Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry
Bioorganic & Medicinal Chemistry 14 (2006) 6940–6953

Optimized design and synthesis of chemical dimerizer substrates for detection of glycosynthase activity via chemical complementation

Haiyan Tao, Pamela Peralta-Yahya, Hening Lin[†] and Virginia W. Cornish[*]

Department of Chemistry, Columbia University, 3000 Broadway, New York, NY 10027, USA

Received 12 January 2006; revised 13 June 2006; accepted 19 June 2006
Available online 25 July 2006

Abstract—Glycosynthases catalyze the formation of a glycosidic bond between a glycosyl fluoride donor substrate and a glycosyl acceptor substrate with high yield, thus providing a valuable approach for the synthesis of carbohydrates and glycoconjugates. Chemical complementation can be used to link glycosynthase activity to the transcription of a reporter gene in vivo, providing a selection for the directed evolution of glycosynthase enzymes with improved properties. In this approach, glycosynthase activity is detected as covalent coupling between a small molecule disaccharide acceptor substrate and a small molecule disaccharide α-fluoro donor substrate. Here we report the optimized design and synthesis of these small molecule substrates. These optimized substrates are shown to give a robust, glycosynthase-dependent transcriptional read-out in the chemical complementation assay. The full synthesis and characterization of these substrates are reported for the first time. These optimized chemical dimerizer substrates should allow the potential of chemical complementation for the directed evolution of glycosynthases with diverse substrate specificities and improved properties to be fully realized.
© 2006 Elsevier Ltd. All rights reserved.

1. Introduction

Recently we demonstrated that chemical complementation can provide a high-throughput assay for the directed evolution of glycosynthase enzymes.[1] Glycosynthases are artificial enzymes derived from retaining glycosidases, which have the ability to catalyze the formation of a glycosidic linkage using a glycosyl fluoride donor.[2] Compared to glycosyltransferases or natural glycosidases, there are several advantages to the use of glycosynthases for carbohydrate synthesis.[3,4] The glycosyl fluoride donors are straightforward to synthesize compared to the nucleotide diphosphate donors required for natural glycosyltransferases. Glycosynthases generally give the product in higher yield than simple use of the glycosidase. Glycosynthases can tolerate a broader range of substrates than natural glycosyltransferases.[3,4] Directed evolution provides an obvious route to improve the activity of these glycosynthase variants and to expand their substrate specificity.[1,5,6] However, the application of directed evolution to this class of enzymes has been limited because the reaction product is not inherently screenable or selectable. The Withers group developed an on-plate endo-cellulase coupled screening assay for the directed evolution of the *Agrobacterium* sp. Abg:E358G glycosynthase. After combining mutations from two rounds of random mutagenesis, a variant was identified with a 27-fold improvement in activity and expanded substrate specificity.[5,6] This screen, however, depends on the identification of a coupling enzyme that will cleave only the product of the glycosynthase catalyzed reaction to release a chromogenic product, and hence may be difficult to extend to glycosynthases with different substrate specificities.[5,6] In addition, screens are limited to smaller library sizes than selections.

The chemical complementation assay developed by our laboratory can be used to provide a general selection strategy for evolving endo-glycosynthases.[1,7] Chemical complementation detects enzyme catalysis of bond formation or cleavage reactions based on covalent coupling of two small molecule ligands in vivo (Fig. 1). The heterodimeric small molecule reconstitutes a transcriptional activator, turning on the transcription of a

Keywords: Chemical complementation; Glycosynthase; Substrate synthesis; High-throughput assay.
* Corresponding author. Tel.: +1 212 854 5209; fax: +1 212 932 1289; e-mail: vc114@columbia.edu
† Present address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA.

0968-0896/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2006.06.034

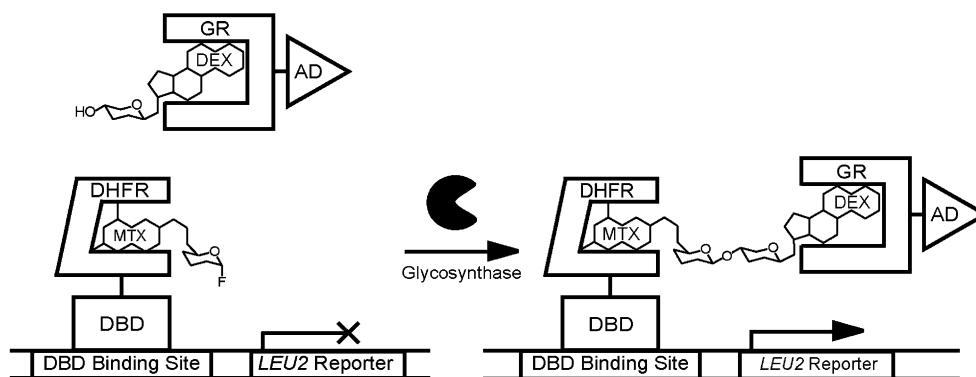


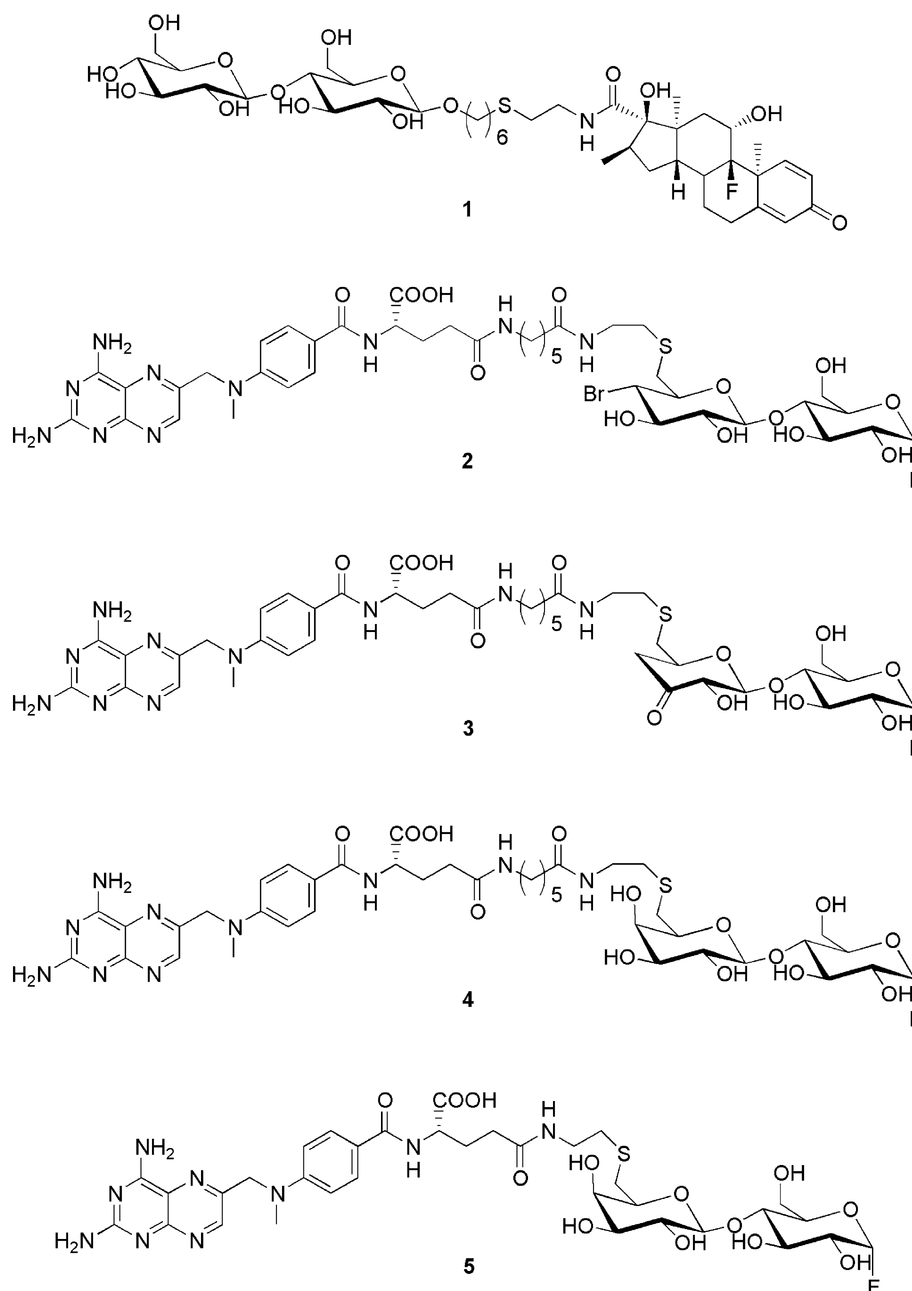
Figure 1. Chemical complementation provides a high-throughput assay for glycosynthase activity. Chemical complementation detects enzyme catalysis of bond formation or cleavage reactions based on covalent coupling of two small molecule ligands. The heterodimeric small molecule reconstitutes a transcriptional activator, turning on the transcription of a downstream reporter gene. Here, a dexamethasone (Dex)–methotrexate (Mtx) yeast three-hybrid system is used. Glycosynthase activity is detected as formation of a glycosidic linkage between a Mtx α-fluoride donor and a Dex acceptor and activation of a *LEU2* reporter gene.

downstream reporter gene. Bond formation is detected as activation of an essential reporter gene; bond cleavage, repression of a toxic reporter gene. To extend the chemical complementation assay to detect glycosynthase activity, a methotrexate-disaccharide-fluoride donor and a dexamethasone-disaccharide acceptor were synthesized, and bond formation was detected in a dexamethasone–methotrexate (Dex–Mtx) yeast three-hybrid system[1] (Fig. 1). In this system, a heterodimeric Dex–Mtx small molecule is synthesized by the glycosynthase, thus dimerizing the hormone-binding domain of the glucocorticoid receptor (GR), which binds to Dex, and dihydrofolate reductase (DHFR), which binds to Mtx. DHFR is fused to a DNA-binding domain (DBD), and GR is fused to a transcription activation domain (AD), such that Dex–Mtx effectively reconstitutes the transcriptional activator (DBD-AD) and increases transcription of a downstream reporter gene. The reporter gene is *LEU2*, which allows for a growth selection in the absence of leucine. Using the *LEU2* selection, we demonstrated that chemical complementation can be used to read-out glycosynthase activity. In addition, a Cel7B:E197S variant with a five-fold increase in glycosynthase activity was isolated from a Glu197 saturation library.[1]

As shown in Scheme 1, we designed the dexamethasone–cellobiose (Dex–Cel) glycosyl acceptor **1** and methotrexate-lactose-fluoride (Mtx-Lac-F) glycosyl donor **4** as the glycosynthase substrates. However, as reported in our first publication, instead of the desired substrate Mtx-Lac-F **4**, we obtained methotrexate-cellobiose-fluoride (Mtx-Cel-F) **2**. The activity of a known glycosynthase, *Humicola insolens* Cel7B:E197A, was detected and a directed evolution experiment was carried out using the Dex-Cel **1** and Mtx-Cel-F **2** substrates.[1] Although substrates Dex-Cel **1** and Mtx-Cel-F **2** worked in the chemical complementation assay, the synthesis of Mtx-Cel-F **2** suffered from poor yields. A side elimination reaction occurred at the final deprotection step in the synthesis of **2**, resulting in the by-product **3** and significantly affecting the overall yield in which **2** could be obtained. Thus, here we report an improved design and synthesis of two methotrexate-lactose-fluoride (Mtx-Lac-F) donor substrates, **4** and **5**, and the first full synthesis and characterization of both the Dex-Cel and Mtx-Lac-F substrates. This improved design and full characterization of the Dex-Cel and Mtx-Lac-F substrates should facilitate further applications of chemical complementation for the directed evolution of glycosynthase enzymes.

1.1. Design and synthesis of the glycosynthase substrates

The glycosynthase substrates must incorporate the Dex and Mtx ligands for use in the chemical complementation assay, yet still be efficient substrates for the Cel7B:E197A glycosynthase variant.[1,7] The structures of the substrates designed for use in this study are shown in Scheme 1. The design was based on the known substrate specificities of the *H. insolens* Cel7B glycosidase and Cel7B:E197A glycosynthase.[8,9] Both kinetic characterization and high-resolution structures of this and related endo-glycosidases suggest that there are five subsites in the active site that accommodate five glucose units, and four of the five subsites (−2, −1, +1, and +2) contribute most of the binding energy.[10,11] In addition, previous studies demonstrated that with lactose fluoride as the donor substrate the acceptor subsites of Cel7B:E197A may accommodate both mono- and disaccharide acceptors and that disaccharides are better acceptors than monosaccharides for the transglycosylation reaction.[8] We therefore decided to use two disaccharide compounds as the substrates for the Cel7B:E197A glycosynthase. Dex–Mtx has proven to be an efficient CID for reconstitution of a transcriptional activator in the corresponding yeast three-hybrid system.[12] In addition, it has been shown that as long as the linker is sufficiently long, the linker has little effect on the activity of the Dex–Mtx CID in the yeast three-hybrid assay.[13] Therefore, we used cellobiose with Dex attached at the anomeric position as the glycosyl acceptor substrate, shown as **1**; and lactosyl fluoride with Mtx attached at the 6′ position of the galactose unit as the glycosyl donor substrates, shown as **4** and **5**. Lactose was chosen to avoid the self-polymerization reaction of the cellobiose donor. Two Mtx-Lac-F donors, **4** and **5**, were tested. The Mtx-Lac-F donor **5** has a shorter

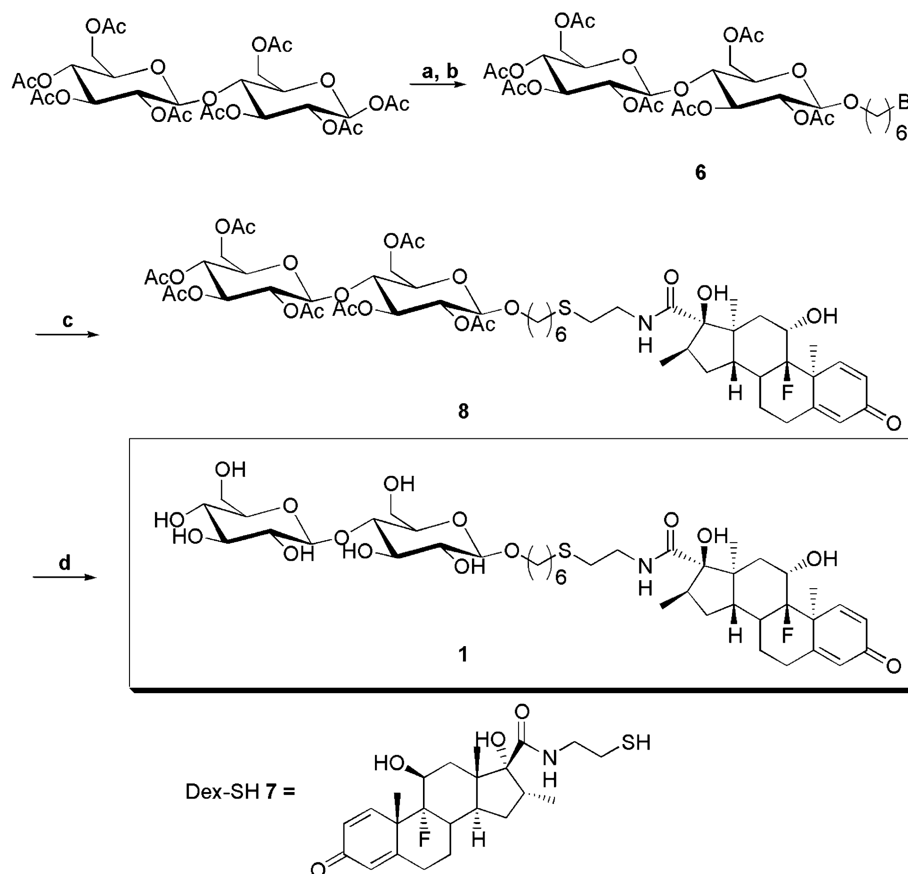


Scheme 1. Structure of the glycosynthase substrates used for the chemical complementation assay.

linker between the Mtx and Lac portions of the substrate and provides a more efficient synthesis. However, based on previous studies on the effect of linker length, we had some concern that the shorter linker of **5** would impair its activity in the yeast three-hybrid assay.[13]

Dex-Cel **1** is essentially as originally reported.[1] Here we report the full synthesis and characterization of this substrate. As shown in Scheme 2, the synthesis of Dex-Cel acceptor **1** was started from the commercially available cellobiose octa-acetate. The treatment of cellobiose octa-acetate with 30% hydrogen bromide in acetic acid for 1 h afforded cellobiosyl-bromide. Then the glycosidation reaction between cellobiosyl-bromide and 6-bromo-1-hexanol was carried out using both mercury cyanide and mercury bromide as activators to give **6** in 26% yield for two steps. Several standard activators were tested in an effort to improve the yield for this reaction, but neither silver triflate nor silver zeolite gave the desired product.[14] The direct coupling of Dex to the glycosyl bromide using either mercury or silver salts as the activator also did not yield any desired product. Therefore, Dex was installed to **6** using the Dex-thiol derivative **7**[13] as the nucleophile and gave **8** in 69% yield. Finally, global deprotection with sodium methoxide removed all acetyl protecting groups, and the Dex-Cel acceptor **1** was prepared in four steps in 18% overall yield.

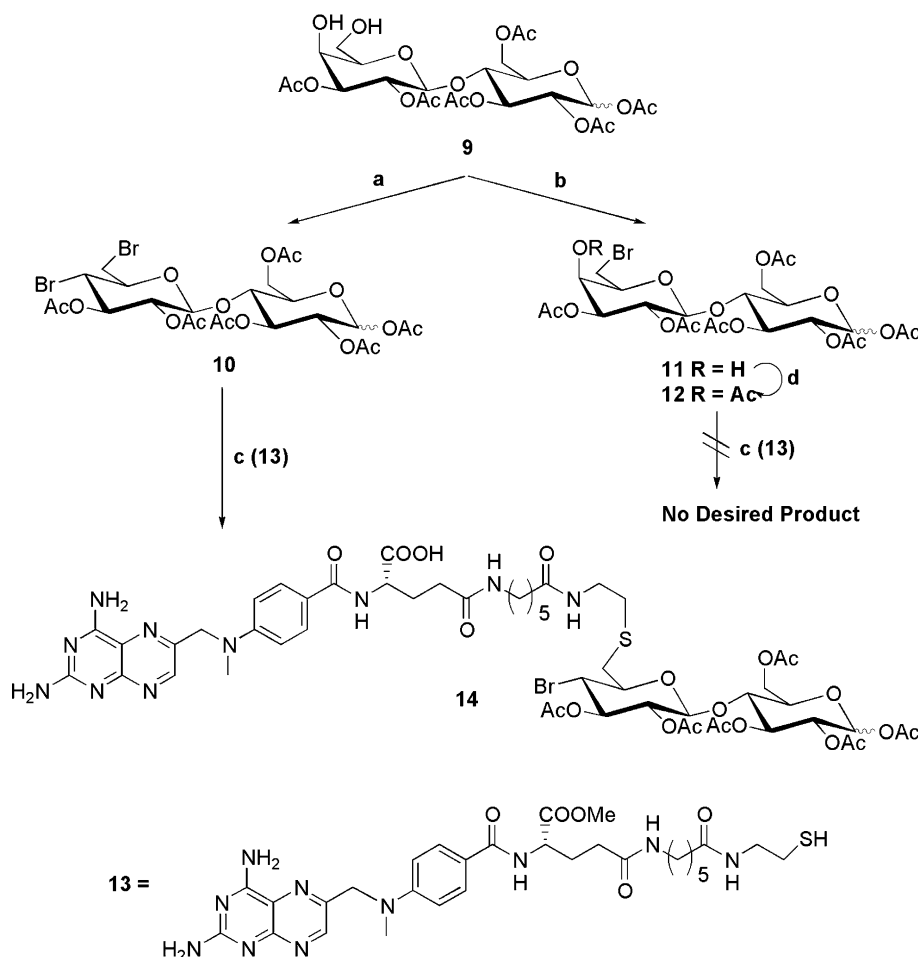
Although we intended Mtx-Lac-F **4** as the donor substrate for the chemical complementation assay, in our original publication the chemical complementation



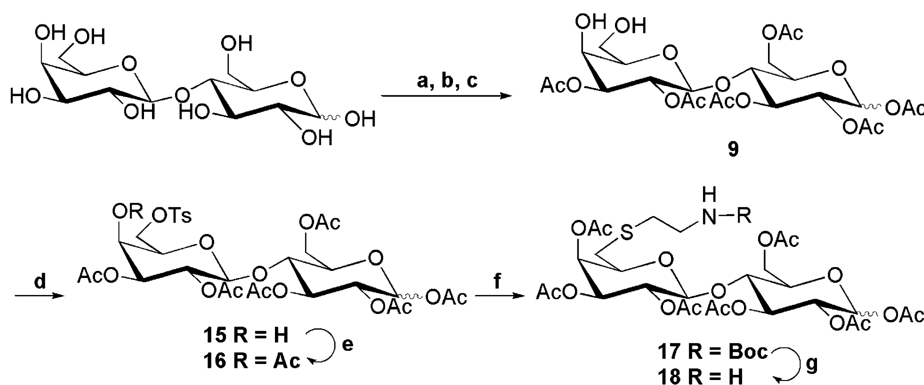
Scheme 2. Synthesis of Dex–Cel **1**. Reagents and condition: (a) 30% HBr in HOAc, 1 h; (b) $HO(CH_2)_6Br$, $Hg(CN)_2$, $HgBr_2$, CH_2Cl_2, 26% for two steps; (c) Dex-SH **7**, MeOH, DIEA, 69%; (d) MeONa, MeOH, 100%.

assay was carried out with Mtx-Cel-F donor **2** instead.[1] The synthesis of Mtx-Cel-F **2** suffered both from a side reaction in the last deprotection step and poor yield, thus the synthesis of the donor substrate had to be modified. As shown in Scheme 3, in our first effort to synthesize Mtx-Lac-F **4**, a disulfide diamine linker was installed to the γ-carboxylic acid group of the Mtx, and then its reduction product, the Mtx-thiol compound **13**, was attempted for coupling with lactose bromoderivative **11** or **12** via a substitution reaction. However, no product was observed using either *N*,*N*-diisopropylethylamine (DIEA) or sodium methoxide as the base. In contrast, Mtx-thiol **13** reacted efficiently with dibromo-substituted lactose derivative**10** when sodium methoxide was used as the base. The difference in reactivity between compounds **11** and **12** compared with **10** can be explained by nonbonding interactions with the 4′ position in the transition state of the S_N2 reaction.[15] The nonbonding interaction from the axial substituent of **11** or **12** is greater than that from the equatorial bromine of **10**. Since the product **14** led to the final product Mtx-Cel-F **2**, which is also a suitable substrate for glycosynthase Cel7B:E197A, we continued the synthesis from compound **14**. However, the compound Mtx-Cel-F **2** was difficult to obtain in pure form as an elimination side-reaction occurred during the last deprotection step. During the final deprotection step, a major by-product, which we tentatively assigned as **3** (Scheme 1), arose. Not surprisingly given their similarity in overall chemical structure, it proved difficult to purify Mtx-Cel-F **2** from by-product **3**. Using reverse-phase high performance liquid chromatography (HPLC), Mtx-Cel-F **2** could be purified. The total yield for the synthesis of compound **2**, however, was significantly lowered by this purification step (<10% for the last step). Mtx-Cel-F **2** was synthesized from two components in eight liner steps in 0.6% overall yield (counted from glutamic acid α-methyl-γ-*tert*-butyl ester).

Thus, as shown in Scheme 4, we came back to our original design to synthesize Mtx-Lac-F **4** as the donor substrate. To carry out the desired substitution reaction, a better leaving group, *p*-toluenesulfonate, was introduced to the 6′ position of lactose derivative **9**, and a *N*-*tert*-butoxycarbonyl-2-aminoethanethiol was used to attack the 6′ position. In addition, a bulky base, potassium *tert*-butoxide, was used as the base to catalyze the substitution reaction. As the result, the lactose derivative **18** was prepared from lactose in 8.9% yield. First, the 4′ and 6′ hydroxyl groups of lactose were first selectively protected as the 4-methoxybenzylidene acetal. Then the remaining hydroxyl groups of lactose were protected using acetic anhydride and pyridine. Hydrolysis of the 4-methoxybenzylidene acetal using 50% trifluoroacetic acid (TFA) in dichloromethane gave **9** in 39% yield. Compound **9** was a mixture of the α- and β-anomers, and the ratio of the α to β anomer product varied. Recrystallization of **9** from ethyl acetate and ether gave



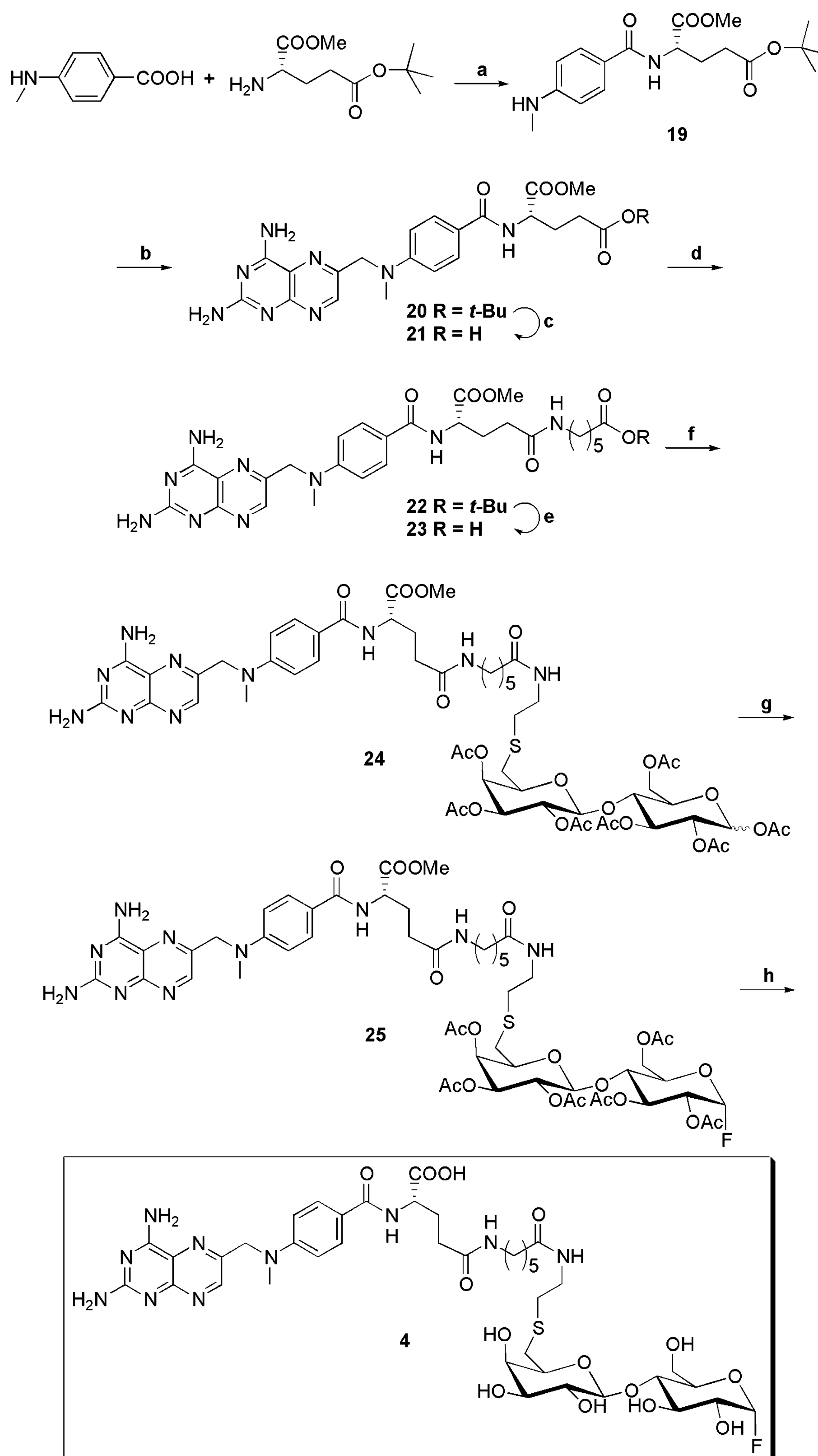
Scheme 3. Synthesis route toward Mtx-Cel-F **2**. Reagents: (a) 4 equiv PPh_3, 2 equiv CBr_4, pyridine; (b) 1 equiv PPh_3, 1 equiv CBr_4, pyridine; (c) 1—NaOMe, DMF, 2—Ac_2O, TEA, DMF; (d) pyridine, Ac_2O.



Scheme 4. Synthesis of lactose derivative **18**. Reagents and conditions: (a) 1-(dimethoxymethyl)-4-methoxy-bezene, *p*-toluenesulfonic acid, DMF, 50 °C; (b) pyridine, Ac_2O; (c) TFA, CH_2Cl_2, 39% for three steps; (d) *p*-tosyl chloride, DABCO, CH_2Cl_2, 65%; (e) pyridine, Ac_2O, 100%; (f) 1—potassium *tert*-butoxide, 2-(Boc-amino) ethanethiol, THF, 2—pyridine, Ac_2O; (g) TFA, CH_2Cl_2, 35% for (f) and (g).

the pure α anomer and was used for characterization. The mixture, however, could be used for the synthesis of the final product. The primary 6′ hydroxyl group of **9** was then selectively activated using one equivalent of *p*-toluenesulfonyl chloride and 1,4-diazabicyclo[2.2.2]octane (DABCO) to give **15**, followed by protection of the 4′ hydroxyl group of **15** using acetic anhydride and pyridine to give **16**. The toluenesulfonate group of **16** was then substituted by *N*-*tert*-butoxycarbonyl-2-aminoethanethiol to give **17**. Deprotection of the Boc group of **17** using TFA gave the lactose derivative **18**.

As shown in Scheme 5, α-methyl protected Mtx **21** was prepared essentially as original reported,[12,16–18] from commercially available glutamic acid α-methyl-γ-*tert*-butyl ester and 4-(methylamino) benzoic acid. *N*,*N*′-dicyclohexylcarbodiimide (DCC) was used as



Scheme 5. Synthesis of Mtx-Lac-F **4**. Reagents: (a) DCC, DIEA, CH_2Cl_2, 53%; (b) bromomethyl 2,4-pteridinediamine HBr–isopropanol complex, DMAc, 75%; (c) TFA, CH_2Cl_2, 76%; (d) PyBOP, *tert*-butyl 6-aminohexanoate, DIEA, CH_2Cl_2, DMF, 65%; (e) TFA, CH_2Cl_2, 93%; (f) PyBOP, CH_2Cl_2, DMF, DIEA, compound **18**, 60%; (g) HF in pyridine, 62%; (h) LiOH, MeOH, H_2O, 55%.

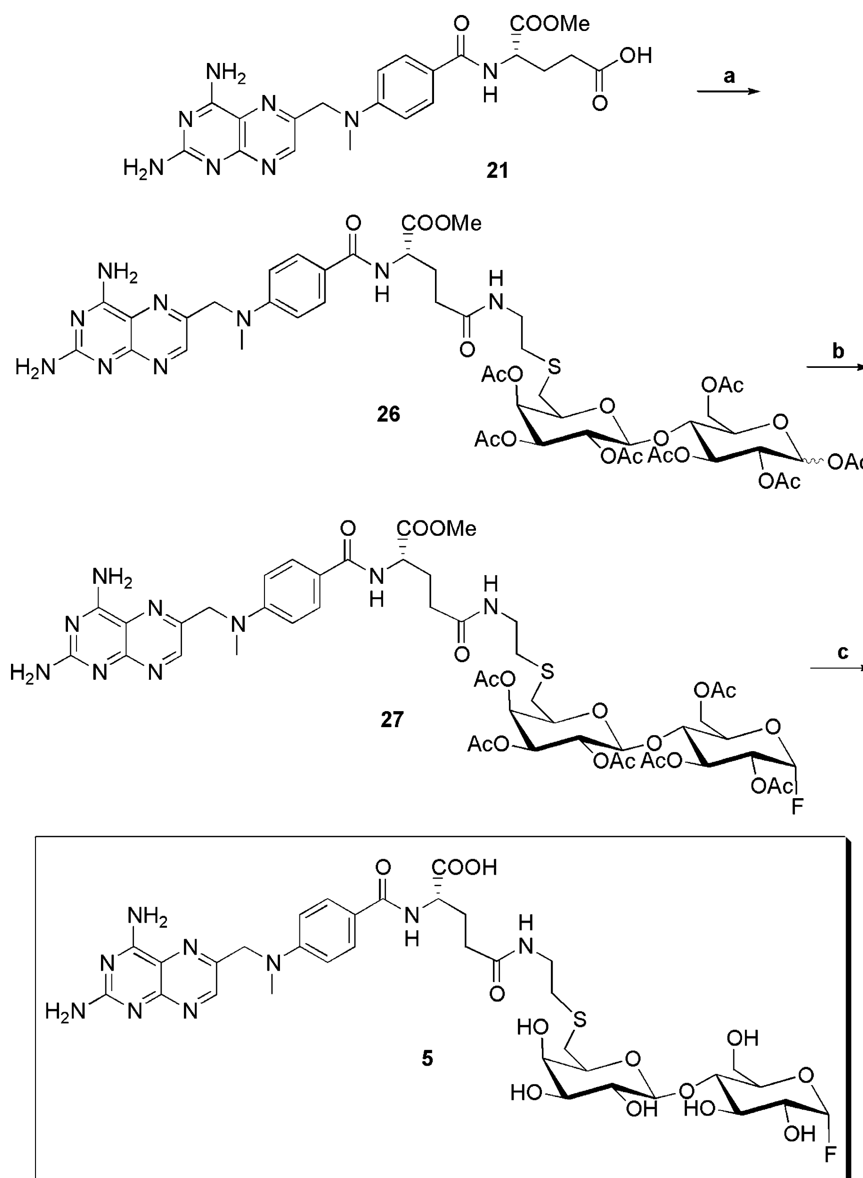
the coupling reagent in the synthesis of **19** to avoid the coupling reaction between the carboxylic acid group of glutamic acid α-methyl-γ-*tert*-butyl ester with the secondary amine of 4-methylamino benzoic acid.[16] Then α-methyl-γ-*tert*-butyl protected Mtx **20** was prepared by a substitution reaction of **19** to 6-bro-

momethyl-2,4-pteridinediamine. Deprotection of the *tert*-butyl protecting group of **20** with 50% TFA in dichloromethane gave **21** in 30% yield for three steps. The coupling reaction of compound **21** and *t*-butyl 6-aminohexanoate using (benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyBOP) as the coupling reagent, and the following deprotection reaction using TFA resulted in compound **23** in 60% yield for two steps. The compound **24** was then synthesized from **18** and **23** in 60% yield using PyBOP as the coupling reagent. The anomeric *O*-acetyl of **24** was converted to the α-fluoride compound **25** in 62% yield using a hydrogen fluoride–pyridine complex. Global deprotection of **25** using lithium hydroxide gave the desired Mtx-Lac-F donor substrate **4**. Mtx-Lac-F **4** was synthesized from two components in 10 linear steps in 1.8% overall yield (counted from lactose).

Similar to the synthesis route of Mtx-Lac-F **4**, the α-methyl protected Mtx **21** was coupled to the lactose derivative **18** in 49% yield using PyBOP as the coupling reagent (Scheme 6). Then the anomeric *O*-acetyl of **26** was converted to the α-fluoride using a hydrogen fluoride–pyridine complex, yielding compound **27** in 52% yield. Global deprotection of **27** using lithium hydroxide gave the desired Mtx-Lac-F donor substrate **5**. Mtx-Lac-F **5** was synthesized from two components in 10 linear steps in 0.9% overall yield (counted from lactose).

1.2. Chemical dimerizer substrate activities in the chemical complementation assay

Having synthesized the chemical dimerizer substrates Dex-Cel **1** and Mtx-Lac-F **4** and **5**, we next determined whether they were active in the chemical complementation assay. If active, the glycosynthase enzyme should



Scheme 6. Synthesis of Mtx-Lac-F **5**. Reagents: (a) PyBOP, CH_2Cl_2, DMF, DIEA, compound **18**, 49%; (b) HF in pyridine, 52%; (c) LiOH, MeOH, H_2O, 40%.

be able to couple the two substrates, synthesizing Dex-Cel-Lac-Mtx and thus reconstituting the transcriptional activator from the DBD-GR and AD-DHFR fusion proteins, here activating transcription of a *LEU2* reporter gene (Fig. 1), providing a growth advantage to yeast three-hybrid cells in the absence of leucine.[1] The efficacy of the different chemical dimerizer substrates is determined by carrying out a *LEU2* mock selection in the yeast selection strain V1019Y,[19] in which LexA is used as the DBD, and B42 as the AD. Both the DBD-DHFR and AD-GR fusion proteins are expressed from a GAL1 promoter. The DBD-DHFR gene is integrated into the chromosome at the *ade4* locus, and the AD-GR gene is on a 2μ plasmid. The *LEU2* gene is integrated into the chromosome under the control of six tandem LexA operators.[1,20] In the *LEU2* mock selection assay, a mixture of plasmids encoding a 1:10 ratio of the Cel7B:E197A glycosynthase to the Cel7B glycosidase (inactive control) variant was transformed en masse into the yeast three-hybrid strain V1019Y. The resulting transformants were incubated in selective media lacking the appropriate auxotrophs and leucine, and containing the donor and/or acceptor substrates (Fig. 2) and grown at 30 °C. After 7 days of growth, the cells were transferred to non-selective media and harvested the next day. Then the plasmids were extracted from the cells, and the genes encoding Cel7B and Cel7B:E197A were analyzed on a DNA gel following restriction digestion. The Cel7B:E197A glycosynthase gene was engineered to obtain a unique *Nco*I site so that it could be readily distinguished from the Cel7B glycosidase gene. Since total plasmid DNA is analyzed in the mock selection, effectively a very large number of samples are analyzed in this assay. Further, because small differences in activity are amplified during the selection process, this assay is very sensitive.

As shown in Figure 2, the plasmids encoding the glycosynthase gene were enriched when the acceptor substrate Dex-Cel **1** was incubated with any of the three donor substrates **2**, **4**, or **5**. Specifically, when combined with the acceptor substrate Dex-Cel **1**, compound **2** conferred the highest (17-fold) enrichment of glycosynthase; compound **4** conferred a 7-fold enrichment of glycosynthase; and compound **5** conferred a 5-fold enrichment of glycosynthase. In addition, all three donor substrates had background activity, conferring glycosynthase enrichment even when used without Dex-Cel **1**. Mtx-Cel-F **2** gave highest background with 5-fold of enrichment; Mtx-Lac-F **4** gave a moderate background with 3-fold of enrichment; and Mtx-Lac-F **5** gave the lowest background with 1.5-fold of enrichment. As we speculated in our original publication, our hypothesis is that this background results from covalent modification of Mtx-Lac-F with a molecule endogenous to the cell that either acts as a transcriptional activator on its own or recruits other molecules with this function. By comparison, Dex-Cel **1** alone did not show glycosynthase enrichment. One reasonable explanation for the higher activity of Mtx-Cel-F **2** compared to the other two donor substrates is that it is a closer structural mimic of the natural cellulose substrate, with an equatorial bromine at the 4′ position as opposed to the axial hydroxyl group at the 4′ position of lactose, and hence a more efficient substrate for the Cel7B:E197A glycosynthase. The mock selection assay data establish unambiguously that compounds **1**, **4**, and **5** are suitable substrates for detecting glycosynthase activity in the chemical complementation selection system. If the background fold-enrichment with donor substrate alone were subtracted from the fold-enrichment with donor and acceptor substrates, donors **4** and **5** would be similar in activity. However, these experiments cannot be used to estimate the amount of fold-enrichment in the presence of **1** + **4** or **1** + **5** that is due to **4** or **5**, respectively, alone. Thus, we can only conclude that **4** is as or more active than **5**. The relative ease of synthesis and biological activity of these substrates set the stage for now using chemical complementation for the directed evolution of glycosynthase catalysts with a range of substrate specificities.

2. Conclusion

In conclusion, we report the full design, synthesis, and activity of optimized chemical dimerizer substrates for the directed evolution of glycosynthase enzymes via

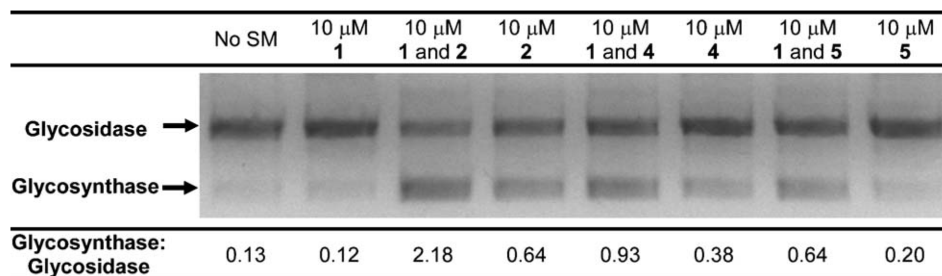


Figure 2. Dex–Cel glycosyl acceptor **1** and Mtx-Lac-F glycosyl donors **2**, **4**, and **5** are suitable substrates for detecting glycosynthase activity in the chemical complementation assay. A mixture of plasmids encoding a 1:10 ratio of the Cel7B:E197A glycosynthase to the Cel7B glycosidase (inactive control) was transformed en masse into the yeast three-hybrid strain V1019Y. The resulting transformants were incubated in selective media lacking the appropriate auxotrophs and leucine, and containing the donor and/or acceptor substrates as indicated and grown at 30 °C. After 7 days of growth, the cells were transferred to non-selective media and harvested the next day. Then the plasmids were extracted from the cells, and the genes encoding Cel7B and Cel7B:E197A were analyzed on a DNA gel following restriction digestion. The Cel7B:E197A glycosynthase gene was engineered to obtain a unique *Nco*I site so that it could be readily distinguished from the Cel7B glycosidase gene. The picture was taken using a Canon SD400 digital camera, and the intensity of DNA bands was quantified using ImageQuant™ (molecule dynamics) without calibration.

chemical complementation. The new synthesis route of the glycosyl donor substrates Mtx-Lac-F **4** and **5** solved a side elimination reaction that occurred in the final step of the synthesis of glycosyl donor substrate Mtx-Cel-F **2**, thus allowing us to prepare both substrates on a large scale for directed evolution experiments. The establishment of an efficient synthetic route and the full characterization of the Dex-Cel **1** and Mtx-Lac-F **4** and **5** chemical dimerizer substrates provide not only efficient substrates for the directed evolution of improved glycosynthase variants with cellobiose substrate specificity, but also set the stage for the directed evolution of glycosynthase enzymes with a range of substrate specificities.

3. Experimental

3.1. General methods

Unless otherwise noted reagents were purchased from Aldrich and used without further purification. Glutamic acid α-methyl ester γ-*tert*-butyl ester hydrochloric acid was purchased from Advanced Chemtech. The dexamethasone derivative **7** and the 1:1 2-propanol complex with 6-bromomethyl 2,4-pteridinediamine monohydrobromide were synthesized as described.[12,13] Mtx-Cel-F **2**, Mtx-Lac-F **4** and **5** were purified on a Waters delta 600 HPLC equipped with a Vydac C18 reverse-phase column using acetonitrile gradients. Nuclear magnetic resonance (NMR) spectra were recorded on a Bruker 300MHz, 400MHz or 500MHz Fourier transform NMR spectrometer. High resolution mass spectra (HRMS) were recorded on a JMS-HX110A mass spectrometer. Low resolution MS were recorded on a JMS-LC mate mass spectrometer. Standard protocols for molecular biology and yeast genetics were used. *Taq* polymerase was purchased from Promega (Madison, WI). The dNTPs used in the polymerase chain reaction (PCR) were purchased from Amersham Biosciences. Oligonucleotides were purchased from Invitrogen. Restriction enzyme *Nco*I was purchased from New England Biolabs (Beverly, MA). The transformation of yeast was carried out by electroporation using a Bio-Rad genepulser® Xcell. Yeast plasmids were prepared using EZNA kit (Omega Bio-Tek). Ultraviolet–visible measurements were taken using a Molecular Devices Spectramax 384. The picture of DNA gel was taken using a Canon SD400 digital camera and was analyzed using ImageQuant™ software (Molecule Dynamics). The yeast strains and the plasmids used in this study are listed in Table 1. The compound **2** used for the mock selection assay contains ca. 20% by-product **3** (estimate based on ^{1}H NMR).

3.2. Synthesis of 6

To D-cellobiose octaacetate (2.00 g, 5.14 mmol) in a 50 mL round-bottomed flask was added 30% HBr in HOAc (10 mL). The reaction was protected from light and stirred at rt for 1 h. The reaction mixture was diluted with CH_2Cl_2 (60 mL) and poured into ice/water (60 mL) in a separation funnel. The organic phase was collected and washed with water (1× 50 mL) and satd aq $NaHCO_3$ (2× 50 mL). After drying over anhydrous Na_2SO_4, the organic phase was concentrated to give a white solid. $R_f = 0.4$ in 2:1 EtOAc:hexanes. The product was lyophilized and was used directly in the next reaction without further purification. The product from bromination reaction (630 mg, 0.90 mmol), mercury cyanide (230 mg, 0.90 mmol), and mercury bromide (32 mg, 0.090 mmol) were dissolved in CH_2Cl_2 (3 mL) and 6-bromo-1-hexanol (130 mg, 1.10 mmol) was added to the reaction mixture. The reaction was protected from light and stirred at rt overnight. The reaction mixture was filtered through celite. The filter cake was washed with CH_2Cl_2 (3× 5 mL), and the filtrate was concentrated and applied to a silica gel column. The product (310 mg) was eluted using 2:1 hexanes/EtOAc in 26% yield for two steps: $R_f = 0.4$ in 1:1 hexanes/EtOAc; ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.15 (m, 2H), 5.07 (t, $J = 9.7$ Hz, 1H), 4.90 (m, 2H), 4.50 (m, 2H), 4.44 (d, $J = 8.0$ Hz, 1H), 4.37 (dd, $J = 4.4$, 12.5 Hz, 1H), 4.08 (dd, $J = 5.0$, 12.0 Hz, 1H), 4.03 (dd, $J = 2.2$, 12.5 Hz, 1H), 3.83 (m, 1H), 3.76 (t, $J = 9.6$ Hz, 1H), 3.65 (m, 1H), 3.57 (m, 1H), 3.35–3.49 (m, 3H), 1.98–2.12 (m, 21H), 1.84 (m, 2H), 1.57 (m, 2H), 1.25–1.48 (m, 4H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 170.7, 170.5, 170.4, 170.0, 169.7, 169.4, 169.2, 101.2, 101.0, 77.0, 73.4, 73.1, 73.0, 72.4, 72.0, 70.4, 68.2, 62.4, 62.0, 34.4, 33.2, 29.8, 28.4, 25.6, 21.6, 21.4; MS (FAB^+) *m/z* 799.2 $(M+H)^+$.

3.3. Synthesis of 8

Compound **6** (218 mg, 0.27 mmol), Dex-SH (132 mg, 0.30 mmol) and DIEA (232 μL, 1.36 mmol) were dissolved in degassed MeOH (1.0 mL). The reaction mixture was stirred at rt overnight. The reaction mixture was concentrated, and the product was purified by silica gel column chromatography using 40:1 CH_2Cl_2/MeOH. The product (218 mg) was obtained as a white solid in 69% yield: $R_f = 0.5$ in 10:1 CH_2Cl_2/MeOH; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.22 (d, $J = 10.1$ Hz, 1H), 6.89 (t,

Table 1. Strains and plasmids used in this study

Name	Description	Source/reference
Strains	Genotype	
V1019Y	*MATα trp1 ura3 6lexAop-LEU2 ade4::P$_{gal1}$-LexA-eDHFR(HIS3) GAL*$^+$ pBC398 pMW112	1,19
Plasmids	Details	
pHL1262	P_{met25}-*Cel7B* 2 μ *URA3* pRS ori amp^R	1
pHL1263	P_{met25}-*Cel7B:E197A* 2 μ *URA3* pRS ori amp^R	1

J = 5.8 Hz, 1H), 6.33 (dd, *J* = 1.8, 10.1 Hz, 1H), 6.11 (s, 1H), 5.16 (q, 2H), 5.06 (t, *J* = 9.6 Hz, 1H), 4.90 (m, 2H), 4.50 (m, 2H), 4.44 (d, *J* = 7.9, 1H), 4.37 (m, 2H), 4.02–4.15 (m, 2H), 3.82 (m, 1H), 3.77 (t, *J* = 9.5 Hz, 1H), 3.67 (m, 1H), 3.35–3.60 (m, 4H), 3.18 (m, 1H), 2.66 (t, *J* = 6.4 Hz, 2H), 2.60 (dd, *J* = 4.9, 12.7 Hz, 1H), 2.52 (t, *J* = 7.3 Hz, 2H), 2.29–2.45 (m, 4H), 1.98–2.24 (m, 22H), 1.75–1.91 (m, 2H), 1.47–1.63 (m, 9H), 1.21–1.37 (m, 4H), 1.12 (s, 3H), 0.94 (d, *J* = 7.2 Hz, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 186.8, 173.0, 170.7, 170.6, 170.4, 170.1, 169.9, 169.5, 169.3, 167.1, 153.0, 129.9, 125.2, 102.0, 101.1, 101.0, 100.3, 87.4, 77.0, 73.4, 73.1, 73.0, 72.8, 72.4, 72.1, 70.5, 68.2, 62.4, 62.0, 49.1, 48.9, 48.6, 44.4, 39.0, 36.8, 35.8, 35.1, 34.9, 32.9, 32.5, 32.1, 31.7, 30.1, 29.8, 29.0, 28.0, 26.0, 23.6, 23.6, 21.6, 21.4, 21.3, 21.2, 17.9, 15.2; HRMS (FAB$^+$) *m*/*z* 1156.4835 ($(M+H)^+$, $C_{55}H_{79}FNO_{22}S^+$ requires 1156.4799).

3.4. Synthesis of 1

Compound **7** (20 mg, 0.017 mmol) was dissolved in MeOH (0.9 mL) and NaOMe in MeOH (4.4 M, 5.0 μL, 0.022 mmol) was added. The reaction mixture was stirred at rt for 40 min. Then Dowex-200 acidic resin (200 mg) was added, and the stirring was continued for 5 more minutes. The resin was removed by filtration and washed with MeOH. The filtrate was concentrated to give a thin film on the side of the flask. The film was washed with plenty of CH_2Cl_2 and EtOAc to remove any soluble impurities and then dried in vacuo. The product **1** (15 mg) was obtained as a white solid in quantitative yield: ^{1}H NMR (400 MHz, CD_3OD) δ 7.45 (d, *J* = 10.1 Hz, 1H), 6.32 (dd, *J* = 1.9, 10.1 Hz, 1H), 6.10 (s, 1H), 4.43 (d, *J* = 7.8 Hz, 1H), 4.30 (d, *J* = 7.8 Hz, 1H), 4.26 (d, *J* = 11.2 Hz, 1H), 3.90 (m, 4H), 3.68 (dd, *J* = 5.4, 11.8 Hz, 1H), 3.63–3.20 (m, 14H), 3.14 (m, 1H), 2.74 (m, 1H), 2.66 (m, 2H), 2.60 (t, *J* = 7.3 Hz, 2H), 2.36–2.57 (m, 2H), 2.22 (m, 2H), 1.90 (m, 1H), 1.77 (q, *J* = 11.5 Hz, 1H), 1.40–1.70 (m, 13H), 1.23 (m, 1H), 1.10 (s, 3H), 0.91 (d, *J* = 7.4 Hz, 3H); ^{13}C NMR (100 MHz, CD_3OD) δ 188.6, 175.2, 170.9, 155.8, 129.5, 124.8, 104.4, 104.0, 88.0, 80.6, 80.0, 77.7, 76.3, 76.3, 74.8, 74.3, 73.2, 72.8, 71.3, 70.7, 62.4, 61.8, 50.5, 50.2, 44.8, 40.0, 36.8, 36.3, 35.9, 35.7, 33.5, 32.5, 32.3, 30.8, 30.7, 29.7, 28.9, 26.8, 23.7, 17.9, 15.3; HRMS (FAB$^+$) *m*/*z* 862.4033 ($(M+H)^+$, $C_{41}H_{64}FNO_{15}S^+$ requires 862.4059).

3.5. Synthesis of 9

Lactose (10.8 g, 31.4 mmol), 4-methoxybenzaldehyde dimethyl acetal (10.7 mL, 11.5 g, 62.9 mmol), and toluene-*p*-sulfonic acid monohydrate (597 mg, 3.14 mmol) were dissolved in anhydrous DMF (60 mL). The reaction vessel was placed in a 50 °C water bath and connected to an aspirator via a drying tower. The reaction mixture was stirred at 50 °C until all the solids were dissolved. Then the water bath was removed, 300 mL CH_2Cl_2 was added to the mixture, and white precipitate was formed and filtered. To the precipitate, acetic anhydride (120 mL) and pyridine (100 mL) were added. After stirred at rt overnight, solvents were removed in vacuo. To the residue was added CH_2Cl_2 (150 mL). The organic layer was washed with 1:1 brine:1 N HCl (2× 100 mL), brine (3× 100 mL), dried over $MgSO_4$, and concentrated. To the residue, CH_2Cl_2 (8 mL) and TFA (8 mL) were added. After stirring at rt for 1 h, $CH_2Cl_2$and TFA were removed in vacuo. The residue was dissolved in CH_2Cl_2 and applied to a silica gel column. 7.24 g of compound **9** was eluted with 25:1 CH_2Cl_2/MeOH in 39% yield as white solid: $R_f = 0.2$ in 25:1 CH_2Cl_2/MeOH; ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.23 (d, *J* = 3.7 Hz, 0.64H), 5.68 (d, *J* = 8.2 Hz, 0.37H), 5.40 (t, *J* = 9.8 Hz, 0.63H), 5.21 (m, 1.36H), 5.02 (m, 1H), 4.85 (m, 1H), 4.45 (m, 2H), 4.10 (m, 2H), 4.02 (m, 0.70H), 3.85 (m, 3.3H), 3.60 (m, 1H), 3.32 (br, 2H), 1.90–2.20 (m, 18H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 170.4, 170.1, 170.0, 169.6, 169.3, 168.6, 168.5, 101.0, 100.9, 91.3, 88.7, 75.8, 74.4, 73.6, 73.5, 73.3, 72.8, 70.6, 70.3, 69.9, 69.6, 69.5, 69.1, 67.1, 61.9, 61.6, 61.4, 21.0, 20.9, 20.9, 20.8, 20.7, 20.6; recrystallization of compound **9** in 1:2 ethyl acetate and ether gave pure α-anomer as white needle: ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.25 (d, *J* = 3.6 Hz, 1H), 5.45 (t, *J* = 9.6 Hz, 1H), 5.22 (dd, *J* = 8.0, 10.0 Hz, 1H), 5.04 (dd, *J* = 3.7, 10.2 Hz, 1H), 4.90 (dd, *J* = 3.1, 10.2 Hz, 1H), 4.51 (d, *J* = 7.9 Hz, 1H), 4.47 (d, *J* = 12.0 Hz, 1H), 4.09 (m, 3H), 3.89 (m, 3H), 3.59 (t, *J* = 5.03 Hz, 1H), 2.93 (d, *J* = 4.3 Hz, 1H), 2.53 (dd, *J* = 4.4, 8.0 Hz, 1H), 2.02–2.18 (m, 18H), The 2.93 and 3.53 peak were removed by treatment of CD_3OD; ^{13}C NMR (100 MHz, $CDCl_3$) δ 170.1, 169.8, 169.6, 169.2, 168.9, 101.0, 88.9, 75.7, 74.4, 73.4, 70.6, 70.1, 69.6, 69.3, 67.9, 62.2, 61.7, 21.1, 21.0, 21.0, 20.9, 20.7. MS (FAB$^+$) *m*/*z* 595.1862 ($(M+H)^+$, $C_{24}H_{35}O_{17}$ requires 595.1874).

3.6. Synthesis of 15, 16

Compound **9** (2.22 g, 3.73 mmol), *p*-tosyl chloride (745 mg, 3.91 mmol), and DABCO (439 mg, 3.91 mmol) were added to a round-bottomed flask. At 0 °C, anhydrous CH_2Cl_2 (40 mL) was added to the bottle. The reaction mixture was stirred at 0 °C for 4 h and white precipitate formed, then MeOH (10 mL) was added to the reaction, and the mixture was concentrated under vacuum. The residue was purified using a silica gel column (1:1 hexanes/EtOAc) to give 1.81 g of **15** in 65% yield as white solid: $R_f = 0.4$ in 1:2 hexanes/EtOAc; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.81 (d, *J* = 8.3 Hz, 2H), 7.40 (d, *J* = 8.3 Hz, 2H), 6.26 (d, *J* = 3.7 Hz, 1H), 5.41 (t, *J* = 9.8 Hz, 1H), 5.17 (dd, *J* = 8.0, 10.2 Hz, 1H), 5.00 (dd, *J* = 3.7, 10.3 Hz, 1H), 4.88 (dd, *J* = 3.3, 10.2 Hz, 1H), 4.46 (m, 2H), 4.27 (m, 1H), 4.00–4.09 (m, 4H), 3.82 (m, 2H), 2.87 (d, *J* = 5.2 Hz, 1H), 2.48 (s, 3H), 2.02–2.15 (m, 18H), the 2.87 peak was removed by treatment of CD_3OD; ^{13}C NMR (100 MHz, $CDCl_3$) δ 170.1, 169.8, 169.6, 169.5, 169.0, 168.6, 145.2, 132.2, 129.8, 127.7, 100.7, 88.8, 75.4, 73.1, 72.2, 70.6, 69.4, 69.3, 69.2, 67.1, 66.3, 61.4, 21.8, 21.0, 21.0, 20.9, 21.8, 20.6; MS (FAB$^+$) *m*/*z* 749.1966 ($(M+H)^+$, $C_{31}H_{41}O_{19}S$ requires 749.1963). To Compound **15** (300 mg, 0.40 mmol), Ac_2O (10 mL) and pyridine (10 mL) were added, and the reaction mixture was stirred at rt overnight. The mixture was concentrated and dried in vacuo to give compound **16** in 100% yield as white solid: $R_f = 0.4$

in 1:1 hexanes:EtOAc; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.77 (d, J = 8.3 Hz, 2H), 7.38 (d, J = 8.0 Hz, 2H), 6.26 (d, J = 3.7 Hz, 1H), 5.45 (t, J = 9.7 Hz, 1H), 5.34 (d, J = 2.7 Hz, 1H), 5.09 (dd, J = 7.9, 10.4 Hz, 1H), 5.00 (dd, J = 3.7, 10.3 Hz, 1H), 4.95 (d, J = 3.4 Hz, 1H), 4.92 (dd, J = 3.2, 10.4 Hz, 1H), 4.48 (d, J = 7.9 Hz, 1H), 4.43 (dd, J = 2.0, 12.0 Hz, 1H), 4.12 (d, J = 4.2 Hz, 1H), 4.09 (d, J = 3.8 Hz, 1H), 4.00 (m, 2H), 3.93 (d, J = 6.3 Hz, 1H), 3.83 (t, J = 9.7 Hz, 1H), 2.47 (s, 3H), 1.95–2.18 (m, 21H); MS (EI$^+$) m/z 789.1962 ((M−H)$^+$, $C_{33}H_{41}O_{20}S^+$ requires 789.1912).

3.7. Synthesis of 17, 18

At −78 °C, a 1.0 M solution of potassium *tert*-butoxide solution in THF (2.87 mL, 2.87 mmol) was diluted with THF (10 mL) and degassed using argon. A solution of 2-(Boc-amino) ethanethiol (580 μL, 609 mg, 3.44 mmol) in THF (3 mL) was added successively, degassed, and stirred for 30 min. A solution of compound **16** (794 mg, 1.00 mmol) in THF (10 mL) was added and degassed, and the mixture was gradually warmed to rt and stirred for 2 h. A mixture of acetic anhydride (12 mL) and pyridine (10 mL) was added to the reaction and the reaction mixture was stirred at rt overnight. The reaction mixture was quenched with water (20 mL) and extracted with ethyl acetate (2× 40 mL). The organic layer was washed with brine (2× 40 mL), dried over Na_2SO_4, and then purified using a silica gel column (3:1 CH_2Cl_2/EtOAc) to give 369 mg product as white solid: R_f = 0.4 (3:1 CH_2Cl_2/EtOAc); MS (FAB$^+$) m/z 796.5 (M+H)$^+$. The product is a mixture of compound **17** and an unknown impurity. To the bottle of the product in CH_2Cl_2 (15 mL) was added TFA (5.0 mL). After stirring at rt for 1 h, CH_2Cl_2 and TFA were removed under vacuum. The residue was dissolved in CH_2Cl_2 and purified using a silica gel column (from 40:1 CH_2Cl_2/MeOH to 10:1 CH_2Cl_2/MeOH with 0.5% TEA) to give 244 mg compound **18** in 35% yield for two steps as white solid: R_f = 0.1 in CH_2Cl_2/MeOH; ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.91 (br, 3H), 6.23 (d, J = 3.5 Hz, 0.36H), 5.69 (d, J = 8.1 Hz, 0.64H), 5.44 (m, 1.15H), 5.25 (t, J = 9.0 Hz, 0.54H), 5.03 (m, 2.72H), 4.50 (d, J = 7.7 Hz, 1H), 4.43 (d, J = 11.1 Hz, 1H), 3.80–4.13 (m, 4H), 3.19 (m, 4H), 2.90 (m, 2H), 2.73 (m, 1H), 2.58 (m, 1H), 1.96–2.18 (m, 21H); HRMS (FAB$^+$) m/z 696.2162 ((M+H)$^+$, $C_{28}H_{42}NO_{17}S^+$ requires 696.2173).

3.8. Synthesis of 19

Glutamic acid α-methyl ester γ-*tert*-butyl ester hydrochloric acid (1.00 g, 3.94 mmol), 4-methylamino benzoic acid (0.835 g, 5.50 mmol), and DCC (1.140 g, 5.50 mmol) were dissolved in anhyd CH_2Cl_2 (10 mL). DIEA (2.40 mL, 13.80 mmol) was added slowly. The reaction mixture was stirred at rt overnight, and then diluted with CH_2Cl_2 (20 mL), filtered through Celite, and washed with CH_2Cl_2. The filtrate was concentrated and then purified using a silica gel column (from 4:1 hexanes/ethyl acetate to 2:1 hexanes:ethyl acetate) to give 0.73 g of **19** in 53% yield as white solid: R_f = 0.2 in 1:1 hexanes:EtOAc; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.69 (d, J = 8.8 Hz, 2H), 6.87 (d, J = 7.4 Hz, 1H), 6.57 (d, J = 8.8 Hz, 2H), 4.76 (m, 1H), 3.76 (s, 3H), 2.87 (s, 3H), 2.38 (m, 2H), 2.22 (m, 1H), 2.07 (m, 1H), 1.42 (s, 9H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 172.6, 172.3, 166.7, 151.8, 128.7, 121.3, 111.1, 80.8, 52.5, 52.3, 31.8, 30.3, 28.2, 27.3.

3.9. Synthesis of 20

Compound **19** (1.04 g, 2.96 mmol) and 6-bromomethyl 2,4-pteridinediamine monohydrobromide (1:1 compound with 2-propanol) (0.94 g, 2.37 mmol) were dissolved in dimethyl acetamide (DMAc) (8 mL). The reaction mixture was stirred at 55 °C for 6 h. The filtrate was concentrated and then purified using a silica gel column 20:1 CH_2Cl_2/MeOH to give 0.93 g of **20** in 75% yield as orange solid: R_f = 0.2 in 10:1 CH_2Cl_2/MeOH; ^{1}H NMR (300 MHz, CD_3OD) δ 8.97 (s, 1H), 7.75 (d, J = 8.7 Hz, 2H), 6.91 (d, J = 8.8 Hz, 2H), 4.96 (s, 2H), 4.59 (dd, J = 5.2, 9.3 Hz, 1H), 3.73 (s, 3H), 3.29 (s, 3H), 2.38 (t, J = 7.2 Hz, 2H), 2.19 (m, 1H), 2.05 (m, 1H), 1.42 (s, 9H); ^{13}C NMR (75 MHz, CD_3OD) δ 173.1, 172.9, 169.0, 163.8, 156.6, 152.1, 151.4, 149.6, 149.5, 145.8, 129.4, 129.3, 122.4, 112.4, 80.9, 55.9, 52.7, 39.3, 31.8, 28.1, 26.5.

3.10. Synthesis of 21

To compound **20** (500 mg, 0.955 mmol) in CH_2Cl_2 (2.0 mL) was added TFA (2.0 mL). After stirring at rt for 1 h, CH_2Cl_2 and TFA were removed in vacuo. The residue was dissolved in CH_2Cl_2 and purified using a silica gel column (from 20:1 CH_2Cl_2/MeOH to 5:1 CH_2Cl_2/MeOH) to give 339 mg **21** in 76% yield as orange solid: R_f = 0.2 in 5:1 CH_2Cl_2/MeOH; ^{1}H NMR (400 MHz, CD_3OD) δ 8.58 (s, 1H), 7.75 (d, J = 8.9 Hz, 2H), 6.83 (d, J = 8.9 Hz, 2H), 4.86 (s, 2H), 4.59 (dd, J = 5.1, 9.2 Hz, 1H), 3.74 (s, 3H), 3.24 (s, 3H), 2.45 (t, J = 7.2 Hz, 2H), 2.24 (m, 1H), 2.10 (m, 1H); ^{13}C NMR (100 MHz, CD_3OD) δ 174.5, 171.8, 167.7, 162.3, 156.4, 150.6, 150.5, 147.8, 145.9, 128.0, 121.1, 120.1, 110.4, 54.6, 51.9, 50.9, 37.9, 29.7, 25.5.

3.11. Synthesis of 22

To a solution of Compound **21** (25 mg, 0.053 mmol) and PyBOP (28 mg, 0.054 mmol) in DMF (1 mL) was added a solution of *tert*-butyl 6-aminohexanoate (27 mg, 0.144 mmol) in anhydrous CH_2Cl_2 (2 mL) and DIEA (40 μL, 30 mg, 0.23 mmol). The mixture was stirred at rt overnight and then concentrated under vacuum. The residue was purified using a silica gel column (20:1 CH_2Cl_2/MeOH to 10:1 CH_2Cl_2/MeOH) to give 22 mg of 23 in 65% yield as orange solid: R_f = 0.4 in 10:1 CH_2Cl_2/MeOH; ^{1}H NMR (400 MHz, CD_3OD) δ 8.55 (s, 1H), 7.74 (d, J = 9.0 Hz, 2H), 6.84 (d, J = 9.0 Hz, 2H), 4.82 (s, 2H), 4.54 (q, 1H), 3.72 (s, 3H), 3.23 (s, 3H), 3.11 (t, J = 7.0 Hz, 2H), 2.32 (m, 2H), 2.16 (m, 4H), 1.51 (m, 2H), 1.41 (m, 11H), 1.27 (m, 2H); HRMS (FAB$^+$) m/z 638.3438 ((M+H)$^+$, $C_{31}H_{44}N_9O_6^+$ requires 638.3415).

3.12. Synthesis of 23

To compound **22** (13 mg, 0.020 mmol) in CH_2Cl_2 (2.0 mL) was added TFA (1.0 mL). After stirring at rt for 3 h, CH_2Cl_2 and TFA were removed in vacuo. The residue was dissolved in CH_2Cl_2 and purified using a silica gel column (from 5:1 CH_2Cl_2/MeOH to 2:1 CH_2Cl_2/MeOH) to give 11 mg **23** in 93% yield as orange solid: $R_f = 0.2$ in 5:1 CH_2Cl_2/MeOH; ^1H NMR (400 MHz, 80% CD_3OD + 20% $CDCl_3$) δ 8.55 (s, 1H), 7.74 (d, $J = 9.0$ Hz, 2H), 6.83 (d, $J = 9.0$ Hz, 2H), 4.82 (s, 2H), 4.54 (q, $J = 4.6$ Hz, 1H), 3.72 (s, 3H), 3.23 (s, 3H), 3.11 (t, $J = 7.0$ Hz, 2H), 2.32 (m, 2H), 2.21 (m, 3H), 2,09 (m, 1H), 1.56 (m, 2H), 1.42 (m, 2H), 1.29 (m, 2H); MS (FAB^+) *m/z* 582.3 ($(M+H)^+$, $C_{27}H_{36}N_9O_6^+$ requires 582.3).

3.13. Synthesis of 24

To a solution of compound **23** (35 mg, 0.060 mmol) and PyBOP (35 mg, 0.067 mmol) in DMF (1 mL) was added a solution of compound **18** (36 mg, 0.052 mmol) in anhydrous CH_2Cl_2 (2 mL) and DIEA (40 µL, 30 mg, 0.23 mmol). The mixture was stirred at rt overnight and then concentrated under vacuum. The residue was purified using a silica gel column (20:1 CH_2Cl_2/MeOH to 10:1 CH_2Cl_2/MeOH) to give 39 mg of **24** in 60% yield as orange solid: $R_f = 0.1$ in 15:1 CH_2Cl_2/MeOH; ^1H NMR (500 MHz, CD_3OD) δ 8.55 (d, 1H), 7.74 (d, $J = 8.8$ Hz, 2H), 6.85 (d, $J = 8.6$ Hz, 2H), 6.22 (d, $J = 3.7$ Hz, 0.53H), 5.74 (d, $J = 8.3$ Hz, 0.45H), 5.47 (t, $J = 3.5$ Hz, 1H), 5.38 (t, 0.53H), 5.28 (t, 0.44H), 5.09 (m, 1H), 4.97 (m, 2H), 4.84 (s, 2H), 4.60 (dd, $J = 2.8$, 7.5 Hz, 1H), 4.54 (m, 1H), 4.46 (t, 1H), 4.14 (m, 1H), 4.07 (m, 0.52H), 3.92 (m, 2.45H), 3.72 (s, 3H), 3.24 (d, 3H), 3.11 (m, 2H), 2.77 (m, 1H), 2.65 (m, 3H), 1.91–2.34 (m, 28H), 1.56 (t, $J = 7.1$ Hz, 2H), 1.44 (t, $J = 7.2$ Hz, 11H), 1.27 (m, 2H); MS (FAB^+) *m/z* 1259.3 $(M+H)^+$.

3.14. Synthesis of 25

Compound **24** (5 mg, 0.0040 mmol) was placed in a 1.5 mL microcentrifuge tube. 70% HF-pyridine (100 µL) was added and the reaction mixture was shaken on an orbital shaker at rt for 2.0 h. The reaction mixture was then diluted with CH_2Cl_2 (30 mL) and poured into iced brine (30 mL). The organic layer was separated from the aqueous layer, washed with iced brine (1× 30 mL), cold satd $NaHCO_3$ (30 mL), and then dried and concentrated. The residue was purified by silica gel column with 20:1 to 15:1 CH_2Cl_2/MeOH to give 3.0 mg compound **25** in 62% yield as orange solid: $R_f = 0.4$ in 10:1 CH_2Cl_2/MeOH; ^1H NMR (500 MHz, CD_3OD) δ 8.57 (s, 1H), 7.75 (d, $J = 9.0$ Hz, 2H), 6.86 (d, $J = 9.0$ Hz, 2H), 5.72 (dd, 1H), 5.47 (d, 1H), 5.36 (t, 1H), 5.09 (m, 1H), 4.85 (m, 57H), 4.66 (d, $J = 7.9$ Hz, 1H), 4.55 (m, 2H), 4.14 (m, 4H), 3.72 (s, 3H), 3.25 (s, 3H), 3.12 (m, 2H), 2.77 (m, 1H), 2.65 (m, 3H), 1.91–2.34 (m, 25H), 1.28–1.58 (m, 6H); MS (FAB^+) *m/z* 1219.5 $(M+H)^+$.

3.15. Synthesis of 4

Compound **25** (3.0 mg, 0.0025 mmol) was dissolved in MeOH (0.5 mL) and cooled with ice. 60 µL of 1.67 mM lithium hydroxide hydrate aq (0.10 mmol) was slowly added to the MeOH solution. The reaction mixture was then allowed to warm to rt. After stirring for 6 h, MeOH (10 mL), water (2 mL), and 40 µL of 10% AcOH aq solution were added to the reaction mixture, and then the mixture was concentrated to dryness. The mixture was purified by HPLC using a C_{18} reverse phase column with a gradient from 6% CH_3CN aq to 15% CH_3CN aq in 50 min. 1.3 mg of compound **4** was obtained in 55% yield as orange solid: ^1H NMR (500 MHz, CD_3OD) δ 8.61 (s, 1H), 7.77 (d, $J = 9.0$ Hz, 2H), 6.90 (d, $J = 9.0$ Hz, 2H), 5.55 (dd, $J_1 = 48.6$ Hz, 1H), 4.58 (s, 3H), 4.49 (m, 1H), 4.41 (d, 1H), 3.84 (m, 5H), 3.68 (t, 2H), 3.47, (m, 5H), 3.47 (s, 3H), 3.27 (s, 3H), 3.19 (s, 2H), 3.10 (t, 2H), 2.83 (d, $J = 7.0$ Hz, 1H), 2.69 (m, 2H), 2.09–2.27 (m, 6H), 1.31–1.59 (m, 6H); MS (ESI^+) *m/z* 953.5 $(M+H)^+$.

3.16. Synthesis of 26

To a solution of compound **21** (25 mg, 0.053 mmol) and PyBOP (28 mg, 0.054 mmol) in DMF (1 mL) was added a solution of compound **18** (27 mg, 0.039 mmol) in anhydrous CH_2Cl_2 (2 mL) and DIEA (40 µL, 30 mg, 0.23 mmol). The mixture was stirred at rt overnight and then concentrated under vacuum. The residue was purified using a silica gel column (20:1 CH_2Cl_2/MeOH to 10:1 CH_2Cl_2/MeOH) to give 22 mg of **26** in 49% yield as orange solid: $R_f = 0.8$ in 5.7:1 CH_2Cl_2/MeOH; ^1H NMR (500 MHz, CD_3OD) δ 8.57 (s, 1H), 7.74 (d, $J = 8.8$ Hz, 2H), 6.86 (t, 2H), 6.22 (d, $J = 3.7$ Hz, 0.46H), 5.74 (d, $J = 8.3$ Hz, 0.52H), 5.45 (d, $J = 3.3$ Hz, 1H), 5.38 (t, $J = 9.8$ Hz, 0.45H), 5.28 (t, $J = 9.0$ Hz, 0.56H), 5.09 (m, 1H), 4.97 (m, 2H), 4.84 (s, 2H), 4.66 (m, 1H), 4.57 (m, 1H), 4.45 (t, $J = 8.4$ Hz, 1H), 4.14 (m, 1H), 4.05 (m, 0.53H), 3.92 (m, 2.54H), 3.73 (s, 3H), 3.25 (s, 3H), 2.74 (m, 1H), 2.28 (m, 3H), 1.89–2.35 (m, 26H); MS (FAB^+) *m/z* 1145.7 $(M+H)^+$.

3.17. Synthesis of 27

Compound **26** (10 mg, 0.0087 mmol) was placed in a 1.5 mL microcentrifuge tube. 70% HF-pyridine (100 µL) was added and the reaction mixture was shaken on an orbital shaker at rt for 2.0 h. The reaction mixture was then diluted with CH_2Cl_2 (30 mL) and poured into iced brine (30 mL). The organic layer was separated from the aqueous layer, washed with iced brine (1× 30 mL), cold satd $NaHCO_3$ (30 mL), and then dried and concentrated. The residue was purified by silica gel column with 20:1 to 15:1 CH_2Cl_2:MeOH to give 5 mg compound **27** in 52% yield as orange solid: $R_f = 0.1$ in 15:1 CH_2Cl_2/MeOH; ^1H NMR (500 MHz, CD_3OD) δ 8.56 (s, 1H), 7.75 (d, $J = 8.8$ Hz, 2H), 6.86 (d, $J = 8.8$ Hz, 2H), 5.69 (dd, $J = 2.7$, 53.3 Hz, 1H), 5.45 (d, $J = 3.3$ Hz, 1H), 5.35 (t, $J = 9.75$ Hz, 1H), 5.09 (dd, $J = 3.5$, 10.4 Hz, 1H), 4.96 (m, 2H), 4.67 (d, $J = 7.8$ Hz, 1H), 4.58 (m, 1H), 4.50 (d, $J = 11.2$ Hz, 1H), 4.17 (dd, $J = 5.3$, 12.2 Hz), 4.07 (m, 1H), 3.95 (m, 2H), 3.73 (s, 3H), 3.25

(s, 3H), 2.74 (m, 1H), 2.60 (m, 3H), 1.90–2.36 (m, 24H); MS (FAB^+) *m*/*z* 1105.4 $(M+H)^+$.

3.18. Synthesis of 5

Compound **27** (4.0 mg, 0.0036 mmol) was dissolved in MeOH (0.5 mL) and cooled with ice. Lithium hydroxide hydrate (1.5 mg, 0.036 mmol) was dissolved in water (0.05 mL), and the solution was slowly added to the MeOH solution. The reaction mixture was then allowed to warm to rt. After stirring for 6 h, MeOH (10 mL), water (2 mL), and 20 μL of 10% AcOH aq solution were added to the reaction mixture, and then the mixture was concentrated to dryness. The mixture was purified by HPLC using a C_{18} column with a gradient from 6% CH_3CN aq to 15% CH_3CN aq in 50 min. 1.2 mg of compound **5** was obtained in 40% yield as orange solid: ^1H NMR (500 MHz, CD_3OD) δ 8.58 (s, 1H), 7.73 (d, $J = 9.0$ Hz, 2H), 6.87 (d, $J = 9.0$ Hz, 2H), 5.50 (dd, ($J_1 = 53.5$ Hz, 1H), 4.55 (s, 3H), 4.48 (m, 1H), 4.38 (d, $J = 6.5$ Hz, 1H), 3.82 (m, 6H), 3.64 (m, 3H), 3.44, (m, 7H), 3.27 (s, 3H), 3.19 (s, 1H), 2.77 (m, 2H), 2.58 (m, 2H), 2.02–2.40 (m, 4H); HRMS (ESI^+) *m*/*z* 840.3054 ($(M+H)^+$, $C_{34}H_{47}FN_9O_{13}S^+$ requires 840.2998).

3.19. Mock selection assay

The plasmids for mock library selection were prepared from *Escherichia coli* strains HL1262 (pHL1262 encoding Cel7B glycosidase) and HL1263 (pHL1263 encoding Cel7B:E197A glycosynthase)[1] using QIAprep® maxiprep kits, and plasmid concentration was determined by UV absorption at 260 nm. Then the plasmid mix was transformed to yeast strain V1019Y using high efficiency yeast transformation protocol.[21] Fifty microliters of transformances was incubated in 50 mL SC media containing 2% glucose, lacking uracil, histidine, and tryptophan at 30 °C for 3 days. The cell suspension was then centrifuged at room temperature using a Sorvail RT7 Plus centrifuge at 2000 rpm for 5 min. The cell pellet was re-suspended in 5.0 mL of 10% glycerol. To set up mock selections, 20 μL of the cell suspension was added into 13 mL Falcon tubes containing 2 mL SC media containing 2% galactose, 2% raffinose, and indicated small molecules, lacking uracil, histidine, tryptophan, methionine, and leucine. The tubes were incubated in a rotary shaker at 200 rpm and 30 °C. On the 7th day, 300 μL of cell culture was transferred into 3 mL non-selective synthetic complete media containing 2% glucose, lacking uracil, histidine, and tryptophan, and incubated for another day. The resulted cells were collected for plasmid preparation using EZNA kit (Omega Bio-Tek). The plasmid mixture from each selection experiment was amplified using *Taq* polymerase with primers VWC9845′ (5′GCA TAC GTC ACT AGT ATG GCT CGC GGT ACC GCT CT3′) and VWC9863′, (5′GCA TAC GTC CCC GGG TTA ATG GTG ATG GTG ATG GTG CTG AAC CTC CTG GTA GGT C3′). The PCR program used was as follows: step 1, 94 °C, 5 min; step 2, 94 °C, 0.5 min, 50 °C, 0.5 min, 72 °C, 3 min, 30 cycles; step 3, 72 °C, 10 min. The PCR product was then purified and quantified using UV absorption at 260 nm. Equal amount of PCR product from each selection experiment was digested with *Nco*I for 3 h and loaded onto an ethidium bromide stained 1% agarose gel. Since only Cel7B:E197A gene can be digested, colonies encoding Cel7B glycosidase can be easily distinguished from those encoding Cel7B E197A.

Acknowledgments

The authors thank Dr. J. Cherry and Dr. C. Ward from Novozymes for providing the Cel7B genes and the Cel7-B:E197A glycosynthase. We are grateful for financial support for this work from the National Science Foundation (CHE-0350183) and National Institutes of Health (GM62867).

Supplementary data

The ^1H NMR, ^{13}C-NMR, COSY and HSQC spectra of compounds synthesized in this paper. Supplementary data associated with this article including calculation details and predicted activities can be found, in the online version, at doi:10.1016/j.bmc.2006.06.034.

References and notes

1. Lin, H.; Tao, H.; Cornish, V. W. *J. Am. Chem. Soc.* **2004**, *126*, 15051–15059.
2. Mackenzie, L. F.; Wang, Q.; Warren, R. A. J.; Withers, S. G. *J. Am. Chem. Soc.* **1998**, *120*, 5583–5584.
3. Daines, A. M.; Maltman, B. A.; Flitsch, S. L. *Curr. Opin. Chem. Biol.* **2004**, *8*, 106–113.
4. Sears, P.; Wong, C. H. *Science* **2001**, *291*, 2344–2350.
5. Kim, Y. W.; Lee, S. S.; Warren, R. A.; Withers, S. G. *J. Biol. Chem.* **2004**, *279*, 42787–42793.
6. Mayer, C.; Jakeman, D. L.; Mah, M.; Karjala, G.; Gal, L.; Warren, R. A.; Withers, S. G. *Chem. Biol.* **2001**, *8*, 437–443.
7. Baker, K.; Bleczinski, C.; Lin, H.; Salazar-Jimenez, G.; Sengupta, D.; Krane, S.; Cornish, V. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16537–16542.
8. Fort, S.; Boyer, V.; Greffe, L.; Davies, G. M. O.; Christiansen, L.; Schulein, M.; Cottaz, S.; Driguez, H. *J. Am. Chem. Soc.* **2000**, *122*, 5429–5437.
9. Schou, C.; Rasmussen, G.; Kaltoft, M. B.; Henrissat, B.; Schulein, M. *Eur. J. Biochem.* **1993**, *217*, 947–953.
10. Ducros, V. M.; Tarling, C. A.; Zechel, D. L.; Brzozowski, A. M.; Frandsen, T. P.; von Ossowski, I.; Schulein, M.; Withers, S. G.; Davies, G. J. *Chem. Biol.* **2003**, *10*, 619–628.
11. MacKenzie, L. F.; Sulzenbacher, G.; Divne, C.; Jones, T. A.; Woldike, H. F.; Schulein, M.; Withers, S. G.; Davies, G. J. *Biochem. J.* **1998**, *335 Pt 2*, 409–416.
12. Lin, H.; Abida, W. M.; Sauser, R. T.; Cornish, V. W. *J. Am. Chem. Soc.* **2000**, *122*, 4247–4248.
13. Abida, W. M.; Carter, B. T.; Althoff, E. A.; Lin, H.; Cornish, V. W. *Chembiochem* **2002**, *3*, 887–895.
14. Toshima, K.; Tatsuta, K. *Chem. Rev.* **1993**, *93*, 1503–1531.
15. Capon, B. *Chem. Rev.* **1969**, *69*, 407–498.
16. Hart, B. P.; Haile, W. H.; Licato, N. J.; Bolanowska, W. E.; McGuire, J. J.; Coward, J. K. *J. Med. Chem.* **1996**, *39*, 56–65.
17. Kokuryo, Y.; Kawata, K.; Nakatani, T.; Kugimiya, A.; Tamura, Y.; Kawada, K.; Matsumoto, M.; Suzuki, R.; Kuwabara, K.; Hori, Y.; Ohtani, M. *J. Med. Chem.* **1997**, *40*, 3280–3291.

18. Piper, J. R.; Montgomery, J. A. *J. Org. Chem.* **1977**, *42*, 208–211.
19. Baker, K.; Sengupta, D.; Salazar-Jimenez, G.; Cornish, V. W. *Anal. Biochem.* **2003**, *315*, 134–137.
20. Licitra, E. J.; Liu, J. O. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 12817–12821.
21. Colby, D. W.; Kellogg, B. A.; Graff, C. P.; Yeung, Y. A.; Swers, J. S.; Wittrup, K. D. *Methods Enzymol.* **2004**, *388*, 348–358.