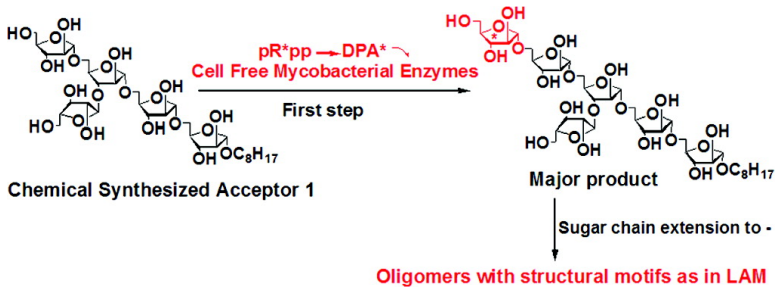
Article

Characterization of a Distinct Arabinofuranosyltransferase in *Mycobacterium smegmatis*

Jian Zhang, Kay-Hooi Khoo, Sz-Wei Wu, and Delphi Chatterjee

J. Am. Chem. Soc., **2007**, 129 (31), 9650-9662• DOI: 10.1021/ja070330k • Publication Date (Web): 14 July 2007



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 3 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 07/14/2007

Characterization of a Distinct Arabinofuranosyltransferase in *Mycobacterium smegmatis*

Jian Zhang,[†] Kay-Hooi Khoo,[‡] Sz-Wei Wu,[‡] and Delphi Chatterjee*,[†]

Contribution from the Mycobacteria Research Laboratories, Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado 80523, and Institute of Biological Chemistry, Academia Sinica, Nankang, Taipei 115, Taiwan

Received January 16, 2007; E-mail: delphi@lamar.colostate.edu

Abstract: The D-arabinans in *Mycobacterium* are essential, extraordinarily complex entity comprised of D-arabinofuranose residues which are rarely found in nature. Despite the well-recognized importance of the mycobacterial arabinan, delineation of the arabinosylation process has been severely hampered due to lack of positively identified arabinosyltransferases. Identification of genes involved in arabinan biosynthesis entailed the use of ethambutol (EMB), a first-line antituberculosis agent that is known to inhibit cell wall arabinan synthesis. The three genes (*embA*, *embB*, and *embC*) encode novel membrane proteins, implicated as the only known mycobacterial arabinosyltransferases to this date. We have now adapted a multifaceted approach involving development of convenient arabinosyltransferase assay using novel synthetic acceptors to identify arabinosyltransferase/s that will be distinct from the Emb proteins. In our present work, *Mycobacterium smegmatis* mc^2155 (WTMsm) was used as a model to study the biosynthesis of cell wall arabinan. In an in vitro assay, we demonstrate that transfer of only α-Ara*f* had occurred from decaprenylphosphoryl-D-arabinofuranose (DPA) on a newly synthesized branched acceptor [α-D-Ara*f*]$_2$-3,5-α-D-Ara*f*-(1→5)-α-D-Ara*f*-(1→5)-α-D-Ara*f* (**1**) with an octyl aglycon. Higher molecular weight (up to Ara_{10}) oligomers were also detected in a parallel reaction using cold phosphoribosepyrophosphate (pRpp). Matrix-assisted laser desorption ionization time-of-flight tandem mass spectrometry (MALDI-TOF MS/MS) analysis of these products revealed that isomeric products were formed and initiation and elongation of arabinan can occur either on the 5-arm or 3-arm of the branched 3,5-α-D-Ara*f*. Individual *embA*, *embB*, and *embC* knockout strains retained this α-1,5 arabinosyltransferase activity, and the activity was partially inhibited by ethambutol. This particular enzyme function is distinct from the function of the Emb proteins.

Introduction

One of the most prominent components of the mycobacterial cell wall is D-arabinan, a constituent present in two distinct settings, arabinogalactan (AG) and lipoarabinomannans (LAM). In the chemical setting of mycolylarabinogalactan−peptidogalactan complex (mAGP), AG forms an integral part of the cell wall proper, whereas LAM is distributed free on the cell surface as well as anchored in the plasma membrane. The latter is implicated as an important virulence factor for the bacterium. Structurally, the D-arabinofurans are unique as they consist of biologically rare sugar, D-arabinofuranose (Ara*f*) and are almost unknown outside the member of the *Actinomycetales*. There may be as many as 60−70 Ara*f* residues in AG and LAM. The arabinan composite has no repeating units, although the nonreducing end of the arabinan in both AG and LAM has been well characterized.[1] Functionally, the arabinan in AG tethers the unusual branched mycolic acids (80 carbon long chain lipids) to peptidoglycan. The mycolic acids are attached to the 5-position of the nonreducing terminal *β*-D-Ara*f* and the penultimate 2-α-D-Ara*f*.[2−4] Seminal studies on AG and LAM were conducted by us and others in the late 1980s and 1990s.[2,5−19] These studies are now leading to biosynthetic pathway determination. The fundamentals of how this polysac-

[†] Colorado State University.
[‡] Institute of Biological Chemistry, Academia Sinica.

(1) McNeil, M. R.; Robuck, K. G.; Harter, M.; Brennan, P. J. *Glycobiology* **1994**, *4*, 165−173.
(2) Daffe, M.; McNeil, M.; Brennan, P. J. *Carbohydr. Res.* **1993**, *249*, 383−398.
(3) McNeil, M. R. In *Genetics of Bacterial Polysaccharides*; Goldberg, J. B., Ed.; CRC Press: Boca Raton, FL, 1999; pp 207−223.
(4) McNeil, M. R.; Brennan, P. J. *Res. Microbiol.* **1991**, *142*, 451−463.
(5) Besra, G. S.; Khoo, K.-H.; McNeil, M. R.; Dell, A.; Morris, H. R.; Brennan, P. J. *Biochemistry* **1995**, *34*, 4257−4266.
(6) Chatterjee, D.; Bozic, C. M.; McNeil, M.; Brennan, P. J. *J. Biol. Chem.* **1991**, *266*, 9652−9660.
(7) Chatterjee, D.; Hunter, S. W.; McNeil, M.; Brennan, P. J. *J. Biol. Chem.* **1992**, *267*, 6228−6233.
(8) Chatterjee, D.; Khoo, K. H. *Glycobiology* **1998**, *8*, 113−120.
(9) Chatterjee, D.; Lowell, K.; Rivoire, B.; McNeil, M.; Brennan, P. J. *J. Biol. Chem.* **1992**, *267*, 6234−6239.
(10) Chatterjee, D.; Roberts, A. D.; Lowell, K.; Brennan, P. J.; Orme, I. M. *Infect. Immun.* **1992**, *60*, 1249−1253.
(11) Daffe, M.; Brennan, P. J.; McNeil, M. *J. Biol. Chem.* **1990**, *265*, 6734−6743.
(12) Hunter, S. W.; Gaylord, H.; Brennan, P. J. *J. Biol. Chem.* **1986**, *261*, 12345−12351.
(13) McNeil, M.; Daffe, M.; Brennan, P. J. *J. Biol. Chem.* **1990**, *265*, 18200−18206.
(14) McNeil, M.; Daffe, M.; Brennan, P. J. *J. Biol. Chem.* **1991**, *266*, 13217−13223.
(15) McNeil, M.; Wallner, S. J.; Hunter, S. W.; Brennan, P. J. *Carbohydr. Res.* **1987**, *166*, 299−308.
(16) Khoo, K. H.; Douglas, E.; Azadi, P.; Inamine, J. M.; Besra, G. S.; Mikusová, K.; Brennan, P. J.; Chatterjee, D. *J. Biol. Chem.* **1996**, *271*, 28682−28690.

 10.1021/ja070330k CCC: $37.00

charide is assembled—if on a lipid carrier, growing from the reducing or nonreducing end, built en bloc, assembled on an enzyme complex resembling the polyketides, adding one residue at a time—is not known. A biosynthetic intermediate has never been isolated nor any functionally active arabinosyltransferases identified.

The Ara*f* residues of the D-arabinan originate from the pentose phosphate pathway/hexose monophosphate shunt, and the immediate precursor (donor of the polymerized arabinan) is decaprenylphosphoryl-D-arabinofuranose (DPA). The biosynthetic pathway for the DPA formation has been recently elucidated.[20] Phosphoribosepyrophosphate (pRpp) is converted to 5-phosphodecaprenylribose (DPPR) which is then dephosphorylated to form decaprenylphosphoryl-D-ribose (DPR); this in turn is epimerized to form DPA.[20−22] Identification of the *emb* operon[23,24] has provided some understanding toward the assembly of the arabinan; however, currently there is no experimental data on the role or function of these predicted arabinosyltransferases in *M. tuberculosis*, but work has been carried out in the related organism *M. smegmatis*. Gene disruptions for individual *emb* genes followed by biochemical characterization of the constituents in the mutant strains led to the striking conclusion that synthesis of AG−arabinan was affected by *embA* and *embB* genes disruption[25] and that synthesis of LAM ceased following *embC* inactivation.[26] All three mutants were viable, and the crucial terminal Ara_6 motif which is the template for mycolylation in AG[14] was altered in both *embA*- and *embB*-deleted strains. In brief, knockouts of *embB* and *embA* produced bacteria with defective AG, in which the number of terminal β-D-Ara*f*(1→2)-α-D-Ara*f* reduced drastically. This suggested the EmbA and EmbB are involved in the formation of the terminal Ara_6 motif in AG. In contrast, in the *embC* knockout mutant, AG was normal but no LAM was synthesized; thus, the effect of this gene deletion was different and more drastic. Thus, two similar polysaccharides utilized two sets of *emb* genes belonging to the same operon.

What became clear from these experimental evidences was that the Emb proteins were required for the arabinan synthesis and worked not only on the nonreducing end of the arabinan but were also involved in the assembly of the internal arabinan as disruption of EmbC led to depletion of arabinan in LAM. Only recently, two other enzymes have been identified as an arabinofuranosyltransferases. One of these enzymes, AftA, is responsible for recognition of the galactan and donating the initial key α-Ara*f* residues for further elaboration by the Emb protein(s).[27] AftB, on the other hand is involved in the terminal β-capping of arabinan in *Corynebacterianeae*.[28]

However, further research relying on the use of active enzymes from bacterial preparations was severely restricted. Therefore, chemical synthesis of well-defined complex D-Ara*f*-based glycoconjugates became essential for significant progress in this field.[29−31] Many acceptors have been synthesized and used in mycobacterial cell-free assay, notably use of α-D-Ara*f*(1→5)-α-D-Ara*f* that converted into organic solvent soluble trisaccharides with membrane preparations from wild-type *M. smegmatis*[29,32] and Araβ1→2Araα1→5Araα1→5Araα1→5Araα1→ that converted into mature nonreducing terminus Ara_6.[31] It was shown that the latter conversion was dependent on the EmbA and EmbB proteins.[31] The current work describes the synthesis and use of a branched internal acceptor (**1**) and examines its effectiveness as an acceptor for novel arabonisyltransferase independent of the Emb proteins.

Experimental Section

General Methods. Radioactive p[^{14}C]Rpp (300 mCi/mmol) was generated from uniformly labeled D-[^{14}C]glucose (American Radiolabeled Chemicals Inc.) as described.[21] Nonradioactive pRpp, ATP, MOPS, and 2-mercaptoethanol were purchased from Sigma. Percoll (sterile) and Biogel P-2 were obtained from Pharmacia. Protein concentrations were determined using the BCA assay (Pierce). All chemicals were purchased from Sigma/Aldrich and were used without further purification except pRpp. Thin-layer chromatography (TLC) was performed on silica gel G60 (EM Science) aluminum-backed plates. Column chromatography was performed using silica gel (70−230) mesh. Unless otherwise stated all reactions were performed at room temperature under nitrogen. Drying after organic extractions was done over anhydrous Na_2SO_4. ^{1}H NMR spectra were recorded at 300 or 400 MHz, and ^{13}C NMR spectra were recorded at 75 or 100 MHz. Fast atom bombardment-mass spectrometry (FAB-MS) was performed on VG Autospec mass spectrometers. High pH anion exchange chromatography (HPAEC) was performed on a Dionex LC system fitted with a Dionex Carbopac PA-1 column and detected with a pulse-amperometric detector (PAD) or a radiometric detector (Beta Ram; IN/US Systems, Tampa, FL).

***p*-Cresyl 2,3-di-*O*-Benzyl-5-*O*-*t*-butyldiphenylsilyl-1-thio-α-D-arabinofuranoside (3).** To a solution of **2**[31] (2.00 g, 5.15 mmol) in dry pyridine (20 mL) at room temperature was added dropwise benzoyl chloride (1.93 mL, 16.7 mmol). After 1.5 h of stirring, the reaction mixture was concentrated. The residue was dissolved in CH_2Cl_2, washed with brine, dried (Na_2SO_4), concentrated, and the residue was used for the next step without further purification. To a solution of the benzoylated residue (3.90 g, 5.57 mmol) in dry CH_2Cl_2 (20 mL) at 0 °C was added *p*-thiocresol (761 mg, 6.13 mmol). After stirring for 10 min, $BF_3{\cdot}Et_2O$ (848 μL, 6.69 mmol) was added dropwise. The reaction mixture was stirred for 1 h at 0 °C, and then Et_3N (450 μL) was added. The mixture was diluted with CH_2Cl_2, washed with a saturated solution of $NaHCO_3$, dried (Na_2SO_4), concentrated, and purified by chromatography on silica gel (hexane/EtOAc, 8:1) to give **3** (0.92 g, 46.0% in

(17) Torrelles, J. B.; Khoo, K. H.; Sieling, P. A.; Modlin, R. L.; Zhang, N.; Marques, A. M.; Treumann, A.; Rithner, C. D.; Brennan, P. J.; Chatterjee, D. *J. Biol. Chem.* **2004**, *279*, 41227−41239.
(18) Treumann, A.; Xidong, F.; McDonnell, L.; Derrick, P. J.; Ashcroft, A. E.; Chatterjee, D.; Homans, S. W. *J. Mol. Biol.* **2002**, *316*, 89−100.
(19) Turnbull, W. B.; Shimizu, K. H.; Chatterjee, D.; Homans, S. W.; Treumann, A. *Angew. Chem., Int. Ed.* **2004**, *43*, 3918−3922.
(20) Mikusova, K.; Huang, H.; Yagi, T.; Holsters, M.; Vereecke, D.; D'Haeze, W.; Scherman, M. S.; Brennan, P. J.; McNeil, M. R.; Crick, D. C. *J. Bacteriol.* **2005**, *187*, 8020−8025.
(21) Scherman, M. S.; Kalbe-Bournonville, L.; Bush, D.; Xin, Y.; Deng, L.; McNeil, M. *J. Biol. Chem.* **1996**, *271*, 29652−29658.
(22) Mills, J. A.; McNeil, M. R.; Belisle, J. T.; Jacobs, W. R., Jr.; Brennan, P. J. *J. Bacteriol.* **1994**, *176*, 4803−4808.
(23) Belanger, A. E.; Besra, G. S.; Ford, M. E.; Mikusova, K.; Belisle, J. T.; Brennan, P. J.; Inamine, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11919−11924.
(24) Telenti, A.; Philipp, W. J.; Sreevatsan, S.; Bernasconi, C.; Stockbauer, K. E.; Wieles, B.; Musser, J. M.; Jacobs, W. R., Jr. *Nat. Med.* **1997**, *3*, 567−570.
(25) Escuyer, V. E.; Lety, M. A.; Torrelles, J. B.; Khoo, K. H.; Tang, J. B.; Rithner, C. D.; Frehel, C.; McNeil, M. R.; Brennan, P. J.; Chatterjee, D. *J. Biol. Chem.* **2001**, *276*, 48854−48862.
(26) Zhang, N.; Torrelles, J. B.; McNeil, M. R.; Escuyer, V. E.; Khoo, K. H.; Brennan, P. J.; Chatterjee, D. *Mol. Microbiol.* **2003**, *50*, 69−76.
(27) Alderwick, L. J.; Seidel, M.; Sahm, H.; Besra, G. S.; Eggeling, L. *J. Biol. Chem.* **2006**, *281*, 15653−15661.
(28) Seidel, M.; Alderwick, L. J.; Birch, H. L.; Sahm, H.; Eggeling, L.; Besra, G. S. *J. Biol. Chem.* **2007**, *282*, 14729−14740.
(29) Ayers, J. D.; Lowary, T. L.; Morehouse, C. B.; Besra, G. S. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 437−442.
(30) Joe, M.; Sun, D.; Taha, H.; Completo, G. C.; Croudace, J. E.; Lammas, D. A.; Besra, G. S.; Lowary, T. L. *J. Am. Chem. Soc.* **2006**, *128*, 5059−5072.
(31) Khasnobis, S.; Zhang, J.; Angala, S. K.; Amin, A. G.; McNeil, M. R.; Crick, D. C.; Chatterjee, D. *Chem. Biol.* **2006**, *13*, 787−795.
(32) Lee, R. E.; Brennan, P. J.; Besra, G. S. *Glycobiology* **1997**, *7*, 1121−1128.

two steps): R_f 0.55 (hexane/EtOAc, 4:1). ^{1}H NMR (300 MHz, $CDCl_3$): δ 8.15 (d, 2 H, J = 8.1 Hz), 8.01 (d, 2 H, J = 8.1 Hz), 7.76−7.15 (m, 20 H), 5.77−5.70 (m, 3 H), 4.66 (dd, 1 H, J = 4.5, 9.2 Hz), 4.08 (d, 2 H, J = 4.5 Hz), 2.37 (s, 3 H), 1.09 (s, 9 H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 165.8, 165.6, 138.2, 136.0, 135.9, 135.1, 133.7, 133.5, 133.3, 133.1, 130.3, 130.2, 130.1, 130.0, 129.6, 129.3, 128.7, 128.6, 128.0, 91.6, 83.4, 82.6, 77.9, 63.8, 27.0, 21.1, 19.6.

3,5-*O*-(1,1,3,3-Tetraisopropyldisiloxane-1,3-diyl)-D-arabinofuranose (4). To a solution of d-arabinose (2.00 g, 13.3 mmol) in dry pyridine (20 mL) at 0 °C was added dropwise $(Cl(i\text{-}Pr)_2Si)_2O$ (4.7 mL, 14.7 mmol). After stirring for 2 h at room temperature, the reaction mixture was concentrated under reduced pressure. The residue was purified by chromatography on silica gel (hexane/EtOAc, 4:1 → 2:1) to give **4** (3.09 g, 59.4%): R_f 0.34 (hexane/EtOAc, 2:1). ^{1}H NMR (300 MHz, $CDCl_3$): δ 5.24 (d, 0.7 H, J = 4.5 Hz), 5.20 (d, 0.3 H, J = 3.0 Hz), 4.28−3.64 (m, 5 H), 1.10−0.93 (m, 28 H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 101.5, 94.8, 82.4, 81.4, 78.3, 78.0, 76.6, 73.0, 65.1, 62.0, 17.8, 17.7, 17.7, 17.6, 17.4, 17.4, 17.3, 17.3, 17.2, 13.9, 13.7, 13.4, 13.4, 13.3, 13.2, 13.0, 13.0, 12.7.

***p*-Cresyl-2-*O*-benzoyl-3,5-*O*-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-1-thio-α-D-arabinofuranoside (5).** To a solution of **4** (1.00 g, 2.55 mmol) in dry pyridine (10 mL) was added dropwise benzoyl chloride (0.65 mL, 5.61 mmol) at room temperature. After 1.5 h of stirring, the reaction mixture was concentrated. The residue was dissolved in CH_2Cl_2, washed with brine, dried (Na_2SO_4), concentrated, and the residue was used for the next step without further purification. To a solution of the benzoylated residue (1.53 g) in dry CH_2Cl_2 (10 mL) at 0 °C *p*-thiocresol (348 mg, 2.81 mmol) was added. After stirring for 10 min, $BF_3{\cdot}Et_2O$ (388 μL, 3.06 mmol) was added dropwise. The reaction mixture was stirred for 1.5 h at 0 °C before Et_3N (422 μL) was added. The mixture was diluted with CH_2Cl_2, washed with a saturated solution of $NaHCO_3$, dried (Na_2SO_4), concentrated, and purified by chromatography on silica gel (hexane/EtOAc, 10:1) to give **5** (0.66 g, 43.1% in two steps): R_f 0.67 (hexane/EtOAc, 4:1). ^{1}H NMR (300 MHz, $CDCl_3$): δ 8.06 (m, 2 H), 7.53−7.41 (m, 5 H), 7.12 (d, 2H, J = 8.1 Hz), 5.62 (dd, 1 H, J = 3.9, 5.4 Hz), 5.49 (d, 1 H, J = 3.9 Hz), 4.59 (dd, 1 H, J = 5.4, 8.1 Hz), 4.27−4.23 (m, 1 H), 4.14−4.07 (m, 2 H), 2.33 (s, 3 H), 1.17−0.94 (m, 28 H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 165.7, 133.8, 133.6, 132.4, 130.4, 130.0, 129.9, 129.0, 128.7, 89.9, 83.4, 81.1, 75.7, 61.5, 21.4, 17.7, 17.6, 17.3, 17.2, 13.7, 13.4, 13.1, 12.7.

Octyl 2,3-Di-*O*-benzyl-α-D-arabinofuranoside (6). Compound **6** was synthesized as described:[31] R_f 0.09 (hexane/EtOAc, 4:1). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.41−7.28 (m, 10 H), 5.05 (s, 1 H), 4.64−4.51 (m, 4 H), 4.16−4.14 (m, 1 H), 4.05 (d, 1 H, J = 0.9 Hz), 4.01 (dd, 1 H, J = 3.0, 6.3 Hz), 3.87 (dd, 1 H, J = 2.7, 12.0 Hz), 3.72 (dd, 1 H, J = 2.7, 6.8 Hz), 3.68 (m, 1 H), 3.41 (m, 1 H), 1.94 (br, 1 H), 1.61 (m, 2 H), 1.30 (m, 10 H), 0.90 (m, 3 H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 138.1, 137.6, 128.7, 128.7, 128.2, 128.1, 128.0, 106.5, 88.3, 82.9, 82.0, 72.5, 72.2, 68.0, 62.4, 32.1, 29.8, 29.6, 29.5, 26.4, 22.9, 14.4.

Octyl 5-*O*-(2,3-Di-*O*-benzoyl-5-*O*-*t*-butyldiphenylsilyl-α-D-arabinofuranosyl)-2,3-di-*O*-benzyl-α-D-arabinofuranoside (7). To a solution of alcohol **6** (178 mg, 0.40 mmol) and donor **3** (309 mg, 0.44 mmol) in dry CH_2Cl_2 (4 mL) powder 4 Å molecular sieves (0.5 g) was added. The reaction mixture was cooled to 0 °C and treated with *N*-iodosuccinimide (NIS, 135 mg, 0.60 mmol) and silver triflate (AgOTf, 21 mg, 0.080 mmol). After stirring for 30 min at this temperature, the reaction mixture was quenched with Et_3N (400 μL) and filtered through Celite. The filtrate was washed successively with a saturated solution of $Na_2S_2O_3$, water, and brine. The organic phase was dried (Na_2SO_4) and concentrated. The residue was purified by chromatography on silica gel (hexane/EtOAc, 10:1) to give **7** (377 mg, 92.0%): R_f 0.49 (hexane/EtOAc, 4:1). ^{1}H NMR (300 MHz, $CDCl_3$): δ 8.06−7.95 (m, 4 H), 7.72−7.18 (m, 26 H), 5.62 (d, 1 H, J = 3.9 Hz), 5.54 (m, 1 H), 5.31 (d, 1 H, J = 3.9 Hz), 5.05 (s, 1 H), 4.84−4.40 (m, 4 H), 4.31 (dd, 1 H, J = 3.0, 6.8 Hz), 4.06−3.92 (m, 4 H), 3.75−3.68 (m, 2 H), 3.41−3.35 (m, 1 H), 1.57 (m, 2 H), 1.26 (m, 10 H), 1.03 (s, 9 H), 0.87 (m, 3 H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 165.8, 165.5, 135.9, 133.5, 133.4, 130.2, 129.9, 129.7, 129.5, 128.6, 128.2, 128.1, 127.9, 106.3, 106.0, 88.7, 83.5, 82.5, 80.0, 77.7, 72.4, 72.3, 68.0, 66.5, 63.7, 32.1, 29.8, 29.6, 29.5, 27.0, 22.9, 19.5, 14.4.

Octyl 5-*O*-(2,3-Di-*O*-benzoyl-α-D-arabinofuranosyl)-2,3-di-*O*-benzyl-α-D-arabinofuranoside (8). To a solution of **7** (257 mg, 0.35 mmol) in THF (10 mL) at 0 °C was added *n*-Bu_4NF (183 mg, 0.70 mmol). After stirring at room temperature for 2 h, the reaction mixture was diluted with CH_2Cl_2 and washed with a saturated solution of $(NH_4)_2SO_4$ and brine. The organic phase was dried (Na_2SO_4) and concentrated. The residue was purified by chromatography on silica gel (hexane/EtOAc, 4:1 → 2:1) to give **8** (95 mg, 48.2%): R_f 0.54 (hexane/EtOAc, 2:1). ^{1}H NMR (300 MHz, $CDCl_3$): δ 8.09−8.04 (m, 4 H), 7.61−7.23 (m, 16 H), 5.63 (d, 1 H, J = 1.2 Hz), 5.41 (dd, 1 H, J = 0.6, 4.5 Hz), 5.34 (s, 1 H), 5.07 (s, 1 H), 4.68−4.38 (m, 3 H), 4.25−4.19 (m, 2 H), 3.98−3.91 (m, 2 H), 3.78−3.67 (m, 2 H), 3.44−3.36 (m, 1 H), 2.30 (br, 1 H), 1.61−1.54 (m, 2 H), 1.29 (m, 10 H), 0.90 (m, 3 H). ^{13}C NMR (75 MHz, $CDCl_3$):δ 166.4, 165.4, 138.1, 137.8, 133.8, 133.7, 130.2, 130.1, 129.3, 128.8, 128.7, 128.7, 128.6, 128.3, 128.2, 127.9, 106.2, 105.9, 88.8, 84.1, 83.3, 81.8, 79.8, 78.1, 72.4, 68.0, 66.5, 62.5, 32.1, 29.8, 29.6, 29.5, 26.4, 22.9, 14.4.

Octyl 5-*O*-(2,3-Di-*O*-benzoyl-5-*O*-(2-*O*-benzoyl-3,5-*O*-(1,1,3,3-tetraisopropyldisiloxane-1,3-diyl)-α-D-arabinofuranosyl)-α-D-arabinofuranosyl)-2,3-di-*O*-benzyl-α-D-arabinofuranoside (9). Alcohol **8** (90 mg, 0.11 mmol) and donor **5** (114 mg, 0.19 mmol) were coupled using NIS (44 mg, 0.17 mmol) and AgOTf (10 mg, 0.038 mmol) in CH_2Cl_2 (3 mL) for 30 min as described for preparation of **7**. Purification by chromatography (hexane/EtOAc, 8:1) gave **9** (100 mg, 69.0%): R_f 0.33 (hexane/EtOAc, 4:1). ^{1}H NMR (300 MHz, $CDCl_3$): δ 8.15−8.02 (m, 6 H), 7.62−7.23 (m, 19 H), 5.60−5.54 (m, 3 H), 5.34 (s, 1 H), 5.16 (d, 1 H, J = 1.5 Hz), 5.07 (s, 1 H), 4.63−4.48 (m, 4 H), 4.39 (dd, 1 H, J = 4.2, 8.1 Hz), 4.25−3.87 (m, 8 H), 3.76−3.71 (m, 3 H), 3.44−3.36 (m, 2 H), 1.61−1.57 (m, 2 H), 1.29 (m, 10 H), 1.13−0.77 (m, 3 H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 165.9, 165.5, 165.5, 138.2, 137.9, 133.6, 133.5, 133.4, 130.2, 130.2, 130.0, 129.9, 129.5, 128.8, 128.7, 128.6, 128.6, 128.3, 128.1, 128.0, 106.2, 106.0, 105.9, 88.8, 84.2, 83.3, 82.4, 81.8, 81.4, 79.9, 77.9, 76.1, 72.4, 72.3, 68.0, 67.3, 66.5, 61.6, 31.1, 29.8, 29.7, 29.6, 26.4, 22.9, 17.7, 17.6, 17.2, 17.1, 14.4, 13.7, 13.4, 13.1, 12.6.

Octyl 5-*O*-(2,3-Di-*O*-benzoyl-5-*O*-(2-*O*-benzoyl-α-D-arabinofuranosyl)-α-D-arabinofuranosyl)-2,3-di-*O*-benzyl-α-D-arabinofuranoside (10). Compound **9** (95 mg, 0.075 mmol) was deprotected in a solution of *n*-Bu_4NF (59 mg, 0.23 mmol) in THF (1 mL) as described for preparation of **8**. Purification by chromatography (hexane/EtOAc, 2:1 → 1:1) gave **10** (69 mg, 89.8%): R_f 0.10 (hexane/EtOAc, 2:1). ^{1}H NMR (300 MHz, $CDCl_3$): δ 8.10−8.00 (m, 6 H), 7.61−7.25 (m, 19 H), 5.61 (d, 1 H, J = 1.2 Hz), 5.55 (dd, 1 H, J = 0.6, 5.1 Hz), 5.40 (s, 1 H), 5.37 (s, 1 H), 5.14 (dd, 1 H, J = 0.6, 3.0 Hz), 5.08 (s, 1H), 4.64−4.46 (m, 4 H), 4.34−4.30 (m, 1 H), 4.26−4.22 (m, 1 H), 4.19−4.14 (m, 2 H), 4.08−4.04 (m, 2 H), 3.98−3.90 (m, 3 H), 3.81−3.69 (m, 3 H), 3.44−3.37 (m, 2 H), 1.61−1.55 (m, 2 H), 1.31−1.29 (m, 10 H), 0.92−0.88 (m, 3 H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 166.9, 166.2, 165.5, 138.1, 137.9, 133.9, 133.8, 133.7, 130.3, 130.1, 130.0, 130.0, 129.3, 129.3, 129.2, 128.8, 128.7, 128.7, 128.6, 128.4, 128.3, 127.9, 106.3, 105.9, 105.4, 88.8, 86.5, 84.5, 83.4, 82.3, 82.0, 79.9, 77.6, 72.4, 68.0, 66.6, 65.9, 62.4, 32.1, 29.8, 29.6, 29.5, 26.4, 22.9, 14.4.

Octyl 5-*O*-(2,3-Di-*O*-benzoyl-5-*O*-(2-*O*-benzoyl-3,5-di-*O*-(2-*O*-benzoyl-3,5-di-*O*-benzyl-α-D-arabinofuranosyl)-α-D-arabinofuranosyl)-α-D-arabinofuranosyl)-2,3-di-*O*-benzyl-α-D-arabinofuranoside (12). Alcohol **10** (30 mg, 0.029 mmol) and donor **11** (38 mg, 0.070 mmol) were coupled using NIS (19 mg, 0.083 mmol) and AgOTf (5.4 mg, 0.021 mmol) in CH_2Cl_2 (3 mL) for 2 h as described for preparation of **7**. Purification by chromatography (hexane/EtOAc, 6:1 → 4:1) gave

12 (48 mg, 88.1%): R_f 0.44 (hexane/EtOAc, 2:1). ^{1}H NMR (300 MHz, $CDCl_3$): δ 8.17−7.88 (m, 10 H), 7.59−7.06 (m, 45 H), 5.68 (d, 1 H, J = 4.2 Hz), 5.61 (d, 1 H, J = 0.6 Hz), 5.53 (d, 1 H, J = 1.5 Hz), 5.49 (s, 1 H), 5.43 (s, 2 H), 5.42 (s, 1 H), 5.33 (s, 1 H), 5.27 (d, 1 H, J = 0.9 Hz), 5.07 (d, 1 H, J = 1.5 Hz), 4.70−4.21 (m, 18 H), 4.15−4.03 (m, 3 H), 4.00−3.92 (m, 3 H), 3.86 (d, 2 H, J = 5.7 Hz), 3.75−3.69 (m, 2 H), 3.61−3.37 (m, 6 H), 1.60−1.53 (m, 2 H), 1.29 (m, 10 H), 0.92−0.88 (m, 3 H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 165.9, 165.7, 165.6, 165.4, 165.1, 138.3, 138.2, 138.2, 138.0, 137.9, 133.6, 133.4, 133.3, 133.3, 130.2, 130.2, 130.1, 130.0, 129.8, 129.4, 129.3, 128.7, 128.7, 128.6, 128.5, 128.5, 128.4, 128.4, 128.3, 128.2, 128.1, 128.0, 128.0, 127.8, 127.7, 127.7, 127.6, 127.6, 126.5, 106.2, 106.1, 106.0, 105.6, 88.8, 83.5, 83.4, 83.3, 83.1, 83.0, 82.6, 82.4, 81.9, 81.8, 81.6, 80.8, 80.0, 77.5, 77.1, 76.8, 73.5, 73.5, 72.4, 72.2, 71.9, 71.8, 69.4, 69.2, 68.0, 66.4, 65.4, 65.1, 32.1, 29.8, 29.7, 29.5, 26.4, 22.9, 14.4. FAB-MS m/z 1852.3 $[M + H]^+$, 1874.3 $[M + Na]^+$, 1984.2 $[M + Cs]^+$.

Octyl 5-*O*-(5-*O*-(3,5-Di-*O*-(α-D-arabinofuranosyl)-α-D-arabinofuranosyl)-α-D-arabinofuranosyl)-α-D-arabinofuranoside (1). A solution of **12** (46 mg, 0.025 mmol) in dry CH_2Cl_2 (0.60 mL) and dry MeOH (1.8 mL) was treated with 0.1 M MeONa in MeOH (40 μL). After stirring for 4 h at room temperature, the mixture was neutralized with Amberlite IR-120 (H^+) resin, filtered, and concentrated. The residue was purified by chromatography on silica gel (hexane/EtOAc, 1:1 → 2:1). The debenzoylated residue (28 mg) was dissolved in AcOH/H_2O (4:1, 2.5 mL) and hydrogenolyzed over 10% Pd/C (30 mg) for 16 h. The mixture was filtered through Celite, concentrated, and the residue was purified by chromatography on silica gel ($CHCl_3$/MeOH, 4:1 → 3:1) to give **1** (15 mg, 76.5% in two steps): R_f 0.70 ($CHCl_3$/MeOH/H_2O, 10:10:3). ^{1}H NMR (400 MHz, $CDCl_3$): δ 5.13 (d, 1 H, J = 1.2 Hz), 5.10 (s, 1 H), 5.07 (d, 1 H, J = 1.2 Hz), 5.06 (d, 1 H, J = 1.2 Hz), 4.99 (d, 1 H, J = 2.4 Hz), 4.31−4.26 (m, 2 H), 4.19 (ddd, 1 H, J = 2.8, 4.4, 8.8 Hz), 4.15−4.06 (m, 6 H), 4.04−3.67 (m, 17 H), 3.58−3.52 (m, 1 H), 1.60−1.57 (m, 2 H), 1.33−1.26 (m, 10 H), 0.86−0.83 (m, 3 H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 107.5, 107.5, 107.4, 107.3, 107.2, 84.1, 84.0, 82.4, 82.3, 81.8, 81.8, 81.3, 81.0, 81.0, 80.9, 79.2, 76.7, 76.6, 76.5, 68.7, 66.9, 66.5, 66.4, 61.2, 61.2, 31.2, 28.7, 28.5, 28.4, 25.3, 22.1, 13.5. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) m/z 813.3 $[M + Na]^+$, 829.3 $[M + K]^+$

Preparation of Mycobacterial Membranes and Cell Wall Enriched Fraction (P60). Wild-type and all mutants of *M. smegmatis* mc^{2}155 cells were grown in 7H9 containing oleic acid−albumin−dextrose−catalase (OADC) to midlog phase (OD 0.7−0.8). Mutants were grown with kanamycin (50 μg/mL). Cells (approximately 8−10 g wet weight) were harvested by centrifugation, washed, and resuspended in ice-cold buffer A containing 50 mM MOPS (pH 8.5), 5 mM 2-mercaptoethanol, and 10 mM $MgCl_2$. Cells were disrupted by 6−7 passages through a French press at 1500 psi. The suspension was centrifuged at 27 000*g* for 60 min at 4 °C. The cell wall pellet was removed, and the supernatant was recentrifuged at 200 000*g* for 2 h at 4 °C. The supernatant was discarded, and the pellet of enzymatically active membrane was gently resuspended in 500 μL of buffer A; the protein concentrations of the membrane fractions were typically between 15 and 20 mg/mL. The 27 000*g* pellet was suspended in 10 mL of buffer A, and Percoll was added to achieve a 60% suspension. The suspension was mixed and centrifuged at 27 000*g* for 60 min at 4 °C. The resulting flocculent, white layer was collected and washed three times with buffer A at 12 000*g* to yield P60. This fraction was resuspended in 1 mL of buffer A to yield a protein concentration of 5−6 mg/mL.

Arabinosyltransferase Assays. Typical reaction mixtures contained buffer A, 62.5 μM ATP, 3.8 μM p[^{14}C]Rpp (500 000 disintegrations per minute (dpm)), acceptor **1** (300 μM), membranes (0.5 mg), and P60 (0.3 mg) in a total volume of 200 μL. The reaction mixtures were incubated at 37 °C for 1 h and then terminated by adding 200 μL of ethanol. The resulting mixture was centrifuged at 14 000*g*, and the supernatants were loaded onto prepacked strong anion exchange (SAX) columns (Burdick and Jackson). The columns were eluted with 2 mL of water. The eluate was evaporated to dryness and partitioned between the two phases (1:1) of water-saturated 1-butanol and water. The 1-butanol fractions were measured for radioactivity incorporation by liquid scintillation counting. Assays were carried out under conditions where synthesis was linear with respect to protein concentration. In the ethambutanol inhibition assay, the membrane and P60 were preincubated with ethambutanol at concentration of 0−100 μg/mL for 30 min, then the acceptor and pRpp were added for further incubation. Values reported are averages of duplicate reactions from representative experiments.

Analytical Procedures. For TLC analysis of the enzymatically synthesized, radiolabeled product, an aliquot of the 1-butanol extract was dried under air and the residue was reconstituted in Milli-Q water (10 μL) for analysis by silica gel TLC. The plates were developed in $CHCl_3$/MeOH/1 M NH_4OAc/NH_4OH/H_2O (180:140:9:9:23) followed by autoradiography at −70 °C using Biomax MR film (Kodak), and compound **1** was visualized using α-naphthol spray reagent. For analysis of the sugar composition of the radiolabeled product, the water layer after 1-butanol extraction were purified by a Biogel P-2 column, eluted with Milli-Q water, and approximately 2000 dpm of the purified product was hydrolyzed in 200 μL of 2 M trifluoroacetic acid (TFA) at 120 °C for 2 h. TFA was removed under a stream of air, and the hydrolysate was analyzed on a silica gel TLC plate developed in pyridine/ethyl acetate/acetic acid/water (5:5:1:3) followed by autoradiography as described above. Radioactive spots were identified by cochromatography with standard radioactive arabinose and galactose from radioactive [^{14}C]-labeled AG.

Radiolabeled products from the enzyme assays were subjected to digestion with *Cellulomonas* endoarabinanase isolated from *C. gelida*.[1] In this case, approximately 5000 dpm of the 1-butanol soluble material were dried under nitrogen, reconstituted in 10 μL of water, sonicated, and treated with endoarabinanase overnight at 37 °C. The digestion products were analyzed by HPAEC. The oligosaccharides were detected with a radiometric detector. The acceptor **1** was also digested with *Cellulomonas* endoarabinanase and subjected to HPAEC with a PAD detector. The digested residues were acetylated by acetic anhydride and pyridine and subjected to MALDI-TOF MS analysis.

Arabinosyltransferase Assays with Nonradioactive pRpp. A large-scale nonlabeled arabinosyltransferase reaction was performed to generate product for MALDI-TOF MS analysis. In order to do so, commercially available pRpp (SIGMA) was further purified by Dionex HPAEC; fractions corresponding to peaks detected by radioactive pRpp were pooled and stored in −20 °C.

The reaction mixtures contained buffer A, 62.5 μM ATP, 2.0 mM pRpp, acceptor **1** (1.0 mM), membranes (0.6 mg), and P60 (0.4 mg) in a total volume of 160 μL. The workup procedures were exactly the same as for the radioactive reactions. The eluate after SAX column was purified by preparative TLC, excising the band that corresponded to the radioactive product. The bands recovered from the plate were extracted with Milli-Q water (3 × 1 mL), pooled, dried, methylated, and subjected to MALDI-TOF MS analysis.

MALDI-MS and MS/MS Analyses. MALDI-TOF mass spectrometry analysis was performed using an UltraFlex TOF/TOF (Bruker Daltonics, Billurica, MA) in which case the peracetylation or permethylation derivatives in acetonitrile were mixed 1:1 with 2,5-dihydroxybenzoic acid (DHB) matrix (10 mg/mL in water) for spotting onto the target plate. MALDI TOF/TOF CID MS/MS sequencing of the permethyl derivatives using DHB as matrix was performed on a 4700 Proteomics Analyzer (Applied Biosystems) as described.[33]

(33) Lee, A.; Wu, S. W.; Scherman, M. S.; Torrelles, J. B.; Chatterjee, D.; McNeil, M. R.; Khoo, K. H. *Biochemistry* **2006**, *45*, 15817−15828.

Results and Discussion

Approach. Through organic synthesis and development of a convenient arabinosyltransferase assay[31] we have successfully shown that use of suitably linked D-arabinose-based oligosaccharides can be utilized in a mycobacterial cell-free system to identify any particular enzyme function. The strategy has been extremely helpful to demonstrate the function of the Emb proteins,[25,26,34] in which case, studies have been limited to examining biochemical events as a consequence of individual gene disruption experiments. Thus, when we synthesized a linear Araβ1→2Araα1→5Araα1→5Araα1→5Araα1→ with an octyl aglycon as an arabinosyl acceptor in cell-free assays with membrane and particulate cell wall as the enzyme source and 5-phosphoribose diphosphate p[^{14}C]Rpp as the arabinose donor a single product was formed. Analysis unequivocally showed that two arabinofuranosyl residues were added at the tertiary →5Araα1→ of the synthetic glycan thereby indicating that the EmbA and EmbB proteins function together for the transfer of the terminal disaccharide Araβ1→2Araα1 on the 3-arm of the penultimate →5Araα1→ on a linear substrate.

In AG, the nonreducing termini invariably form a well-defined [β-D-Ara*f*(1→2)-α-D-Ara*f*]$_2$-3,5-α-D-Ara*f*(1→5)-α-D-Ara*f* motif,[1] of which the terminal β-Ara*f* and the penultimate α-2-Ara*f* serve as the anchoring points for the mycolic acids.[14] Two such characteristic terminal Ara$_6$ motifs with additional α-5-Ara*f* residues at the reducing end are assembled into a unique Ara$_{22}$-mer, the largest structurally defined arabinan unit to date.[5,33] In the arabinan of LAM, additional linear α-5-Ara*f* extensions occur mostly into stretches of −[α-D-Ara*f*(1→5)-α-D-Ara*f*]$_n$− with critically spaced α3,5-Ara*f* branched sites.[34] We reasoned that if the acceptor is modified in a manner such that it represents the internal branched assembly upon which α-5-Ara*f* can be extended, it would allow us to identify this particular enzyme.

Synthesis of Acceptor 1. 1. Strategy. We used the well-established thioglycoside glycosylation method[35] in all chemical synthesis. The building blocks with participating groups were employed to get the α-anomer in high yield. Glycolipid **10** was set as a precursor to add two monosaccharides in 3- and 5-OHs at one time.

2. Synthesis of Building Blocks. Four building blocks (**3**, **5**, **6**, **11**) were used in this work. Compounds **6** and **11** were known,[31] and compounds **3** and **5** were prepared as in Scheme 1.

To access **3**, we started from **2**,[31] which was protected by benzoylation and then reacted with *p*-thiocresol and boron trifluoride to give α-anomer **3** in 43.8% yield over the two steps. Thioglycoside **5** was synthesized in three steps from D-arabinose, which was, first, treated with (Cl(*i*-Pr)$_2$Si)$_2$O to get 3,5-OHs protected block **4**. Compound **4** was benzoylated and thiocresolated as preparation of **3**.

3. Glycosylation and Deprotection. After all building blocks were in hand, conditions for assembly of the pentasaccharide were explored (Scheme 2). Coupling of **6** with thioglycoside **3**, upon activation with NIS and AgOTf, the disaccharide **7** was obtained in 92.0% yield. Compound **8** obtained after deprotection of **7**, was coupled with building block **5** to give trisaccharide **9** in 69.0% yield. In subsequent steps, compound **9** was deprotected, and the resulting diol **10** were coupled with 2.4 equiv of **11** to get pentasaccharide **12**. Deprotection of **12** was achieved by treatment with a catalytic amount of sodium methoxide followed by hydrogenolysis affording pentasaccharide **1** in overall 76.5% yield.

The anomeric stereochemistry of the glycosyl residue in **1** was determined by both ^{1}H and ^{13}C NMR spectroscopy. The ^{13}C NMR spectrum showed five α-anomeric carbons at δ 107.6, 107.5, 107.4, 107.3, and 107.2 ppm (Figure 1A). In the ^{1}H NMR spectrum (Figure 1B), five anomeric hydrogens (δ 5.13, 5.10, 5.07, 5.06, 4.99 ppm) appeared as singlet or doublet ($J_{H1,H2}$ = 0−2.0 Hz), characteristic of α-arabinofuranosides. The HSQC spectrum (data not shown) correlated δ ^{1}H 5.13 and 4.99 ppm with δ ^{13}C 107.2 and 107.3 ppm, which were assigned to be t-α-Ara*f*→3 and t-α-Ara*f*→5, respectively. MALDI-TOF mass spectral analysis yielded a positive ion at *m*/*z* 813.3 [M + Na]$^+$ and 829.3 [M + K]$^+$ corroborating for $C_{33}H_{58}O_{21}$ (Figure 1C). Linkage analysis (GC/MS, Figure 1D) indicated that **1** had t-Ara*f*, 5-Ara*f*, and 3,5-Ara*f*, in the ratio of 1:1:0.5 in GC (data not shown).

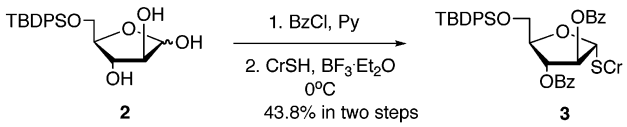
Enzymatic Assays. Compound **1** was tested for its ability to act as an acceptor substrate for arabinosyltransferase activity in a cell-free assay using p[^{14}C]Rpp (which is converted in situ to DPA) as an indirect arabinosyl donor.[21] When a total crude cell lysate of wild-type *M. smegmatis* was incubated with p[^{14}C]-Rpp, a low level of radioactivity was detected in the 1-butanol extract.

When a mixture of the membrane and cell wall enriched fraction (P60) was incubated with p[^{14}C]Rpp and compound **1**, a reasonable amount (1−2% of the p[^{14}C]Rpp used) of radiolabeled material was found in the 1-butanol extract with a significant amount of radioactivity remaining in the water layer. When compound **1** was omitted from the reaction mixture trace radioactivity was detected in the 1-butanol layer and a reduced amount of radioactivity associated with the water layer. This indicated that the product formed had an affinity for organic solvent despite addition of sugar(s). TLC analysis revealed formation of a major slower migrating component (band I, Figure 2A) along with some minor bands (band II, Figure 2A) that were absent in the control reaction. Incorporation of radioactivity into the 1-butanol layer was also linear with respect to protein up to a concentration of 0.8 mg/mL (Supporting Information Figure S1). There was no significant increase of activity with increasing protein concentrations above this amount. TLC analyses of the 1-butanol layer isolated at different protein concentration level corroborated that the reaction was optimal at 0.8 mg/mL protein concentration (Figure 2B). The concentration of the indirect donor of arabinose residues (p[^{14}C]Rpp) used was 3.8 μM, and consequently the concentration of the direct arabinose donor (DPA) was probably less. In a typical reaction, product was extracted with ethanol and excess p[^{14}C]Rpp removed by passing through an anion exchange column (SAX). The eluate was dried, and the residue was partitioned between water-saturated 1-butanol and water. TLC analysis revealed the product formation when the enzyme source was used from either the *embA*, *embB*, or *embC* knockout strains indicating that the biosynthetic step in the product formed utilizing this particular acceptor was independent of the Emb proteins (Figure 3A). A reduced amount of product (40%) was noted in the reaction with *embA* knock-

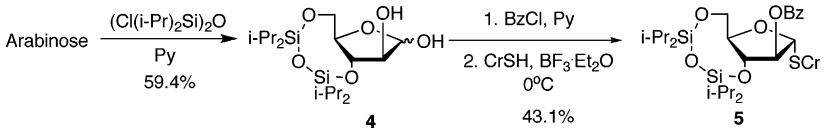
(34) Shi, L.; Berg, S.; Lee, A.; Spencer, J. S.; Zhang, J.; Vissa, V.; McNeil, M. R.; Khoo, K. H.; Chatterjee, D. *J. Biol. Chem.* **2006**, *281*, 19512−19526.
(35) Garegg, P. J. *Adv. Carbohydr. Chem. Biochem.* **1997**, *52*, 179−205.

Scheme 1. Chemical Synthesis of Building Blocks

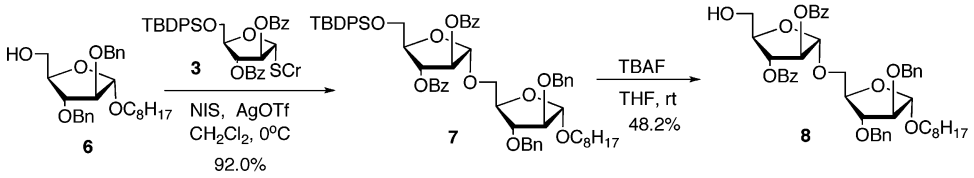
Synthesis of building block **3**

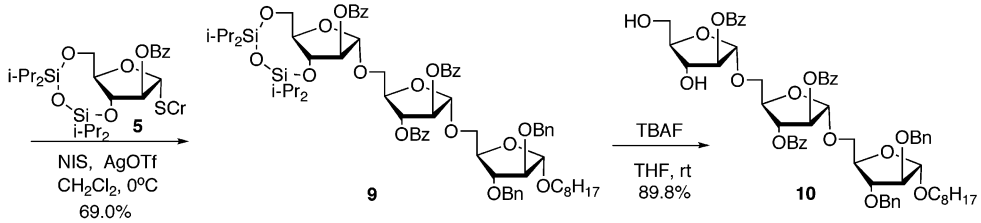


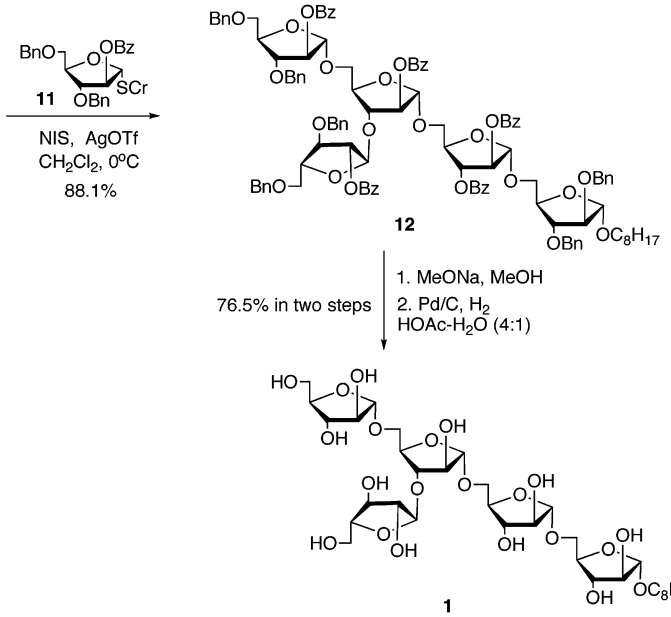
Synthesis of building block **5**



Scheme 2. Chemical Synthesis of Acceptor **1**







out strain (lane 2, Figure 3A), and a normal amount of product was detected from the *embA*-complemented strain (data not shown), which indicated that inactivation of *embA* somewhat affected this arabinosyltransferase activity. When membrane and P60 from the wild-type strain were preincubated with ethambutol (EMB) at a concentration of 50 μg/mL for 30 min, a reduced level (65%) of product formation was observed (Figure 3B and Supporting Information Figure S2). If P60 was absent from the assay, the biosynthetic products were not detectable in TLC (Supporting Information Figure S3). So P60 is necessary in the reaction, and the reason is still unknown.

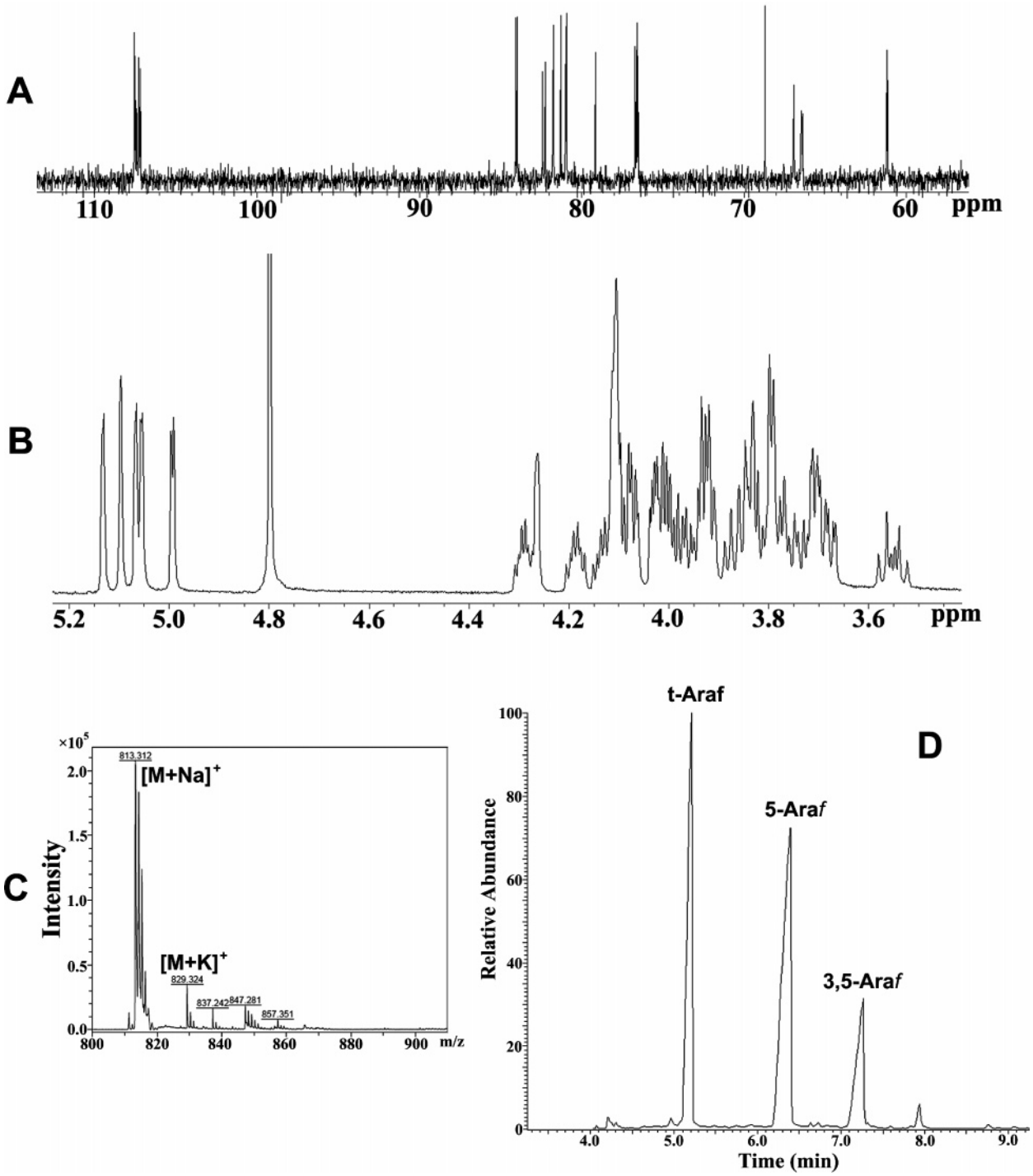


Figure 1. Structural characterization of acceptor **1** by ^{13}C NMR (A), ^{1}H NMR (B), MALDI-TOF MS (C), and glycosyl linkage analysis (D). For linkage analysis, the sample was per-O-methylated, hydrolyzed using 2 M TFA, reduced and per-O-acetylated. The resulting partially per-O-methylated, per-O-acetylated glycosyl derivatives were analyzed by GC/MS.

Characterization of the Arabinosyltransferase Reaction Product. To confirm that [^{14}C]arabinose had been added to the acceptor an aliquot of the purified material (2000 dpm) was hydrolyzed with 2 M TFA and the hydrolysate was subjected to TLC. Autoradiography showed that the radioactivity was exclusively associated with arabinose (Figure 4A).

The major product (band I) was excised from a TLC plate, and the migration pattern of the purified product was compared to that of two separate synthetic standards, Araβ1→2Araα1→5-(Araα1→3)Araα1→5Araα1→5Araα1−OC_8H_{17} (hexasaccharide) and [Araβ1→2Araα1]$_2$→3,5Araα1→5Araα1→5Araα1−OC_8H_{17} (heptasaccharide) (Zhang, J. and Chatterjee, D., unpublished work, see the Supporting Information Figure S4). The TLC mobility of the band corresponding to the major product was closer to the synthetic hexamer although not superimposible suggesting that only one arabinosyl residue has been donated by pRpp resulting onto formation of perhaps a hexamer in band I. *C. gelida* endoarabinanase has been used extensively to digest LAM and AG and define the nonreducing ends.[1,36] When this enzyme was used to digest the acceptor **1**,

(36) Chatterjee, D.; Khoo, K.-H.; McNeil, M. R.; Dell, A.; Morris, H. R.; Brennan, P. J. *Glycobiology* **1993**, *3*, 497−506.

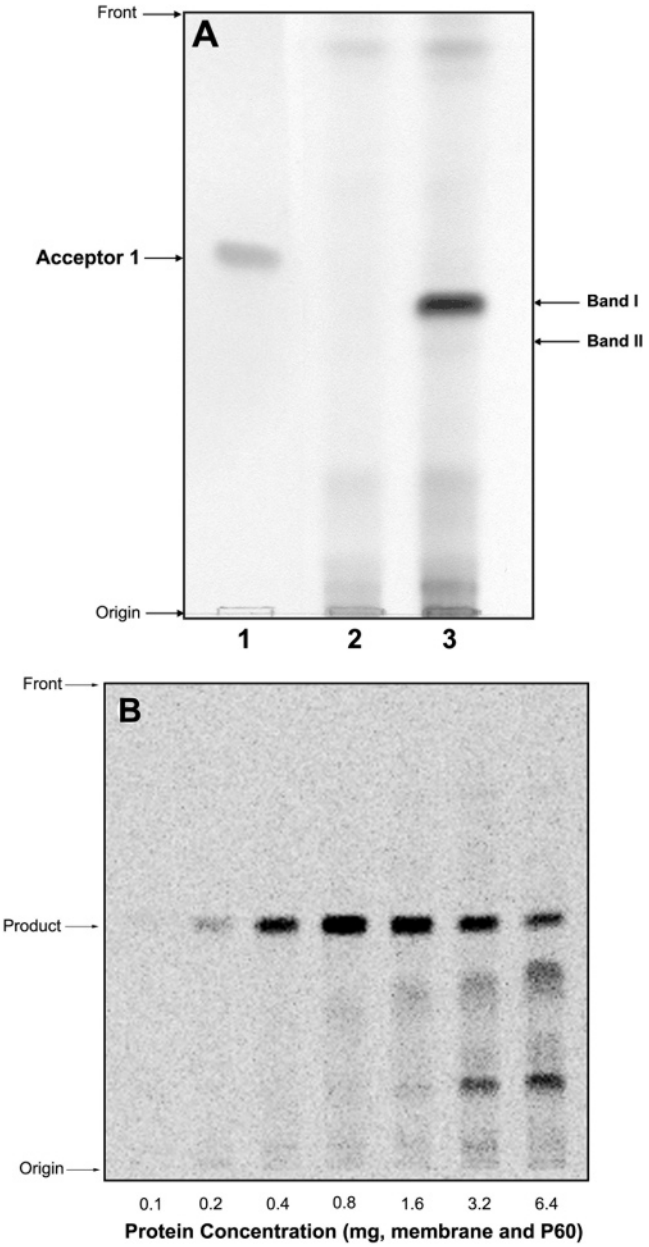


Figure 2. TLC analysis of biosynthetic products (A) and effect of protein concentration (B). (A) Acceptor **1** was visualized after spraying with α-naphthol (lane 1); the radioactive product without (lane 2) or with (lane 3) acceptor was revealed by autoradiography. An aliquot of the product labeled with p[^{14}C]Rpp from the eluate after the SAX column followed by 1-butanol extraction was applied to the TLC plate, developed in $CHCl_3$/MeOH/1 M $NH_4OAc/NH_4OH/H_2O$ (180:140:9:9:23), and exposed for 4 days. The migration of the enzymatically synthesized product is slower than acceptor **1**. Major band I and minor band II were detected from radioactive TLC. (B) Incorporation of radioactivity into the product relative to protein concentration. Seven sets of reactions were performed, each having different protein (membrane + P60) concentrations (0.1, 0.2, 0.4, 0.8, 1.6, 3.2, and 6.4 mg). The ratio of membrane to P60 was maintained as in the basic arabinosyltransferase assay. All other reaction conditions were identical to the basic arabinosyltransferase assay. Same aliquots of products from different protein concentration were loaded to TLC.

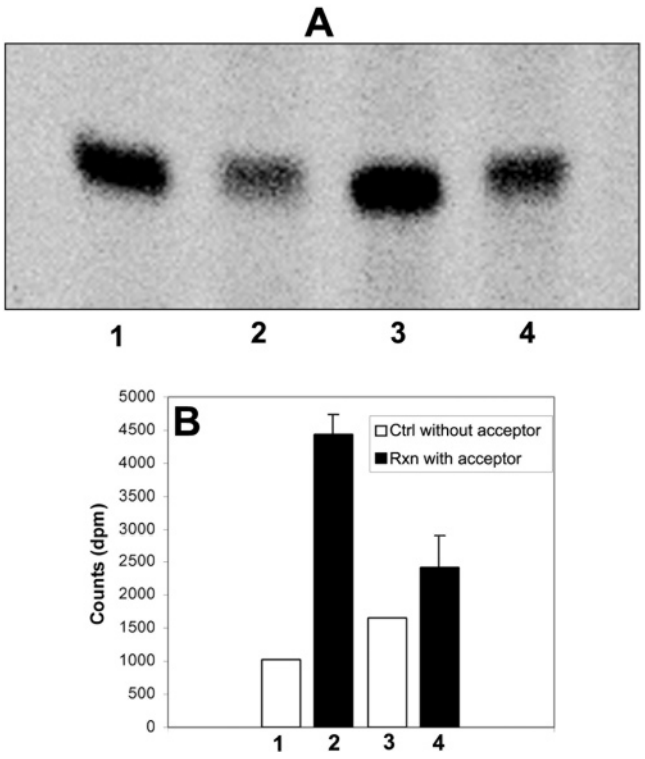


Figure 3. Analysis of biosynthetic products from reactions with mutant strains (A) and ethambutol (EMB) inhibition (B). (A) The reactions were set as *M. smegmatis* WT (lane 1); *embA* knockout mutant (lane 2); *embB* knockout mutant (lane 3); *embC* knockout mutant (lane 4). (B) Incorporation of radioactivity into the product in the presence of EMB (lanes 3 and 4) and absence of EMB (lanes 1 and 2). The enzymes (membrane and P60) were preincubated with EMB (50 μg/mL) for 30 min, followed by addition of acceptor and pRpp.

branched Ara_4 (33.27 min) was obtained (peak a, Figure 4B) whose structure was confirmed using MALDI-TOF mass spectrometry (inset of Figure 4B), plus trace amounts of Ara_5 (peak b, Figure 4B). In contrast, when the de novo synthesized product was subjected to endoarabinanase digestion followed by Dionex HPAEC using the radioactive detector, the digested material yielded radioactive Ara_2 (peak c, Figure 4C), and presumably some Ara_5 (peak d, Figure 4C) and Ara_2 eluted with the same retention time as authentic Ara_2 released from radiolabeled AG.[37] If the Ara*f* residue incorporated was β-Ara*f*, no Ara_2 formation would occur. Therefore, we reasoned that addition of one α-Ara*f* had occurred in this major band.

In order to characterize the product further, commercially available unlabeled pRpp was used in 2-fold excess in one set of reactions as described in the Experimental Section. After removing excess unreacted pRpp with the SAX column, the workup was modified by inclusion of preparative TLC followed by excision of the band corresponding to the radiolabeled hexamer. Along with this, other slower migrating bands were also recovered. The components were dried, methylated, and analyzed by MALDI-TOF mass spectrometry. In addition to the $[M + Na]^+$ ion of *m*/*z* 1127 corresponding to an Ara_6 (Figure 5), ions (*m*/*z* 1287, 1447, 1607, and 1767) corresponding to Ara_7, Ara_8, Ara_9, and Ara_{10} were also observed (Figure 5), albeit in visibly smaller quantities. The aqueous layer was also examined by TLC and autoradiography but did not reveal any products that would correspond to elongation of chain length beyond Ara_{10}.

To define more specifically the branching pattern of the Ara oligomers synthesized, the permethyl derivatives of the major products, Ara_6 and Ara_7, were subjected to high-energy CID

(37) Xin, Y.; Huang, Y.; McNeil, M. R. *Biochim. Biophys. Acta* **1999**, *1473*, 267−271.

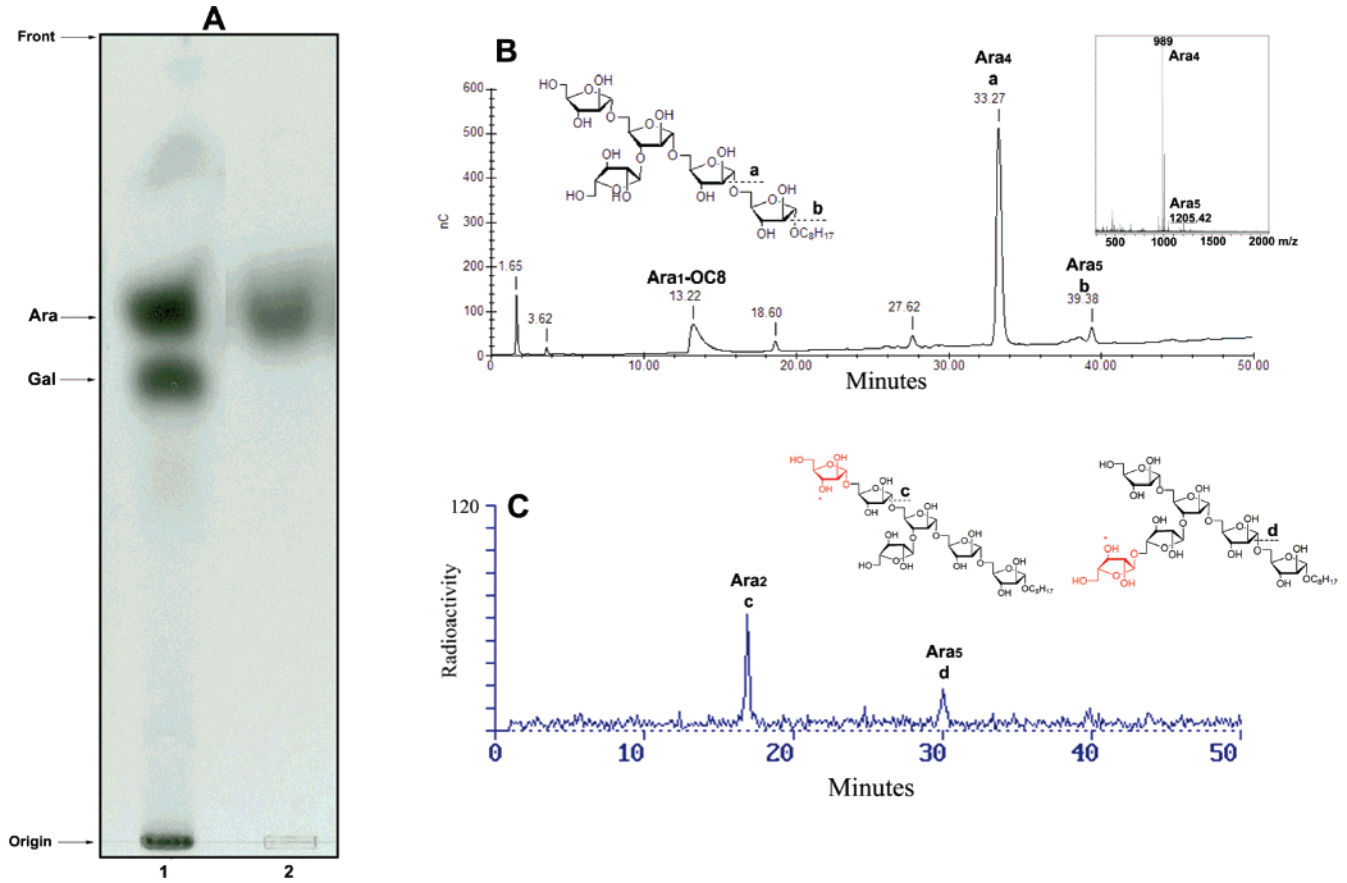


Figure 4. Structural analysis of biosynthetic product. (A) An amount of 2000 dpm of radioactive AG (lane 1) and the product (lane 2) were hydrolyzed with 2 M TFA, the acid removed by evaporation, and applied to TLC. The chromatogram was developed in pyridine/ethyl acetate/acetic acid/water (5:5:1:3) and subjected to autoradiography. (B) Acceptor **1** (10 μg) was digested with the *Cellulomonas* enzyme and subjected to HPAEC. The peaks were confirmed after per-O-acetylation and MALDI-TOF mass spectral analyses (inset) to be major Ara_4 and minor Ara_5. (C) Purified product was digested with *Cellulomonas* and subjected to HPAEC with a radioactive detector. Red labels represent newly incorporated radioactive arabinose added on the acceptor.

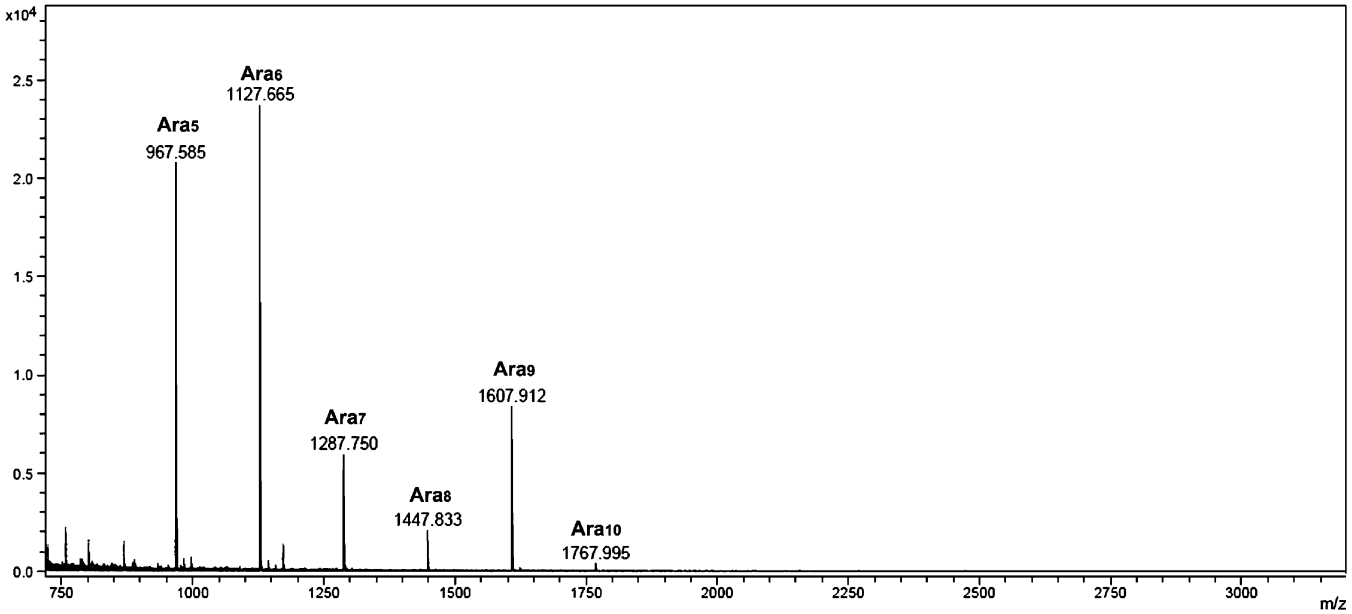


Figure 5. MALDI-TOF MS profile of products from reaction utilizing unlabeled pRpp. The band that corresponded to the radioactive hexamer (band I in Figure 2A) and the regions corresponding to larger oligomers (band II in Figure 2A) were recovered and permethylated.

MS/MS analysis on a MALDI-TOF/TOF, along with the starting material, namely, the acceptor substrate Ara_5 (compound 1). The MS/MS fragmentation pattern, which has been previously established through analysis of arabinan oligomers derived from AG,[33] was further validated here against the synthetic Ara_5 acceptor. Importantly, under the instrumental conditions employed, the most sequence informative ions were the $^{1,4}X$ ions. This ion series is often sufficient to define the presence of

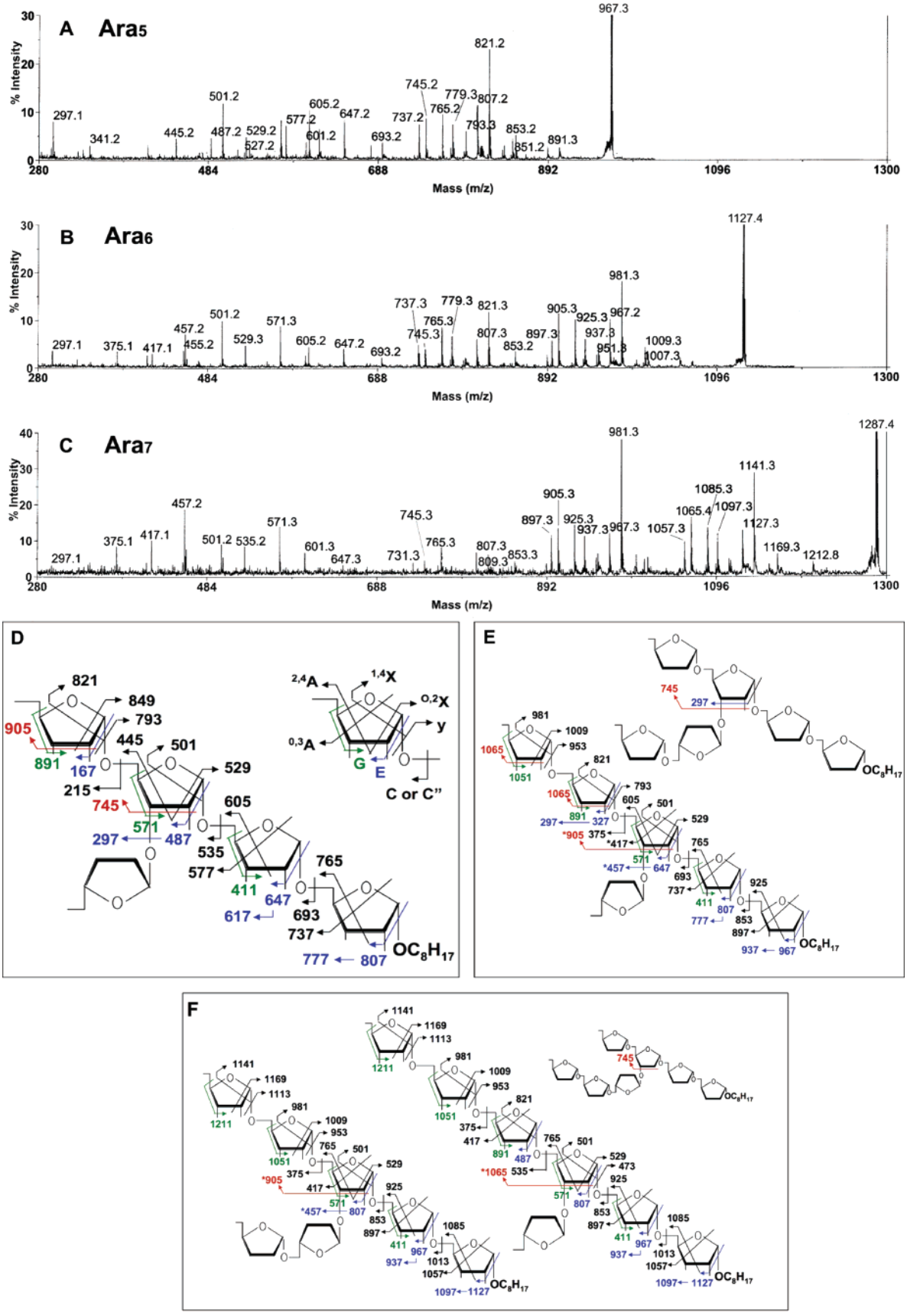


Figure 6. MS/MS analysis of per-O-methylated acceptor (A and D), biosynthetic Ara_6 (B and E), and Ara_7 (C and F). The origins of the fragmentation pattern (ref 33) and the adopted nomenclature are inserted in panel D. The E-30 u ion at *m*/*z* 457 clearly establishes that a structure with Ara_2 on the 5-arm and Ara_1 on the 3-arm exists. As expected, the $^{0,4}A$ ion at *m*/*z* 417 can be readily detected, whereas the elimination of an Ara residue from the C3 (or C2) position gave the F ion at *m*/*z* 905. However, the F ion corresponding to elimination of two Ara residues from C3 is also present at *m*/*z* 745, and thus the other isomeric structure with Ara_2 on the 3-arm and Ara_1 on the 5-arm is also implicated by the MS/MS data.

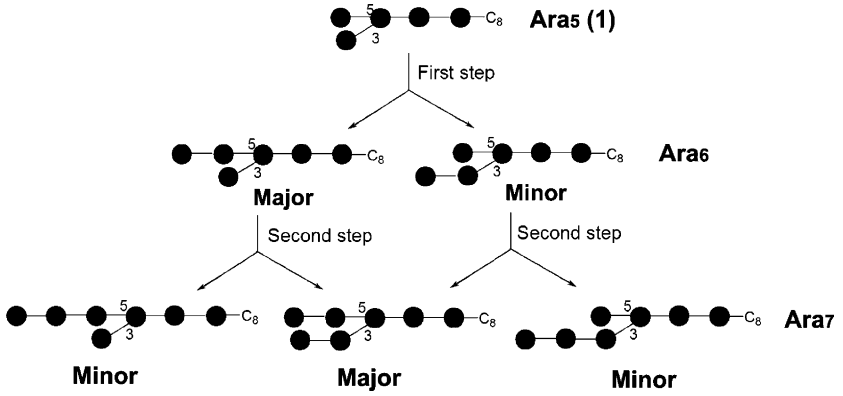


Figure 7. Putative biosynthetic steps involved in the conversion of Ara_5 (acceptor **1**) to Ara_7.

branching which is indicated by the absence or significant drop in intensity of one or more of the ions in series. Further confirmation and linkage assignment could then be derived from identifying other characteristic ion series, particularly the reducing end ^{O,2}X and G ions and the nonreducing end ^{O,3}A, ^{2,4}A, and E ions (Figure 6D).

For the synthetic Ara_5, the branching pattern is clearly defined (Figure 6, parts A and D). ^{1,4}X ions corresponding to loss of one (*m*/*z* 821) and three (*m*/*z* 501) Ara but not two (*m*/*z* 661) Ara residues could be detected. All other prominent ions produced are consistent with the expected linkages and cleavages. The previously observed but not validated E-30 u ions are also readily observed which could now be rationalized as deriving from concerted elimination of substituents at C1, C2, and C3. Thus, while the E ions at *m*/*z* 807 and 647 were accompanied by the minus 30 u ions at *m*/*z* 777 and 617, respectively, the corresponding E-30 u ion for *m*/*z* 487 was not found. Instead, *m*/*z* 297 was detected which vindicates the presence of a single Ara residue at the C3 position of the branched 3,5-Ara.

For the most abundant Ara_6 product (Figure 6, parts B and E), the conspicuous absence of the ^{1,4}X ion deriving from loss of three Ara residues (*m*/*z* 661), while the corresponding ones deriving from loss of one, two, and four Ara residues (*m*/*z* 981, 821, and 501, respectively) were readily detected in high abundance, is a clear indication that the third Ara residue from the nonreducing end is branched. As suggested by the endoarabinanase digestion data described above (Figure 4C), one arm is substituted by Araα1-5Araα1− and the other is a single Ara substitution. This branching pattern is supported by a full complement of fragment ions detected, as schematically depicted in Figure 6E, which also indicated the presence of both isomeric products.

For Ara_7 (Figure 6, parts C and F), although the ^{1,4}X ion corresponding to loss of three Ara residues is present at *m*/*z* 821, the significant drop in its intensity relative to that of *m*/*z* 981 (loss of two Ara) is in favor of an isomer with Ara_2 on both arms, accompanied by a lesser amount of an isomer with Ara_1 and Ara_3 on either arm. In agreement with this interpretation, an ^{O,3}A was detected at *m*/*z* 417 but not *m*/*z* 577, indicating that the isomer with Ara_2 and not Ara_3 attached to the 5-arm is the major form. Likewise, the C ion implicating an Ara_3 motif is barely detectable at *m*/*z* 535. The E ions corresponding to a stretch of Ara_3 and Ara_5 are present at *m*/*z* 487 and 807, respectively. Importantly, the latter was accompanied not by a minus 30 u ion or further loss of an Ara at *m*/*z* 617, but further loss of two Ara instead, to give a prominent ion at *m*/*z* 457. Collectively, these and other ions (as assigned and drawn in Figure 6F) indicate the presence of both isomers and that the one with a symmetrical distribution of Ara_2 on both arms dominates over the one with an asymmetrical distribution of Ara_1 and Ara_3 on either arm.

These cumulative results indicate that the enzyme responsible for addition of the Ara*f* residue on the branched acceptor **1** is distinct from the Araβ1→2Araα tranferase activity shown by the EmbA and EmbB proteins, or the AftA which is responsible for transferring Ara*f* residues on the galactan backbone, and it is presumably an α-(1→5)-Ara*f*, and this can also further chain extend subsequent to the addition of the first Ara*f* on the branch. The putative biosynthetic steps from Ara_5 to Ara_7 are indicated in Figure 7.

The crucial structural attributes such as substitution of terminal Ara*f* residues with mycolic acids in AG and mannose capping in LAM make the arabinan a valid drug target. Inhibition of the arabinan formation is expected to inhibit growth of the organism, and it is not found in Gram-negative or Gram-positive bacteria. Numerous attempts to interrupt the *emb* genes have repeatedly failed in *M. tuberculosis* suggesting that these genes are essential in the pathogenic organism. Moreover, enzymes responsible for the formation of DPA, the sole arabinose donor, are required for the growth of *M. tuberculosis*.[38] As the arabinosyltransferases do not exist in nature, direct homology search through genome/proteome database are unlikely to give any positive identification results. Consequently, in order to identify the enzymes involved in the arabinan assembly we designed complex oligoarabinoside acceptors of median length and used these in cell-free assays. An important criterion for designing an acceptor was that it would not be a substrate of recognition by the following implicated and experimentally proven arabinosyltransferases (Figure 8), Rv3792 (AftA), Rv3793 (EmbC), Rv3794 (EmbA), Rv3795 (EmbB), and Rv3805c (AftB). When used with the enzyme source from WTMsm, the acceptor **1** (analog of the precursor in Figure 8), described in this present study, yielded one major product. The product upon degradation with the *Cellulomas* arabinanase gave a distinct dimer (Ara_2) which could arise only if an α-Ara*f* residue had been transferred to the substrate. This was a deviation from our previous work,[31] where using the linear

(38) Sassetti, C. M.; Boyd, D. H.; Rubin, E. J. *Mol. Microbiol.* **2003**, *48*, 77−84.

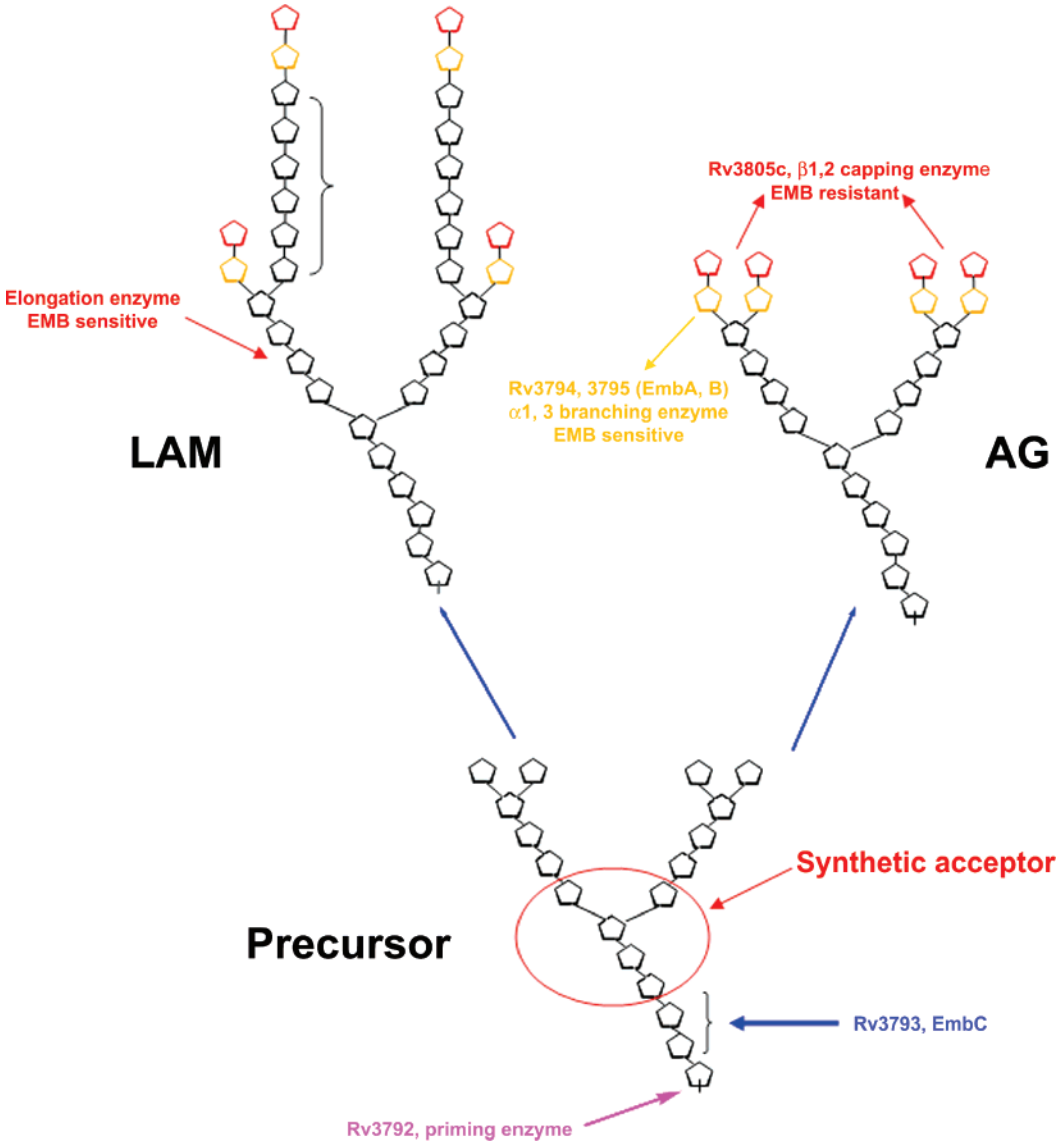


Figure 8. Experimentally validated arabinosyltransferases identified to date for assembly of arabinan of LAM or AG.

acceptor *β*-D-Ara*f*(1→2)-α-D-Ara*f*(1→5)-α-D-Ara*f*(1→5)-α-D-Ara*f* (1→5)-Ara*f*, a branched heptamer was obtained suggesting that the terminal Ara_6 in AG and LAM is formed by transferring *β*-D-Ara*f* to the 2-position of a nonreducing linear α-D-Ara*f*(1→5)-α-Ara*f* chain. On the contrary, when the acceptor is already branched, sequential addition of α-Ara*f* occurs without any chain termination with the function of this elongation enzyme (Figure 8). Our assumption was further corroborated by results obtained from the experiment that utilized cold pRpp. In this case, acceptor-dependent products were obtained whose molecular ions corresponded to Ara_{10}, showing five arabinose residues have been added. In LAM, it is now known that a linear extended motif exists and *β*-capping can occur only after the optimal length has been synthesized. In *M. smegmatis* this extended motif comprises eight Ara*f* after which *β*-capping occurs. The enzymatic products $Ara_{7-}Ara_{10}$ were obtained in extremely low abundance, and we believe that this could be due to the optimal K_m of the enzyme having been reached utilizing this particular acceptor and higher concentrations of the hexamer will be required for further extension. Future work will investigate the elongation of the arabinan and serve as a molecular probe for enzymatic studies.

Conclusions

A primary bottleneck in studying the biogenesis of D-arabinofuran is the intractability of information on genes encoding arabinosyltransferases. In this study, using a synthetic arabinose-based acceptor, we have been able to identify a novel arabinosyltransferase activity in mycobactertial membrane preparations that is dissimilar from the activities described in the literature. Our data shows that this activity is the internal α-D-Ara*f*(1→5)-transferase and is initiated when an α3,5 branch point is introduced in the acceptor. Heterogeneity in the products obtained from the enzymatic reactions is analogous to the heterogeneity present in the native arabinan structure in LAM. Thus, this strategy offers unlimited potential in identifying new arabinosyltransferase that seems to be logically involved in the assembly of a very complex structure and circumvents the problem of using the conventional way of identifying enzymes.

Acknowledgment. We thank Anita G. Amin for generously providing p[^{14}C]Rpp and Dr. Mike McNeil for helpful discussions. This work was supported by Grant AI 37139 from the National Institutes of Health to D.C.; a Taiwan NSC Grant 95-2311-B-001-031 to K.-H.K. High-energy CID MALDI MS/MS analyses were performed at the National Core Facilities for Proteomics located at the Institute of Biological Chemistry, Academia Sinica, supported by a Taiwan NSC Grant (95-3112-B-001-014).

Supporting Information Available: Complete ^{1}H, ^{13}C, and HSQC NMR spectra of compounds **1**−**12**, a graph showing protein concentration dependency of the enzymatic reaction, and three TLCs depicting ethambutol inhibition assay, size of major biosynthetic product, and the requirement of P60 particulate fraction in the cell-free assay. This material is available free of charge via the Internet at http://pubs.acs.org.

JA070330K