

# Chemoenzymatic elaboration of monosaccharides using engineered cytochrome P450<sub>BM3</sub> demethylases

Jared C. Lewis<sup>a</sup>, Sabine Bastian<sup>a</sup>, Clay S. Bennett<sup>b,1</sup>, Yu Fu<sup>b</sup>, Yuichi Mitsuda<sup>a,2</sup>, Mike M. Chen<sup>a</sup>, William A. Greenberg<sup>b</sup>, Chi-Huey Wong<sup>b,c</sup>, and Frances H. Arnold<sup>a,3</sup>

<sup>a</sup>Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125; <sup>b</sup>Department of Chemistry, The Scripps Research Institute, La Jolla, CA, 92037; and <sup>c</sup>Genomics Research Center, Academia Sinica, Nankang, Taipei 115, Taiwan

Contributed by Frances H. Arnold, August 10, 2009 (sent for review July 2, 2009)

**Polysaccharides comprise an extremely important class of biopolymers that play critical roles in a wide range of biological processes, but the synthesis of these compounds is challenging because of their complex structures. We have developed a chemoenzymatic method for regioselective deprotection of monosaccharide substrates using engineered *Bacillus megaterium* cytochrome P450 (P450<sub>BM3</sub>) demethylases that provides a highly efficient means to access valuable intermediates, which can be converted to a wide range of substituted monosaccharides and polysaccharides. Demethylases displaying high levels of regioselectivity toward a number of protected monosaccharides were identified using a combination of protein and substrate engineering, suggesting that this approach ultimately could be used in the synthesis of a wide range of substituted mono- and polysaccharides for studies in chemistry, biology, and medicine.**

biocatalysis | demethylation | evolution | P450 | polysaccharide

Complex carbohydrates comprise an extremely important class of biopolymers that play critical roles in a number of biological processes, including inflammation, cancer metastasis, and bacterial and viral infection (1). Unlike nucleic acids and proteins, oligosaccharides can be highly branched, and their biosynthesis proceeds in a non-template-directed fashion to produce complex mixtures of closely related structures that are difficult, if not impossible, to separate (2). Because subtle changes in the structures of these compounds can alter their properties dramatically (3), detailed studies of oligosaccharide-mediated events require the use of homogeneous synthetic material (4). Oligosaccharide synthesis traditionally has been accomplished through the installation and removal of various orthogonally reactive protecting groups to differentiate each position of a monosaccharide substrate, and a similar procedure is used to synthesize glycosides and substituted monosaccharides in which 1 hydroxyl group is used as a handle for the introduction of small molecules or other functional groups (5). Despite numerous advances in the synthesis of these materials during the past decade, their preparation remains an arduous task (6).

Researchers have sought to simplify the preparation of suitably protected monosaccharides with 1-pot synthetic procedures, but these approaches require considerable synthetic expertise and are limited in terms of substrate scope (7). Elimination of protecting groups altogether also has been demonstrated using enzymes to affect the glycosylation of various substrates (8). The utility of this approach could be expanded greatly by identifying or engineering additional enzymes (9), but the preparation of substituted monosaccharides lies beyond its scope. Chemoenzymatic regioselective deprotection of a globally protected monosaccharide is a potentially practical alternative to these approaches. This approach has been demonstrated using lipases to deprotect peracylated monosaccharides, but the range of products reported has been limited thus far to those accessible with known enzymes, and no attempt has been made to improve or alter the biocatalyst activity or regioselectivity (10). Although careful control of reaction conditions permits regioselective acyl

migration to produce other products (11), non-selective migration remains problematic, and the generality of the lipase method seems to be limited.

The generality and utility of a chemoenzymatic method for the synthesis of monosaccharide derivatives would be expanded greatly if each site of a given substrate could be deprotected using a unique catalyst. This approach would require not only a protecting group resistant to migration under the deprotection conditions but also a catalyst capable of removing this group with high levels of chemo- and regioselectivity. We hypothesized that cytochrome P450-catalyzed heteroatom demethylation might provide a solution to both these problems. Although chemical removal of methyl groups requires harsh and unselective conditions, this enzyme-catalyzed reaction proceeds via insertion of oxygen into a methyl C-H bond to generate a hemiacetal intermediate that spontaneously decomposes to release formaldehyde and an alcohol under ambient conditions (Fig. 1A) (12). One of our groups (F.H.A.) has used variants of *Bacillus megaterium* cytochrome P450<sub>BM3</sub> (BM3) to hydroxylate a wide range of non-natural substrates with high levels of regioselectivity (13). If this enzyme could be engineered to accept permethylated monosaccharides, the challenging task of regioselectively demethylating these substrates could be possible. Because BM3 is amenable to protein engineering, the activity, regioselectivity, and substrate scope of these reactions could be optimized using directed evolution (14) to generate BM3 demethylases for the production of any desired monosaccharide derivative.

## Results

We therefore screened libraries of BM3 variants on hand in our laboratory to determine the scope of their reactivity toward a number of common pentamethyl hexoses (Fig. 1C). A compilation library comprised of 96 BM3 variants previously engineered for a variety of applications was used to survey functional enzymes rapidly (15). Because many of these variants were related along an evolutionary lineage by mutation of 1–3 residues, we also used a targeted library of 1,024 BM3 variants having up to 10 mutations in the active site of the parent variant, 9–10A. Demethylase activity first was determined using a high-

Author contributions: J.C.L., W.A.G., C.-H.W., and F.H.A. designed research; J.C.L., S.B., C.S.B., Y.M., and M.M.C. performed research; J.C.L., C.S.B., and Y.F. contributed new reagents/analytic tools; J.C.L., S.B., Y.M., and M.M.C. analyzed data; and J.C.L. and F.H.A. wrote the paper.

The authors declare no conflict of interest.

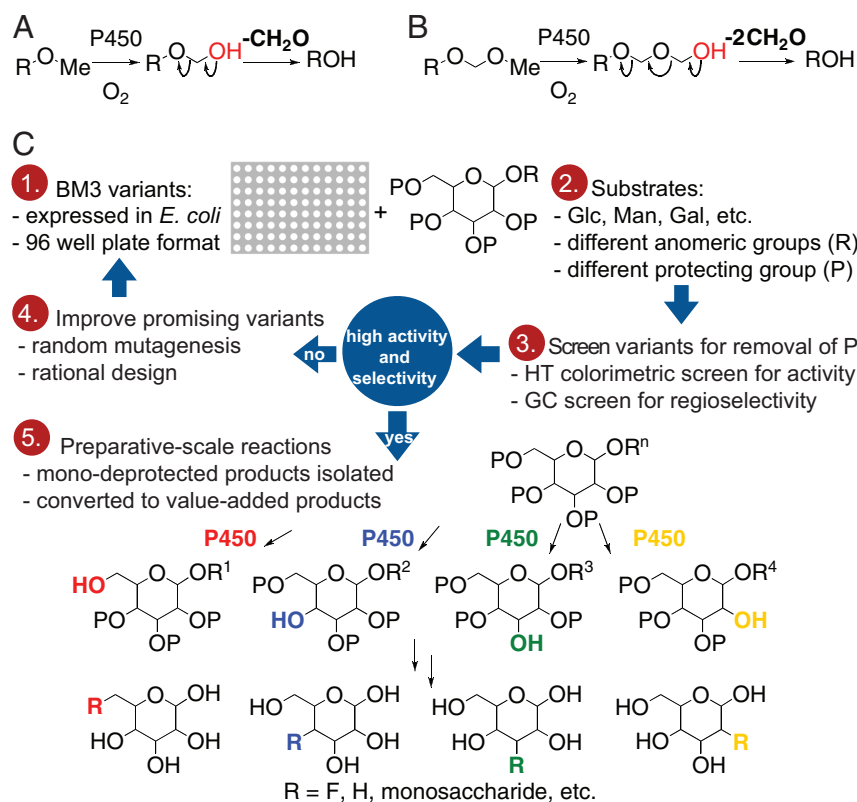
Data deposition: The atomic coordinates have been deposited in the Cambridge Structural Database, Cambridge Crystallographic Data Centre, Cambridge CB2 1EZ, United Kingdom (CSD reference no. 673244).

<sup>1</sup>Present address: Department of Chemistry, Pearson Chemistry Laboratory, Tufts University, Medford MA, 02155.

<sup>2</sup>Present address: Discovery Research Laboratories, Shionogi and Co. Ltd., 2-5-1 Mishima, Setu, Osaka 566-0022, Japan.

<sup>3</sup>To whom correspondence should be addressed. E-mail: frances@cheme.caltech.edu.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0908954106/DCSupplemental](http://www.pnas.org/cgi/content/full/0908954106/DCSupplemental).



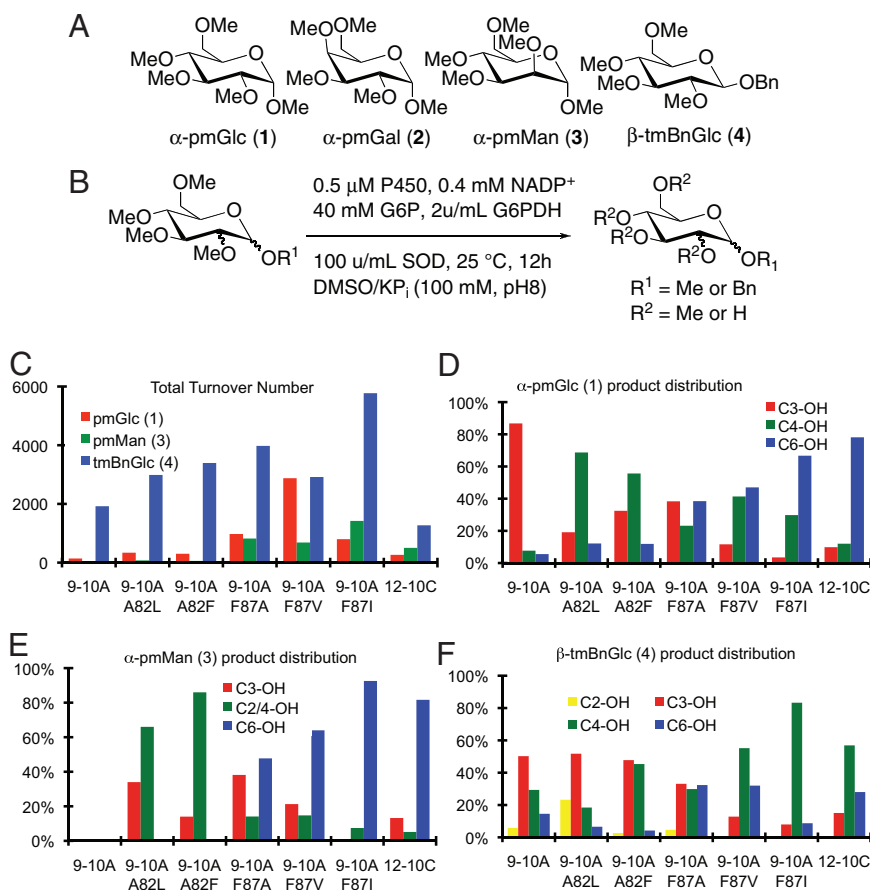
**Fig. 1.** (A) P450-catalyzed removal of methyl (Me) and (B) methoxymethyl (MOM) groups. (C) Method used to screen and optimize activity of BM3 demethylases. Libraries of BM3 variants were screened for activity on Me- or MOM-protected substrates. The activity of BM3 variants in cell lysate was assayed by colorimetric detection of formaldehyde ( $CH_2O$ ) released during the deprotection, and regioselectivity was determined by GC analysis of the reaction mixtures. Promising enzymes were optimized by directed evolution to provide enzymes suitable for preparative scale. Mono-demethylated products were converted to a variety of value-added products.

throughput (HT) colorimetric assay for formaldehyde released during the reactions (Fig. 1A; see [supporting information \(SI Appendix\)](#)). The most active enzymes then were re-expressed, purified, and used to catalyze the same reaction under controlled conditions to ensure the accuracy and precision of catalyst loadings. Finally, the reaction mixtures were analyzed by gas chromatography (GC) to establish the regioselectivity of demethylation. Promising variants were optimized by incorporation of random mutations and screening to provide total turnover numbers (TTN, *ca.* 1,000) and regioselectivities (> 80%) sufficient for practical gram-scale preparation of regioselectively deprotected monosaccharides. This approach also was used to identify variants capable of removing the methoxymethyl (MOM) group from MOM-protected monosaccharides to enable additional synthetic applications (Fig. 1B).

Initially, 4 substrates, 1,2,3,4,6-pentamethyl- $\alpha$ -glucopyranoside (**1**), 1,2,3,4,6-pentamethyl- $\alpha$ -galactopyranoside (**2**), 1,2,3,4,6-pentamethyl- $\alpha$ -mannopyranoside (**3**), and 1-benzyl-2,3,4,6-tetramethyl- $\beta$ -glucopyranoside (**4**) (Fig. 2A) were examined. The last of these compounds, **4**, was included to determine if a substrate-engineering approach (16) could lead to alternative regioselectivities, and authentic samples of each of the possible mono-demethylation products for **1–4** were synthesized to serve as GC standards for product identification. We first screened the compilation library for activity on these substrates (Fig. 2B). Pleasingly, we found that a number of enzymes in this small library were active on substrates **1**, **3**, and **4** in the HT assay, and GC analysis of these same reactions conducted with purified enzymes under optimized conditions (17) revealed that, except for C-1, nearly every position of the monosaccharide substrates investigated was accessible (Fig. 2 C–F). This broad range of

regioselectivities contrasts with the narrow scope of existing chemoenzymatic methods and suggested that BM3-catalyzed demethylation could provide a general approach to monosaccharide elaboration if the yields and regioselectivities observed for these enzymes could be improved using directed evolution. Furthermore, the significant differences in both activity and regioselectivity between the methyl- and benzyl-protected glucose substrates **3** and **4** indicated that varying the anomeric substituent could be used to identify suitable enzyme/substrate combinations without the need for protein engineering. Indeed, variant 9–10A F87I catalyzed the demethylation of both **3** and **4** with suitably high yield (> 1,000 TTN) and regioselectivity (> 80%) to warrant investigation of these reactions on a preparative scale, as discussed later.

Although these results were very promising, we were concerned about the lack of reactivity observed with galactose substrate **2**, especially given the frequency with which this monosaccharide is found in natural systems (1). To address this issue, we examined the activities of the compilation library enzymes on 4 additional tetramethyl galactose derivatives with a variety of anomeric substituents in the  $\beta$  configuration, including methyl (Me, **5**), benzyl (Bn, **6**), benzoyl (Bz, **7**), and acetyl (Ac, **8**). Once again, 9–10A F87I displayed high demethylation activity, this time on the 1- $\beta$ -benzyl galactose substrate **6**, and nearly quantitative conversion and high selectivity for the 4-position was confirmed by NMR spectroscopy. This result further established the utility of substrate engineering to achieve highly selective demethylation reactions and confirmed that derivatives of 3 of the most common hexoses were compatible with our enzymatic demethylation method.



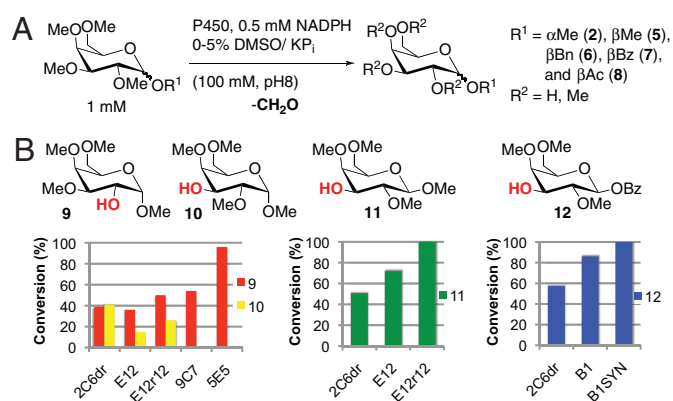
**Fig. 2.** (A) Structures of substrates used in initial screens. (B) Optimized reaction conditions for demethylation reactions. (C) TTN for enzymes (see *SI Appendix* for sequences) identified in HT screen on substrates 1, 3, and 4 and (D–F) product distributions (% of total product) for demethylation of substrates 1, 3, and 4 by BM3 variants determined by GC analysis of crude product mixtures.

To broaden the generality of the method further, we explored the activities of the active-site library enzymes on  $\alpha$ - and  $\beta$ -pentamethyl galactose substrates 2 and 5 because galactose derivatives had proven most resistant to demethylation by the small set of enzymes surveyed thus far (Fig. 3A). A number of hits were obtained in the HT assay, and 8 of the most active variants were purified and re-screened against substrates 2, 5, 6,

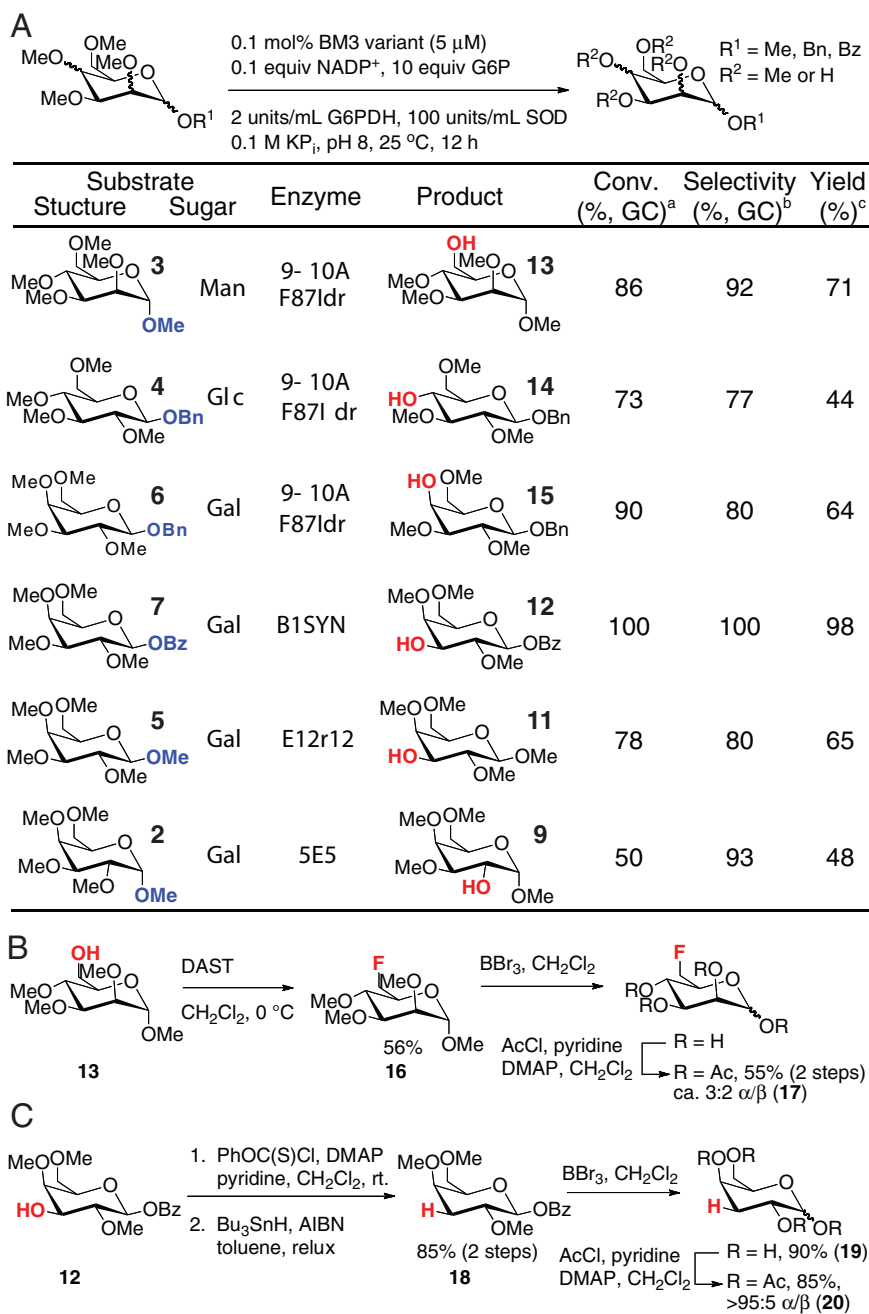
7, and 8. Colorimetric analysis of formaldehyde produced in these reactions confirmed the activity observed on 2 and 5 but also indicated that 7 is a compatible substrate (see *SI Appendix*). Six variants catalyzed the demethylation of these substrates with moderate efficiency (*ca.* 30% conversion based on formaldehyde concentration), indicating that active-site engineering could extend further the scope of the reaction.

GC analysis of the crude reaction mixtures obtained with 2 revealed that variant 2C6dr provided the highest conversion of substrate, producing a 1:1 mixture of O-2 (39%) and O-3 (41%) demethylation (compounds 9 and 10). Although the activity of 2C6dr on 5 and 7 was slightly lower under these reaction conditions, we observed nearly exclusive O-3 demethylation with these compounds to give 11 (50%) and 12 (57%), as confirmed by NMR spectroscopy and x-ray analysis, respectively. These results indicated that the remaining 2 secondary hydroxyl positions of galactose could be accessed, but both the conversion and regioselectivity of these transformations required improvement for practical synthetic applications.

A combination of random and site-directed mutagenesis of the gene for 2C6dr led to the identification of variant E12r12, which had improved efficiency on substrates 2 (50% increase in regioselectivity) and 5 (100% conversion) because of active-site mutation A180V (see *SI Appendix*). Variant 5E5 was obtained from an additional round of random mutagenesis of the gene for E12r12 and recombination of improved variants, and this enzyme provided high regioselectivity in the conversion of 2 to 9 (96%, single isomer) because of active-site mutation V87G. A similar engineering effort led to the generation of B1SYN, which



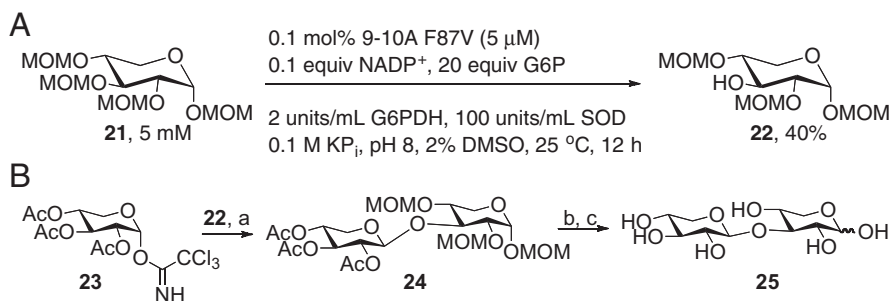
**Fig. 3.** (A) Reaction conditions used for colorimetric analysis. (B) Yields for conversion of 2 to 9 + 10, 5 to 11, and 7 to 12 determined by GC analysis of crude reaction mixtures using indicated BM3 variants. The 1 mM (0.1 mol%) BM3 variant was used except for the conversion of 7 to 12 where 0.5 mM (0.05 mol%) B1SYN were used.



**Fig. 4.** (A) Regioselective deprotection of substrates 11–15. <sup>a</sup>, Conversion of starting material determined by GC analysis of crude reaction mixture. <sup>b</sup>, Percent of desired product relative to additional products determined by GC analysis of crude reaction extracts. <sup>c</sup>, Isolated yield of pure product. <sup>d</sup>, 2 mM substrate used; 0.05 mol% B1SYN used. (B) Fluorination and global deprotection of 13. (C) Deoxygenation and global deprotection of 12.

had improved activity on 7 dramatically and enabled quantitative demethylation of this substrate even at 0.05 mol% enzyme loading. Fig. 3B shows the improvement in the conversion of 2 to a mixture of 9 and 10, of 5 to 11, and of 7 to 12 using the evolved BM3 variants. In each case, the activity or regioselectivity of 2C6dr was improved sufficiently by mutations identified from 1 or 2 rounds of error-prone PCR and screening to enable efficient demethylation of the 2- and 3-positions of suitable  $\beta$ -galactose derivatives. Given the range of activities and regioselectivities observed for BM3 variants identified from screening both the compilation and active-site libraries (Fig. 2 C–F), directed evolution should provide a rapid means for optimizing demethylases well beyond the examples shown here.

Next, the synthetic utility of the engineered demethylases was demonstrated by their application in preparative-scale reactions (Fig. 4A). The high conversions and regioselectivities observed previously in small-scale reactions translated well into the 100-mg scale investigated, and the desired products, 9–15, were isolated in high yield using 0.05–0.1 mol% enzyme loading. These monodemethylated compounds could be converted to a variety of biologically relevant products using established chemical transformations. Fluorination frequently is used to modify the pharmacokinetic properties of molecules for biological studies or for PET imaging (18). The Arnold group previously reported the tandem demethylation/fluorination of a variety of small molecules (15), and this procedure proved applicable for the fluorination of 13 to



**Fig. 5.** (A) Regiospecific removal of 3-MOM group from substrate **21**. (B) Synthesis of QS21-A disaccharide **25**. a, TMSOTf, 4 Å MS, 88%; b, NaOMe; c, 2 M HCl, 55% (2 steps)

provide **16**. The latter compound could be demethylated in quantitative yield and per-acetylated to provide **17** in 55% yield over 2 steps. Deoxy sugars also are important components of a variety of small-molecule therapeutics (19), and deoxygenation of **12** using Barton conditions proceeded smoothly to provide **18** in good yield (20). This product was demethylated in excellent yield (**19**) and was acetylated (**20**) to facilitate handling and characterization.

To augment the utility of our method further, we explored the regiospecific deprotection of MOM-protected monosaccharides in the context of a chemoenzymatic disaccharide synthesis (Fig. 1B). The ability to remove the MOM group is highly valuable for the synthesis of complex molecules, because the conditions required for its removal are much milder than those required for removal of methyl groups and therefore are compatible with sensitive functionality (21). A number of substrates with differing sugar scaffolds, anomeric substituents, and anomeric configurations were synthesized, and the compilation library was screened for activity on these substrates. High activity was observed on a number of MOM-protected pentoses without the need for any protein engineering, and GC analysis of reactions catalyzed by the most active variants revealed that 9-10A F87V catalyzed the removal of a single MOM group from α-xylose substrate **21**. Importantly, the product, **22**, could be isolated in moderate yield on a preparative scale, and removal of the 3-MOM group was confirmed by NMR spectroscopy (Fig. 5A). Compound **22** could be glycosylated with trichloroacetimidate donor **23** to give disaccharide **24** in high yield (22). Global deprotection of **24** under mild conditions provided disaccharide **25**, a component of the promising immune adjuvant, QS-21 (Fig. 5B) (23).

## Discussion

Regiospecific deprotection of monosaccharide substrates using engineered P450<sub>BM3</sub> demethylases provides a highly efficient means to access valuable intermediates that can be converted to a wide range of substituted monosaccharides and polysaccharides. Demethylases with activity on a number of monosaccharides and different positions on a given monosaccharide were identified using a combination of protein and substrate engineering, suggesting that this procedure ultimately could be used to functionalize regiospecifically any monosaccharide at any desired position. The ability of these enzymes (i.e., 9-10A F87V) to catalyze efficient demethylation of various substrates with high regiospecificity should greatly simplify the expansion of the method to additional substrates.

These results also highlight the efficiency with which BM3 variants can insert oxygen into the C-H bonds of methyl ethers to catalyze heteroatom demethylation under conditions far more mild than those associated with chemical methods. Furthermore, the ability of the demethylases to situate substrates in a particular orientation to favor hydroxylation of a particular methyl group over 3 or 4 others located within only a few angstroms indicates the exquisite control imparted by supramolecular interactions.

Unlike synthetic supramolecular catalysts (24), the activity and selectivity of BM3 variants can be tuned readily to suit a particular application. Such control enables efficient hydroxylation of C-H bonds lacking any significant stereoelectronic differentiation and augments the means by which synthetic chemists can target otherwise indistinguishable C-H bonds (25).

All these factors result in a procedure that greatly simplifies the synthesis of selectively deprotected monosaccharide derivatives compared with conventional procedures that require extensive synthetic manipulation. The extension of this technology to additional synthetically useful protecting groups, such as the MOM ether, renders the method compatible with sensitive glycosidic linkages and greatly expands the scope of the reaction. We anticipate that this chemoenzymatic approach could greatly simplify the synthesis of a wide range of polysaccharides for studies in chemistry, biology, and medicine.

## Materials and Methods

**Experimental Procedures and Characterization.** Complete experimental procedures and characterization can be found in the online *SI Appendix*.

**BM3 Expression and Purification.** For preparative bioconversions, P450 BM3 demethylases were used in purified form. Enzyme batches were prepared as follows. We inoculated 2 L TB<sub>amp</sub> with an overnight culture (100 mL, LB<sub>amp</sub>) of recombinant *E. coli* DH5α cells harboring a pCWoRI plasmid encoding for the P450 variant under the control of Plac promoter. At an OD<sub>600</sub> of 1.8 (ca. 3–4 h), the incubation temperature was reduced to 25 °C (30 min), and the cultures were induced by adding isopropyl-beta-D-thiogalactopyranoside to a final concentration of 0.1 mM. The cultures were allowed to continue at this temperature for another 24 h. After the cells were harvested by centrifugation (4 °C, 15 min, 3000 × g), the cell pellet was resuspended in 25 mM Tris-HCl buffer (pH 8.0), and cells were disrupted by sonication (4 × 1 min, 50% duty cycle). Cell debris was removed by centrifugation for 20 min at 4 °C and 20,000 × g; the resulting cell lysate was loaded onto a Q resin, and the column was washed with 3 column volumes of 25 mM Tris-HCl (pH 8.0), 150 mM NaCl. Bound protein was eluted with 25 mM Tris-HCl (pH 8.0), 340 mM NaCl and was concentrated using Millipore Centricon tubes. After buffer exchange with 100 mM KP<sub>i</sub> (pH 8.0), protein samples were frozen and stored at –80 °C. Protein concentration was determined in duplicate from CO-difference spectra as previously described. Yields typically ranged between 100 and 500 mg protein/L, depending on the variant.

**Preparative-Scale Bioconversions.** Potassium phosphate buffer (100 mM, pH 8) and a solution of the desired substrate in phosphate buffer or DMSO (1 equiv, 5 mM final concentration) was added to a 100 × 50 mm crystallizing dish (i.e., ChemGlass #CG-8276–100). A solution of NADP<sup>+</sup> (0.1 equiv, 0.5 mM final concentration), glucose-6-phosphate (10 equiv, 50 mM final concentration), and glucose-6-phosphate dehydrogenase (2 units/mL final reaction volume) was added. A solution of the appropriate BM3 variant (0.001 equiv, 5 mM final concentration) was added, the dish was covered loosely with aluminum foil, and the contents were stirred using a magnetic stir bar. The moderate stir rate was used to allow efficient O<sub>2</sub> transfer to the reaction mixture while avoiding foaming. The reaction progress could be monitored by analyzing aliquots of the reaction mixture as described for the medium-scale bioconversions. After 12 h, the pH of the reaction mixture was adjusted to 3–4, and the solution was

