Supporting Information for:

Functionally Validated Proteome-Wide Bioinformatic Annotation of the Monotopic Phosphoglycosyl Transferase Family

Theo Durand, [a,b] Greg J. Dodge, [a,c] Roxanne P. Siuda, [d,e] Hugh R. Higinbotham,[a] Christine A. Arbour,[a] Soumi Ghosh,[a] Karen N. Allen,[d] Barbara Imperiali[a,*]

[a]Department of Biology and Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

[b]Imperial College London, Exhibition Rd, South Kensington, London SW7 2AZ, UK

[c]Current address Biogen, 225 Binney Street, Cambridge MA 02139, USA

[d]Department of Chemistry, Boston University, 590 Commonwealth Ave, Boston MA 02215, USA

[e] Dept. of Pharmacology Physiology, and Biophysics, Boston University Chobanian & Avedisian School of Medicine, 72 E Concord St L-630D, Boston, MA 02215, USA

*Corresponding author: imper@mit.edu

Theo Durand	0009-0004-8131-581X	Dept. of Biology, MIT, Cambridge MA 02139, USA. Dept. of Life Sciences, ICL, London SW7 2AZ, UK
Greg J. Dodge	0000-0002-6555-8350	Biogen, 225 Binney Street, Cambridge MA 02139, USA
Roxanne P. Siuda	0000-0001-7387-8429	Dept. of Chemistry, BU and Dept. of Pharmacology, Physiology and Biophysics, BUSM
Hugh R. Higinbotham	0000-0002-9418-7958	Dept. of Biology and Dept. of Physics, MIT
Christine A. Arbour	0000-0001-6056-296X	Dept. of Biology and Dept. of Chemistry, MIT
Soumi Ghosh	0000-0001-6101-0147	Dept. of Biology, MIT
Karen N. Allen	0000-0001-7296-0551	Dept. of Chemistry, BU
Barbara Imperiali	0000-0002-5749-7869	Dept. of Biology and Dept. of Chemistry, MIT

Table of contents

Table S1: Comprehensive table of smPGT assignments based on literature precedent

Organism	PGT Name	Uniprot	Putative Substrate(s)	E-71 cluster	Glycoconjugate	Reference
R. etli	WreU	Q2K1T1	KdgNAc	5	O-antigen/LPS	(1)
F. tularensis schu S4	WbtB	Q5NEZ2	KdgNAc/ QuiNAc	4	O-antigen/LPS	(2, 3)
S. suis serotype 9	Cps9F	Q9RG41	KdgNAc	4	CPS9	(4)
S. pneumoniae serotype 5		Q7WVW9	KdgNAc	4	CPS5	(5)
S. aureus	Cap5M	P95706	KdgNAc/ FucNAc	50	Type 5 CPS	(6, 7)
S. saprophyticus		Q4A111	KdgNAc	50	CPS	(8)
F. nucleatum strain 25586		Q8R6F8	QuiNAc	132	O-antigen/LPS	(9)
C. botulinum		A0A6B4JHH5	FucNAc4N	10	N/A	
B. fragilis	WcfS	Q93QV6	FucNAc4N	10	CPSA	(10)
F. nucleatum ATCC 23726		D5RFH4	FucNAc4N	10	O-antigen/LPS	(11)
F. nucleatum ATCC 51191		F9EN55	FucNAc4N	10	O-antigen/LPS	(12)
F. nucleatum ATCC 10953		A5TVH5	FucNAc4N	10	O-antigen/LPS	(13)
F. nucleatum HM-997 or CTI-07		A0A829KYS8	FucNAc4N	10	O-antigen/LPS	(14)
F. nucleatum HM-994 or CTI03		ERT37135.1	FucNAc4N		O-antigen/LPS	(14)
F. nucleatum MJR 7757		A0A133NN30	FucNAc4N	10	O-antigen/LPS	(15)
S. sonnei		Q3YTB3, Q9S0U8	FucNAc4N	17	Phase I polysaccharide	(16, 17)
P. alcalifaciens		A0A346CL59M 9P0X8, M9P183	FucNAc4N	17	O-antigen O22 and O8	(18-21)
B. bronchiseptica ATCC BAA-588,		A0A0R4J6E5	FucNAc4N	17	Endotoxin/LPS	(22, 23)
S. mitis strain B6		D3H7E6	FucNAc4N	6	Cell wall PS/LTA IV	(24, 25)
S. pneumoniae serotype 4		A0A0H2URP3	FucNAc4N	6	LTA IV	(26, 27)
S. suis serotype 7	Cps7F	Q9RFX2	FucNAc4N	6	LTA IV	(28)
S. pneumoniae serotype 1		A0A0H2ZRD5	FucNAc4N	6	LTA IV	(29)
H. parainfluenzae T3T1		E1W1Z5	FucNAc4N		LPS	(30)
H. parainfluenzae strain 20		R9WQP1	FucNAc4N	10	LPS	(31)
B. fragilis strain 638R		E1WLL6	FucNAc4N	21	PS1	(32, 33)
R. meliloti	ExoY	Q02731	Gal	3	Succinoglycan	(34)
R. fredii	ExoY2	G9AFL7	Gal	3	LPS	(35)
M. japonicum		*Q98C89*	Gal	3	LPS	(35)

A. baumannii B8300	ItrA4	A0A2I8CW33	Gal	3	CPS	(36)
A. baumannii B850	ItrA4	A0A7S8F8L0	Gal	3	CPS	(37)
L. rhamnosus GG	EpsE	C1J9J2, A0A809NCL8	Gal	14	EPS	(38)
R. leguminsarum	PssA	Q52856	Glc	18	EPS	(35)
R. legum. Bv trifolii	PssA	B2Z9S5/A0A9X5D0X6	Glc	18	EPS	(35)
L. johnsonii FI9785	EpsE	D0R4M7	Glc	2	EPS	(39)
V. parahaemo-lyticus serotype O3:K6	SypR	Q87PP3	Unknown	20	Structural symbiosis PS	
C. difficile	CD2783	Q183M0	GalNAc	9	PS-II	(40, 41)
V. vulnificus M06-24	WbfU	A0A4Q7IER7	D-KdgNAc/ L-KdgNAc	7	CPS	(42)
A. fischeri		Q5E8F5	D-KdgNAc/ L-KdgNAc	7	O-antigen/LPS	(43)

Table S2: Common and alternate names of biochemically-modified UDP-sugar substrates for monoPGTs

Sugar, name used in this work	**Alternative names**
KdgNAc	2-acetamido-2,6-dideoxy-D-xylose-hexos-4-ulose (**Sug**) (4, 44) 2-acetamido-2,6-dideoxy-D-xylo-4-hexulose (**ADHexu**) (3) 4-keto sugar (45)
FucNAc4N	2-acetamido-4-amino-2,4,6-trideoxy-β-galactopyranoside (14) 4-*N*-D-FucNAc 2-acetamido-4-amino-2,4,6-trideoxy-D-galactose (16) D-FucNAc4N: 2-acetamido-4-amino-2,4-dideoxy-D-fucose (46) α-D-Fuc*p*NAc4NR (R is partial acetylation of the sugar) 2,4-diamino-2,4,6-trideoxydeoxy-D-galactose (12) 4-amino-D-FucNAc (47) 2-acetamido-4-amino-2,4,6-trideoxy-D-galactose (**AATGal**) (48) 2-acetamido-4-amino-6-deoxygalactopyranose (**AADGal**) (10)
diNAcBac	*N,N*′-diacetylbacillosamine 2,4-diacetamido-2,4,6-trideoxy-D-glucose (49) 2,4-diNAc: 2,4-diacetamido-2,4,6-trideoxy-α-D-glucopyranose (50) di-*N*-acetyl D-bacillosamine (**D-Bac**) (51) 2,4-diamino-2,4,6-trideoxyhexoses (**DATDH**) (52) 2,4-diacetamido-2,4,6-trideoxy-D-glucose (**QuiNAc4NAc**) (53) 2-acetamido-4-acylamino-2,4,6-trideoxy-D-glucose (**D-QuiNAc4NR**) (54)

Table S3: The 95% representative node (repnode) SSN cluster count for assignment of unique SmPGT sequences. The percent of SmPGT sequence space predicted based on our assignments is: whole network node number divided by the number of nodes in the assigned cluster = 44.3%.

Clusters assigned at 95% repnode	Number of Nodes
Whole network	24677
Cluster 1	4514
Cluster 2	1373
Cluster 3	645
Cluster 4	753
Cluster 5	705
Cluster 6	367
Cluster 7	601
Cluster 8	477
Cluster 9	295
Cluster 10	367
Cluster 17	242
Cluster 18	138
Cluster 21	175
Cluster 28	105
Cluster 31	101
Cluster 38	76

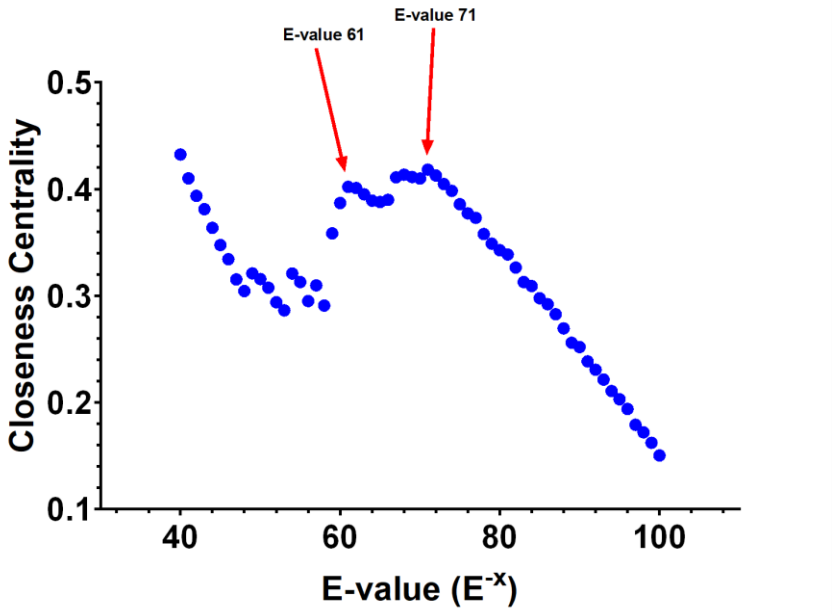


Figure S1: Closeness Analysis of SSN showing inflection points at E-value 61 and 71. Closeness centrality was calculated for 75% repnode networks (sequences above a 75% identity collapsed into single nodes) across E-values.

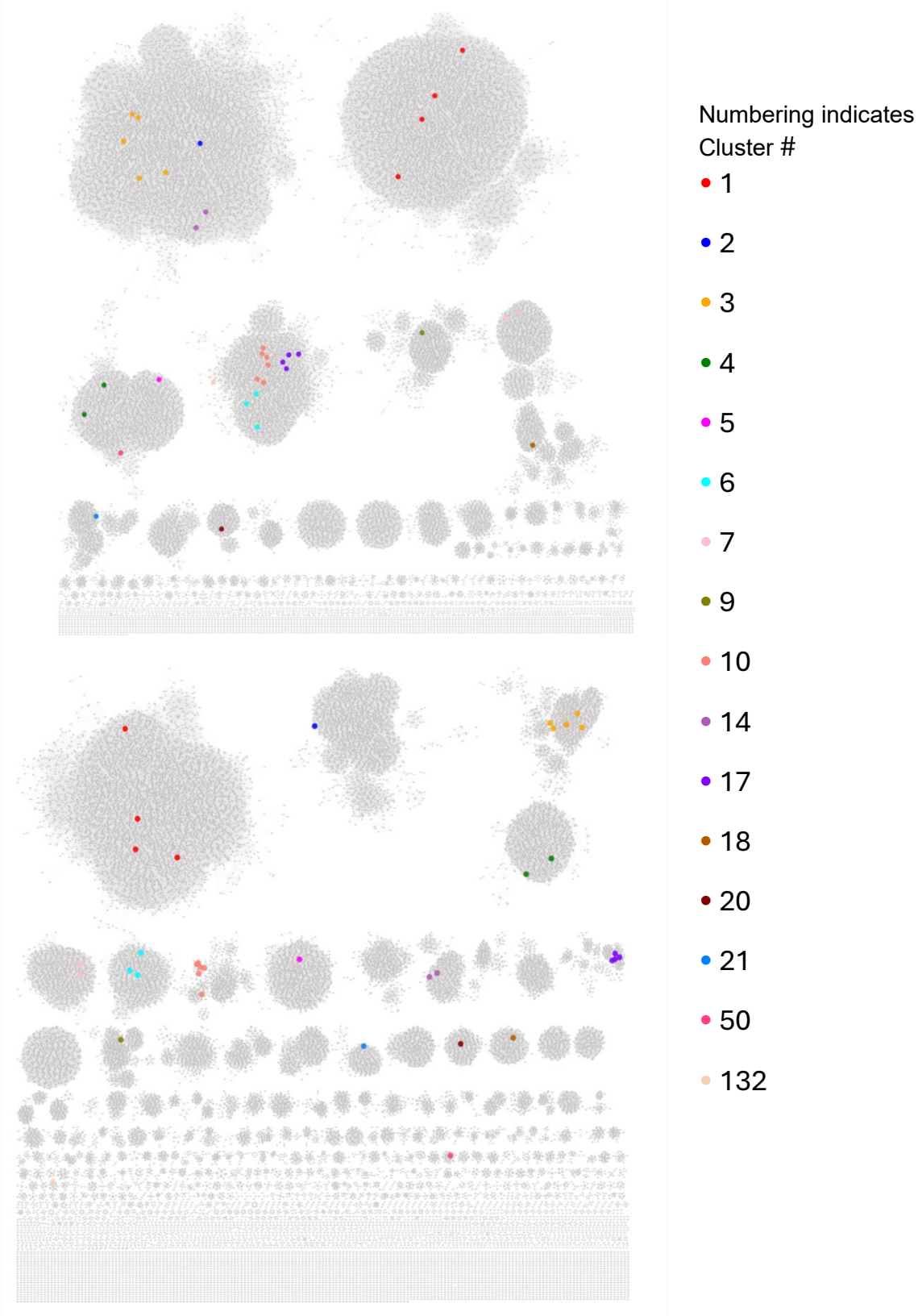


Figure S2: SSN Positions of SmPGTs presented in the manuscript. **Top**: E-61 network, **Bottom**: E-71 network. SmPGTs with putative assignment based on Table 1 and S2 are highlighted and color-coded in the E-61 and E-71 networks.

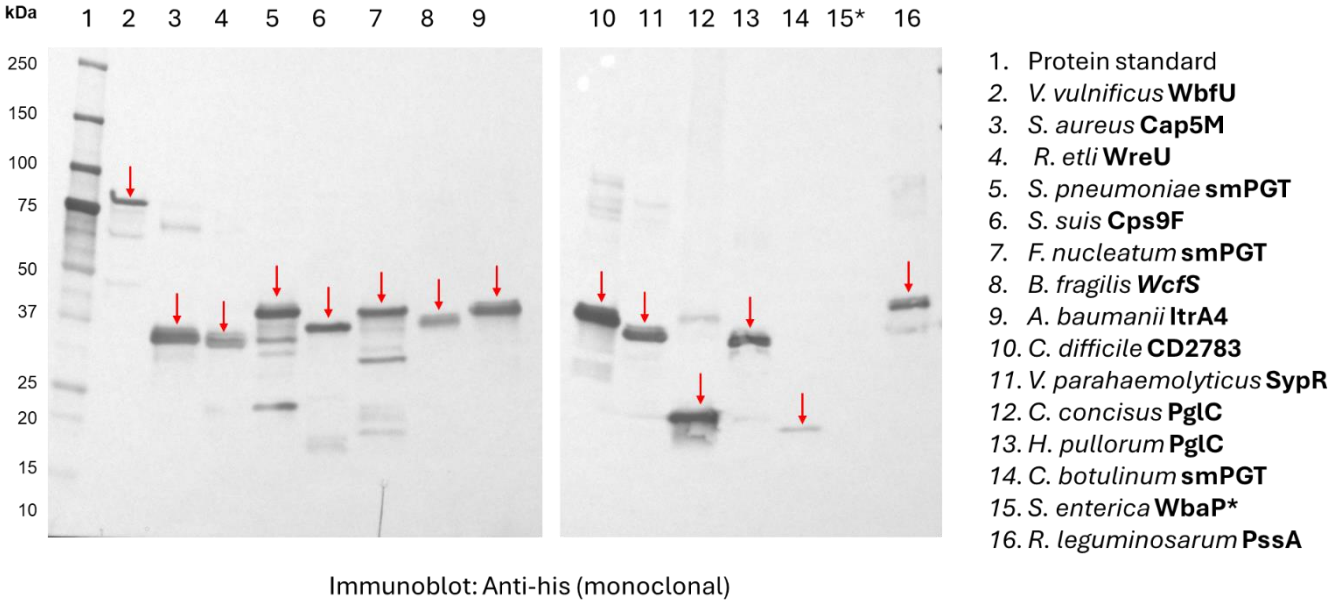


Figure S3: Expression of the SmPGTs in unfractionated cell envelope fractions (CEF) by western blot analysis: The CEF of SmPGTs (10 µg total protein, based on 280 nm measurement) were separated by gel electrophoresis, transferred to a nitrocellulose membrane, and probed with monoclonal anti-his antibody (LifeTein). Red arrows indicate the presence of protein at expected molecular weights. **S. enterica* WbaP was not observed on the western blot as it was expressed with a twin Strep-tag in the CEF.

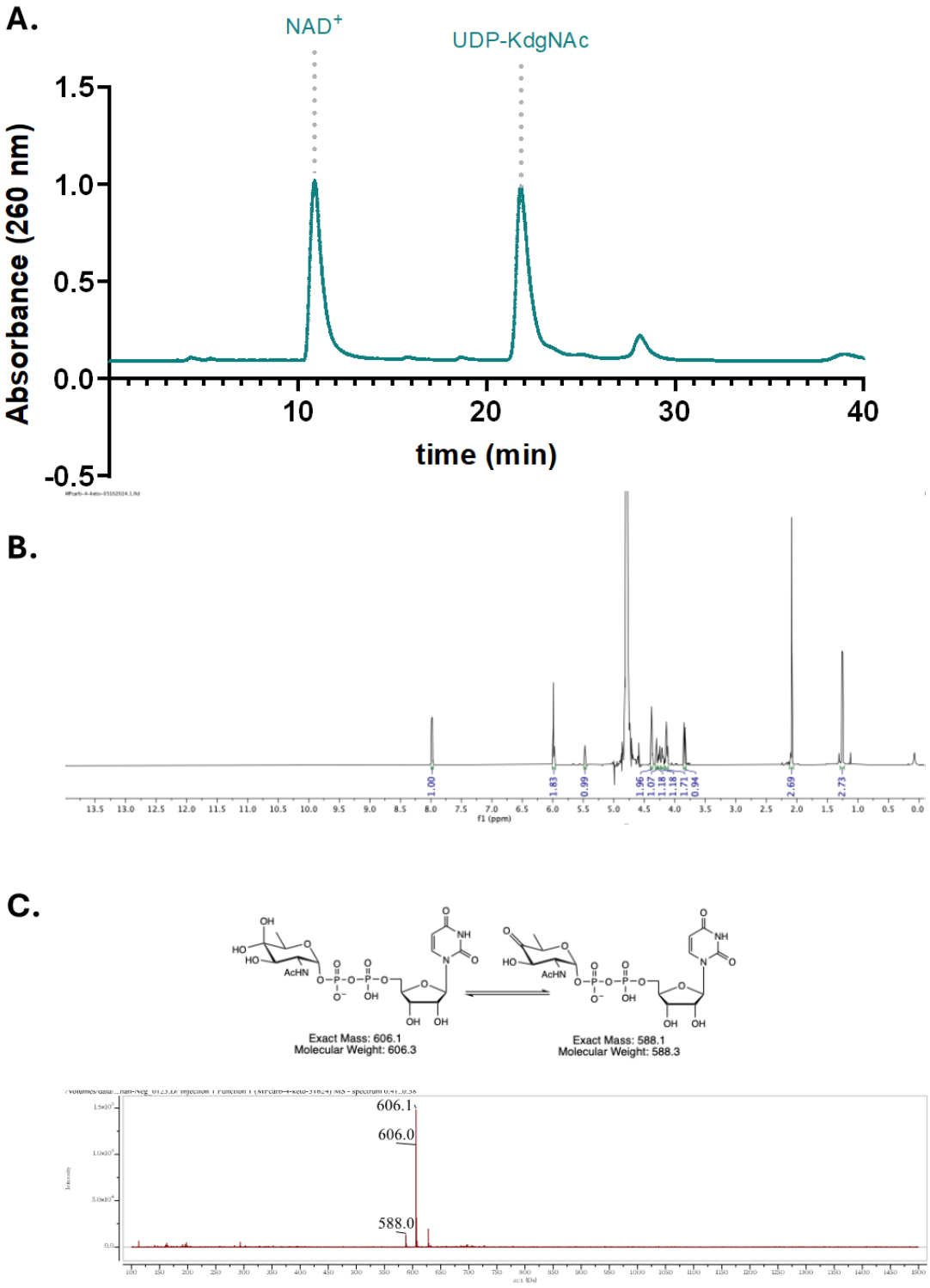


Figure S4: Characterization of UDP-KdgNAc from chemoenzymatic synthesis. **A**. A 16 µL injection of 20mM UDP-KdgNAc reaction material using anion-exchange HPLC with gradient A. The first peak is hypothesized to be NAD^+ based on equimolar addition of NAD^+ and UDP-KdgNAc in the reaction mixture. UDP-KdgNAc was isolated over multiple rounds of purification and elution fractions were lyophilized. **B**. ^1H NMR spectrum (600 MHz, D_2O) of 1.5 mg of lyophilized elution fractions, confirming identity of the isolated peak as UDP-KdgNAc based on comparison with published spectra (55). **C.** LRMS ESI(-) of lyophilized elutions and exact molecular weight predictions of the hydrated (left) and un-hydrated (right) UDP-KdgNAc species which exist in equilibrium in solution. LRMS ESI(-) confirms the presence of both species in agreement with NMR.

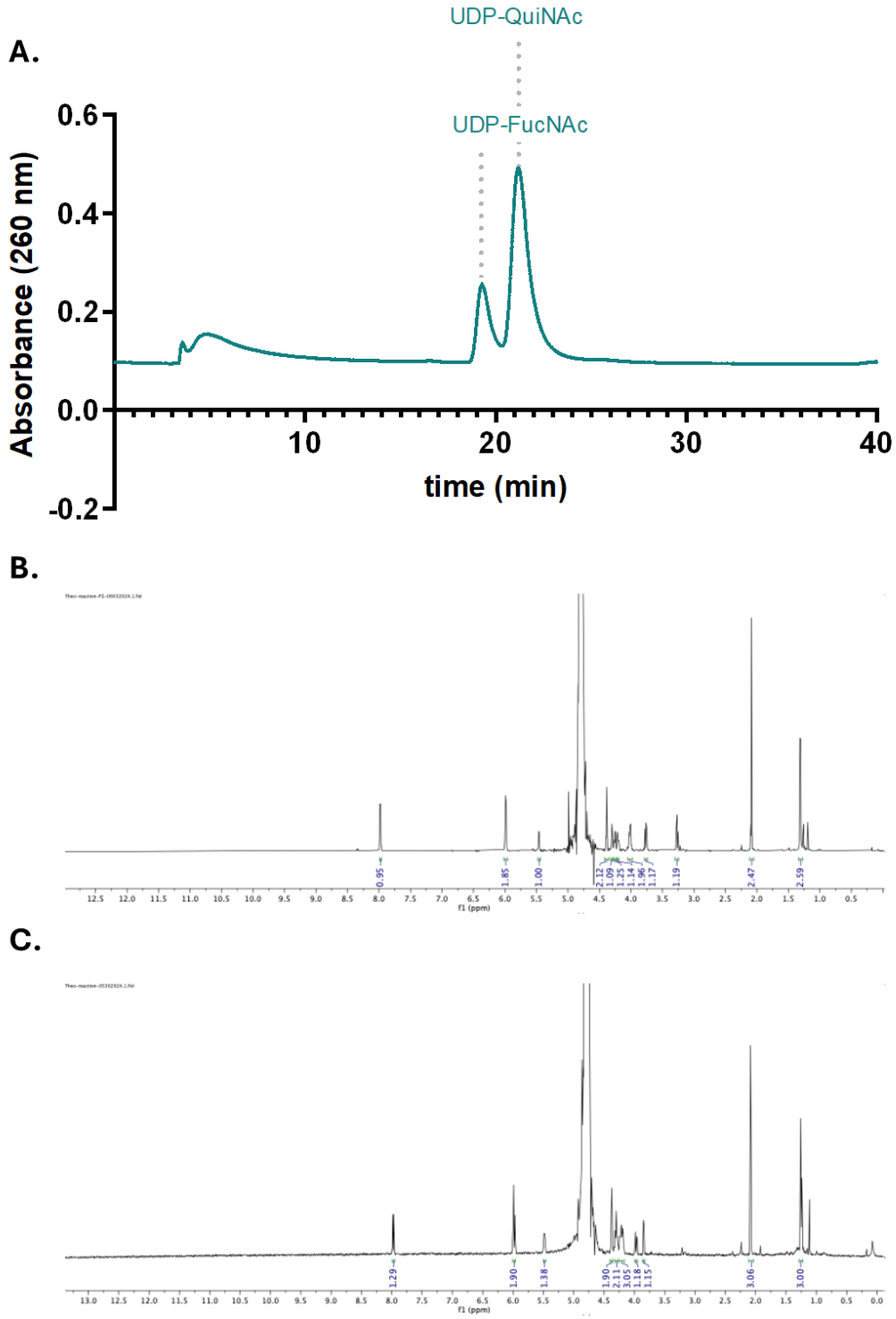


Figure S5: Characterization of UDP-FucNAc and UDP-QuiNAc from chemoenzymatic synthesis. **A**. A 400 µL injection of UDP-QuiNAc/UDP-FucNAc reduction reaction using anion-exchange HPLC with gradient B. The broad peak at retention time ~5 min is hypothesized to be residual acetone. The major peaks between 18-23 min were individually isolated over several rounds of purification and lyophilized for analysis generating 0.06 mg of peak 1 and 0.18 mg of peak 2. **B**. ^{1}H NMR spectrum (600 MHz, D_2O) of 0.18 mg of lyophilized peak 2 which was compared to published spectra to identify as UDP-QuiNAc (55). **C**. ^{1}H NMR spectrum (600 MHz, D_2O) of 0.06 mg of lyophilized peak 1 which was compared to published spectra to identify as UDP-FucNAc (56).

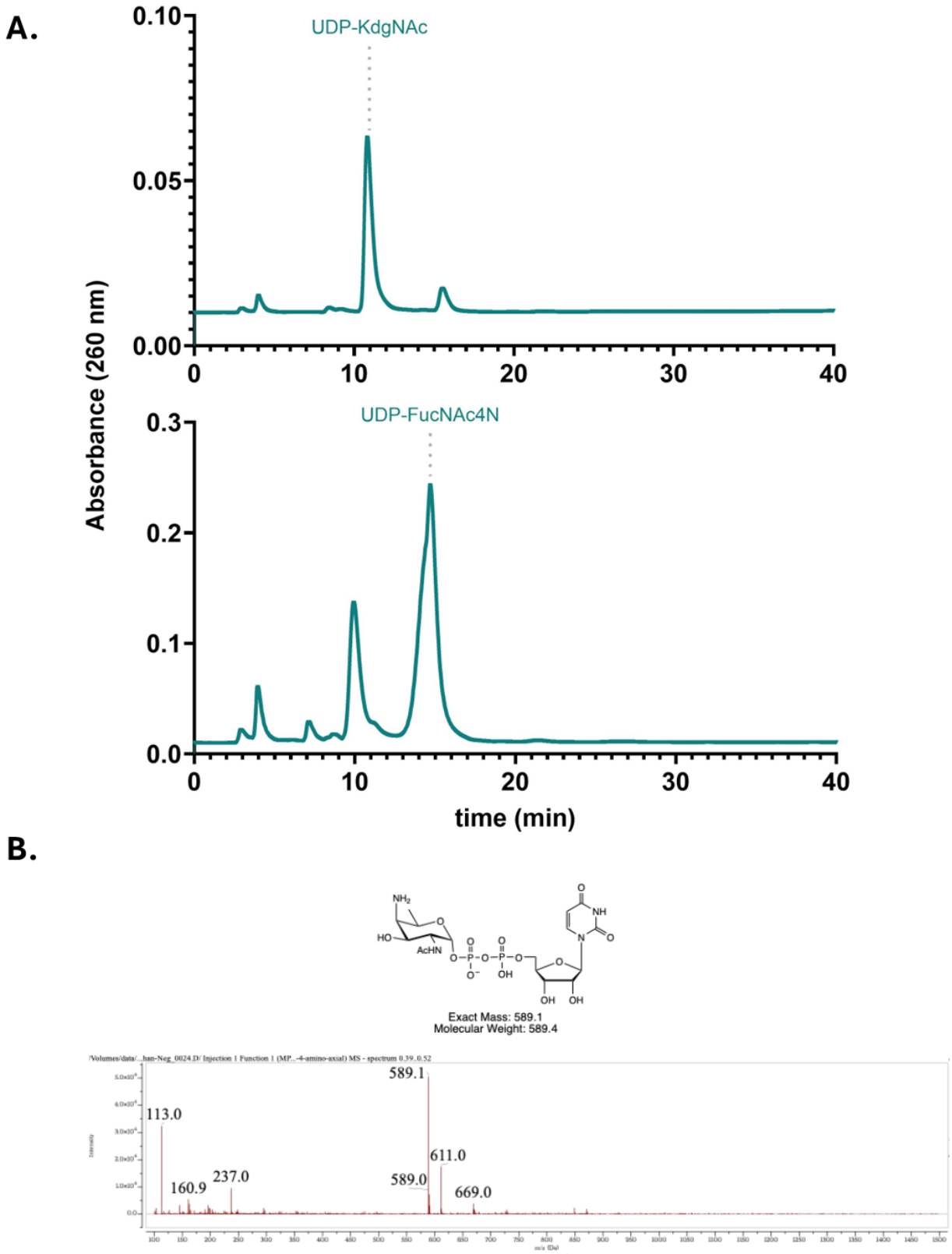


Figure S6: Characterization of UDP-FucNAc4N from chemoenzymatic synthesis. **A**. top: A 10 µL injection of WcfR reaction starting material by anion-exchange HPLC on gradient C, bottom: A 100 µL injection of WcfR reaction after incubation with gradient C. **B**. LRMS ESI(-) of WcfR reaction elution fractions and exact molecular weight predictions of the UDP-FucNAc4N, 611.0 m/z peak is thought to be the FucNAc4N sodium adduct.

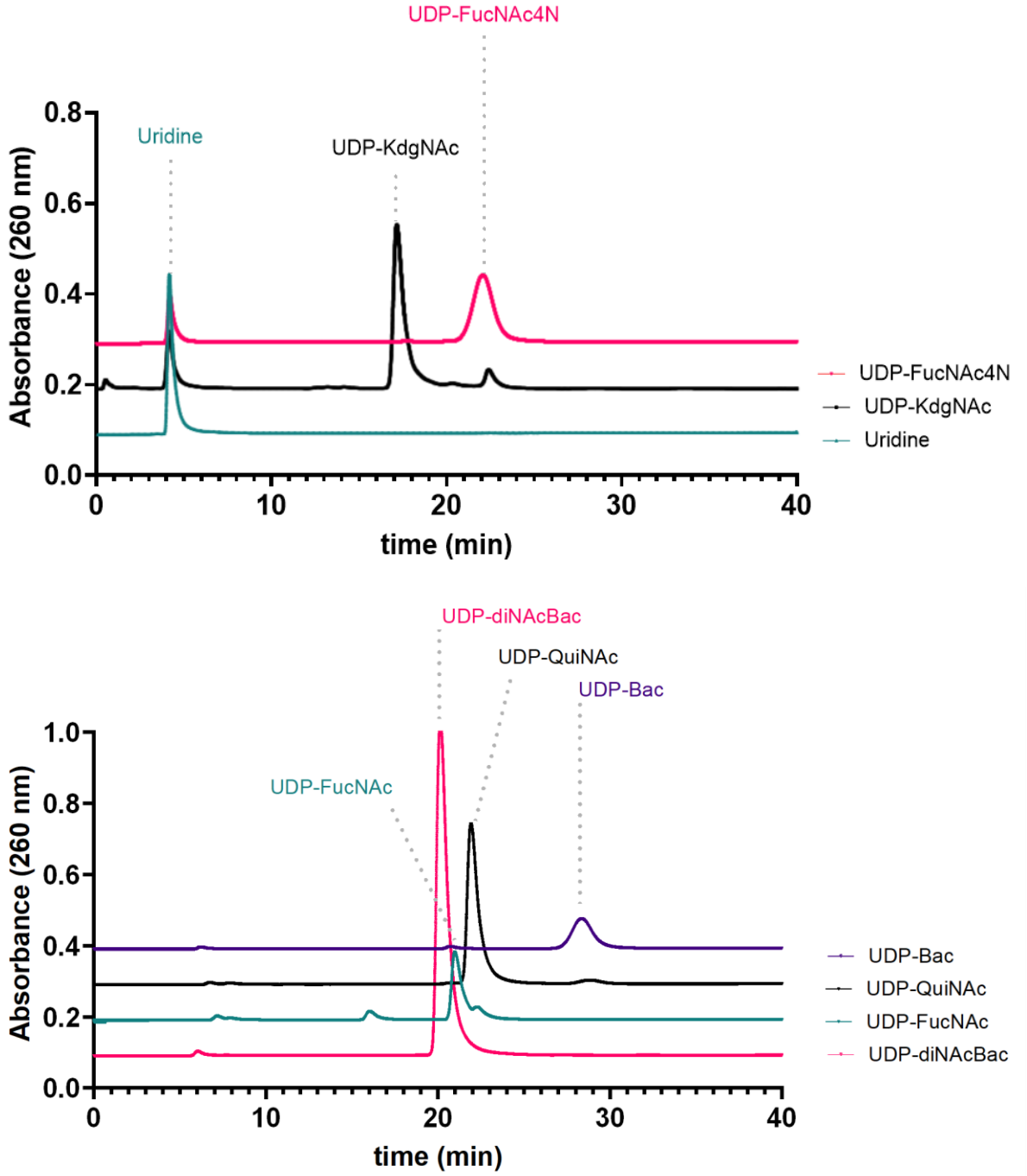


Figure S7: Calf Intestinal Alkaline Phosphatase (CIAP) treatment for removal of nucleotide contaminants. **Top**. A 200 µL injection of 500 µM CIAP treated UDP-FucNAc4N and UDP-KdgNAc and 5 µL injection of 9 mM uridine on Gradient C. **Bottom**. Injections of CIAP-treated materials, 250 µL of 1mM diNAcBac, 200 µL of 250 µM FucNAc, 250 µL of 500 µM of UDP-QuiNAc and 200 µL of 500 µM UDP-Bac on Gradient C.

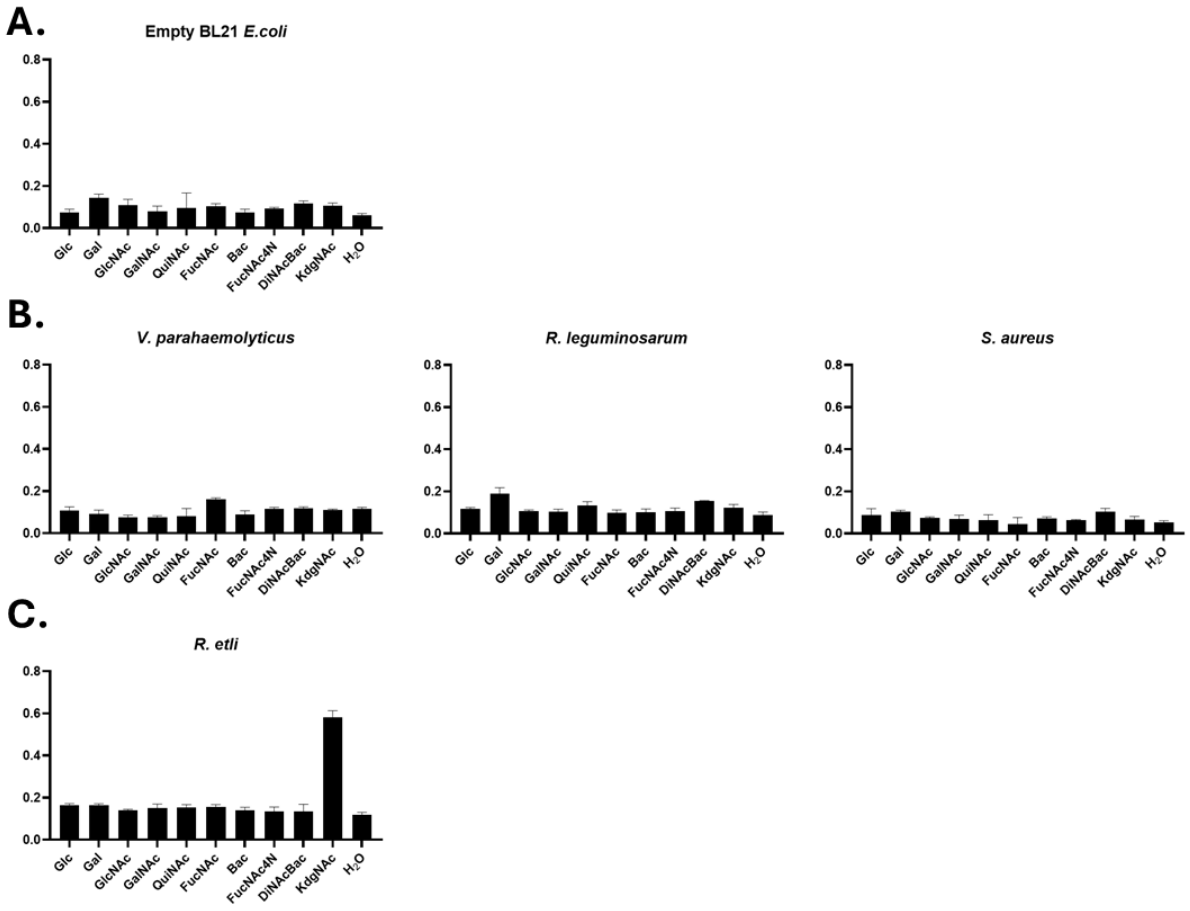


Figure S8: CEF-Glo control data. **A**. BL21 *E. coli* empty plasmid CEF controls. **B**. CEF-Glo screens of SmPGTs with no identified UDP-sugar substrate. **C**. Positive control with *R. etli* WreU.

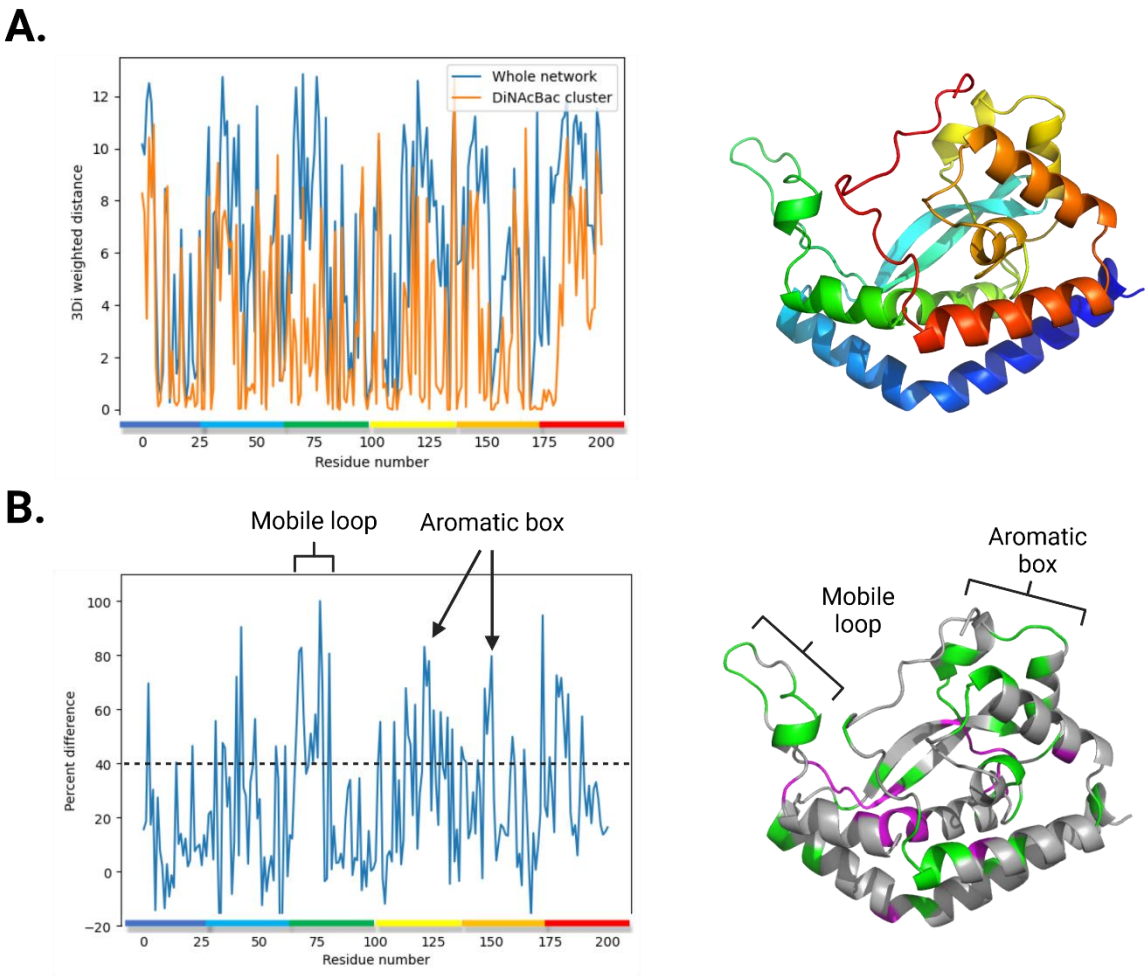


Figure S9: Example of structural conservation analysis for chosen smPGT. **A**. Euclidean dissimilarity for 3Di alignments at *C. concisus* PglC residues. Whole network alignment (blue) shows higher dissimilarity than the substrate cluster alignment (orange), particularly at prominent structural features like the mobile loop. Colors above residue numbers map to the protein structure on the right for visualization. **B**. Percent difference between curves in panel **A** highlights distinct substrate-specific structural features. Black dotted line shows 40% threshold used to color green residues on the lower right structure. Highlighted mobile loop and aromatic box correlate with previous studies on PglC (57, 58).

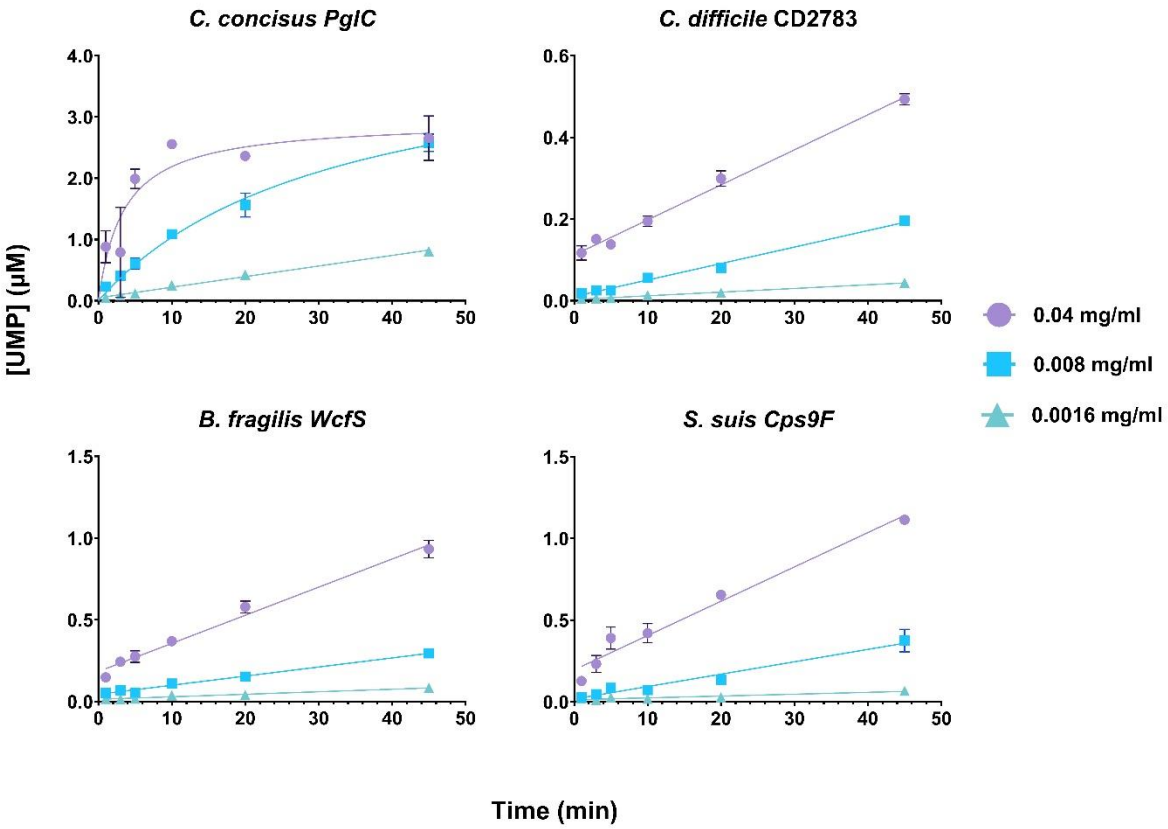


Figure S10: Kinetic analysis of selected SmPGTs in CEF.

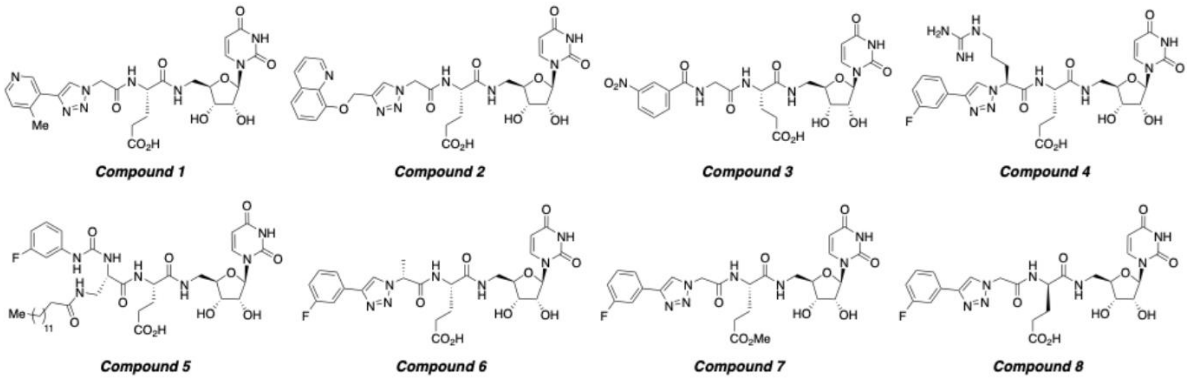


Figure S11: Structures of nucleoside analogs used for inhibitor screening.

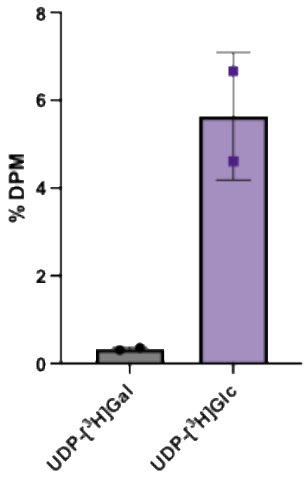


Figure S12: Radioactivity-based substrate assay with CEF containing overexpressed *R. leguminosarum* PssA. Assays were carried out as previously reported (59) with commercial UDP-[^{3}H]Hex sugar donors. Activity is reported as the percentage of disintegrations per minute (% DPM) in the organic layer normalized to the total disintegrations per minute per quench point. Error bars are given for mean ± standard deviation (SD), n = 2.

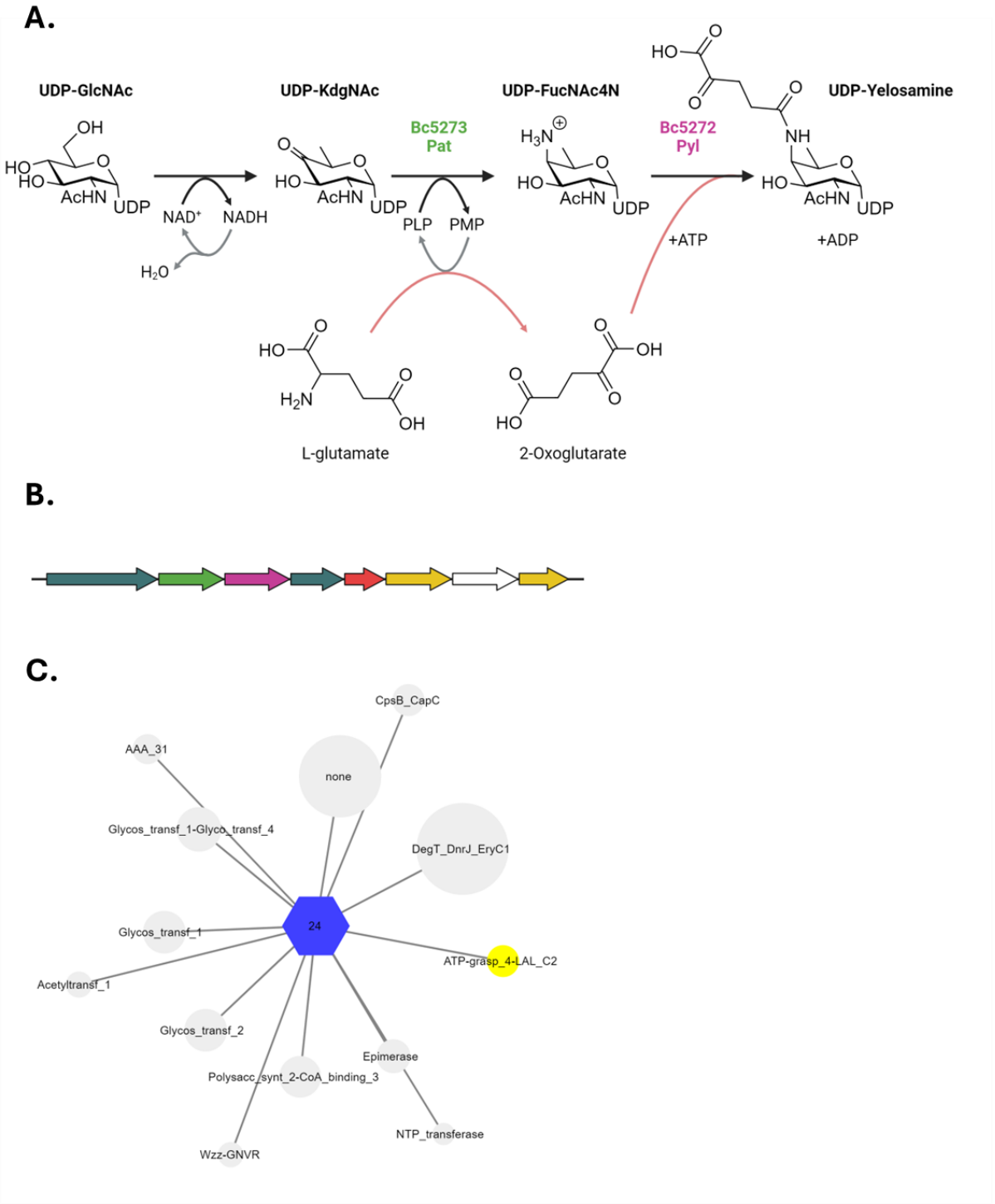


Figure S13: UDP-Yelosamine biosynthesis and operon analysis: **A**. Proposed biosynthetic route for generation of UDP-Yelosamine (47). **B**. Biosynthetic operon for secondary cell wall polysaccharide generation in *Bacillus cereus* ATCC 14579. SmPGT is shown in red, glycosyl transferases in yellow, UDP-biosynthesis enzymes in teal, with Pat and Pyl enzymes shown with same coloring as A., genes with unknown function are in white. **C**. Genome Neighbourhood Network of cluster 24, containing SmPGT shown in B., as hub-nodes and neighbouring Pfam familys as spoke-nodes. Highlighted in yellow is the frequent occurrence of the ATP-grasp (ATP-grasp_4-LAL_C2) domain protein (Pyl) within the cluster. The DegT_DnrJ_EryC1 aminotransferase Pfam family, of which Pat is a member, is also highly represented.

Supplementary Methods

Typical chemical synthesis scheme of nucleoside analogs

Uridine loading with BAL resin.

5'-Aminouridine was attached to BAL resin using a modified protocol (60, 61). BAL resin (Advanced Chemtech, 100-200 mesh, 0.6–1.2 mmol/g, 1% DVB) was weighed into a reaction vessel containing a frit (Torviq 10 mL Luer Lock Fritted Syringe from Fisher, Catalog No. NC9299151). The resin was swelled in 1% AcOH/DMF (1 g resin/10 mL) for 1 h. The solvent was removed, and the resin was treated with a 0.2 M mixture of 2',3' acetal-protected 5'-amino uridine (62) (2 equiv) in 1% AcOH/DMF. The reaction was agitated for 1 h (ambient temperature). Then a mixture of $NaBH_3CN$ (2 equiv) in MeOH (0.45 M) was added directly to the slurry of resin and agitated overnight. *Note: A closed reaction vessel is not recommended as production of gas is observed after the addition of $NaBH_3CN$. After 18 h, the solvent mixture was evacuated and the resin was rinsed with DMF, CH_2Cl_2, MeOH, and H_2O (2 x 5 mL each). The resin, in a slurry of H_2O, was frozen and lyophilized to provide the dried resin.

3-NO$_2$Ph (Compound 3). *(S)-5-((((2R,3S,4R,5R)-5-(2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)amino)-4-(2-(3-nitrobenzamido)acetamido)-5-oxopentanoic acid.* Resin (150 mg, loading: 0.53 mmol/g, 0.08 mmol, 1 equiv) was weighed into a 10 mL fritted syringe. The resin was swelled for 15 min in CH_2Cl_2 (5 mL). Then a mixture of Fmoc-Glu(OtBu)-OH (5 equiv), HBTU (5 equiv), and Hünig's base (10 equiv) in DMF (5 mL) was added to the resin. The mixture was agitated for 2 h (ambient temperature). After this time, the solvent was removed, and the resin was rinsed with DMF (2 x). Then a mixture of 20% piperidine/DMF (v/v) was added to the resin and agitated for 20 min. The solvent was removed, and the resin was rinsed with DMF (2 x). A mixture of Fmoc-Gly-OH (5 equiv), HBTU (5 equiv), and Hünig's base (10 equiv) in DMF (5 mL) was added to the resin. The mixture was agitated for 2 h (ambient temperature). Then a mixture of 20% piperidine/DMF (v/v) was added to the resin and agitated for 20 min. The solvent was removed, and the resin was rinsed with DMF (2 x). Then a mixture of carboxylic acid (5 equiv), HBTU (5 equiv), and Hünig's base (10 equiv) in DMF (10 mL) was added to the resin. The mixture was agitated for 2 h (ambient temperature). After this time, the solvent was removed, and the resin was rinsed with DMF and CH_2Cl_2 (2 x each). The crude product was cleaved from the resin with 2 mL TFA/TIPS/H_2O (95:2.5:2.5) for 2 h. The TFA cleavage was concentrated in volume under a stream of N_2 and the crude product was precipitated with –20 ºC diethyl ether. The resulting slurry was centrifuged, and the TFA/ether supernatant was decanted to yield the crude pellet. The pellet was resuspended in MeCN/H_2O and purified by preparative RP-HPLC (Luna 5 µm C_{18}(2) 100 Å, 250 x 21.2 mm Phenomenex column) with a gradient of 20-75% B over 25 min, flow rate: 10 mL/min [solvents A: H_2O (0.1% TFA), B: MeCN (0.1% TFA)]. The purified product was transferred to 50 mL

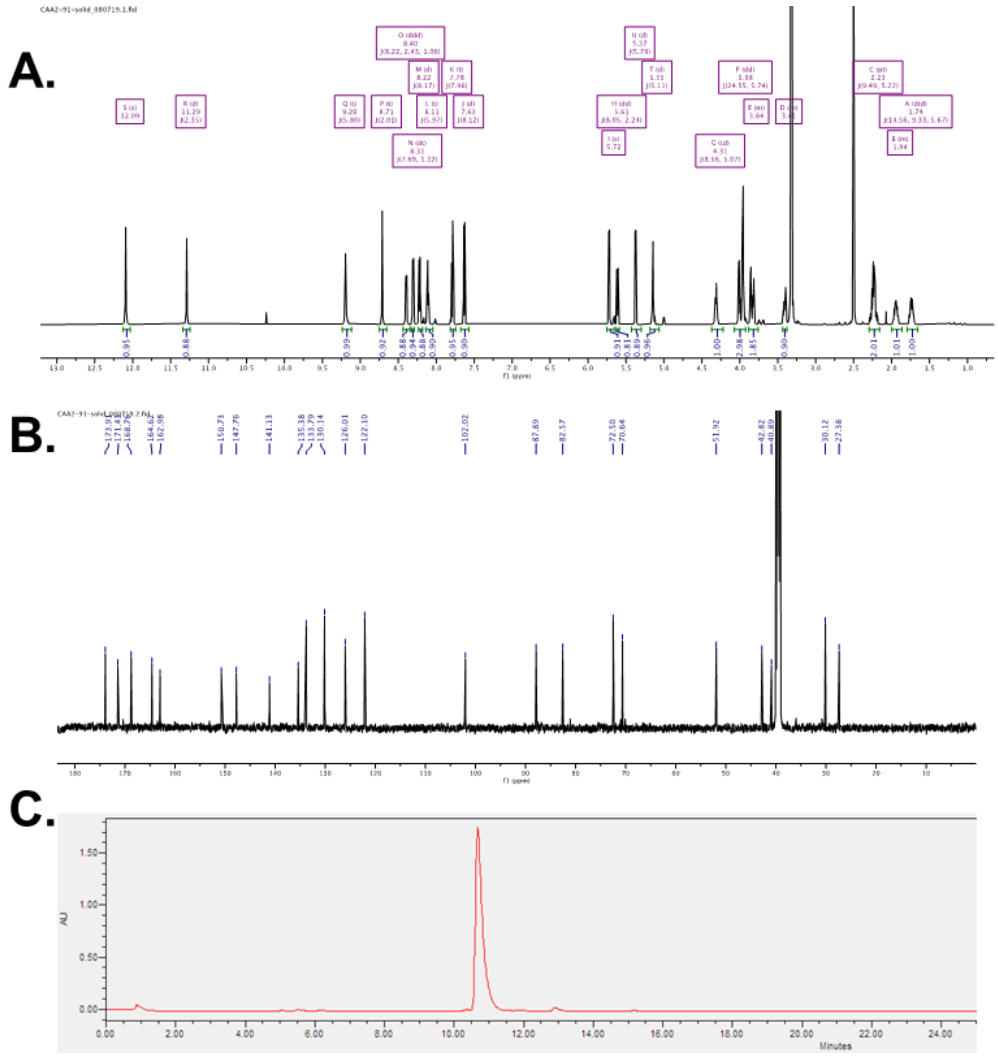
centrifugation tubes, frozen in LN_2, and lyophilized to yield a fluffy white solid (18% yield, 8.2 mg).

^{1}H NMR (600 MHz, DMSO-*d*$_6$) δ 12.09 (s, 1H), 11.29 (d, *J* = 2.4 Hz, 1H), 9.20 (t, *J* = 5.8 Hz, 1H), 8.71 (t, *J* = 2.0 Hz, 1H), 8.40 (ddd, *J* = 8.2, 2.4, 1.1 Hz, 1H), 8.31 (dt, *J* = 7.7, 1.3 Hz, 1H), 8.22 (d, *J* = 8.2 Hz, 1H), 8.11 (t, *J* = 6.0 Hz, 1H), 7.78 (t, *J* = 8.0 Hz, 1H), 7.63 (d, *J* = 8.1 Hz, 1H), 5.72 (s, 1H), 5.61 (dd, *J* = 8.0, 2.2 Hz, 1H), 5.37 (d, *J* = 5.8 Hz, 1H), 5.15 (d, *J* = 5.1 Hz, 1H), 4.31 (td, *J* = 8.6, 5.1 Hz, 1H), 3.98 (dd, *J* = 24.6, 5.7 Hz, 3H), 3.88 – 3.80 (m, 2H), 3.48 – 3.37 (m, 1H), 2.23 (pt, *J* = 9.5, 5.2 Hz, 2H), 1.98 – 1.90 (m, 1H), 1.74 (dtd, *J* = 14.6, 9.3, 5.7 Hz, 1H).

^{13}C NMR (151 MHz, DMSO) δ 173.9, 171.4, 168.8, 164.6, 163.0, 150.7, 147.8, 141.1, 135.4, 133.8, 130.1, 126.0, 122.1, 102.0, 87.9, 82.6, 72.5, 70.6, 51.9, 42.8, 40.9, 30.1, 27.4.

HPLC R$_t$: 10.8 min (260 nm).

HRMS (ESI$^+$) m/z: $[M+H]^+$ Calc'd for $C_{23}H_{27}N_6O_{12}^+$ 579.1681; found 579.1690.



Characterization of compound **3**. **A:** ^{1}H NMR spectrum (600 MHz, DMSO-d_6) of compound **3**. **B:** ^{13}C NMR spectrum (151 MHz, DMSO- d_6) of compound **3**. **C:** Crude RP-HPLC UV (260 nm) spectrum of compound **3**.

References

1. T. Li, K. D. Noel, Synthesis of N-acetyl-d-quinovosamine in *Rhizobium etli* CE3 is completed after its 4-keto-precursor is linked to a carrier lipid. *Microbiology* **163**, 1890-1901 (2017).
2. J. L. Prior *et al.*, Characterization of the O antigen gene cluster and structural analysis of the O antigen of *Francisella tularensis* subsp. tularensis. *J. Med. Microbiol.* **52**, 845-851 (2003).
3. T. Li, Coordination of Primer Sugar Synthesis with O-antigen Initiation in *Rhizobium Etli* CE3. (2014).
4. E. Vinogradov *et al.*, Structure determination of *Streptococcus suis* serotype 9 capsular polysaccharide and assignment of functions of the cps locus genes involved in its biosynthesis. *Carbohydr. Res.* **433**, 25-30 (2016).
5. P.-E. Jansson, B. Lindberg, U. Lindquist, Structural studies of the capsular polysaccharide from *Streptococcus pneumoniae* type 5. *Carbohydr. Res.* **140**, 101-110 (1985).
6. M. Rausch *et al.*, Coordination of capsule assembly and cell wall biosynthesis in *Staphylococcus aureus*. *Nat. Commun.* **10**, 1404 (2019).
7. S. Visansirikul, S. A. Kolodziej, A. V. Demchenko, *Staphylococcus aureus* capsular polysaccharides: a structural and synthetic perspective. *Org. Biomol. Chem.* **18**, 783-798 (2020).
8. S. Park *et al.*, Characterization of the structure and biological functions of a capsular polysaccharide produced by *Staphylococcus saprophyticus*. *J. Bacteriol.* **192**, 4618-4626 (2010).
9. E. Vinogradov, F. St. Michael, A. D. Cox, The structure of the LPS O-chain of *Fusobacterium nucleatum* strain 25586 containing two novel monosaccharides, 2-acetamido-2,6-dideoxy-l-altrose and a 5-acetimidoylamino-3,5,9-trideoxy-gluco-non-2-ulosonic acid. *Carbohydr. Res.* **440-441**, 10-15 (2017).
10. A. Z. Mostafavi, J. M. Troutman, Biosynthetic assembly of the *Bacteroides fragilis* capsular polysaccharide A precursor bactoprenyl diphosphate-linked acetamido-4-amino-6-deoxygalactopyranose. *Biochemistry* **52**, 1939-1949 (2013).
11. E. Vinogradov, F. St Michael, A. D. Cox, Structure of the LPS O-chain from *Fusobacterium nucleatum* strain ATCC 23726 containing a novel 5,7-diamino-3,5,7,9-tetradeoxy-l-gluco-non-2-ulosonic acid presumably having the d-glycero-l-gluco configuration. *Carbohydr. Res.* **468**, 69-72 (2018).
12. P. Garcia-Vello *et al.*, Structure of the O-Antigen and the Lipid A from the Lipopolysaccharide of *Fusobacterium nucleatum* ATCC 51191. *ChemBioChem* **22**, 1252-1260 (2021).
13. E. Vinogradov, F. St. Michael, K. Homma, A. Sharma, A. D. Cox, Structure of the LPS O-chain from *Fusobacterium nucleatum* strain 10953, containing sialic acid. *Carbohydr. Res.* **440-441**, 38-42 (2017).
14. E. Vinogradov, F. St Michael, A. D. Cox, Structure of the lipopolysaccharide O-antigens from *Fusobacterium nucleatum* strains HM-994, HM-995, HM-997. *Carbohydr. Res.* **522**, 108704 (2022).
15. E. Vinogradov, F. St Michael, A. D. Cox, Structure of the LPS O-chain from *Fusobacterium nucleatum* strain MJR 7757 B. *Carbohydr. Res.* **463**, 37-39 (2018).
16. D.-Q. Xu, J. O. Cisar, N. Ambulos Jr, D. H. Burr, D. J. Kopecko, Molecular cloning and characterization of genes for *Shigella sonnei* form IO polysaccharide:

proposed biosynthetic pathway and stable expression in a live *Salmonella* vaccine vector. *Infect. Immun.* **70**, 4414-4423 (2002).
17. L. Kenne, B. Lindberg, K. Petersson, E. Katzenellenbogen, E. Romanowska, Structural studies of the O-specific side-chains of the *Shigella sonnei* phase I lipopolysaccharide. *Carbohydr. Res.* **78**, 119-126 (1980).
18. B. Liu *et al.*, Genetic analysis of the O-antigen of *Providencia alcalifaciens* O30 and biochemical characterization of a formyltransferase involved in the synthesis of a Qui4N derivative. *Glycobiology* **22**, 1236-1244 (2012).
19. O. G. Ovchinnikova, A. Rozalski, B. Liu, Y. A. Knirel, O-antigens of bacteria of the genus *Providencia*: Structure, serology, genetics, and biosynthesis. *Biochem. (Mosc.)* **78**, 798-817 (2013).
20. F. V. Toukach *et al.*, Structure of the O-polysaccharide of *Providencia alcalifaciens* O8 containing (2S,4R)-2,4-dihydroxypentanoic acid, a new non-sugar component of bacterial glycans. *Carbohydr. Res.* **343**, 2706-2711 (2008).
21. O. G. Ovchinnikova *et al.*, Structure of the O-Polysaccharide of *Providencia alcalifaciens* O22 Containing D-Glyceramide 2-Phosphate. *Eur. J. Org. Chem.* **2012**, 3500-3506 (2012).
22. A. Preston *et al.*, Complete structures of *Bordetella bronchiseptica* and *Bordetella parapertussis* lipopolysaccharides. *J. Biol. Chem.* **281**, 18135-18144 (2006).
23. M. Caroff, J.-R. Brisson, A. Martin, D. Karibian, Structure of the *Bordetella pertussis* 1414 endotoxin. *FEBS Lett.* **477**, 8-14 (2000).
24. N. Gisch *et al.*, Commensal *Streptococcus mitis* produces two different lipoteichoic acids of type I and type IV. *Glycobiology* **31**, 1655-1699 (2021).
25. N. Bergström, P.-E. Jansson, M. Kilian, U. B. Skov Sørensen, Structures of two cell wall-associated polysaccharides of a *Streptococcus mitis* biovar 1 strain. *Eur. J. Biochem.* **267**, 7147-7157 (2000).
26. H. S. Seo, R. T. Cartee, D. G. Pritchard, M. H. Nahm, A New Model of Pneumococcal Lipoteichoic Acid Structure Resolves Biochemical, Biosynthetic, and Serologic Inconsistencies of the Current Model. *J. Bacteriol.* **190**, 2379-2387 (2008).
27. C. Draing *et al.*, Comparison of Lipoteichoic Acid from Different Serotypes of *Streptococcus pneumoniae*. *J. Biol. Chem.* **281**, 33849-33859 (2006).
28. G. Goyette-Desjardins *et al.*, Structure determination of *Streptococcus suis* serotypes 7 and 8 capsular polysaccharides and assignment of functions of the cps locus genes involved in their biosynthesis. *Carbohydr. Res.* **473**, 36-45 (2019).
29. M. L. Gening, E. A. Kurbatova, N. E. Nifantiev, Synthetic Analogs of *Streptococcus pneumoniae* Capsular Polysaccharides and Immunogenic Activities of Glycoconjugates. *Russ. J. Bioorg. Chem.* **47**, 1-25 (2021).
30. R. E. B. Young *et al.*, *Haemophilus parainfluenzae* expresses diverse lipopolysaccharide O-antigens using ABC transporter and Wzy polymerase-dependent mechanisms. *Int. J. Med. Microbiol.* **303**, 603-617 (2013).
31. V. Vitiazeva, B. Twelkmeyer, R. Young, D. W. Hood, E. K. H. Schweda, Structural studies of the lipopolysaccharide from H*aemophilus parainfluenzae* strain 20. *Carbohydr. Res.* **346**, 2228-2236 (2011).

32. W. M. Kalka-Moll *et al.*, Immunochemical and biological characterization of three capsular polysaccharides from a single *Bacteroides fragilis* strain. *Infect. Immun.* **69**, 2339-2344 (2001).
33. Y. Wang, W. M. Kalka-Moll, M. H. Roehrl, D. L. Kasper, Structural basis of the abscess-modulating polysaccharide A2 from *Bacteroides fragilis*. *Proc. Natl. Acad. Sci. USA* **97**, 13478-13483 (2000).
34. T. L. Reuber, G. C. Walker, Biosynthesis of succinoglycan, a symbiotically important exopolysaccharide of *Rhizobium meliloti*. *Cell* **74**, 269-280 (1993).
35. S. Acosta-Jurado, F. Fuentes-Romero, J.-E. Ruiz-Sainz, M. Janczarek, J.-M. Vinardell, Rhizobial Exopolysaccharides: Genetic Regulation of Their Synthesis and Relevance in Symbiosis with Legumes. *Int. J. Mol. Sci.* **22**, 6233 (2021).
36. S. y. N. Senchenkova *et al.*, A novel ItrA4 d-galactosyl 1-phosphate transferase is predicted to initiate synthesis of an amino sugar-lacking K92 capsular polysaccharide of *Acinetobacter baumannii* B8300. *Res. Microbiol.* **172**, 103815 (2021).
37. J. Roshini, L. P. P. Patro, S. Sundaresan, T. Rathinavelan, Structural diversity among *Acinetobacter baumannii* K-antigens and its implication in the in silico serotyping. *Front. Microbiol.* **14** (2023).
38. S. Lebeer *et al.*, Identification of a gene cluster for the biosynthesis of a long, galactose-rich exopolysaccharide in *Lactobacillus rhamnosus* GG and functional analysis of the priming glycosyltransferase. *Appl. Environ. Microbiol.* **75**, 3554-3563 (2009).
39. E. Dertli *et al.*, Structure and biosynthesis of two exopolysaccharides produced by *Lactobacillus johnsonii* FI9785. *J. Biol. Chem.* **288**, 31938-31951 (2013).
40. Z. Ma *et al.*, *Clostridioides difficile* cd2775 encodes a unique mannosyl-1-phosphotransferase for polysaccharide II biosynthesis. *ACS Infect. Dis.* **6**, 680-686 (2020).
41. M. Chu *et al.*, A *Clostridium difficile* cell wall glycopolymer locus influences bacterial shape, polysaccharide production and virulence. *PLoS Path.* **12**, e1005946 (2016).
42. G. Reddy *et al.*, Purification and determination of the structure of capsular polysaccharide of *Vibrio vulnificus* M06-24. *J. Bacteriol.* **174**, 2620-2630 (1992).
43. D. M. Post *et al.*, O-antigen and core carbohydrate of *Vibrio fischeri* lipopolysaccharide: composition and analysis of their role in Euprymna scolopes light organ colonization. *J. Biol. Chem.* **287**, 8515-8530 (2012).
44. K. Sackett *et al.*, Identification of a Novel Keto Sugar Component in *Streptococcus pneumoniae* Serotype 12F Capsular Polysaccharide and Impact on Vaccine Immunogenicity. *J. Immunol.* **210**, 764-773 (2023).
45. M. J. Morrison, B. Imperiali, Biosynthesis of UDP-N, N'-diacetylbacillosamine in *Acinetobacter baumannii*: Biochemical characterization and correlation to existing pathways. *Arch. Biochem. Biophys.* **536**, 72-80 (2013).
46. B. Liu *et al.*, Structure and genetics of *Shigella* O antigens. *FEMS Microbiol. Rev.* **32**, 627-653 (2008).
47. S. Hwang *et al.*, The biosynthesis of UDP-d-FucNAc-4N-(2)-oxoglutarate (UDP-Yelosamine) in *Bacillus cereus* ATCC 14579: Pat and Pyl, an aminotransferase and an ATP-dependent Grasp protein that ligates 2-oxoglutarate to UDP-4-amino-sugars. *J. Biol. Chem.* **289**, 35620-35632 (2014).

48. T. Behr, W. Fischer, J. Peter-Katalinić, H. Egge, The structure of pneumococcal lipoteichoic acid: Improved preparation, chemical and mass spectrometric studies. *Eur. J. Biochem.* **207**, 1063-1075 (1992).
49. M. J. Morrison, B. Imperiali, The renaissance of bacillosamine and its derivatives: pathway characterization and implications in pathogenicity. *Biochemistry* **53**, 624-638 (2014).
50. N. B. Olivier, M. M. Chen, J. R. Behr, B. Imperiali, In vitro biosynthesis of UDP-N, N ‘-diacetylbacillosamine by enzymes of the *Campylobacter jejuni* general protein Glycosylation system. *Biochemistry* **45**, 13659-13669 (2006).
51. P. Luong, A. Ghosh, K. D. Moulton, S. S. Kulkarni, D. H. Dube, Synthesis and Application of Rare Deoxy Amino l-Sugar Analogues to Probe Glycans in Pathogenic Bacteria. *ACS Infect. Dis.* **8**, 889-900 (2022).
52. O. Vasquez, A. Alibrandi, C. S. Bennett, De Novo Synthetic Approach to 2, 4-Diamino-2, 4, 6-trideoxyhexoses (DATDH): Bacterial and Rare Deoxy-Amino Sugars. *Org. Lett.* **25**, 7873-7877 (2023).
53. A. S. Riegert, N. M. Young, D. C. Watson, J. B. Thoden, H. M. Holden, Structure of the external aldimine form of PglE, an aminotransferase required for N, N'-diacetylbacillosamine biosynthesis. *Protein Sci.* **24**, 1609-1616 (2015).
54. A. A. Kasimova *et al.*, *Acinetobacter baumannii* K20 and K21 capsular polysaccharide structures establish roles for UDP-glucose dehydrogenase Ugd2, pyruvyl transferase Ptr2 and two glycosyltransferases. *Glycobiology* **28**, 876-884 (2018).
55. T. Li, L. Simonds, E. Kovrigin, K. Noel, In vitro biosynthesis and chemical identification of UDP-N-acetyl-d-quinovosamine (UDP-d-QuiNAc). *J. Biol. Chem.* **289**, 18110-18120 (2014).
56. P. Illarionov, V. Torgov, I. Hancock, V. Shibaev, A novel synthesis of N-acetyl-α-d-fucosamine 1-phosphate and uridine 5"-diphospho-N-acetyl-α-d-fucosamine. *Russ. Chem. Bull.* **50**, 1303-1308 (2001).
57. A. J. Anderson, G. J. Dodge, K. N. Allen, B. Imperiali, Co-conserved sequence motifs are predictive of substrate specificity in a family of monotopic phosphoglycosyl transferases. *Protein Sci.*, e4646 (2023).
58. A. Majumder *et al.*, Synergistic computational and experimental studies of a phosphoglycosyl transferase membrane/ligand ensemble. *J. Biol. Chem.* **299** (2023).
59. K. B. Patel, E. Ciepichal, E. Swiezewska, M. A. Valvano, The C-terminal domain of the *Salmonella enterica* WbaP (UDP-galactose: Und-P galactose-1-phosphate transferase) is sufficient for catalytic activity and specificity for undecaprenyl monophosphate. *Glycobiology* **22**, 116-122 (2012).
60. D. Carbajo, A. El-Faham, M. Royo, F. Albericio, Optimized stepwise synthesis of the API Liraglutide using BAL resin and pseudoprolines. *ACS Omega* **4**, 8674-8680 (2019).
61. C. A. Arbour, B. Imperiali, Backbone-Anchoring, Solid-Phase Synthesis Strategy To Access a Library of Peptidouridine-Containing Small Molecules. *Org. Lett.* **24**, 2170-2174 (2022).
62. A. Babič, S. Gobec, C. Gravier-Pelletier, Y. Le Merrer, S. Pečar, Synthesis of 1-C-linked diphosphate analogues of UDP-N-Ac-glucosamine and UDP-N-Ac-muramic acid. *Tetrahedron* **64**, 9093-9100 (2008).