Supporting Information

Processive Carbohydrate Polymerases Mediate Bifunctional Catalysis Using a Single Active Site

John F. May, Matthew R. Levengood, Rebecca A. Splain, Christopher D. Brown, & Laura L. Kiessling

Table S1 GlfT2 homologs used in the multiple sequence alignment.

Abbreviation	GenBank Accession Number or NCBI Reference Sequence	Source of sequence (bacterial species)
Nbr	BAF46769.1	*Nocardia brasiliensis*
Nf1	YP_116420.1	*Nocardia farcinica* IFM 10152
Rr1	YP_704064.1	*Rhodococcus* sp. RHA1
Ser	YP_001107843.1	*Saccharopolyspora erythraea* NRRL 2338
Bde	ZP_02917795.1	*Bifidobacterium dentium* ATCC 27678
Ban	ZP_02964173.1	*Bifidobacterium animalis* subsp. lactis HN019
Bl1	NP_695433.1	*Bifidobacterium longum* NCC2705
Mth	NP_218325.1	***Mycobacterium tuberculosis* H37Rv**
Mtc	NP_338467.1	*Mycobacterium tuberculosis* CDC1551
Mma	YP_001853631.1	*Mycobacterium marinum* M
Mul	YP_908340.1	*Mycobacterium ulcerans* Agy99
Mle	NP_301191.1	*Mycobacterium leprae* TN
Mav	YP_879503.1	*Mycobacterium avium* 104
Map	NP_959146.1	*Mycobacterium avium* subsp. paratuberculosis K-10
Mjl	YP_001073654.1	*Mycobacterium* sp. JLS
Mmc	YP_642179.1	*Mycobacterium* sp. MCS
Mva	YP_956420.1	*Mycobacterium vanbaalenii* PYR-1
Mgi	YP_001132430.1	*Mycobacterium gilvum* PYR-GCK
Msm	YP_890616.1	*Mycobacterium smegmatis* str. MC2 155
Mab	YP_001700925.1	*Mycobacterium abscessus*
Rr2	YP_704009.1	*Rhodococcus* sp. RHA1
Nf2	YP_116383.1	*Nocardia farcinica* IFM 10152
Cur	YP_001799546.1	*Corynebacterium urealyticum* DSM 7109
Cje	YP_249900.1	*Corynebacterium jeikeium* K411
Cg1	NP_602073.1	*Corynebacterium glutamicum* ATCC 13032
Cg2	YP_227122.1	*Corynebacterium glutamicum* ATCC 13032
Cg3	BAC00276.1	*Corynebacterium glutamicum* ATCC 13032
Cgr	YP_001139691.1	*Corynebacterium glutamicum* R
Cef	NP_739324.1	*Corynebacterium efficiens* YS-314
Cdi	NP_940506.1	*Corynebacterium diphtheriae* NCTC 13129
Se1	YP_001107844.1	*Saccharopolyspora erythraea* NRRL 2338

Se2	YP_001102444.1	*Saccharopolyspora erythraea* NRRL 2338
Rsa	YP_001624767.1	*Renibacterium salmoninarum* ATCC 33209
Afb	YP_832172.1	*Arthrobacter* sp. FB24
Aau	YP_948391.1	*Arthrobacter aurescens* TC1
Kra	YP_001363614.1	*Kineococcus radiotolerans* SRS30216
Krh	YP_001854751.1	*Kocuria rhizophila* DC2201
Njs	YP_922603.1	*Nocardioides* sp. JS614
Cms	YP_001709406.1	*Clavibacter michiganensis* subsp. Sepedonicus
Cmm	YP_001221762.1	*Clavibacter michiganensis* subsp. michiganensis NCPPB 382
Bl2	NP_695432.1	*Bifidobacterium longum* NCC2705
Sen	ZP_02701557.1	*Salmonella enterica* subsp. enterica serovar Newport str. SL317
Seh	ZP_02665507.1	*Salmonella enterica* subsp. enterica serovar Heidelberg str. SL486
Sew	ZP_02832148.1	*Salmonella enterica* subsp. enterica serovar Weltevreden str. HI_N05-537
Sep	YP_001589014.1	*Salmonella enterica* subsp. enterica serovar Paratyphi B str. SPB7
Ses	ZP_02663925.1	*Salmonella enterica* subsp. enterica serovar Schwarzengrund str. SL480
Sev	ZP_02706803.1	*Salmonella enterica* subsp. enterica serovar Virchow str. SL491
Sej	ZP_02653520.1	*Salmonella enterica* subsp. enterica serovar Javiana str. GA_MM04042433
Rba	ZP_01014392.1	*Rhodobacterales bacterium* HTCC2654
Gox	YP_191889.1	*Gluconobacter oxydans* 621H

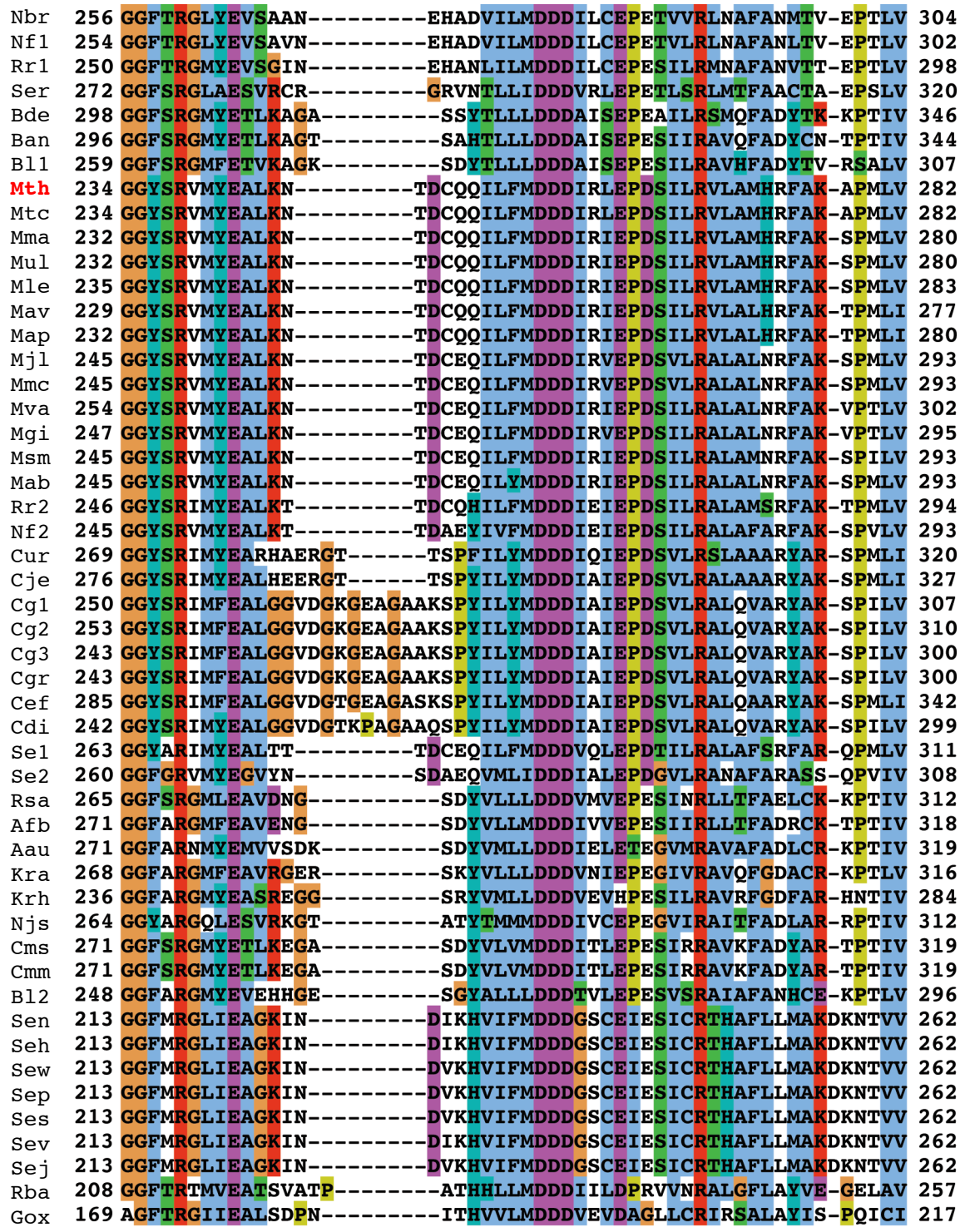

Figure S1. DDD motif is conserved among GlfT2 variants. Sequences are labeled according to Table S1, and the label for GlfT2 is in red. The DDD motif (Asp256–Asp258 for GlfT2) is invariant among these homologs, and the nearby residues are conserved. This alignment is colored using ClustalW2 coloring to indicate conservation.

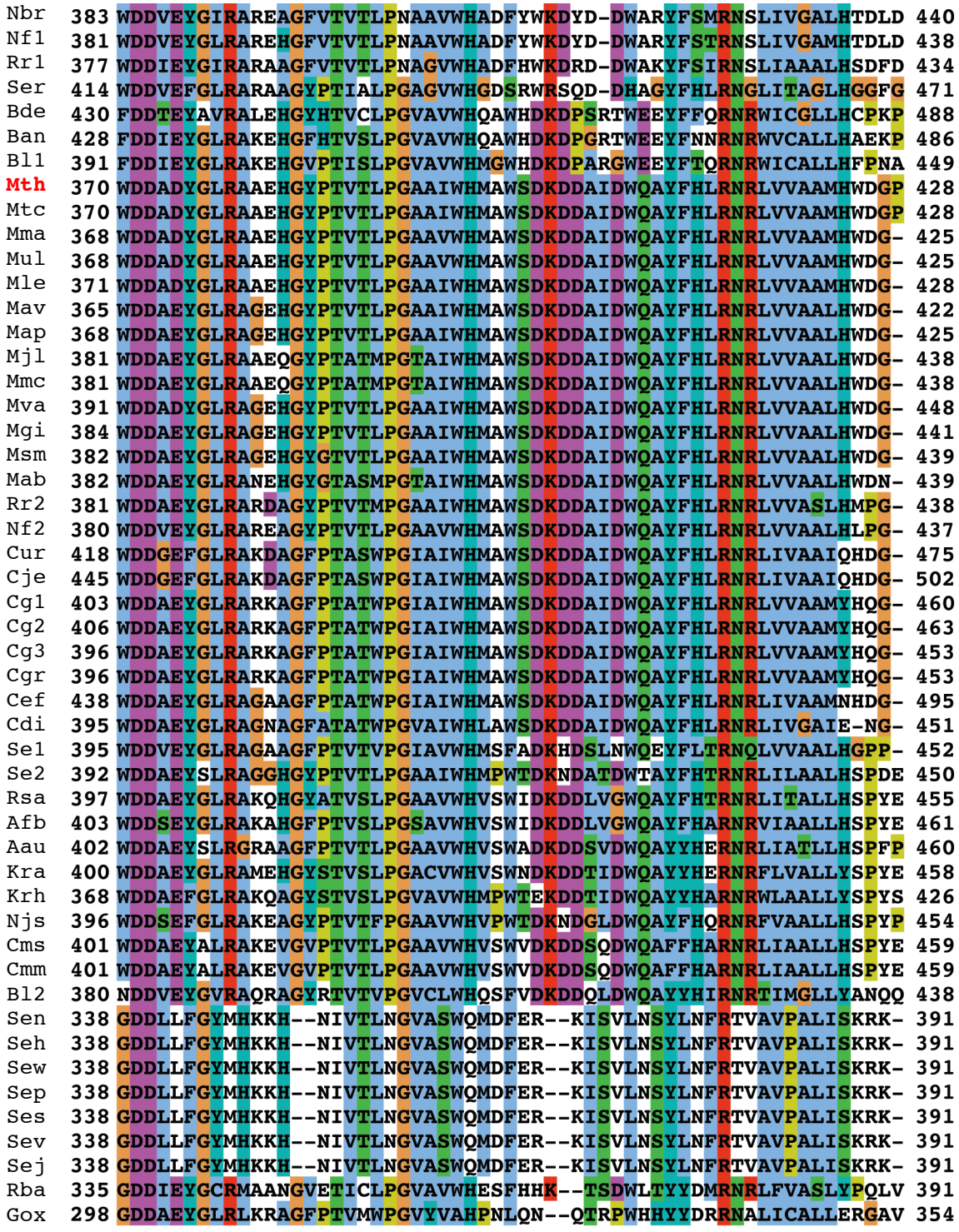

Figure S2. DDA motif is conserved among GlfT2 variants. Sequences are labeled according to Table S1, and the label for GlfT2 is in red. The Asp residues in the DDA motif (Asp371–Ala373 for GlfT2) are invariant among these homologs, and nearby residues are conserved. This alignment is colored using ClustalW2 coloring to indicate conservation.

```
GlfT2 160 ---------- ---------- ---------- ----ANIAVG IPTFNRPADC
2z86A 111 PLDWPSDLTL PPLPESTNDY VWAGKRKELL II--DGLSIV IPTYNRAKIL

GlfT2 176 VNALRELTAD -PLVDQVIGA VIVPDQGERK VRDHPDFPAA AARLGSRLSI
2z86A 166 AITLACLCNQ KTIYD---YE VIVADDGSK- --EN--IEEI VREFESLLNI

GlfT2 225 HDQPNLGGSG GYSRVMYEAL KNTDCQQILF MDDDIRLEPD SILRVLAMHR
2z86A 208 KYVRQKDYGY QLCAVRNLGL RAAKYNYVAI LDCDMAPNPL WVQSYMELLA

GlfT2 275 FAKAPMLVGG QMLNLQE-PS HLHIMGEVVD RSIFMWTAAP HAEYDHDFAE
2z86A 258 VDDNVALIGP RKYIDTSKHT Y----LDFLS QKSLINEIPE SV--DWRIEH

GlfT2 324 YPLNDNNSRS KLLHRRIDVD YNGWWTCMIP RQVAEELGQP LPLFIKWDDA
2z86A 317 FKN-TDN--- -LRLCNTPFR FFSGGNVAFA KKWLFRAGWF DEEFTHWGGE

GlfT2 374 --DYGLRAAE HGYPTVTLPG AAIWHMA--- ---------- ----------
2z86A 362 DNEFGYRLYR EGCYFRSVEG AMAYHQEPPQ LLQQKVPYFY RKKEKIESAT
```

Figure S3. Alignment used for homology model of GlfT2. This alignment was generated by the SWISS-MODEL server and used as the basis for the homology model shown in Figure 3B. This region of the GlfT2 sequence corresponds to its GT2 domain. The DDD and DDA motifs in GlfT2 are highlighted in red. This alignment showed 14.8% identity between the two sequences with an e-value of 3.6×10^{-9}.

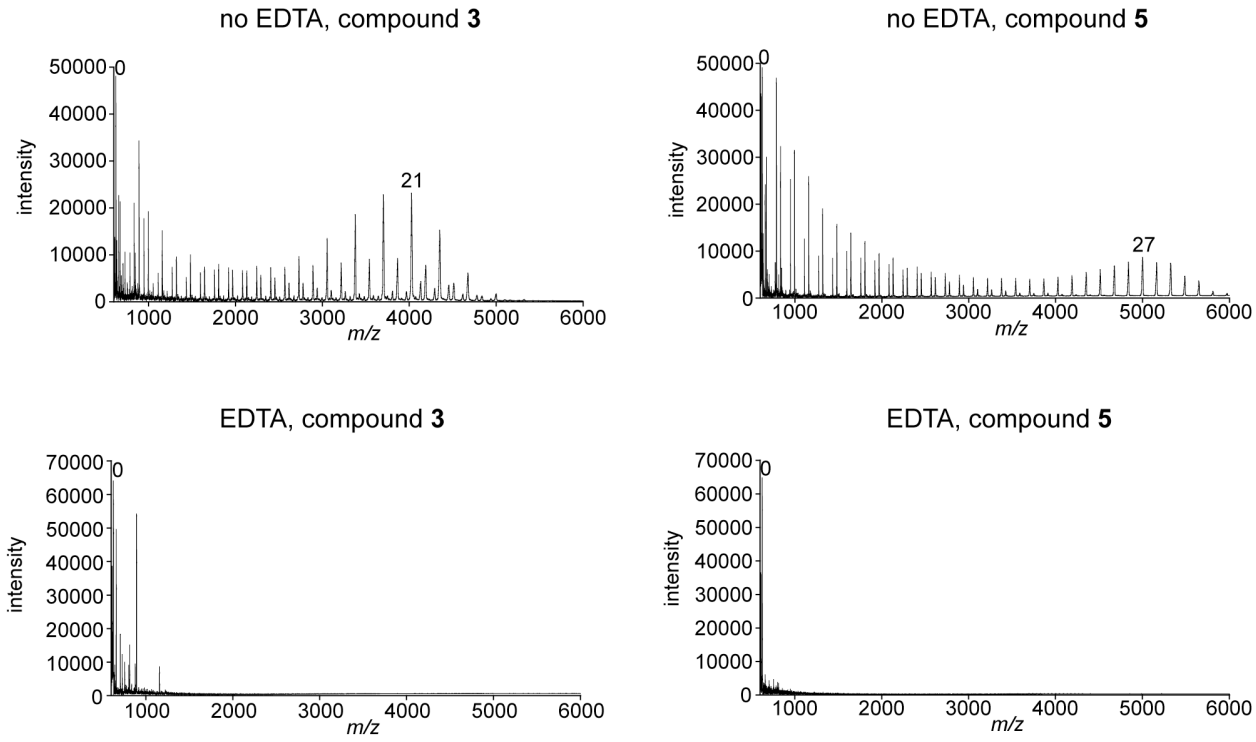

Figure S4. GlfT2 catalysis is metal-dependent. Spectra from MALDI-TOF MS analysis of GlfT2 reactions are shown. Products corresponding to addition of *n* Gal*f* residues to compound **3** (left) or compound **5** (right) are observed from reactions of GlfT2, acceptor, and UDP-Gal*f* only when metal (Mg^{2+}; "no EDTA") is present. In the presence of EDTA and absence of added metal, no products are observed from reactions with either acceptor (bottom spectra).

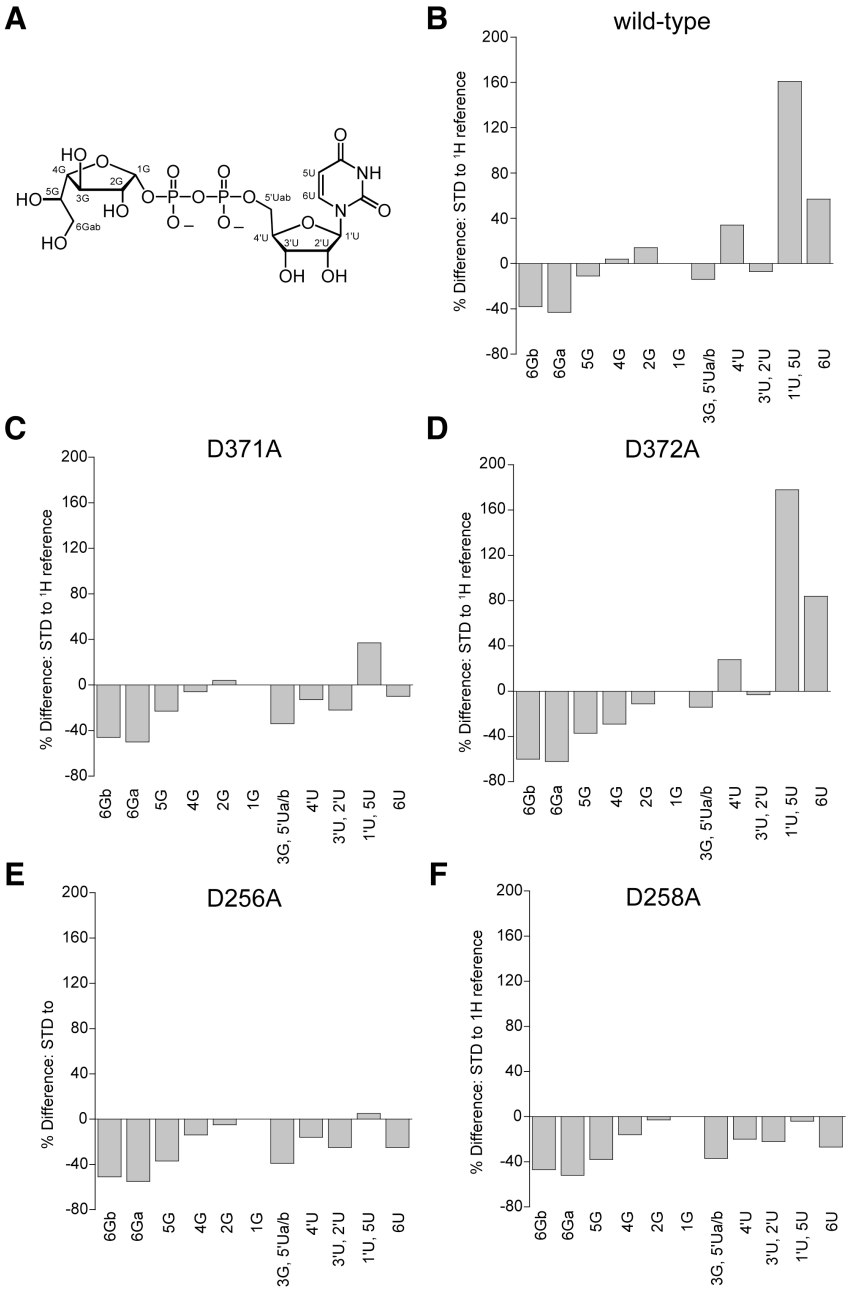

Figure S5. Analysis of STD NMR spectra for GlfT2 variants binding to UDP-Gal*f*. Addition of wild-type GlfT2 (B), D371A GlfT2 (C), D372A GlfT2 (D), D256A GlfT2 (E), or D258A GlfT2 (F) to a solution containing UDP-Gal*f* results in saturation transfer difference NMR signals at the indicated protons in the UDP-Gal*f* ligand (A). The signal for each distinct proton is plotted to give an STD NMR profile. Comparison of STD NMR profiles for the variants indicates that D372A and wild-type GlfT2 bind UDP-Gal*f* in a similar manner, D256A and D258A bind UDP-Gal*f* similarly to each other but distinctly from wild-type GlfT2, and D371A binds UDP-Gal*f* in a manner distinct from wild-type GlfT2.

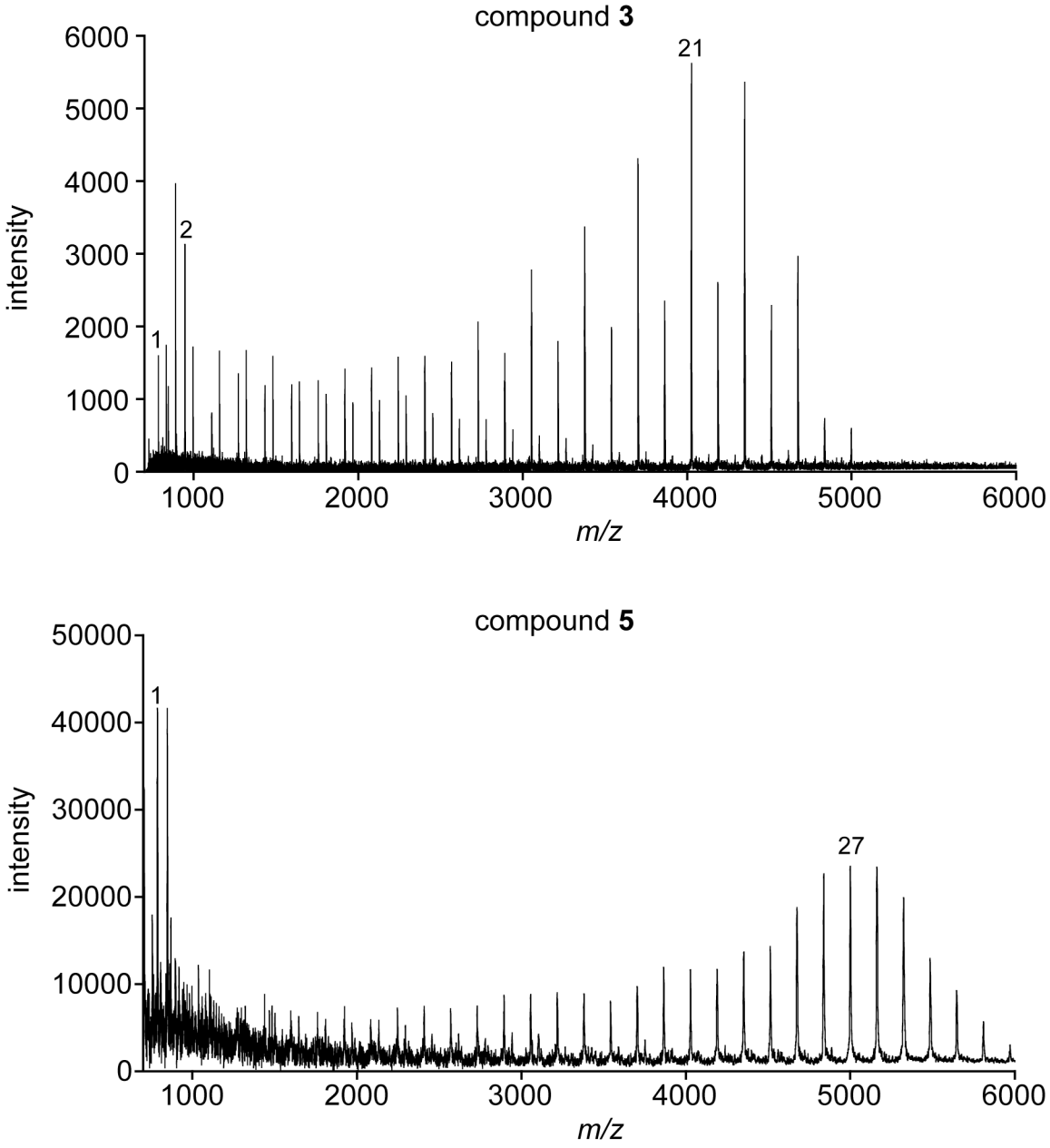

Figure S6. GlfT2 D371E variant shows wild-type polymerization activity. Spectra from MALDI-TOF MS analysis of 20 h incubations of GlfT2 D371E with either compound **3** (top) or compound **5** (bottom) show polymeric products. To orient these spectra with those shown in Figure 5, certain peaks that correspond to $[M+Na]^+$, where M equals the mass of the acceptor plus *n* Gal*f* residues, are labeled with the appropriate value of *n*. Labels are shown only for the short products ($n = 1, 2$) and for the highest-intensity peak of the polymeric products ($n = 21, 27$).

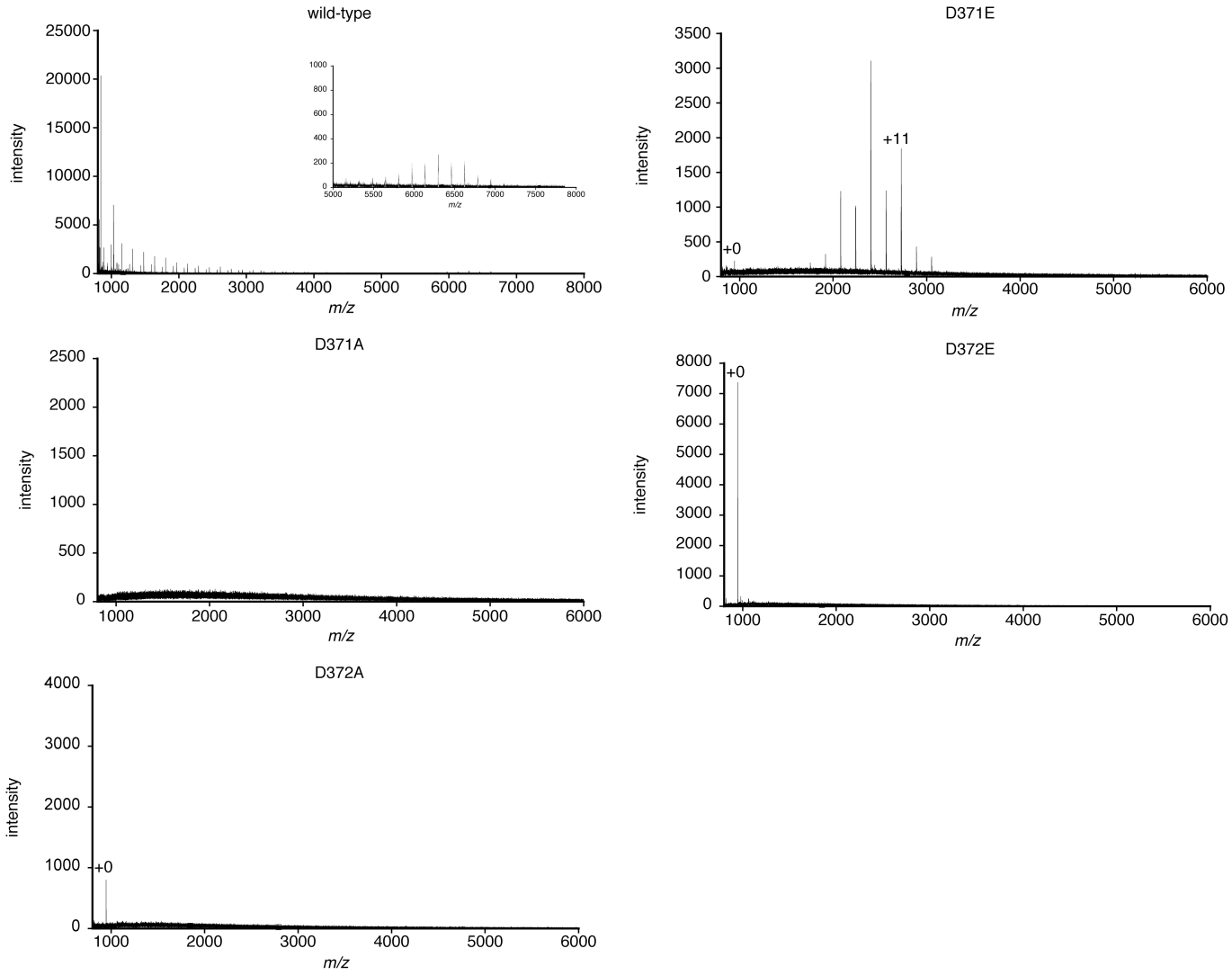

Figure S7. GlfT2 DDA motif variants show impaired activity with tetrasaccharide acceptor substrate. GlfT2 D371A, D372A, and D372E do not promote addition of Gal*f* residues to a tetrasaccharide acceptor that is elongated by wild-type GlfT2 (up to +37 Gal*f* residues) or D371E GlfT2 (up to +17 Gal*f* residues).

A

```
SpHasA  41 -----KVAAV IPSYNEDAES LLETLKSVLA QTYPLSEIYI VDDGSSNTDA
3bcvB    3 LIP--KVSVI VPIYNVE-KY LDQCVQALLA QTLSDIEIIL IDDESP-DNC

SpHasA  86 IQLIEEYVNR EVDICRNVIV HRSLVNKGKR HAQAWAFERS DADVFLTVDS
3bcvB   49 PKICDDYAAQ YPN-----IK VIHKKNAGLG XACNSGLDVA TGEYVAFCDS

SpHasA 136 DTYIYPNALE ELLKSFNDET VYAATGHLNA RNRQTNLLTR LTDIRYDNAF
3bcvB   94 DDYVDSDXYX TXYNVAQKYT CDAVFTFKLY KNKNEI-HTL LKDLIASDPY

SpHasA 186 GVERAAQSLT GNILVCSGPL SIYRREVIIP NLERYKNQTF LGLPVSIGDD
3bcvB  163 AREE------ -RAIQVSAKV VLYRRNLIEK KHLRFVSER- ----ILPSED

SpHasA 236 RCLTNYAIDL GRTVY --
3bcvB  201 LIFNVDVLAN SNIVCVLP
```

B

```
CesA  145 ---------- -------PDE WPTVDIFVPT YNEELSIVRL TVLGSLGIDW
1xhbA  95 NRSLPDVRLE GCKTKVYPDN LPTTSVVIVF HNEAWSTLLR TVHSVINRSP

CesA  178 PPEKVRVHIL DDGRRPE-FA AFAAECG--- ---ANYIARP TNEHAKAGNL
1xhbA 145 RHMIEEIVLV DDASERDFLK RPLESYVKKL KVPVHVIRME QRS-GLIRAR

CesA  221 NYAIGHTDGD YILIFDCDHV PTRAFLQLTM GWMVEDPKIA LMQTPHHFYS
1xhbA 194 LKGAAVSRGQ VITFLDAHCE CTAGWLEPLL ARIKHDRRTV VCPII-DVIS

CesA  271 PDPFQRNLSA GYR------- ------TPPE GNLFYGVVQD GNDFWDATFF
1xhbA 243 DDTFEYMAGS DMTYGGFNWK LNFRWYPVPQ REMDRR-KGD RTLPVRTPTM

CesA  308 CGSCAILRRT AIEQIGGFAT QTV---TEDA HTALKMQRLG WSTAYL ---
1xhbA 292 AGGLFSIDRD YFQEIGTYDA GMDIWGGENL EISFRIWQCG GTLEIVTCSH
```

Figure S8. Sequence alignments used for homology modeling of other GT-2 glycosyltransferases. (A) Alignment of hyaluronan synthase from *Streptococcus pyogenes* (SpHasA; GenBank accession number AAA17984.1) with the sequence of a glycosyltransferase PDB ID: 3BCV. This alignment, generated automatically by SWISS-MODEL, was used for the subsequent homology model of SpHasA, which showed 18.6% identity between the two sequences and gave an e-value of 1.3×10^{-14}. (B) Alignment of cellulose synthase from *Gluconacetobacter xylinus* (CesA; GenBank accession number P21877.3) with the glycosyltransferase UDP-GalNAc:polypeptide α-N-acetylgalactosaminyltransferase-T1 (PDB ID: 1XHB). This alignment was generated automatically in SWISS-MODEL and was used for subsequent homology modeling. This alignment showed 13.1% sequence identity between the two sequences with an e-value of 1.6×10^{-11}.

Table S2. Oligonucleotide primers used for site-directed mutagenesis. For each desired mutation, the primer indicated in the table and its reverse complement (not shown) were used together in the PCR mixture.

GlfT2 D256A	5'— CAGATCCTATTCATGGCCGACGACATCCGCCTC—3'
GlfT2 D256E	5'— CAGATCCTATTCATGGAGGACGACATCCGCCTC—3'
GlfT2 D258A	5'— CTATTCATGGACGACGCCATCCGCCTCGAGCCG—3'
GlfT2 D258E	5'— CTATTCATGGACGACGAGATCCGCCTCGAGCCG—3'
GlfT2 D371A	5'— GCCGTTGTTCATCAAATGGGCCGACGCCGATTACGGCC—3'
GlfT2 D371E	5'— GCCGTTGTTCATCAAATGGGAAGACGCCGATTACGGCC—3'
GlfT2 D372A	5'— GCCGTTGTTCATCAAATGGGACGCCGCCGATTACGGCC—3'
GlfT2 D372E	5'— GCCGTTGTTCATCAAATGGGACGAAGCCGATTACGGCC—3'