MOLECULAR BASIS OF S-LAYER GLYCOPROTEIN GLYCAN BIOSYNTHESIS IN *GEOBACILLUS STEAROTHERMOPHILUS**

Kerstin Steiner[‡,1], René Novotny[‡,2], Daniel B. Werz[§,3], Kristof Zarschler[‡], Peter H. Seeberger[§], Andreas Hofinger[¶], Paul Kosma[¶], Christina Schäffer[‡,4], and Paul Messner[‡,4]

SUPPLEMENTARY METHODS

Materials-All chemicals used were reagent grade and used as supplied except where noted. Dichloromethane (CH_2Cl_2) was purchased from JT Baker and purified by a Cycle-Tainer Solvent Delivery System. Pyridine was refluxed over calcium hydride and distilled prior to use. Analytical thin-layer chromatography was performed on E. Merck silica gel 60 F_{254} plates (0.25 mm). Compounds were visualized by dipping the plates in a cerium sulfate ammonium molybdate solution or a sulfuric acid/methanol solution followed by heating. Liquid chromatography was performed using forced flow of the indicated solvent on Sigma H-type silica (10–40 mm). ^{1}H NMR spectra were obtained on a Varian VXR-300 (300 MHz), Bruker-600 (600 MHz) and are reported in parts per million (δ) relative to $CHCl_3$ (7.26 ppm) or in the case of CD_3OD as solvent relative to TMS (0.00 ppm). Coupling constants (J)[1] are reported in Hertz. ^{13}C NMR spectra were obtained on a *Varian* VXR-300 (75 MHz), Bruker-600 (150 MHz) and are reported in δ relative to $CDCl_3$ (77.0 ppm) as an internal reference or to TMS (0.00 ppm).

Synthesis of acceptor substrate-The synthesis of the β-D-Gal-(1→*O*)-octyl (**I**), α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**II**) and α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) acceptors is shown in Fig. S1. Building block dibutyl 4,6-di-*O*-benzyl-3-*O*-fluorenylmethoxycarbonyl-2-*O*-pivaloyl-β-D-galactopyranosyl phosphate **a** (27) was reacted using common glycosylation conditions with 1-octanol to afford **b**. Fmoc deprotection furnished compound **c** (*n*-octyl 4,6-di-*O*-benzyl-2-*O*-pivaloyl-β-D-galactopyranoside), which served as a starting point for the synthesis of all three substrates. For the synthesis of the trisaccharide (**III**) an (2+1) approach was envisioned. For this approach the dirhamnose building block **f** was synthesized. The MP protected rhamnose unit **d** (28) was glycosylated at 0 °C with another rhamnose fragment to furnish **e**. A removal of the MP group using cerium ammonium nitrate (CAN) in aqueous solution and the subsequent reaction with trichloroacetonitrile yielded the dirhamnose building block **f** in good yield.

A removal of the pivaloyl group of compound **c** using strong basic conditions and the global deprotection of all the benzyl groups afforded β-D-Gal-(1→*O*)-octyl (**I**). For the synthesis of disaccharide (**II)**, compound **c** was glycosylated using a rhamnose building block with Ac in position 2 to ensure α-selectivity. Column chromatography to purify the reation products yielded an inseparable mixture of the desired compound together with (Overman)-rearranged trichloroacetimidate. Therefore, this mixture was subjected to basic deprotection removing acetyl (Ac) and pivaloyl (Piv) moieties. Using this approach disaccharide **g** was obtained as a pure compound. Global deprotection using $Pd(OH)_2$/C and H_2 in methanol/dichloromethane afforded disaccharide (**II**).

Glycosylating agent **f** as well as octyl galactoside **c** were utilized for the assembly of the trisaccharide **h**. A deprotection of all ester moieties followed by a hydrogenation in order to remove all benzyl groups afforded the completely deprotected trisaccharide (**III**). All the deprotected trisaccharide was purified by silica gel chromatography using mixtures of dichloromethane/methanol in order to remove all traces of palladium which could cause problems in biological experiments.

***n*-Octyl 4,6-Di-*O*-benzyl-2-*O*-pivaloyl-β-D-galactopyranoside (c)**-Compound **a** (960 mg, 1.14 mmol) (27) was azeotroped three times with toluene and dried *in vacuo*. Dichloromethane (25 ml) were added and the solution cooled to -40 °C. 1-Octanol (296 mg, 359 µl, 2.28 mmol) and afterwards TMSOTf (253 mg, 206 µl, 1.14 mmol) were added. The mixture was stirred for 1 h and quenched by addition of some drops of pyridine before the solvent was removed *in vacuo*. The resulting crude product was purified by flash chromatography (5:1 hexane/EtOAc) to afford 631 mg (71%) of **b** as a colorless oil. Compound **b** (585 mg, 0.75 mmol) was dissolved in 8 ml of DMF. Piperidine (2 ml) was added and the mixture stirred for 90 min at room temperature. The solvent was removed *in vacuo* and the residue purified by column chromatography (4:1, hexane/EtOAc) to afford 413 mg (99%) of **c** as a

colorless oil: $[\alpha]_D$: +17.0 (c = 0.41, $CHCl_3$). IR (thin film, $CHCl_3$): 3000, 1742, 1440, 1260, 1077 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz) δ 0.90 (t, *J* = 6.4 Hz, 3H), 1.25 (s, 9H), 1.28 (b, 10H), 1.55 (b, 2H), 2.39 (d, *J* = 9.6 Hz, 2H), 3.45 (quart, *J* = 6.9 Hz, 1H), 3.65 (m, 4H), 3.86 (m, 2H), 4.38 (d, *J* = 7.8 Hz, 1H), 4.52 (m, 2H), 4.71 (s, 2H), 5.00 (dd, *J* = 10.2, 7.8 Hz, 1H), 7.25-7.39 (m, 10H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.2, 22.8, 26.1, 27.2, 29.3, 29.5, 29.7, 31.9, 38.9, 68.3, 69.8, 73.4, 73.5, 75.4, 76.5, 101.0, 127.8, 128.0, 128.1, 128.4, 128.4, 137.6, 138.0, 178.5. ESI-MS: *m/z* $[M + Na]^+$ calculated 579.3292, observed 579.3283.

***p*-Methoxyphenyl 2-*O*-Acetyl-3,4-di-*O*-benzyl-α-L-rhamnopyranosyl-(1→3)-2-*O*-acetyl-4-*O*-benzyl-α-L-rhamnopyranoside (e)**-Rhamnose building block **d** (237 mg, 0.589 mmol) (28) and rhamnosyl trichloroacetimidate (406 mg, 0.766 mmol) (26) were azeotroped three times with toluene and dried *in vacuo*. Dichloromethane (9 ml) were added and the solution cooled to 0 °C. TMSOTf (20 mg, 11 µl, 0.09 mmol) was added. The mixture was stirred for 90 min and quenched by addition of some drops of pyridine. Afterwards the solvent was removed *in vacuo*. The resulting crude product was purified by column chromatography (7:3, hexane/EtOAc) to afford 448 mg (98%) of **e** as a colorless oil: $[\alpha]_D$: -45.3 (c = 1.08, $CHCl_3$). IR (thin film, $CHCl_3$): 3005, 2933, 1739, 1585, 1503, 1456, 1369, 1226, 1097, 1046 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz) δ 1.32 (pt, *J* = 7.2 Hz, 6H), 2.15 (s, 3H), 2.17 (s, 3H), 3.47 (t, *J* = 9.6 Hz, 1H), 3.54 (t, *J* = 9.6 Hz, 1H), 3.77 (s, 3H), 3.84 (m, 1H), 3.92 (m, 2H), 4.34 (dd, *J* = 9.3 Hz, 3.3 Hz, 1H), 4.48 (d, *J* = 11.4 Hz, 1H), 4.62-4.68 (m, 3H), 4.85 (d, *J* = 10.8 Hz, 1H), 4.93 (d, *J* = 11.1 Hz, 1H), 5.11 (ps, 1H), 5.32 (m, 1H), 5.37 (ps, 1H), 5.49 (m, 1H), 6.81 (pd, *J* = 9.3 Hz, 2H), 6.97 (pd, *J* = 9.3 Hz, 2H), 7.25-7.39 (m, 15H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 18.0, 18.1, 21.4, 21.2, 55.7, 68.4, 68.8, 69.1, 71.7, 72.1, 75.3, 75.5, 77.6, 79.8, 80.1, 95.9, 99.8, 114.5, 117.6, 127.6, 127.8, 127.8, 127.9, 127.9, 128.3, 128.4, 137.8, 138.3, 149.9, 154.9, 170.0, 170.1. MALDI-MS: *m/z* $[M + Na]^+$ calculated 793.3195, observed 793.3182.

2-*O*-Acetyl-3,4-di-*O*-benzyl-α-L-rhamnopyranosyl-(1→3)-2-*O*-acetyl-4-*O*-benzyl-α-L-rhamnopyranosyl trichloroacetimidate (f)-Disaccharide **e** (420 mg, 0.545 mmol) was suspended in a mixture of acetonitrile and water (20 ml, 1:1). The mixture was stirred for 2 h at room temperature until TLC control experiments showed that no starting material was left. The mixture was poured onto brine and extracted 2x with EtOAc. The combined organic phases were washed twice with water, dried over Na_2SO_4 and concentrated. The resulting crude product was purified by column chromatography (2:1, hexane/EtOAc) to afford 334 mg (92%) of the hemiacteal as a colorless oil. This hemiacetal (285 mg, 0.429 mmol) was dissolved in dichloromethane (5 ml) and trichloroacetonitrile (5 ml). NaH (5 mg) was added and the mixture was stirred for 90 min at room temperature. The solvent was removed *in vacuo*, column chromatography (3:1, hexane/EtOAc) furnished 305 mg (88%) of **f** as a slightly yellow oil: $[\alpha]_D$: -29.0 (c = 0.70, $CHCl_3$). IR (thin film, $CHCl_3$): 2995, 1744, 1672, 1615, 1451, 1369, 1092 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz) δ 1.29 (d, *J* = 6.3 Hz, 3H), 1.36 (d, *J* = 6.0 Hz, 3H), 2.14 (s, 3H), 2.17 (s, 3H), 3.46 (t, *J* = 9.3 Hz, 1H), 3.59 (t, *J* = 9.6 Hz, 1H), 3.82 (m, 1H), 3.88-3.99 (m, 2H), 4.25 (dd, *J* = 9.6 Hz, 3.3 Hz, 1H), 4.51 (d, *J* = 11.1 Hz, 1H), 4.62-4.70 (m, 3H), 4.84 (d, *J* = 11.1 Hz, 1H), 4.92 (d, *J* = 11.1 Hz, 1H), 5.09 (ps, 1H), 5.31 (m, 1H), 5.50 (m, 1H), 6.21 (ps, 1H), 7.26-7.39 (m, 15H), 8.71 (s, 1H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 18.0, 18.1, 21.1, 21.1, 68.9, 69.0, 70.7, 70.7, 71.7, 75.2, 75.7, 76.1, 79.7, 79.8, 90.8, 94.3, 99.7, 127.6, 127.6, 127.9, 128.0, 128.3, 128.3, 128.4, 137.5, 137.8, 138.3, 159.9, 169.7, 169.9. MALDI-MS: *m/z* $[M + Na]^+$ calculated 830.1872, observed 830.1858.

***n*-Octyl β-D-galactopyranoside (I)**-Monosaccharide **c** (255 mg, 0.459 mmol) was dissolved in THF (20 ml), LiOH solution (1 M, 25 ml) and hydrogen peroxide solution (15 ml) were added at -5 °C. The reaction mixture was stirred for 2 days while warming to room temperature. Then methanol (15 ml) and potassium hydroxide (3 M, 30 ml) were added and stirred for 2 days. Hydrochloric acid (1 M) was added until pH 7 was reached. The neutral solution was extracted three times with dichloromethane, dried over $MgSO_4$, filtered and concentrated. Column chromatography (EtOAc) furnished 209 mg (97%) of a colorless solid. This compound (148 mg, 0.263 mmol) was dissolved in methanol (15 ml), $Pd(OH)_2/C$ (15 mg) was added and the Ar atmosphere replaced by a H_2 atmosphere. The reaction mixture was stirred for 12 h. Filtration through a plug of celite removed all solids. Column chromatography using silica gel and dichloromethane/methanol (1:1) in order to remove traces of Pd yielded 75 mg (98%) of **I** as a colorless wax: $[\alpha]_D$: -10.0 (c = 0.13, H_3COH). ^1H NMR (CD_3OD, 300 MHz) δ 0.89 (b, 3H), 1.29-1.37 (b, 10H), 1.61 (m, 2H), 3.47-3.53 (m, 4H), 3.73 (m, *J* = 6.0 Hz, 2H), 3.84 (s, b, 1H), 3.88 (m, 1H), 4.20 (d, *J* = 6.6 Hz, 1H). ^{13}C NMR (CD_3OD, 75 MHz) δ

14.0, 23.3, 26.7, 30.0, 30.2, 30.4, 32.6, 62.0, 69.8, 70.4, 72.1, 74.6, 76.1, 104.5. MALDI-MS: *m/z* [M + Na]$^+$ calculated 315.1778, observed 315.1775.

***n*-Octyl 3,4-di-*O*-benzyl-α-L-rhamnopyranosyl-(1→3)-4,6-di-*O*-benzyl-β-D-galactopyranoside (g)-**Compound **c** (185 mg, 0.332 mmol) and rhamnosyl trichloroacetimidate (228 mg, 0.431 mmol) were azeotroped three times with toluene and dried *in vacuo*. Dichloromethane (6 ml) were added and the solution cooled to 0 °C. TMSOTf (11 mg, 9 µl, 0.05 mmol) was added. The mixture was stirred for 1 h and quenched by addition of some drops of pyridine. Afterwards the solvent was removed *in vacuo*. The resulting crude product was purified by column chromatography (4:1, hexane/EtOAc) to afford a mixture of disaccharide and rearranged trichloroacetimidate. Also, further attempts to separate the compounds failed. Therefore, the mixture was dissolved in THF (35 ml), LiOH solution (1 M) (19 ml) and hydrogen peroxide solution (12 ml) were added at -5 °C. The reaction mixture was stirred for 30 h while warming to room temperature. Then methanol (17 ml) and potassium hydroxide (3 M, 35 ml) were added and stirred for 3 d. Hydrochloric acid (1 M) was added until pH 7 was reached. The neutral solution was extracted three times with dichloromethane, dried over $MgSO_4$, filtered and concentrated. Column chromatography (3:1, hexane/EtOAc) furnished 145 mg (55%) of **g** as a colorless solid: $[\alpha]_D$: -41.2 (c = 0.25, $CHCl_3$). IR (thin film, $CHCl_3$): 3005, 2923, 1600, 1492, 1451, 1082 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz) δ 0.87 (t, *J* = 6.9 Hz, 3H), 1.33 (m, 15H), 1.61 (m, 2H), 2.33 (s, 1H), 2.52 (s, 1H), 3.48 (m, 2H), 3.62 (m, 4H), 3.78-3.92 (m, 7H), 4.20 (m, 2H), 4.46 (m, 2H), 4.52-4.66 (m, 4H), 4.83 (d, *J* = 11.7 Hz, 1H), 4.90 (d, *J* = 11.1 Hz, 1H), 5.38 (s, 1H), 7.23-7.37 (m, 20H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.0, 18.0, 22.5, 25.8, 29.1, 29.3, 29.4, 31.7, 68.1, 68.3, 68.6, 70.1, 71.7, 72.6, 73.5, 73.7, 74.9, 75.1, 75.9, 77.9, 79.5, 79.7, 99.9, 103.1, 127.4, 127.6, 127.7, 127.8, 128.1, 128.2, 128.3, 128.4, 137.7, 137.8, 138.4, 138.5. MALDI-MS: *m/z* [M + Na]$^+$ calculated 821.4235, observed 821.4247.

***n*-Octyl α-L-rhamnopyranosyl-(1→3)-β-D-galactopyranoside (II)-**Disaccharide **g** (103 mg, 0.129 mmol) was dissolved in a mixture of methanol and dichloromethane (30 ml, 2:1), $Pd(OH)_2$/C was added and the Ar atmosphere was replaced by a H_2 atmosphere. The reaction mixture was stirred for 12 h. Filtration through a plug of celite removed all solids. Column chromatography using silica gel and dichloromethane/methanol (10:1 → 3:1) in order to remove traces of Pd yielded 50 mg (quant.) of (**II**) as a colorless solid: $[\alpha]_D$: -32.3 (c = 0.13, H_3COH). ^{1}H NMR (CD_3OD, 300 MHz) δ 0.89 (b, 3H), 1.24-1.36 (b, 13H), 1.61 (m, 2H), 3.31 (s, 2H), 3.39 (m, 1H), 3.51 (m, 2H), 3.63 (m, 1H), 3.71 (m, 3H), 3.92 (m, 3H), 4.23 (d, *J* = 7.5 Hz, 1H), 5.05 (s, 1H). ^{13}C NMR (CD_3OD, 75 MHz) δ 14.0, 17.6, 23.3, 26.7, 30.0, 30.1, 30.4, 32.6, 61.8, 69.6, 69.7, 70.5, 71.7, 73.6, 76.0, 81.2, 103.4, 104.5. MALDI-MS: *m/z* [M + Na]$^+$ calculated 461.2357, observed 461.2357.

***n*-Octyl 2-*O*-acetyl-3,4-di-*O*-benzyl-α-L-rhamnopyranosyl-(1→3)-2-*O*-acetyl-4-*O*-benzyl-α-L-rhamnopyranosyl-(1→3)-4,6-di-*O*-benzyl-2-*O*-pivaloyl-β-D-galactopyranoside (h)-**Galactose building block **c** (100 mg, 0.179 mmol) and rhamnose disaccharide building block **f** (209 mg, 0.258 mmol) were azeotroped three times with toluene and dried *in vacuo*. Dichloromethane (3.5 ml) were added and the solution cooled to 0 °C. TMSOTf (9 mg, 7 µl, 0.04 mmol) was added. The mixture was stirred for 120 min and quenched by addition of some drops of pyridine. Afterwards the solvent was removed *in vacuo*. The resulting crude product was purified by column chromatography (4:1, hexane/EtOAc) to afford 164 mg (76%) of **h** as a colorless oil: $[\alpha]_D$: -25.8 (c = 0.33, $CHCl_3$). IR (thin film, $CHCl_3$): 2933, 2871, 1739, 1369, 1072 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz) δ 0.88 (t, *J* = 6.9 Hz, 3H), 1.24-1.34 (m, 25H), 1.58 (m, 2H), 2.10 (s, 3H), 2.13 (s, 3H), 3.41 (m, 4H), 3.64 (ps, 3H), 4.78-4.88 (m, 5H), 4.07 (dd, *J* = 10.7, 2.2 Hz, 1H), 4.38 (d, *J* = 10.1 Hz, 1H), 4.40-4.68 (m, 7H), 4.80 (d, *J* = 12.3 Hz, 1H), 5.91 (m, 3H), 5.17 (ps, 1H), 5.21 (ps, 2H), 7.24-7.39 (m, 25H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.2, 18.0, 18.2, 21.1, 21.1, 22.8, 26.1, 27.2, 29.3, 29.5, 29.7, 31.9, 38.9, 68.6, 68.7, 68.8, 69.0, 69.7, 71.5, 71.7, 73.5, 73.7, 74.8, 75.0, 75.1, 76.1, 76.2, 77.6, 78.2, 79.8, 80.3, 98.7, 99.1, 101.6, 127.4, 127.4, 127.5, 127.6, 127.7, 127.7, 127.8, 127.8, 128.2, 128.2, 128.3, 128.3, 137.7, 137.8, 137.9, 138.1, 138.5, 169.3, 169.9, 176.5. MALDI-MS: *m/z* [M + Na]$^+$ calculated 1225.6070, observed 1225.6050.

***n*-Octyl α-L-rhamnopyranosyl-(1→3)-α-L-rhamnopyranosyl-(1→3)-β-D-galactopyranoside (III)-**Trisaccharide **h** (135 mg, 0.112 mmol) was dissolved in THF (13 mL), LiOH solution (1 M, 7 ml) and hydrogen peroxide solution (5 ml) were added at -5 °C. The reaction mixture was stirred for 3 d while warming to room temperature. Then methanol (7 ml) and potassium hydroxide (3 M, 13 ml) were added and stirred for 2 d. Hydrochloric acid (1 M) was added until pH 7 was reached. The

neutral solution was extracted three times with dichloromethane, dried over $MgSO_4$, filtered and concentrated. Mass spectrometry still shows a pivaloyl group in the molecule. Therefore, the crude product was dissolved in methanol (10 ml) and NaOMe (500 mg) were added. The mixture was stirred for 12 h and the solvent removed *in vacuo*. Column chromatography (EtOAc) furnished 102 mg (88%) of a colorless solid. This compound (95 mg, 0.092 mmol) was dissolved in a mixture of methanol (10 ml) and dichloromethane (15 mL), $Pd(OH)_2/C$ was added and the Ar atmosphere replaced by a H_2 atmosphere. The reaction mixture was stirred for 12 h. Filtration through a plug of celite removed all solids. Column chromatography using silica gel and dichloromethane/methanol (10:1 $\rightarrow$ 4:1) was performed in order to remove traces of Pd. 38 mg (quant.) of (**III**) as a colorless highly viscous oil were obtained: $[\alpha]_D$: -60.5 (c = 1.25, H_3COH). ^{1}H NMR (CD_3OD, 300 MHz) δ 0.89 (b, 3H), 1.24-1.33 (b, 16H), 1.61 (b, 2H), 3.38 (s, 1H), 3.47-3.57 (m, 4H), 3.62-3.92 (m, 9H), 3.99 (s, b, 1H), 4.07 (s, b, 1H), 4.23 (d, *J* = 7.5 Hz, 1H), 5.02 (s, 1H), 5.04 (s, 1H). ^{13}C NMR (CD_3OD, 75 MHz) δ 14.0, 17.6, 17.6, 23.3, 26.7, 30.0, 30.1, 30.4, 32.6, 61.8, 69.6, 69.7, 70.0, 70.5, 71.5, 71.7, 72.8, 73.7, 76.0, 79.4, 81.1, 103.4, 103.6, 104.6. MALDI-MS: *m/z* $[M + Na]^+$ calculated 607.2936, observed 607.2925.

References included in supplementary methods:

26. Fürstner, A., Müller, T. (1999) *J. Am. Chem. Soc.* **121**, 7814-7821
27. Love, K. R., Seeberger, P. H. (2004) *Angew. Chem. Int. Ed.* **43**, 602-605
28. Werz, D. B., Seeberger, P. H. (2005) *Angew. Chem. Int. Ed.* **44**, 6315-6318

[1]The abbreviations used are: Ac, acetyl; Ar, argon; Bn, benzyl; CAN, cerium ammonium nitrate; EtOAc, ethyl acetate; Fmoc, fluorenylmethoxycarbonyl; H_2, hydrogen; IR, IR spectroscopy; *J*, coupling constant; MP, methoxyphenyl; NaOMe, sodium methoxide; OBu, *O*-butyl; Pd, palladium; Piv, pivaloyl; THF, tetrahydrofuran; TMSOTf, trimethylsilyl trifluoromethanesulfonate.

SUPPLEMENTARY RESULTS

Sequence comparison of the transferases WsaC, WsaD, WsaE and WsaF–WsaC shows homology to the α-1,3-L-rhamnosyltransferases from *Streptococcus pyogenes* MGAS10394 (62) (protein accession number YP_059939) and MGAS6180 (63) (YP_280052) as well as to RgpB from *Lactococcus lactis* and *Streptococcus thermophilus*, RgpBc from *Streptococcus mutans*, and Cps2F from *Streptococcus pneumoniae*, and several other enzymes, all of which have been classified as rhamnosyltransferases involved in cell wall biosynthesis (Fig. S2). WsaD shows high homology to the putative rhamnosyltransferase JexE from *Paenibacillus jamilae*, the glycosyltransferases from *Clostridium beijerincki* (protein accession number ZP_00911019) and *Methanosphaera stadtmanae*, as well as some homology to rhamnosyltransferases from *Shigella dysenteriae* (37) and *Shigella flexneri* (64).

The N-terminal portion of WsaE (aa 70 to 150) revealed homology to methyltransferases *e.g.*, to COG2226.2, UbiE, a menaquinone biosynthesis methyltransferase from *Methanosarcina acetivorans* (65), and to PFAM 08241.1 and PFAM 08242.1 methyltransferase family 11 and 12, respectively, both of which are SAM-dependent methyltransferases (Fig. S2). The central and C-terminal portions contain two glycosyltransferase domains, which are homologous *e.g.,* to the O-antigen biosynthesis protein from *Planctomyces maris* DSM 8797 (ZP_01855298) and *Xanthomonas oryzae* (ABI93188), or to the GT-2 glycosyltransferases from *Burkholderia cenocepacia* MC0-3 (ZP_01563358) and from *Enterococcus faecalis* (66) (ORFde16, AAC35930) (Fig. S3). The first glycosyltransferase domain contains motifs that are typical of inverting glycosyltransferases and contains a conserved DD motif (aa 638-639) and, 53 aa downstream, the DXDD motif (DHDD, aa 691-694). The second glycosyltransferase domain contains ED motifs (aa 970-971 and 981-982) and a DXE motif (aa 1062-1064).

Highest homology for WsaF was found to conserved hypothetical proteins from *Planctomyces maris* (ZP_01855299; 430 aa; E-value: $2e^{-89}$) and from *Streptococcus pneumoniae* (CAI34499; 414 aa, E-value: $8e^{-70}$). Homologies to putative glycosyltransferases, for instance, from *Anabaena variabilis* (protein accession number ABA22956; 406 aa, E-value: $6e^{-20}$) and WbbX (421 aa, E-value: $2e^{-08}$) from *Yersinia enterocolitica* were found as well.

References included in supplementary results:

37. Morona, R., Macpherson, D. F., Van Den, B. L., Carlin, N. I., and Manning, P. A. (1995) *Mol. Microbiol.* **18,** 209-223
62. Banks, D. J., Porcella, S. F., Barbian, K. D., Beres, S. B., Philips, L. E., Voyich, J. M., DeLeo, F. R., Martin, J. M., Somerville, G. A., and Musser, J. M. (2004) *J. Infect. Dis.* **190,** 727-738
63. Green, N. M., Zhang, S., Porcella, S. F., Nagiec, M. J., Barbian, K. D., Beres, S. B., LeFebvre, R. B., and Musser, J. M. (2005) *J. Infect. Dis.* **192,** 760-770
64. Klena, J. D. and Schnaitman, C. A. (1993) *Mol. Microbiol.* **9,** 393-402
65. Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D., Brown, A., Allen, N., Naylor, J., Stange-Thomann, N., DeArellano, K., Johnson, R., Linton, L., McEwan, P., McKernan, K., Talamas, J., Tirrell, A., Ye, W., Zimmer, A., Barber, R. D., Cann, I., Graham, D. E., Grahame, D. A., Guss, A. M., Hedderich, R., Ingram-Smith, C., Kuettner, H. C., Krzycki, J. A., Leigh, J. A., Li, W., Liu, J., Mukhopadhyay, B., Reeve, J. N., Smith, K., Springer, T. A., Umayam, L. A., White, O., White, R. H., Conway, d. M., Ferry, J. G., Jarrell, K. F., Jing, H., Macario, A. J., Paulsen, I., Pritchett, M., Sowers, K. R., Swanson, R. V., Zinder, S. H., Lander, E., Metcalf, W. W., and Birren, B. (2002) *Genome Res.* **12,** 532-542
66. Xu, Y., Murray, B. E., and Weinstock, G. M. (1998) *Infect. Immun.* **66,** 4313-4323

SUPPLEMENTARY TABLES

Table S1: Bacterial strains and plasmids used in this work.

Strains or plasmids	Genotype and relevant characteristics	Sources or references
E. coli DH5α™	F^- φ80*lacZ*M*15* (*lacZYA-argF*) *U169 deoR recA1 endA1 hsdR17* (rk^-, mk^+) *phoA supE44 thi-1 gyrA96 relA1* λ^-	Invitrogen
E. coli BL21 Star (DE3)	F^- *ompT hsdSB* (rB^-mB^-) *gal dcm rne131*	Invitrogen
E. coli C43(DE3)	F^- *ompT hsdSB* (rB^-mB^-) *gal dcm* (DE3) C43	Lucigen (67)
pET28a(+)	*E. coli* expression vector; Km^R	Novagen
pNGB220	pET28a-WsaC; pET28a(+) expressing WsaC (aa 1-324) from *G. stearothermophilus* NRS 2004/3a with N-terminal His_6-tag; Km^R	This study
pNGB221	pET28a-WsaC_I; pET28a(+) expressing WsaC from *G. stearothermophilus* NRS 2004/3a devoid of the C-terminal transmembrane domain (aa 1-280) with N-terminal His_6-tag; Km^R	This study
pNGB230	pET28a-WsaD; pET28a(+) expressing WsaD (aa 1-289) from *G. stearothermophilus* NRS 2004/3a with N-terminal His_6-tag; Km^R	This study
pNGB231	pET28a-WsaD_I; pET28a(+) expressing WsaD from *G. stearothermophilus* NRS 2004/3a devoid of the C-terminal transmembrane domain (aa 1-254) with N-terminal His_6-tag; Km^R	This study
pNGB240	pET28a-WsaE; pET28a expressing WsaE (aa 1-1127) from *G. stearothermophilus* NRS 2004/3a with N-terminal His_6-tag; Km^R	This study
pNGB241	pET28a-WsaE_M; pET28a expressing WsaE_M (aa 1-170) from *G. stearothermophilus* NRS 2004/3a with N-terminal His_6-tag; Km^R	This study
pNGB242	pET28a-WsaE_N; pET28a expressing WsaE_N (aa 1-368) from *G. stearothermophilus* NRS 2004/3a with N-terminal His_6-tag; Km^R	This study
pNGB243	pET28a-WsaE_B; pET28a expressing WsaE_B (aa 368-1127) from *G. stearothermophilus* NRS 2004/3a with N-terminal His_6-tag; Km^R	This study
pNGB244	pET28a-WsaE_C; pET28a expressing WsaE_C (aa 765-1127) from *G. stearothermophilus* NRS 2004/3a with N-terminal His_6-tag; Km^R	This study
pNGB245	pET28a-WsaE_A; pET28a expressing WsaE_A (aa 368-863) from *G. stearothermophilus* NRS 2004/3a with N-terminal His_6-tag; Km^R	This study
pNGB250	pET28a-WsaF; pET28a(+) expressing WsaF (aa 1-413) from *G. stearothermophilus* NRS 2004/3a with N-terminal His_6-tag; Km^R	This study
pNGB200	pET28a-WsaP; pET28a(+) expressing WsaP (aa 1-471) from *G. stearothermophilus* NRS 2004/3a with N-terminal His_6-tag; Km^R	(22)
pNGB261	pET28a-RmlB; pET-28a(+) expressing RmlB from *G. stearotermophilus* NRS 2004/3a with N-terminal His_6-tag; Km^R	This study
pNGB262	pET28a-RmlC; pET-28a(+) expressing RmlC from *G. stearothermophilus* NRS 2004/3a with N-terminal His_6-tag; Km^R	This study
pNGB263	pET28a-RmlD; pET-28a(+) expressing RmlD from *G. stearotermophilus* NRS 2004/3a with N-terminal His_6-tag; Km^R	This study

Table S2: PCR primers used for the amplification of WsaD, WsaC, WsaE, WsaF, RmlB, RmlC and RmlD from *G. stearothermophilus* NRS 2004/3a for the design of different expression constructs.

Primer	Nucleotide sequence (5' → 3')	Orientation
pET-WsaD_for	AATCACCATATGGATATTAGCATTATTATCGTGAAT	forward
pET-WsaD_rev	ATAAGAATCTCGAGTTAACCACCTATTTTTCGAAAAGTGT	reverse
pET-WsaD_I_rev	AATCACTCGAGCTAATGCTTCCTATGGAATAAAAACATC	reverse
pET-WsaC_for	AATCACCATATGGAGATGCCATTGGTTT	forward
pET-WsaC_rev	ATAAGAATCTCGAGCTAATATTTTAACTTTTTAAAAATCCATATTG	reverse
pET-WsaC_I_rev	AATCACTCGAGCTATTGACTAAATTTTTTTCTAGCTTCCG	reverse
pET-WsaE_for	AATCAGCTAGCATGGAGCGTTGTAGAATGAATAA	forward
pET-WsaE _B_for	AATCAGCTAGCATGCGTATTAAGAATAGATTAAAAA	forward
pET-WsaE _C_for	AATCAGCTAGCATGGGCTTTCGAAAAGGTTTTG	forward
pET-WsaE_rev	ATAAGAATCTCGAGCTACGACCTTATTGTTGGAATCAAA	reverse
pET-WsaE_A_rev	ATAAGAATCTCGAGCTAAACCAATGGATAATCCTCTG	reverse
pET-WsaE_N_rev	ATAAGAATCTCGAGCTACATAGACTCAGCTTGGTTTTGC	reverse
pET-WsaE_M_rev	ATAAGAATCTCGAGCTATATCTTCCGTTTCCTCTAGG	reverse
pET-WsaF_for	GGGGTACCCCATATGgTTCAAAAATTAATACAGATATTAAG	forward
pET-WsaF_rev	GGGGTACCCCGAGCTCGAAAATAAACCTACGAATAGAGTCATCA	reverse
RmlB_for	AATCAGCTAGCATGAAAGTATTGATTACCGGC	forward
RmlB_rev	AATCACTCGAGCCTAACTGCCCGTTTGC	reverse
RmlC_for	AATCACCATATGgAAATTATTGAGACTAAGTTTAGTAATG	forward
RmlC_rev	AATCACTCGAGCTGCATTTCCTTCCCTTAATAAG	reverse
RmlD_for	AATCACCATATGgAAATTGTTGTTACGGGGG	forward
RmlD_rev	AATCACTCGAGGATCTATTGTAAATATATCACTTCAAATC	reverse

Triplets corresponding to the initiation and termination codons in the primer sequence are boxed. Lowercase letters indicate changes in the original nucleotides sequence. Artificial restriction sites are underlined.

Table S3: ESI-QTOF MS analysis of octyl-linked products of *in vitro* activity assays.

Sample	$[M+Na]^+_{exp}$	$[M+Na]^+_{theor}$	assignment
substrate (**II**)	461.22	461.24	Rha-Gal-octyl
substrate (**III**)	607.25	607.29	Rha-Rha-Gal-octyl
WsaF + (**II**) + dTDP-β-L-Rha	607.26	607.29	Rha-Rha-Gal-octyl (**IV**)
WsaC + (**III**) + dTDP-β-L-Rha	753.33	753.35	Rha-Rha-Rha-Gal-octyl (**V**)
WsaC + (**III**) + dTDP-β-L-Rha	899.37	899.41	Rha-Rha-Rha-Rha-Gal-octyl (**VI**)
WsaE + (**III**) + dTDP-β-L-Rha	753.33	753.35	Rha-Rha-Rha-Gal-octyl (**VII**)

References included in supplementary tables:

22. Steiner, K., Novotny, R., Patel, K., Vinogradov, E., Whitfield, C., Valvano, M. A., Messner, P., and Schäffer, C. (2007) *J. Bacteriol.* **189,** 2590-2598
67. Miroux, B. and Walker, J. E. (1996) *J. Mol. Biol.* **260,** 289-298

SUPPLEMENTARY FIGURE LEGENDS

Fig. S1. **Reaction scheme** for the synthesis of β-D-Gal-(1→*O*)-octyl (**I**), α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**II**) and α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) used as acceptors in this study.

Fig. S2. **Multiple sequence alignment of the N-terminal region of WsaC** from *G. stearothermophilus* NRS 2004/3a with putative rhamnosyltransferases of two different *Streptococcus pyogenes* strains, RgbB from *Lactococcus lactis* and RgpBc from *Streptococcus mutans*. Conserved amino acids described for inverting transferases are highlighted.

Fig. S3. **Sequence alignment** of the N-terminal region of WsaE from *G. stearothermophilus* NRS 2004/3a, containing the putative methyltransferase domain, with menaquinone biosynthesis methyltransferase (UbiE) from *Methanosarcina acetivorans* C2A.

Fig. S4. **Sequence alignment** of the C-terminal region of WsaE from *G. stearothermophilus* NRS 2004/3a, containing the putative rhamnosyltransferase domains, with O-antigen biosynthesis proteins from *Planctomyces maris* DSM 8797 and (protein accession number ZP_01855298) and *Xanthomonas oryzae* (ABI93188) or the GT-2 glycosyltransferases from *Burkholderia cenocepacia* MC0-3 (ZP_01563358) and *Enterococcus faecalis* (AAC35930).

Fig. S5. A: **Western immunoblot analysis of the expression of WsaC, WsaD and WsaF** in *E. coli* BL21 Star (DE3). Proteins were detected with anti-His-tag antibody. Lanes 1, 6 and 8: Precision Plus Protein™ Standard All Blue (Biorad); lane 2: WsaC (37 kDa); lane 3: WsaC_I (36.5 kDa); lane 4: WsaD (30 kD); lane 5: WsaD_I (31.6 kDa); lane 7: WsaF (50.5 kDa). B: **Western Immunoblot analysis of the expression of WsaE and truncated forms thereof** in *E. coli* BL21 Star (DE3). Proteins were detected with anti-His-tag antibody. Lane 1: Precision Plus Protein™ Standard All Blue (Biorad); lane 2: full length WsaE (135.0 kDa); lane 3: WsaE_A (60.5 kDa; aa 368-863, first rhamnosyltransferase domain); lane 4: WsaE_B (90.8 kDa; aa 368-1127, both rhamnosyltransferase domains); lane 5: WsaE_C (44.1 kDa; aa 765-1127, C-terminal rhamnosyltransferase domain) lane 6: WsaE_M (22.7 kDa; aa 1-170, UbiE domain); lane 7: WsaE_N (46.8 kDa; aa 1-368, UbiE domain plus 200 aa downstream). For details about the truncated forms see Figure 1.

Fig. S6. **TLC pattern of rhamnosyltransferase activity assays using octyl-linked oligosaccharides.** For the reactions crude extract (WsaC and WsaD) or purified enzyme (WsaC_I, WsaE and WsaF) were used and the products were separated on Silica TLC plates with chloroform/methanol/water 65:25:4 as solvent. Lanes 1-7: WsaD; lanes 12-18: WsaC; lanes 20-23: WsaC_I; lanes 24-27: WsaE; lanes 28-31: WsaF; lanes 1, 12, 20, 24 and 28: β-D-Gal-(1→*O*)-octyl (**I**) and dTDP-β-L-Rha; lanes 2, 13, 21, 25 and 29: α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**II**) and dTDP-β-L-Rha; lanes 3, 14, 22, 26 and 30: α-L-Rha-(1→3)-α-L-Rha-(1→3)-β-D-Gal-(1→*O*)-octyl (**III**) and dTDP-β-L-Rha; lanes 4, 8 and 15: (**I**); lanes 5, 9 and 16: (**II**); lanes 6, 10 and 17: (**III**); lanes 7, 11, 18, 23, 27 and 31: dTDP-β-L-Rha. Staining was performed with thymol.

Fig. S7. **(+)ESI-QTOF MS^2 spectrum of the singly charged ions at *m/z* 899.40** corresponding to the product (**VI**) of the *in vitro* assay of WsaC; the loss of rhamnoses can be deduced from the mass difference of ~146 between neighboring Y-ions.

Figure S1

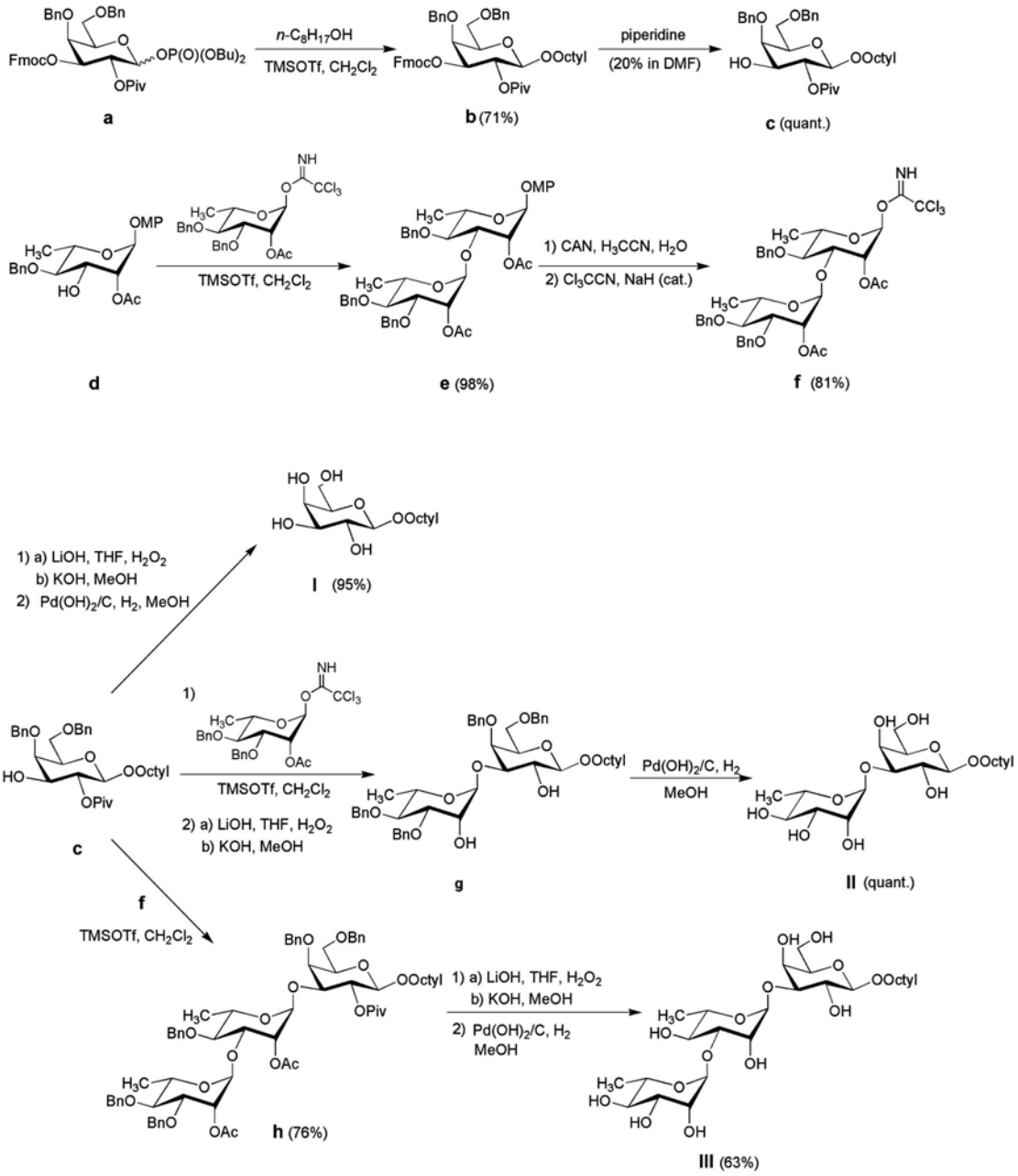


Figure S2

```
StrepPyo10394   ---MNINILLSTYN-GERFLAEQIQSIQRQTVNDWTLLIRDDGSTDGTQDIIRTFVK-ED 55
StrepPyo6180    ---MNINILLSTYN-GERFLAEQIQSIQRQTVNDWTLLIRDDGSTDGTQDIIRTFVK-ED 55
RgpBc           ---MKVNILMSTYN-GQEFIAQQIQSIQKQTFENWNLLIRDDGSSDGTPKIIADFAK-SD 55
RgbBLac         ---MRVNILMSTYN-GEKFVADQIESIQKQTYTDWNLIIRDDGSSDRTCEIVDDFVS-KD 55
WsaC            MEMPLVSIVVATYFPRTDFFEKQLQSLNNQTYENIEIIICDDSANDAEYEKVKKMVENII 60
                 .   :.*:::** .   *. .*::*::.**  :  ::* **.:.*   . :  :...

                                                  DXDD
SP10394         KRIQWINEGQTENLGVIKNFYTLLR-HQKADVYFFSDQDDIWLDNKLEVTLLEAQKHEMT 114
SP6180          KRIQWINEGQTENLGVIKNFYTLLK-HQKADVYFFSDQDDIWLDNKLEVTLLEAQKHEMT 114
RgpBc           ARIRFINADKRENFGVIKNFYTLLK-YEKADYYFFSDQDDVWLPQKLELTLASVEKENNQ 114
RgbBLac         NRIKLIR---AENVGVIKSFHELVTDSNNADFYFFADQDDYWLPEKLSVMLEETKKHDNS 112
WsaC            SRFPCKVIRNEKNVGSNKTFERLTQ-EANGDYICYCDQDDIWLSEKVERLVNHITKHHCT 119
                 *:      . :*.*  *.*  *     :.*   :.**** ** :*:.  :    *..

SP10394         APLLVYTDLKVVTQHLAICHDSMIKT----QSGHANTSLLQELTENTVTGGTMMITHALA 170
SP6180          APLLVYTDLKVVTQHLAICHDSMIKT----QSGHANTSLLQELTENTVTGGTMMITHALA 170
RgpBc           IPLMVYTDLTVVDRDLQVLHDSMIKT----QSHHANTSLLEELTENTVTGGTMMVNHCLA 170
RgbBLac         KPVMYYTDLKVTDRNLNVTSESMIRG----QSDHANTKLVQELTENTVTGGASMINHELA 168
WsaC            ---LVYSDLSLIDENDRIIHKSFKRSNFRLKHVHGDNTFAHLINRNSVTGCAMMIRADVA 176
                   : *:**.:  ..  :  .*: : .   :  *.:..: . :..*:*** : *:   :*

SP10394         EEWTTC--DGLLMHDWYLALLASATGKLVYLDIPTELYRQHDANVLGARTWS--KRMKNW 226
SP6180          EEWTTC--DGLLMHDWYLALLASATGKLVYLDIPTELYRQHDANVLGARTWS--KRMKNW 226
RgpBc           KQWKQCY-DDLIMHDWYLALLAASLGKLIYLDETTELYRQHESNVLGARTWS--KRLKNW 227
RgbBLac         QLWQST--NDIIMHDWYLAIVAAALGELVYIDQPTHLYRQHDSNVLGARTLS--KRIKKW 224
WsaC            KSAIPFPDYDEFVHDHWLAIHAAVKGSLGYIKEPLVWYRIHLGNQIGNQRLVNITNINDY 236
                :      . . ::** :**: *:  *.* *:. .   ** * .* :* :   . ..::.:
```

Figure S3

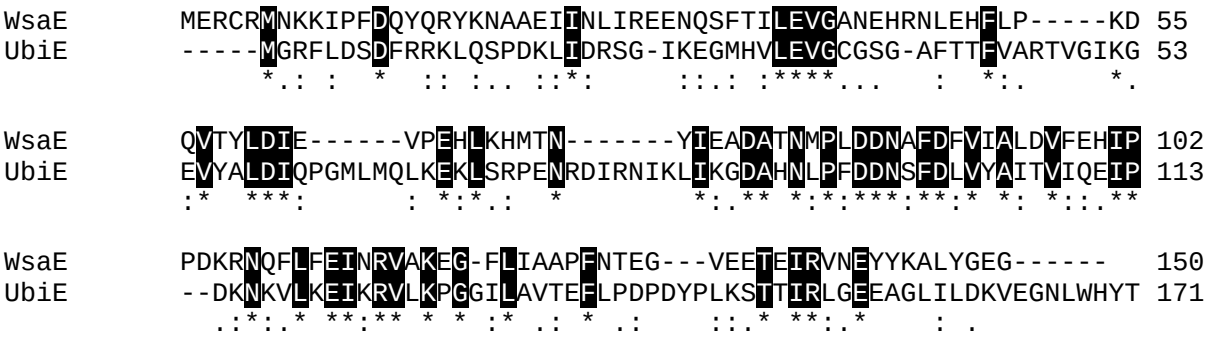

WsaE MERCRMNKKIPFDQYQRYKNAAEIINLIREENQSFTILEVGANEHRNLEHFLP-----KD 55
UbiE -----MGRFLDSDFRRKLQSPDKLIDRSG-IKEGMHVLEVGCGSG-AFTTFVARTVGIKG 53
WsaE QVTYLDIE------VPEHLKHMTN-------YIEADATNMPLDDNAFDFVIALDVFEHIP 102
UbiE EVYALDIQPGMLMQLKEKLSRPENRDIRNIKLIKGDAHNLPFDDNSFDLVYAITVIQEIP 113
WsaE PDKRNQFLFEINRVAKEG-FLIAAPFNTEG---VEETEIRVNEYYKALYGEG------ 150
UbiE --DKNKVLKEIKRVLKPGGILAVTEFLPDPDYPLKSTTIRLGEEAGLILDKVEGNLWHYT 171

Figure S4

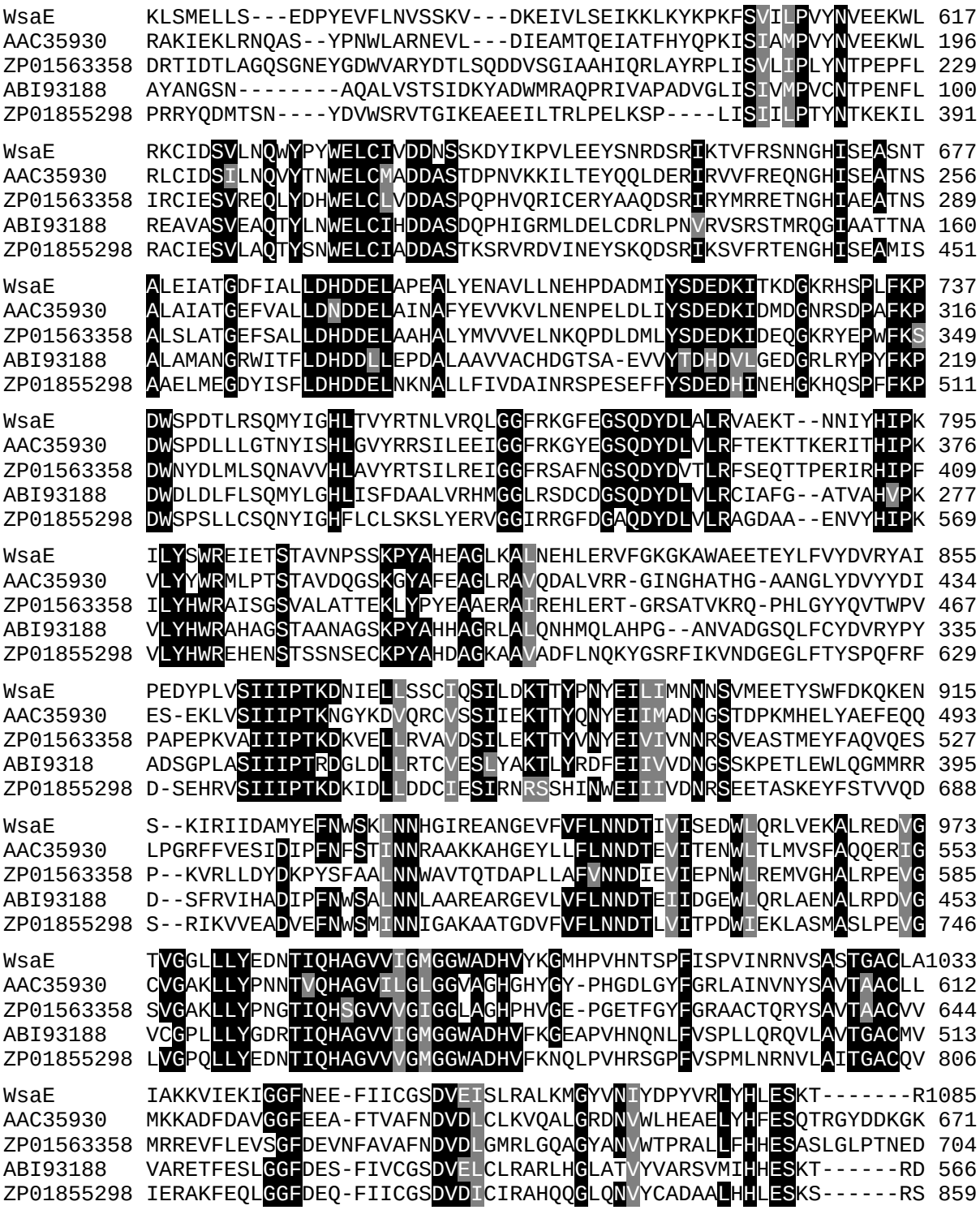

WsaE KLSMELLS---EDPYEVFLNVSSKV---DKEIVLSEIKKLKYKPKFSVILPVYNVEEKWL 617
AAC35930 RAKIEKLRNQAS--YPNWLARNEVL---DIEAMTQEIATFHYQPKISIAMPVYNVEEKWL 196
ZP01563358 DRTIDTLAGQSGNEYGDWVARYDTLSQDDVSGIAAHIQRLAYRPLISVLIPLYNTPEPFL 229
ABI93188 AYANGSN--------AQALVSTSIDKYADWMRAQPRIVAPADVGLISIVMPVCNTPENFL 100
ZP01855298 PRRYQDMTSN----YDVWSRVTGIKEAEEILTRLPELKSP----LISIILPTYNTKEKIL 391
WsaE RKCIDSVLNQWYPYWELCIVDDNSSKDYIKPVLEEYSNRDSRIKTVFRSNNGHISEASNT 677
AAC35930 RLCIDSILNQVYTNWELCMADDASTDPNVKKILTEYQQLDERIRVVFREQNGHISEATNS 256
ZP01563358 IRCIESVREQLYDHWELCLVDDASPQPHVQRICERYAAQDSRIRYMRRETNGHIAEATNS 289
ABI93188 REAVASVEAQTYLNWELCIHDDASDQPHIGRMLDELCDRLPNVRVSRSTMRQGIAATTNA 160
ZP01855298 RACIESVLAQTYSNWELCIADDASTKSRVRDVINEYSKQDSRIKSVFRTENGHISEAMIS 451
WsaE ALEIATGDFIALLDHDDELAPEALYENAVLLNEHPDADMIYSDEDKITKDGKRHSPLFKP 737
AAC35930 ALAIATGEFVALLDNDDELAINAFYEVVKVLNENPELDLIYSDEDKIDMDGNRSDPAFKP 316
ZP01563358 ALSLATGEFSALLDHDDELAAHALYMVVVELNKQPDLDMLYSDEDKIDEQGKRYEPWFKS 349
ABI93188 ALAMANGRWITFLDHDDLLEPDALAAVVACHDGTSA-EVVYTDHDVLGEDGRLRYPYFKP 219
ZP01855298 AAELMEGDYISFLDHDDELNKNALLFIVDAINRSPESEFFYSDEDHINEHGKHQSPFFKP 511
WsaE DWSPDTLRSQMYIGHLTVYRTNLVRQLGGFRKGFEGSQDYDLALRVAEKT--NNIYHIPK 795
AAC35930 DWSPDLLLGTNYISHLGVYRRSILEEIGGFRKGYEGSQDYDLVLRFTEKTTKERITHIPK 376
ZP01563358 DWNYDLMLSQNAVVHLAVYRTSILREIGGFRSAFNGSQDYDVTLRFSEQTTPERIRHIPF 409
ABI93188 DWDLDLFLSQMYLGHLISFDAALVRHMGGLRSDCDGSQDYDLVLRCIAFG--ATVAHVPK 277
ZP01855298 DWSPSLLCSQNYIGHFLCLSKSLYERVGGIRRGFDGAQDYDLVLRAGDAA--ENVYHIPK 569
WsaE ILYSWREIETSTAVNPSSKPYAHEAGLKALNEHLERVFGKGKAWAEETEYLFVYDVRYAI 855
AAC35930 VLYYWRMLPTSTAVDQGSKGYAFEAGLRAVQDALVRR-GINGHATHG-AANGLYDVYYDI 434
ZP01563358 ILYHWRAISGSVALATTEKLYPYEAAERAIREHLERT-GRSATVKRQ-PHLGYYQVTWPV 467
ABI93188 VLYHWRAHAGSTAANAGSKPYAHHAGRLALQNHMQLAHPG--ANVADGSQLFCYDVRYPY 335
ZP01855298 VLYHWREHENSTSSNSECKPYAHDAGKAAVADFLNQKYGSRFIKVNDGEGLFTYSPQFRF 629
WsaE PEDYPLVSIIIPTKDNIELLSSCIQSILDKTTYPNYEILIMNNNSVMEETYSWFDKQKEN 915
AAC35930 ES-EKLVSIIIPTKNGYKDVQRCVSSIIEKTTYQNYEIIMADNGSTDPKMHELYAEFEQQ 493
ZP01563358 PAPEPKVAIIIPTKDKVELLRVAVDSILEKTTYVNYEIVIVNNRSVEASTMEYFAQVQES 527
ABI9318 ADSGPLASIIIPTRDGLDLLRTCVESLYAKTLYRDFEIIVVDNGSSKPETLEWLQGMMRR 395
ZP01855298 D-SEHRVSIIIPTKDKIDLLDDCIESIRNRSSHINWEIIIVDNRSEETASKEYFSTVVQD 688
WsaE S--KIRIIDAMYEFNWSKLNNHGIREANGEVFVFLNNDTIVISEDWLQRLVEKALREDVG 973
AAC35930 LPGRFFVESIDIPFNFSTINNRAAKKAHGEYLLFLNNDTEVITENWLTLMVSFAQQERIG 553
ZP01563358 P--KVRLLDYDKPYSFAALNNWAVTQTDAPLLAFVNNDIEVIEPNWLREMVGHALRPEVG 585
ABI93188 D--SFRVIHADIPFNWSALNNLAAREARGEVLVFLNNDTEIIDGEWLQRLAENALRPDVG 453
ZP01855298 S--RIKVVEADVEFNWSMINNIGAKAATGDVFVFLNNDTLVITPDWIEKLASMASLPEVG 746
WsaE TVGGLLLYEDNTIQHAGVVIGMGGWADHVYKGMHPVHNTSPFISPVINRNVSASTGACLA1033
AAC35930 CVGAKLLYPNNTVQHAGVILGLGGVAGHGHYGY-PHGDLGYFGRLAINVNYSAVTAACLL 612
ZP01563358 SVGAKLLYPNGTIQHSGVVVGIGGLAGHPHVGE-PGETFGYFGRAACTQRYSAVTAACVV 644
ABI93188 VCGPLLLYGDRTIQHAGVVIGMGGWADHVFKGEAPVHNQNLFVSPLLQRQVLAVTGACMV 513
ZP01855298 LVGPQLLYEDNTIQHAGVVVGMGGWADHVFKNQLPVHRSGPFVSPMLNRNVLAITGACQV 806
WsaE IAKKVIEKIGGFNEE-FIICGSDVEISLRALKMGYVNIYDPYVRLYHLESKT-------R1085
AAC35930 MKKADFDAVGGFEEA-FTVAFNDVDLCLKVQALGRDNVWLHEAELYHFESQTRGYDDKGK 671
ZP01563358 MRREVFLEVSGFDEVNFAVAFNDVDLGMRLGQAGYANVWTPRALLFHHESASLGLPTNED 704
ABI93188 VARETFESLGGFDES-FIVCGSDVELCLRARLHGLATVYVARSVMIHHESKT------RD 566
ZP01855298 IERAKFEQLGGFDEQ-FIICGSDVDICIRAHQQGLQNVYCADAALHHLESKS------RS 859
WsaE DSFIPERDFELSAKYYSP-YREIGDPYYNQNLSYNHLIPTIRS 1127
AAC35930 KKKRFEQEKVMMEEKWGP-LIEN-DPFYNPNLTRDIP------ 706
ZP01563358 RRRQFLEECDNFRRIWAD-VIRN-DPFYNPNLTISGGDFRPNF 762
ABI93188 PREIPESDFVRSAQAYSPYREE-GDPFFSPNLDYMASSPRLRG 611
ZP01855298 S-FIPKQDFLMSEIRYAPYRNDKGDPYFNENLDLMSTMPRMLT 903

Figure S5

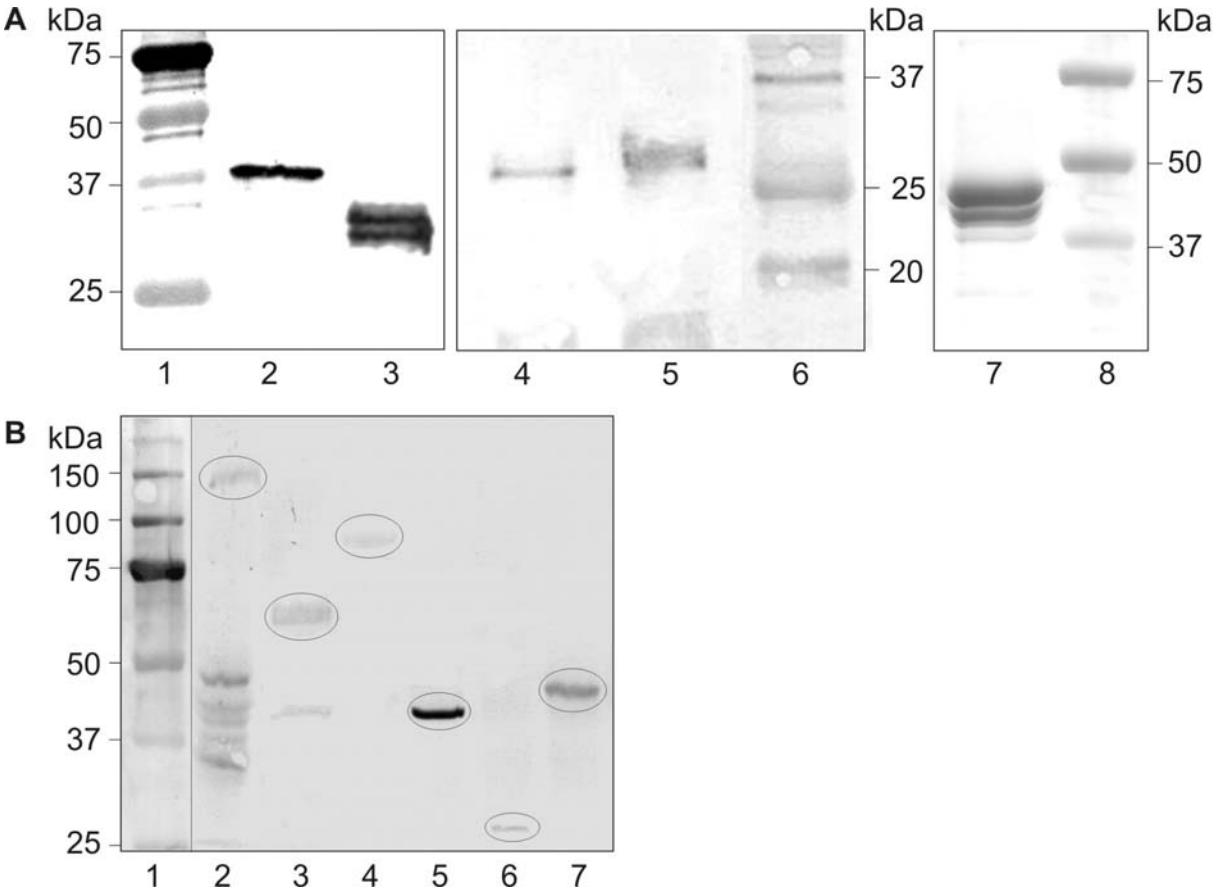

A
kDa
75
50
37
25
1
2
3
kDa
37
25
20
4
5
6
kDa
75
50
37
7
8
B
kDa
150
100
75
50
37
25
1
2
3
4
5
6
7

Figure S6

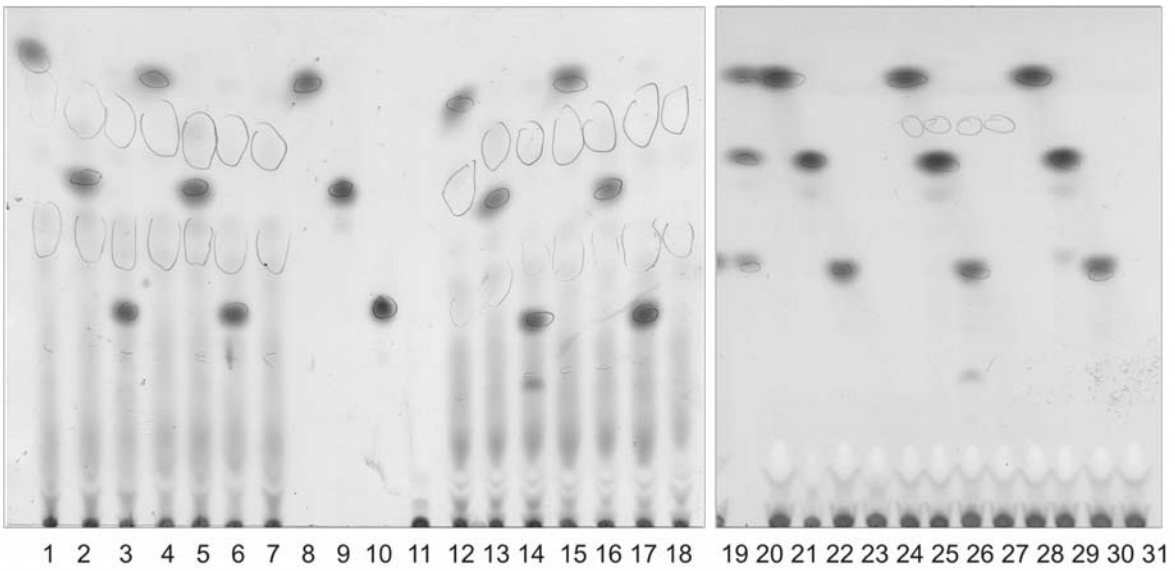
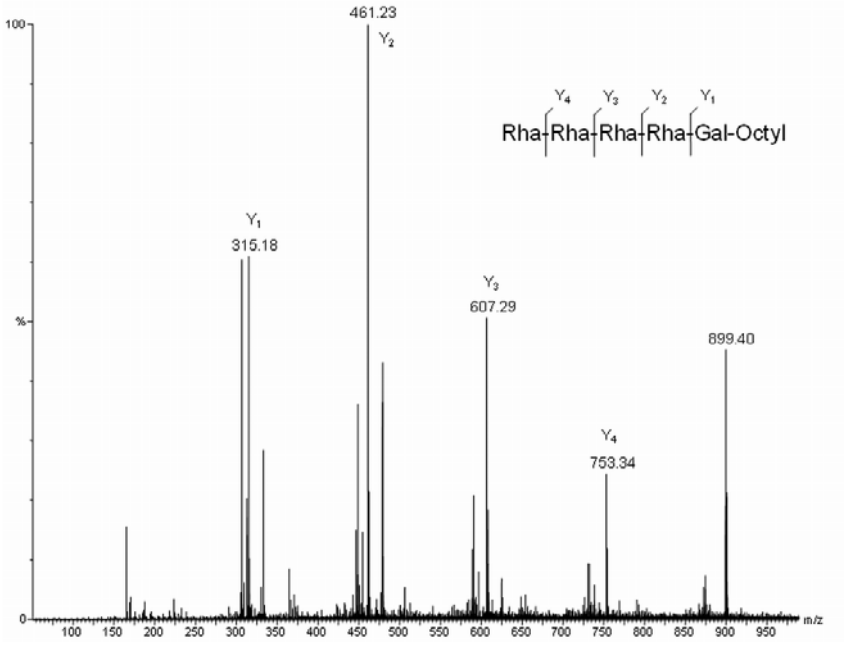

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Figure S7


461.23
Y2
Y4 Y3 Y2 Y1
Rha-Rha-Rha-Rha-Gal-Octyl
Y1
315.18
Y3
607.29
899.40
Y4
753.34
100
%
0
100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950
m/z