiments conducted on in vitro grown *M. tuberculosis* H37Rv analyzed in normal phase HPLC-Q-TOF-MS, likewise identified two terpene nucleoside isomers, with the late eluting form matching the retention time of 1-TbAd. Separate analysis of cell-associated and shed lipids were accomplished by extraction of cell pellets and conditioned supernatants. CID-MS analysis confirmed the expected ions for the late eluting 1-TbAd isomer. CID-MS of the unknown detected ions consistent with a diterpene adenosine structure, but bright ions at *m/z* 408.31, 148.06 were seen only in the early eluting isomer (unknown).

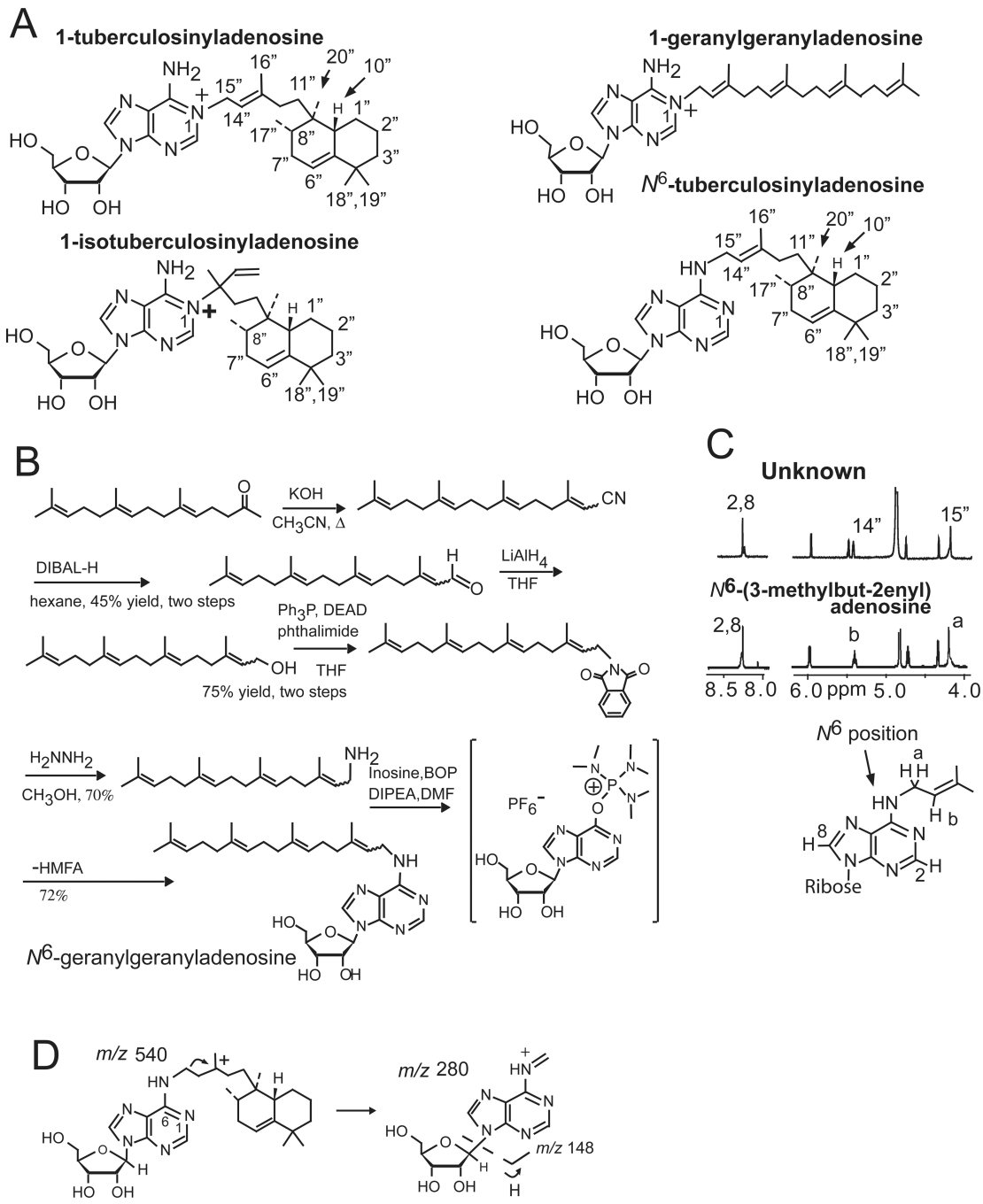


Figure 4.

NMR analysis of the unknown isomer establishes the structure of N^6-tuberculosinyladenosine. (A) The alternative diterpene hypothesis predicts compounds that 1-TbAd and the unknown isomer might differ in the identity of the diterpene coupled to generate 1-geranylgeranyladenosine or 1-isotuberculosinyl adenosine, which are depicted as candidate structures. Alternatively the unknown isomer could be generated by coupling adenosine to tuberculosinol at the N^6 position of adenine to generate N^6-TbAd. (B) N^6-geranylgeranyladenosine was synthesized to provide a 6-linked NMR standard, which is

analyzed with the unknown in supplemental data S1 and S2. (C) NMR (Bruker 800 MHz) of the unknown and standard shows upfield shifts, relative to 1-TbAd (Layre et al.,2014), in the protons of the tuberculosinyl group closest to adenine. These are accounted for by the absence of a positive charge on adenine in the N^6-TbAd. The unknown also compares well in this spectral region with N^6-(3-methyl-2-butenyl)adenosine. The resonances from the tuberculosinyl side chain positions 14" and 15" (labeled with a and b in the corresponding N^6-(3-methyl-2-butenyl)adenosine spectrum) adjacent to the heterocyclic adenine moiety are shifted similarly, relative to 1-TbAd, due to the positive to neutral charge shift. (D) The presence of the *m/z* 148 fragment in N^6-TbAd results from charge-directed fragmentation after protonation of a double bond that can occur with the N^6-but not the analogous 1-TbAd.

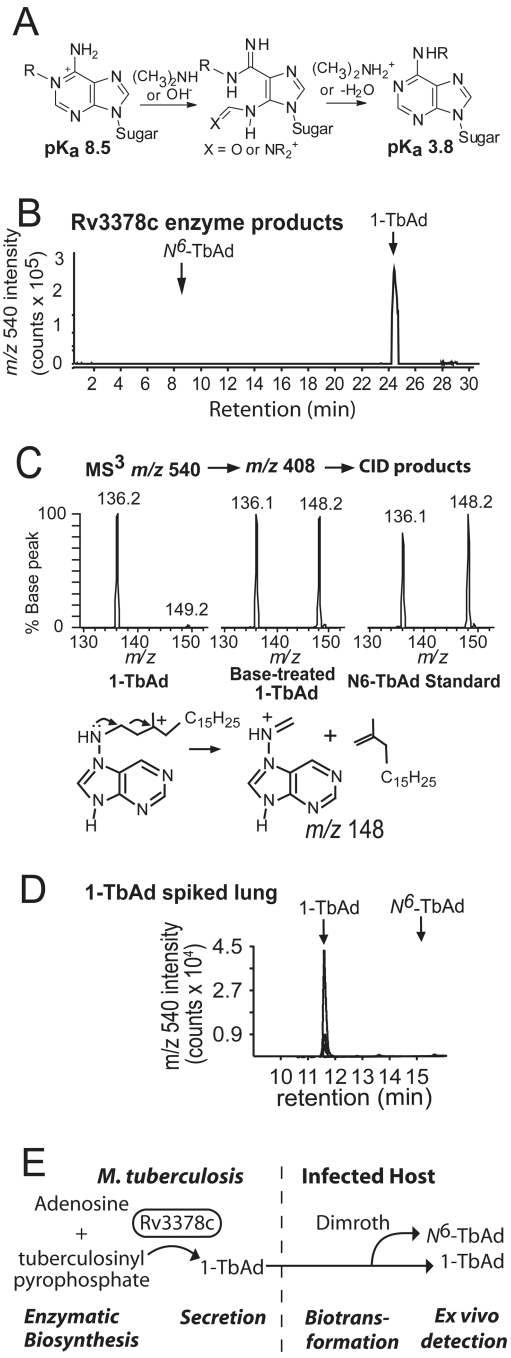


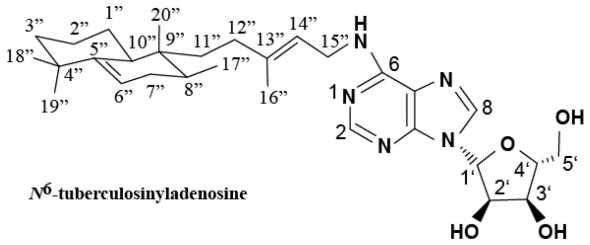
Figure 5.

A two-step mechanism for N^6-TbAd biosynthesis. (A) The Dimroth rearrangement represents a specific and plausible mechanism that accounts for the rearrangement from cationic 1-linked adenosine to neutral N^6-linked adenosine compounds. The pKa values are derived from analogous compounds and illustrate that only the 1-linked form is predominantly charged. (B) Incubation of recombinant Rv3378c, tuberculosinyl pyrophosphate and adenosine yielded a strong signal for the 1-TbAd isomer (~24 min) but no detectable signal at the retention time for N^6-TbAd (~8 min). (C) 1-TbAd was treated

with dimethylamine in methanol to effect the conversion (Ottria et al., 2010). MS^3 analysis of the reaction products detected the diagnostic fragment ion for N^6-TbAd at *m/z* 148.2. (D) Lung lipids were spiked with 1-TbAD, extracted for lipid and analyzed using normal phase HPLC-Q-TOF-MS under conditions that detect both isomers. (E) These data support a two-step model in which 1-TbAd is the predominant or sole product of Rv3378c followed by conversion to N^6-TbAd, which can occur during export of 1-TbAd to the extrabacterial space.

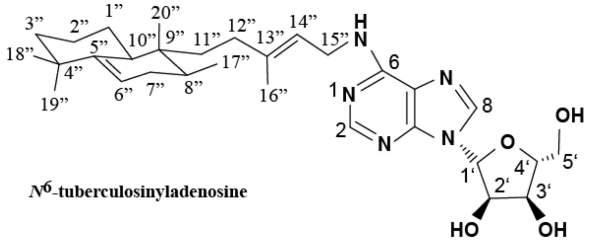
Table 1

NMR analysis. The combination of 1-D and 2-D NMR data lead to the determination of the side group structure as a tuberculosinyl group in N^6 linkage, defining N^6-TbAD. Detailed spectra are shown in Supplemental Data S1. The expected COSY and NOESY cross peaks are seen within the terpene, base and sugar fragments. The only NOESY cross peak between the three fragments is between $H_{1'}$ and H_8. The allylic methylene peak from the two $H_{15''}$ protons is broad rather than the expected doublet due to slow rotation about the C_6-N bond.



N^6-tuberculosinyladenosine

Atom	^{13}Carbon	Hydrogen	COSY	NOESY
2	*	8.24		
8	141.3	8.26		1'
1'	90.8	5.95(d,6.6)	2'	8,2"
2'	75.3	4.74(dd,6.6,5.0)	1',3'	1',3'
3'	72.4	4.32(dd,5.0,2.6)	2',4'	2',5'
4'	88.0	4.17(ddd,2.6,2.4,2.4)	3',5',5'	5',5'
5'	63.3	3.89(dd,12.6,2.4)	4',5'	4'
		3.75(dd,12.6,2.4)	4',5'	3',4'
1α"	28.4	1.77	1",2",2"	1",2",10"
1β"		1.04	1",2",2"	1",2β",20"
2α"	23.0	1.62	1",1",3",3"	1",3",19"
2β"		1.62	1",1",3",3"	1",3"
3α"	41.8	1.41	3",2",2"	2",18",19"
3β"		1.21(ddd,13.1,13.1,4.8)	3",2",2"	2",18"
6"	117.2	5.48	7",7",10"	7",7",18"
7α"	32.5	1.86(br d,17.3)	6",7",8"	6",7",8",17"
7β"		1.77	6",7",8"	6",7",16",17",20"
8"	34.3	1.54	7",7",17"	7",7",10",17"
10"	40.9	2.25(br d,12.9)	1",1",6"	2",7",10",12",19"
11"	35.8	1.58	11",12",12"	8",10",12",16",17",2
		1.42	11",12",12"	8",10",12",16",17",2
12"	33.7	2.00	11",11",12"	10",11",11",14",16"
14"	120.5	5.42	12",15",16"	12",12"
15"	39.3	4.22-4.16	14"	14",16"
16"	16.8	1.80		12",15"
17"	15.2	0.85 (d,6.7)	8"	7",7",8",11",11",20"
18"	30.0	1.06		3",3",6",19"



N^6-tuberculosinyladenosine

Atom	^{13}Carbon	Hydrogen	COSY	NOESY
19"	29.3	1.02		2",3",10",18"
20"	16.4	0.65		1",7",11",17"

* not observed