Elucidating the biosynthetic pathways of volatile organic compounds in *Mycobacterium tuberculosis* through a computational approach

Purva Bhatter[1], Karthik Raman[1]* and Vani Janakiraman[1]*

[1]Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai – 600 036, INDIA

***Corresponding authors. E-mail: vani@iitm.ac.in, kraman@iitm.ac.in, Tel: +91-44-2257-4141**

Key-words: *Mycobacterium tuberculosis*, microbial volatile organic compounds, methyl nicotinate, methyl phenylacetate, methyl *p*-anisate

ABSTRACT

Microbial volatile organic compounds (VOCs) have gained prominence in the recent past for their potential use as disease markers. The discovery of microbial VOCs has benefitted 'difficult to detect' diseases such as tuberculosis (TB). Few of the identified VOCs of *Mycobacterium tuberculosis* (Mtb) are currently being explored for their diagnostic potential. However, very little is known about the biosynthesis of these small lipophilic molecules. Here, we propose putative biosynthetic pathways in *Mycobacterium tuberculosis* to synthesize three VOCs, namely methyl nicotinate, methyl phenylacetate and methyl *p*-anisate, using computational approaches. In particular, we identify S-adenosyl methionine (SAM) transferases that play a crucial role in esterification of the acids to the final product. Our results provide important insights into the specificity of these pathways to Mtb species.

Introduction

Tuberculosis (TB) is caused by an intracellular pathogen, *Mycobacterium tuberculosis* (Mtb). The failure to control the spread of TB can be attributed to several factors. The major known contributors to disease transmission include the lack of an effective vaccine (Andersen and Doherty, 2005), failure of current antibiotic therapies (Dheda *et al.*, 2014), the delayed onset of symptoms and delayed diagnosis. (Sreeramareddy *et al.*, 2014) The current method of diagnosis is microscopic determination of acid fast bacilli in the sputum of patients. Additionally, chest radiographs are also a common practice. The tests are affordable but lack sensitivity and specificity. (van't Hoog, Onozaki and Lonnroth, 2014) In low TB endemic regions, immune-based tests such as Interferon gamma release assay have been found to be specific for assessing exposure to Mtb in both adults and children. (Lighter *et al.*, 2009) However, the tests are less

applicable in high TB endemic regions due to their non-specificity. (Barth, Mudrikova and Hoepelman, 2008) While rapid diagnostic tools such as GenXpert are reliable and sensitive, they are expensive in terms of the infrastructure needed to run the machine and the recurring costs of the consumables. (Parida, 2012) Thus, there is a clear unmet need for a tool that is rapid, sensitive, affordable, high throughput and can detect disease much earlier than the onset of symptoms sparked interest in volatile organic compounds (VOCs).

Most VOCs are lipophilic compounds with low molecular weight and high vapor pressure at ambient temperatures. A number of Mtb infection (or) TB-associated VOCs have been explored for their specificity and sensitivity. (Phillips *et al.*, 2012)(Phillips *et al.*, 2010)(McNerney *et al.*, 2012)(Syhre and Chambers, 2008)(Phillips *et al.*, 2007) However, most of these VOCs are generated by the host as a response to Mtb replication *in vivo. The* use of host-based VOCs as a disease marker always faces the challenge of pre-sensitization in high TB endemic regions, as in the case of immune based assays. Additionally, it is difficult to distinguish the trigger for the response viz. mycobacterial, bacterial, viral or an air pollutant.

The VOCs associated with TB are detected directly from (a) head space of Mtb cultures, (b) via the blood and/or (c) via the exhaled breath. (Sethi, Nanda and Chakraborty, 2013) The principle of the latter is that TB-related changes are reflected as measurable changes in the breath through exchange via lungs. Whilst their presence has been demonstrated both *in vitro* and *in vivo,* their origin in the pathogen remains elusive. We therefore elucidate the putative biosynthetic pathways of Mtb-associated VOCs, as a first step towards understanding the importance of these compounds and their contribution to bacterial physiology. Further, to the best of our knowledge, the homology of proteins in the pathway across species as well as the use of common pathways to generate different compounds has not been studied so far. An understanding of biosynthetic

pathways is important to signify that Mtb VOCs are not just randomly produced compounds but rather ones that may have been shaped and optimized during evolution. The distribution patterns and information arising out of these biosynthetic pathways may also assist in understanding the phylogeny of mycobacterial species in relation to their environments.

In this study, we employ a computational approach to elucidate the biosynthetic pathways of three VOCs that are unique to TB, viz. methyl nicotinate, methyl phenylacetate and methyl *p*-anisate. These VOCs have been discovered earlier. (Syhre and Chambers, 2008)We here systematically elucidate the pathways for biosynthesis of these VOCs using computational analyses, building on publicly available pathway databases and data available from published literature. Further, we analyze the enzymes involved in the pathways for their (a) essentiality for survival of the pathogen *in vivo* (b) potential as drug targets, (c) cellular location viz. cytoplasmic, transmembrane or extracellular, and (d) expression during non-replicating stage of the pathogen.

Methods

Selection of VOCs: A study by Syhre *et al* (2005), identified three VOCs, namely methyl nicotinate, methyl phenylacetate and methyl *p*-anisate as specific to Mtb. (Syhre and Chambers, 2008) The study found that methyl nicotinate was detected in the range of 300–3000 ppt and methyl *p*-anisate at 30–300 ppt in the head space of axenic cultures. (Syhre and Chambers, 2008)(Banday *et al.*, 2011) In addition, the presence of methyl nicotinate (at a concentration of 5 femtomol/mol) has also been demonstrated in the breath of TB patients. (Syhre *et al.*, 2009) However, lower amounts of the other two compounds in the breath renders them unquantifiable. (Suckling and Sagar, 2011) The exclusivity of these compounds to Mtb justifies their preliminary selection. The structure for the three compounds is depicted in Figure 1.

Pathway prediction: The compounds are systematically screened for their presence in the existing databases of KEGG (Kanehisa and Goto, 2000) and MetaCyc. (Caspi *et al.*, 2015) If the presence of the compound and the associated pathway for Mtb were found in the database, the pathway was chosen and each of the enzymes was mapped back to the organism.

When organism and/ or compound specific pathway searches with the databases yielded no hits the search was expanded to published literature for description of the compounds across genus and species. The enzymes catalyzing the formation of the compounds were then assessed for their presence in mycobacterial genus. Upon confirmation of the match, we proposed that mycobacteria uses an approach identical to that observed in the other organism. In the absence of a match from the universal metabolome we used a logical approach for predicting the reactions and looked for the presence of enzymes mediating these reactions in the mycobacterial genus. The specificity of the enzymes for the class of substrates is confirmed through literature. The selected unique pathways and the respective enzymes in the pathway were ascertained through the TubercuList database for their Rv numbers, gene and product names. (Lew *et al.*, 2011)

BLASTP: Close homologues of the Mtb proteins in the non-redundant (NR) database were identified by performing a BLASTP(Altschul *et al.*, 1990) search. We identified homologues of a protein on the basis of sequence identity of 50% or greater, for an alignment length of 50% or more than that of the mycobacterial query protein. We compared the Mtb sequences against sequences for 417 'core' organisms from the STRING database.(Szklarczyk *et al.*, 2011) For each organism, we performed a BLASTP versus all other sequences, to identify the best matching sequences for each protein in an organism, across proteins in all other organisms. Only matches with at least 50% sequence identity across 50% of the alignment length (query cover) were

considered. This cut-off value has been chosen to include all proteins in the database that even slightly match the proteins of the three VOC pathways.

Other databases: The identified proteins are characterized for their properties using available databases. This is to ascertain the essentiality of the pathway in bacterial metabolism and its presence during varied physiological conditions such as hypoxia. The genes are assessed for their essentiality as identified by the Himar-1 based transposon mutagenesis in H37Rv, reconfirmed with *E.coli* PhoA mutants. (Sassetti, Boyd and Rubin, 2003) The potential of these proteins as putative drug targets that are fast acting and less toxic is evaluated using the open source drug discovery (OSDD) database. (Vashisht *et al.*, 2014) The database of Gomez *et al* (2000) is utilized to identify proteins that are secretory in nature. (Gomez, Johnson and Gennaro, 2000) The secretory proteins bearing prominent antigenic determinants of Mtb have been shown to be responsible disease pathogenesis. Some of these proteins mediate cell wall biogenesis whilst others mediate virulence. (Ge *et al.*, 2003) These properties make them suitable biomarkers with diagnostic and prognostic utility. To explore the same, we evaluated the secretory nature of the proteins of the pathways through the use of this literature. The changes in gene expression during latency are determined through the study of Gopinath *et al*, which lends an understanding of gene regulation in Mtb during non-replicating phase (NRP)1 & 2 under hypoxic conditions. (Gopinath *et al.*, 2015) In order to ensure reliability on the results, data mining has been undertaken from more than one database. The selected databases are a combination of evidences from experimental and computational analysis there by increasing the dependability.

Network Analysis: We obtained the protein interactome of Mtb from the STRING database (Version 10). (Szklarczyk *et al.*, 2011) We filtered the interactome to retain only reliable protein interactions and functional associations, with a high-confidence score (STRING score ≥700), for

further analysis. Confidence scores indicate the estimated likelihood that a given interaction is biologically meaningful, specific and reproducible, given the seven channel supporting evidences (Szklarczyk *et al.*, 2016). We analyzed and visualized the network using Cytoscape, version 3.4.2. (Shannon *et al.*, 2003) We analyzed the network characteristics using the NetworkAnalyzer plugin of Cytoscape. (Assenov *et al.*, 2008)

Results

Pathway analysis reveals key enzymes in VOC biosynthesis

The databases of KEGG (Kanehisa and Goto, 2000) and MetaCyc (Caspi *et al.*, 2015) were searched for existing pathways leading to the synthesis of the chosen compounds viz. methyl nicotinate, methyl phenylacetate and methyl *p*-anisate in mycobacterial species. Methyl nicotinate, also known as trigonelline, is a product of the pathway for nicotinate and nicotinamide metabolism. We observed that the synthesis of trigonelline in Mtb proceeds with the conversion of aspartate to nicotinate D-ribonucleotide via quinolinate. Nicotinate D-ribonucleotide has two fates, viz. conversion to nicotinate D-ribonucleoside, nicotinic acid and subsequently nicotinic acid; else conversion to deamino-NAD+ leading to synthesis of nicotinamide. Nicotinamide is then converted to nicotinic acid. However, the absence of a functional enzyme in Mtb mediating the conversion of deamino-NAD+ to nicotinamide has helped us choose the former pathway for trigonelline synthesis. In addition, the KEGG database also shows that tryptophan may contribute to synthesis of methyl nicotinate. However, studies undertaken in other species have proof to the contrary(Joshi and Handler, 1960). When ^{14}C labeled Tryptophan and/ or anthranillic acid-7-^{14}C were used as substrate, no labeled nicotinic acid or trigonelline was synthesized. The final methylation is accomplished by reaction between nicotinic acid and *S*-adenosyl methionine.

The compounds methyl phenylacetate was not present in either KEGG or MetaCyc databases. The compound is known to occur naturally in *Capsicum annum*, *Piper nigrum*, honey and wine. (*methyl phenylacetate*, no date) However, the pathway leading to the formation of the compounds is unknown in all of these cases. For the formation of methyl phenylacetate, we thus posit that the phenylacetic acid may serve as a precursor. The shikimate pathway leads to the formation of

phenylacetate in mycobacterial species. The conversion of phenyl acetate to methyl phenyl acetate via a catechol O-methyl transferase (COMT) is on the basis of the affinity of COMT to aryl acetic acids. (Bourbeau, Allen and Gu, 2010)

Methyl *p*-anisate is known to occur naturally in mushroom, cocoa, and white wine. (*Methyl p-anisate*, no date) An extensive search of literature revealed the formation of Methyl *p*-anisate via the methylation of methyl 4-benzoic acid through a SAM transferase in *Phanerochaete chrysosporium.* (Coulter *et al.*, 1993) Using this as a basis, we identified that O-methyl transferase may be capable of catalyzing the reaction in mycobacterial species. The substrate for methylation viz. 4-hydroxyl benzoate may be formed via the shikimate pathway through conversion of chorismate via the chorismate pyruvate lyase. The catalysis of 4-hydroxy benzoate using a 4-methoxybenzoate monooxygenase is supported by MetaCyc. The enzyme equivalent in Mtb was identified through a BLAST of the 4-methoxybenzoate monooxygenase against the Mtb proteome. The three pathways are depicted in Figure 2. The enzymes catalyzing the reaction and their specificity to the pathway are represented in Table 1.

Interestingly, we also observed that the chosen pathways consist of several proteins essential for the survival of mycobacteria. These proteins are largely intracellular in nature, with very few of them being secretory or membrane associated. The non-secretory nature of these proteins may be one of the reasons limiting their exploration as diagnostic marker. Of all the proteins in the pathways only Rv2537c has been detected in the serum of TB patients using an ELISA (Zhang *et al.*, 2009) (Table 1).

Sequence analyses illustrate exclusivity of VOC biosynthesis enzymes to Mtb

The proteins of the pathway were tested against the non-redundant protein (NR) database using BLASTP(Altschul *et al.*, 1990) with cut off values of 50% query coverage and identity. Our observations indicate the exclusivity of the proteins of the pathway to Mtb. Consequently, to confirm our observation we performed the sequence alignment only for mycobacterial species. Proteomes of 84 mycobacterial strains grouped as saprophytic, non-tuberculous mycobacteria (NTM) and strains of the Mtb complex (MTBC) were assessed for the presence of homologous proteins which form the individual components of the pathway. Each protein from the pathway was queried against functionally identical protein sequences of the 84 strains to ascertain the conservation of the pathway. We found that the pathways were conserved across laboratory adapted strains of Mtb, *viz.* H37Rv, H37Ra, CDC5079, CDC5180, CDC1551, Erdman, and the clinical strains belonging to the Beijing (F11) and EAI-5 lineage. However, they were not conserved across other members of the MTBC such as *M. africanum*, *M. bovis* BCG, BCG Pasteur, BCG Korea, BCG Mexico, BCG Tokyo and *M. leprae*. In addition, the pathways were not conserved across saprophytic strains such as *M. smegmatis* and members of NTM such as *M. fortuitum*, *M.abscessus*, *M. chelonae*, *M. canettii*, and *M. marinum* amongst others.

Network analysis reaffirms the importance of VOC enzymes

We also examined the immediate network neighborhood of the enzymes participating in VOC biosynthesis. We generated this network from the STRING database, as mentioned in Methods. A snapshot of this network is represented in Figure S1. The network consists of 164 nodes, each node representing a protein participating in VOC biosynthesis, or a protein directly interacting with or functionally associated with a VOC biosynthesis enzyme. The list of nodes with values of

degree and rank order is presented in supplementary table 1 (Table S1). From the network analysis, we identified Rv0948c, Rv3838c, Rv3227, Rv3772, Rv2552c, Rv2538c, Rv2540c, Rv2539c, Rv0573c and Rv1595 as proteins of the network that are found to be highly connected to the other proteins within this network. Examining the complete (high-confidence) interactome of Mtb indicates that these proteins belong to the top 25 percentile of the highly connected proteins in the entire network. Amongst the proteins of the pathway with a lower number of links, a substantial number was found to be non-essential for survival. Previous studies have demonstrated that highly connected proteins are three times more likely to be essential than proteins with only a small number of links to other proteins. (Jeong *et al.*, 2001) Amongst these highly connected 10 proteins, 7 are essential for bacterial survival *in vitro* (Sassetti, Boyd and Rubin, 2003) and have consequently have been identified as important drug targets (Vashisht *et al.*, 2014) (Table 2 & Supplementary Figure S1).

Discussion

VOCs are low molecular weight compounds with a distinct odor profile. VOCs generated from compounds of bacterial origin are termed as microbial VOCs. Other than microbes, plant, fungi (molds) and animals are other known biological sources of VOCs. Plant-based VOCs are used as a means of intra-(Ueda, Kikuta and Matsuda, 2014) and inter-species communication. (Holopainen and Blande, 2012)[2][2][2][2][2] Similarly, microbial VOCs (mVOCs) appear to be involved in antagonism, mutualism, intra- and interspecies regulation of cellular and developmental processes, and modification of their surrounding environments. (Bitas *et al.*, 2013) However, knowledge about impact of mVOCs on other organisms of the biosphere is limited. The exploration of mVOCs as a diagnostic marker has received attention but a lot remains to be done for understanding the molecular basis of VOC induction and their physiological roles. This

knowledge will facilitate a basic understanding of the pathogenic consequences of VOC as a secondary metabolite.

We have undertaken a computational approach to elucidate the probable pathways of synthesis of VOCs specific to Mtb. This report is the first attempt ever, to elucidate the biosynthetic pathways of VOCs of mycobacterial origin. Our analyses demonstrate that the three VOCs, methyl nicotinate, methyl phenylacetate, methyl *p*-anisate are derivatives of amino acids. The presence of methyl nicotinate has also been confirmed in the breath of TB patients. (Syhre *et al.*, 2009) It is known that Mtb utilizes aspartate for its growth implying the existence of a pathway to breakdown aspartate and formation of a by-product such as methyl nicotinate. It has been demonstrated that Mtb uses aspartate largely for nitrogen metabolism and very little of it enters the Krebs cycle for carbon metabolism. The presence of aspartate in granulomas, which is known to be central to mycobacterial activity, signifies bacterial virulence through utilization of aspartate for growth and survival within the host. (Gouzy, Poquet and Neyrolles, 2013)[3][3][3][3][3]

While the diagnostic potential of these VOCs has been well-explored, their physiological roles have remained elusive. It is evident that these are secondary metabolites that stem out of amino acid metabolic pathways but their role in bacterial survival is less understood. Trigonelline is a pyridine piperidine alkaloid and alkaloids in general are known to possess antibacterial activity. (Özçelik, Kartal and Orhan, 2011) Methyl *p*-anisate, on the other hand, is a polyphenol. The role of polyphenols in reactive oxygen species (ROS) scavenging activity has already been reported. (Anantachoke *et al.*, 2015)(Kurek-Gorecka *et al.*, 2013)

The biosynthetic pathways for methyl nicotinate, methyl phenylacetate and methyl *p*-anisate are cytoplasmic, are largely comprised of essential proteins for survival *in vivo*, and are conserved

across strains of Mtb. Coupling of VOC based detection with species-specific markers could serve as a more stringent TB detection tool. The unaltered expression of these proteins through the non-replicating phase (Gopinath *et al.*, 2015) may be an evidence for the synthesis of these compounds during latency. It also suggests the use of mVOCs as diagnostic marker during different stages of the disease. However, the non-secretory or cytoplasmic nature of the proteins of these pathways makes them less desirable as diagnostic markers and immune targets.

In sum, the key contribution of our study is the revelation of putative pathways for the biosynthesis of three Mtb VOCs through a combination of metabolic pathway analysis and sequence homology studies. Our study sets the platform for experimental interrogation of these pathways to better understand the biosynthesis of VOCs as well as their utility for diagnosis.

Statement of conflict of interest

The authors declare no conflict of interest

Acknowledgements

The authors thank Karthik Azhagesan for help with sequence analysis. PB acknowledges the institute post-doctoral fellowship from IIT, Madras. VJ acknowledges funding from IIT Madras vide grants BIO/14-15/645/NFSG and BIO/14-15/810/NFIG.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) 'Basic local alignment search tool', *J Mol Biol*. 1990/10/05, 215(3), pp. 403–410. doi: 10.1016/S0022-2836(05)80360-2S0022-2836(05)80360-2 [pii].

Anantachoke, N., Kitphati, W., Mangmool, S. and Bunyapraphatsara, N. (2015) 'Polyphenolic compounds and antioxidant activities of the leaves of Glochidion hypoleucum', *Nat Prod Commun*. 2015/05/01, 10(3), pp. 479–482. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25924533.

Andersen, P. and Doherty, T. M. (2005) 'The success and failure of BCG - implications for a novel tuberculosis vaccine', *Nat Rev Microbiol*. 2005/07/14, 3(8), pp. 656–662. doi: nrmicro1211 [pii]10.1038/nrmicro1211.

Assenov, Y., Ramírez, F., Schelhorn, S.-E., Lengauer, T. and Albrecht, M. (2008) 'Computing topological parameters of biological networks.', *Bioinformatics (Oxford, England)*, 24(2), pp. 282–4. doi: 10.1093/bioinformatics/btm554.

Banday, K. M., Pasikanti, K. K., Chan, E. C., Singla, R., Rao, K. V, Chauhan, V. S. and Nanda, R. K. (2011) 'Use of urine volatile organic compounds to discriminate tuberculosis patients from

healthy subjects', *Anal Chem*. 2011/05/31, 83(14), pp. 5526–5534. doi: 10.1021/ac200265g.

Barth, R. E., Mudrikova, T. and Hoepelman, A. I. M. (2008) 'Interferon-gamma release assays (IGRAs) in high-endemic settings: could they play a role in optimizing global TB diagnostics?', *International Journal of Infectious Diseases*, 12(6), pp. e1–e6. doi: 10.1016/j.ijid.2008.03.026.

Bitas, V., Kim, H.-S., Bennett, J. W. and Kang, S. (2013) 'Sniffing on Microbes: Diverse Roles of Microbial Volatile Organic Compounds in Plant Health', *Molecular Plant-Microbe Interactions*, 26(8), pp. 835–843. doi: 10.1094/mpmi-10-12-0249-cr.

Bourbeau, M. P., Allen, J. G. and Gu, W. (2010) 'Recent Advances in Acetyl-CoA Carboxylase Inhibitors', 45, pp. 95–108. doi: 10.1016/s0065-7743(10)45006-x.

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D. S. and Karp, P. D. (2015) 'The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases', *Nucleic Acids Res*. 2015/11/04. doi: gkv1164 [pii]10.1093/nar/gkv1164.

Coulter, C., Kennedy, J. T., McRoberts, W. C. and Harper, D. B. (1993) 'Purification and Properties of an S-Adenosylmethionine: 2,4-Disubstituted Phenol O-Methyltransferase from Phanerochaete chrysosporium', *Appl Environ Microbiol*. 1993/03/01, 59(3), pp. 706–711. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16348886.

Dheda, K., Gumbo, T., Gandhi, N. R., Murray, M., Theron, G., Udwadia, Z., Migliori, G. B. and Warren, R. (2014) 'Global control of tuberculosis: from extensively drug-resistant to untreatable tuberculosis', *Lancet Respir Med*. 2014/04/11, 2(4), pp. 321–338. doi: 10.1016/S2213-

2600(14)70031-1S2213-2600(14)70031-1 [pii].

Ge, Y., ElNaggar, M., Sze, S. K., Oh, H. Bin, Begley, T. P., McLafferty, F. W., Boshoff, H. and Barry, C. E. (2003) 'Top down characterization of secreted proteins from Mycobacterium tuberculosis by electron capture dissociation mass spectrometry', *Journal of the American Society for Mass Spectrometry*, 14(3), pp. 253–261. doi: 10.1016/s1044-0305(02)00913-3.

Gomez, M., Johnson, S. and Gennaro, M. L. (2000) 'Identification of secreted proteins of Mycobacterium tuberculosis by a bioinformatic approach', *Infect Immun*. 2000/03/18, 68(4), pp. 2323–2327. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10722636.

Gopinath, V., Raghunandanan, S., Gomez, R. L., Jose, L., Surendran, A., Ramachandran, R., Pushparajan, A. R., Mundayoor, S., Jaleel, A. and Kumar, R. A. (2015) 'Profiling the Proteome ofMycobacterium tuberculosisduring Dormancy and Reactivation', *Molecular & Cellular Proteomics*, 14(8), pp. 2160–2176. doi: 10.1074/mcp.M115.051151.

Gouzy, A., Poquet, Y. and Neyrolles, O. (2013) 'A central role for aspartate in Mycobacterium tuberculosis physiology and virulence', *Frontiers in Cellular and Infection Microbiology*, 3. doi: 10.3389/fcimb.2013.00068.

Holopainen, J. K. and Blande, J. D. (2012) 'Molecular Plant Volatile Communication', 739, pp. 17–31. doi: 10.1007/978-1-4614-1704-0_2.

Jeong, H., Mason, S., Barabasi, A. and Oltvai Zn (2001) 'Lethality and Centrality in Protein networks', *Nature*, 411, p. 41.

Joshi, J. G. and Handler, P. (1960) 'Biosynthesis of trigonelline', *J Biol Chem*. 1960/10/01, 235,

pp. 2981–2983. Available at: http://www.ncbi.nlm.nih.gov/pubmed/13790768.

Kanehisa, M. and Goto, S. (2000) 'KEGG: kyoto encyclopedia of genes and genomes', *Nucleic Acids Res*. 1999/12/11, 28(1), pp. 27–30. doi: gkd027 [pii].

Kurek-Gorecka, A., Rzepecka-Stojko, A., Gorecki, M., Stojko, J., Sosada, M. and Swierczek-Zieba, G. (2013) 'Structure and antioxidant activity of polyphenols derived from propolis', *Molecules*. 2013/12/24, 19(1), pp. 78–101. doi: 10.3390/molecules19010078molecules19010078 [pii].

Lew, J. M., Kapopoulou, A., Jones, L. M. and Cole, S. T. (2011) 'TubercuList--10 years after', *Tuberculosis (Edinb)*. 2010/10/29, 91(1), pp. 1–7. doi: 10.1016/j.tube.2010.09.008S1472-9792(10)00111-3 [pii].

Lighter, J., Rigaud, M., Eduardo, R., Peng, C. H. and Pollack, H. (2009) 'Latent Tuberculosis Diagnosis in Children by Using the QuantiFERON-TB Gold In-Tube Test', *Pediatrics*, 123(1), pp. 30–37. doi: 10.1542/peds.2007-3618.

McNerney, R., Mallard, K., Okolo, P. I. and Turner, C. (2012) 'Production of volatile organic compounds by mycobacteria', *FEMS Microbiology Letters*, 328(2), pp. 150–156. doi: 10.1111/j.1574-6968.2011.02493.x.

Methyl p-anisate (no date) *SIGMA-ALDRICH*. doi: W267902 ALDRICH.

methyl phenylacetate (no date) *The Good scents company*. Available at: http://www.thegoodscentscompany.com/data/rw1008431.html (Accessed: 20 May 2016).

Özçelik, B., Kartal, M. and Orhan, I. (2011) 'Cytotoxicity, antiviral and antimicrobial activities of

alkaloids, flavonoids, and phenolic acids', *Pharmaceutical Biology*, 49(4), pp. 396–402. doi: 10.3109/13880209.2010.519390.

Parida, S. K. (2012) 'True facets of TB diagnosis in 2012: Hypes and realities', *European Journal of Microbiology and Immunology*, 2(4), pp. 275–281. doi: 10.1556/EuJMI.2.2012.4.5.

Phillips, M., Basa-Dalay, V., Blais, J., Bothamley, G., Chaturvedi, A., Modi, K. D., Pandya, M., Natividad, M. P., Patel, U., Ramraje, N. N., Schmitt, P. and Udwadia, Z. F. (2012) 'Point-of-care breath test for biomarkers of active pulmonary tuberculosis', *Tuberculosis (Edinb)*. 2012/06/01, 92(4), pp. 314–320. doi: 10.1016/j.tube.2012.04.002S1472-9792(12)00079-0 [pii].

Phillips, M., Basa-Dalay, V., Bothamley, G., Cataneo, R. N., Lam, P. K., Natividad, M. P., Schmitt, P. and Wai, J. (2010) 'Breath biomarkers of active pulmonary tuberculosis', *Tuberculosis (Edinb)*. 2010/03/02, 90(2), pp. 145–151. doi: 10.1016/j.tube.2010.01.003S1472-9792(10)00015-6 [pii].

Phillips, M., Cataneo, R. N., Condos, R., Ring Erickson, G. A., Greenberg, J., La Bombardi, V., Munawar, M. I. and Tietje, O. (2007) 'Volatile biomarkers of pulmonary tuberculosis in the breath', *Tuberculosis*, 87(1), pp. 44–52. doi: 10.1016/j.tube.2006.03.004.

Sassetti, C. M., Boyd, D. H. and Rubin, E. J. (2003) 'Genes required for mycobacterial growth defined by high density mutagenesis', *Mol Microbiol*. 2003/03/27, 48(1), pp. 77–84. doi: 3425 [pii].

Sethi, S., Nanda, R. and Chakraborty, T. (2013) 'Clinical Application of Volatile Organic Compound Analysis for Detecting Infectious Diseases', *Clinical Microbiology Reviews*, 26(3), pp. 462–475. doi: 10.1128/cmr.00020-13.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) 'Cytoscape: a software environment for integrated models of biomolecular interaction networks', *Genome Res*. 2003/11/05, 13(11), pp. 2498–2504. doi: 10.1101/gr.123930313/11/2498 [pii].

Sreeramareddy, C. T., Qin, Z. Z., Satyanarayana, S., Subbaraman, R. and Pai, M. (2014) 'Delays in diagnosis and treatment of pulmonary tuberculosis in India: a systematic review', *The International Journal of Tuberculosis and Lung Disease*, 18(3), pp. 255–266. doi: 10.5588/ijtld.13.0585.

Suckling, D. M. and Sagar, R. L. (2011) 'Honeybees Apis mellifera can detect the scent of Mycobacterium tuberculosis', *Tuberculosis*, 91(4), pp. 327–328. doi: 10.1016/j.tube.2011.04.008.

Syhre, M. and Chambers, S. T. (2008) 'The scent of Mycobacterium tuberculosis', *Tuberculosis (Edinb)*. 2008/02/26, 88(4), pp. 317–323. doi: 10.1016/j.tube.2008.01.002S1472-9792(08)00004-8 [pii].

Syhre, M., Manning, L., Phuanukoonnon, S., Harino, P. and Chambers, S. T. (2009) 'The scent of Mycobacterium tuberculosis – Part II breath', *Tuberculosis*, 89(4), pp. 263–266. doi: 10.1016/j.tube.2009.04.003.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., Jensen, L. J. and von Mering, C. (2011) 'The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored', *Nucleic Acids Res*. 2010/11/04, 39(Database issue), pp. D561-8. doi: 10.1093/nar/gkq973gkq973 [pii].

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A.,

Doncheva, N. T., Roth, A., Bork, P., Jensen, L. J. and von Mering, C. (2016) 'The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible.', *Nucleic acids research*. Oxford University Press, p. gkw937. doi: 10.1093/nar/gkw937.

Ueda, H., Kikuta, Y. and Matsuda, K. (2014) 'Plant communication', *Plant Signaling & Behavior*, 7(2), pp. 222–226. doi: 10.4161/psb.18765.

van't Hoog, A. H., Onozaki, I. and Lonnroth, K. (2014) 'Choosing algorithms for TB screening: a modelling study to compare yield, predictive value and diagnostic burden', *BMC Infectious Diseases*, 14(1), p. 532. doi: 10.1186/1471-2334-14-532.

Vashisht, R., Bhat, A. G., Kushwaha, S., Bhardwaj, A. and Brahmachari, S. K. (2014) 'Systems level mapping of metabolic complexity in Mycobacterium tuberculosis to identify high-value drug targets', *J Transl Med*. 2014/10/12, 12, p. 263. doi: 10.1186/s12967-014-0263-5s12967-014-0263-5 [pii].

Zhang, G., Zhang, L., Zhang, M., Pan, L., Wang, F., Huang, J., Li, G., Yu, J. and Hu, S. (2009) 'Screening and Assessing 11 Mycobacterium tuberculosis Proteins as Potential Serodiagnostical Markers for Discriminating TB Patients from BCG Vaccinees', *Genomics, Proteomics & Bioinformatics*, 7(3), pp. 107–115. doi: 10.1016/S1672-0229(08)60039-X.

Figure Legends

Figure 1: 2-D molecular structures of the three volatile organic compounds (VOCs) under study.

[1a] National Center for Biotechnology Information. PubChem Compound Database; CID=7151, https://pubchem.ncbi.nlm.nih.gov/compound/7151 (accessed Sept. 27, 2016).

[1b] National Center for Biotechnology Information. PubChem Compound Database; CID=7559, https://pubchem.ncbi.nlm.nih.gov/compound/7559 (accessed Sept. 27, 2016).

[1c] National Center for Biotechnology Information. PubChem Compound Database; CID=8499, https://pubchem.ncbi.nlm.nih.gov/compound/8499 (accessed Sept. 27, 2016).

Figure 2: Pathways leading to the synthesis of the compounds (a) Methyl nicotinate (b) Methyl phenylacetate (c) Methyl *p*-anisate. The enzyme mediating the reaction is indicated by the Rv number as determined from the TubercuList database. The reaction ID as determined from KEGG is represented on the left hand side of the arrow.

Figure S1: Immediate neighbourhood of the enzymes in the pathways involved in the biosynthesis of the chosen VOCs**.** Nodes are sized based on their degree in this network; the network has one large connected component with the majority of important nodes, and a few other connected components that are small. The figure was generated using Cytoscape.

Table 1: List of proteins involved in the pathways for synthesis of methyl nicotinate, methyl phenyl acetate and methyl *p*-anisate.

Rv#	**Gene name**	**Product name**	**Methyl nicotinate**	**Methyl phenylacetate**	**Methyl *p*-anisate**
Rv1595	nadB	L-aspartate oxidase	+	–	–
Rv1594	nadA	Quinolinate synthetase	+	–	–
Rv1596	nadC	nicotinate-nucleotide pyrophosphatase	+	–	–
Rv0573c		nicotinate phosphoribosyltransferase	+	–	–
Rv1703c		catechol O-methyl transferase	+	+	–
Rv2178c	aroG	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase	–	+	+
Rv2538c	aroB	3-dehydroquinate synthase	–	+	+
Rv2537c	aroD	3-dehydroquinate dehydratase	–	+	+
Rv2552c	aroE	shikimate 5-dehydrogenase	–	+	+
Rv2539c	aroK	Shikimate kinase	–	+	+

Rv3227	aroA	3-phosphoshikimate 1-carboxyvinyltransferase	–	+	+
Rv2540c	aroF	Chorismate synthase	–	+	+
Rv0948c		Chorismate mutase	–	+	–
Rv3838c	pheA	Prephenate dehydratase	–	+	–
Rv3772		histidinol-phosphate aminotransferase	–	+	–
Rv3170	aofH	monoamine oxidase	–	+	–
Rv1263		Amidase	–	+	–
Rv1703c		catechol O-methyl transferase	–	+	–
Rv2949c		Chorismate pyruvate lyase	–	–	+
Rv0766c		Cytochrome p450	–	–	+
Rv1153c		O-methyl transferase	–	–	+

-: Indicates absence of gene in the respective pathway

+: Indicates presence of gene in the respective pathway

Cells highlighted in gray indicate new proteins identified in the pathways.

Table 2: Analysis of the genes/proteins as essential for survival *in vivo*, non-toxic drug target, secretory protein, biomarker and regulation during latency

Rv#	**Secretory proteins[2]**	**Upregulated[3]**	**Essentiality prediction[4]**	**Diagnostic[5]**
Rv1595	-	-	+	-
Rv1594	-	-	-	-
Rv1596	-	-	-	-
Rv0573c	-	-	-	-
Rv1703c	-	-	-	-
Rv2178c[1]	-	-	+	-
Rv2538c[1]	-	-	+	-
Rv2537c[1]	-	-	+	+
Rv2552c[1]	-	-	+	-
Rv2539c[1]	-	-	+	-
Rv3227[1]	-	-	+	-
Rv2540c[1]	-	NRP1	+	-
Rv0948c	-	-	-	-
Rv3838c[1]	-	-	+	-
Rv3772	-	-	-	-
Rv3170	+	-	-	-
Rv1263	-	-	-	-

Rv#	Secretory proteins[2]	Upregulated[3]	Essentiality prediction[4]	Diagnostic[5]
Rv1703c	-	-	-	-
Rv2949c	-	-	-	-
Rv0766c	-	-	-	-
Rv1153c	-	-	-	-

- indicates the absence of the respective trait

+Identifies the association of the gene for the respective trait

NRP1: non-replicating phase-1 attained at 192h. None of the proteins of the pathways were found to be down regulated during latency

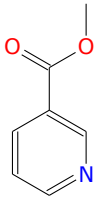
[1]genes/proteins identified as non-toxic drug targets through a systems biology approach

[2]secretory proteins through a computational and experimental approach

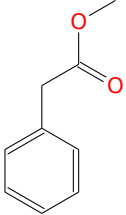
[3]genes up-regulated or down regulated during latency

[4]genes essential for *Mycobacterium tuberculosis* survival *in vivo*

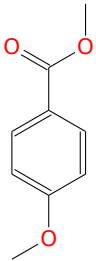
[5]genes/proteins for their role as a biomarker through available literature



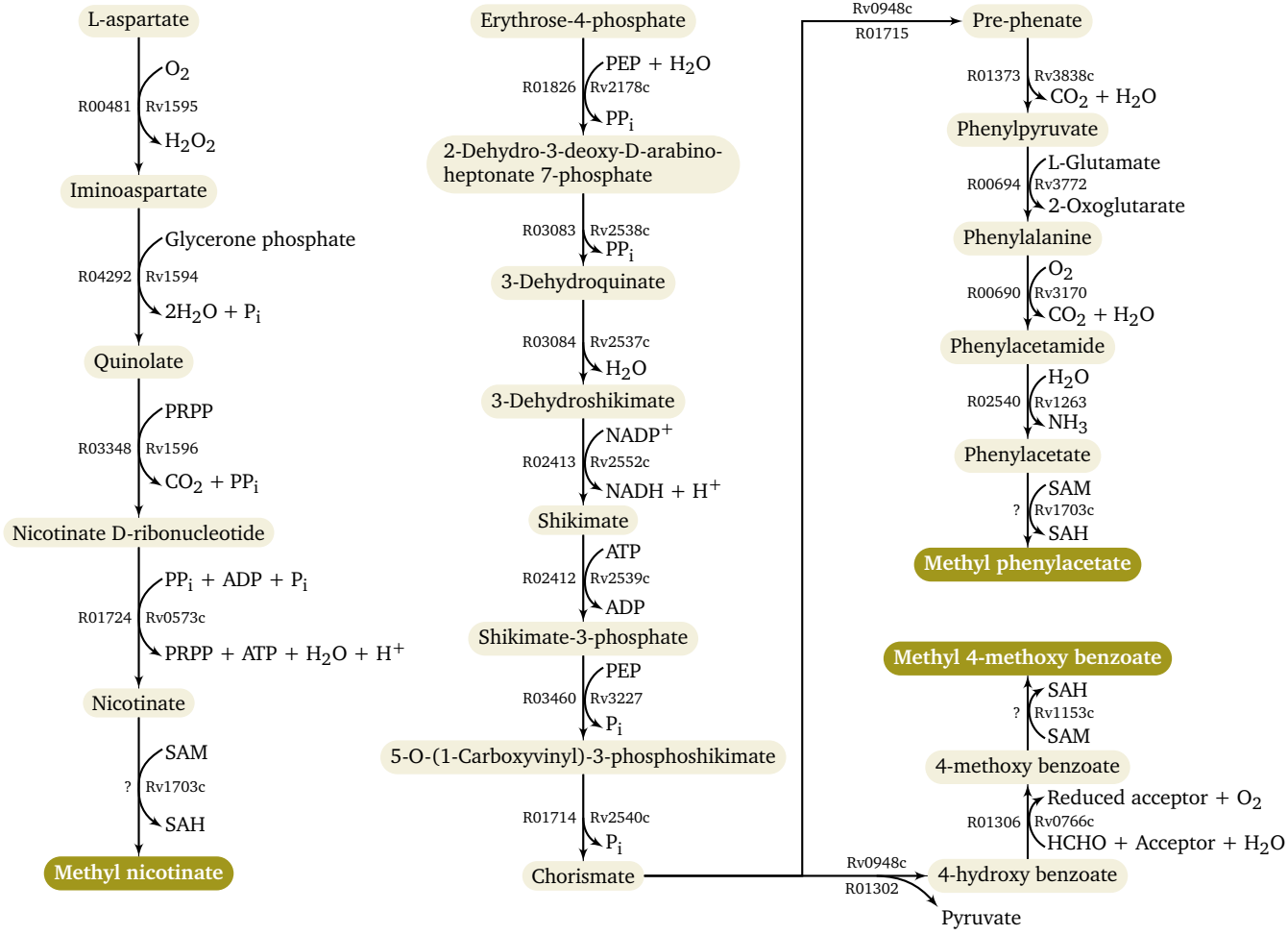
(a) Methyl nicotinate



(b) Methyl phenylacetate



(c) Methyl *p*-anisate


L-aspartate
R00481
Rv1595
O_2
H_2O_2
Iminoaspartate
R04292
Rv1594
Glycerone phosphate
$2H_2O + P_i$
Quinolate
R03348
Rv1596
PRPP
$CO_2 + PP_i$
Nicotinate D-ribonucleotide
R01724
Rv0573c
$PP_i + ADP + P_i$
$PRPP + ATP + H_2O + H^+$
Nicotinate
?
Rv1703c
SAM
SAH
Methyl nicotinate
Erythrose-4-phosphate
R01826
Rv2178c
PEP + H_2O
PP_i
2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate
R03083
Rv2538c
PP_i
3-Dehydroquinate
R03084
Rv2537c
H_2O
3-Dehydroshikimate
R02413
Rv2552c
$NADP^+$
$NADH + H^+$
Shikimate
R02412
Rv2539c
ATP
ADP
Shikimate-3-phosphate
R03460
Rv3227
PEP
P_i
5-O-(1-Carboxyvinyl)-3-phosphoshikimate
R01714
Rv2540c
P_i
Chorismate
Rv0948c
R01715
Pre-phenate
R01373
Rv3838c
$CO_2 + H_2O$
Phenylpyruvate
R00694
Rv3772
L-Glutamate
2-Oxoglutarate
Phenylalanine
R00690
Rv3170
O_2
$CO_2 + H_2O$
Phenylacetamide
R02540
Rv1263
H_2O
NH_3
Phenylacetate
?
Rv1703c
SAM
SAH
Methyl phenylacetate
Rv0948c
R01302
Pyruvate
4-hydroxy benzoate
R01306
Rv0766c
Reduced acceptor + O_2
HCHO + Acceptor + H_2O
4-methoxy benzoate
?
Rv1153c
SAH
SAM
Methyl 4-methoxy benzoate