

## Supporting Information

### Processive Carbohydrate Polymerases Mediate Bifunctional Catalysis Using a Single

### Active Site

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

**Table S1** GltT2 homologs used in the multiple sequence alignment.

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

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

Nbr	256	GGFTRGLYEVSAAAN-----EHADVILMDDDILCEPETVVRLNAFANMTV-EPTLV	304
Nf1	254	GGFTRGLYEVS AVN-----EHADVILMDDDILCEPETVLRNNAFANLTV-EPTLV	302
Rr1	250	GGFTRGMYEVS GIN-----EHANLILMDDDILCEPESILRMNAFANVTT-EPTLV	298
Ser	272	GGFSRGLAESVRCR-----GRVNTLLIDDDVRLEPETLSRLMTFAACTA-EPSLV	320
Bde	298	GGFSRGMYEETLKAGA-----SSYTLILLDDDAISEPEAILRSMQFADYTK-KPTIV	346
Ban	296	GGFSRGMYEETLKAGT-----SAHTLLLDLDDAISEPEIIRAVQFADYCN-TPTIV	344
Bl1	259	GGFSRGMFETVKAGK-----SDYTLILLDDDAISEPESILRAVHFADYTV-RSALV	307
<b>Mth</b>	234	GGYSRVMYEALKN-----TDCQQILFMDDDIRLEPDSILRVLAMHRFAK-APMLV	282
Mtc	234	GGYSRVMYEALKN-----TDCQQILFMDDDIRLEPDSILRVLAMHRFAK-APMLV	282
Mma	232	GGYSRVMYEALKN-----TDCQQILFMDDDIRIEPDSILRVLAMHRFAK-SPMLV	280
Mul	232	GGYSRVMYEALKN-----TDCQQILFMDDDIRIEPDSILRVLAMHRFAK-SPMLV	280
Mle	235	GGYSRVMYEALKN-----TDCQQILFMDDDIRIEPDSILRVLAMHRFAK-SPMLV	283
Mav	229	GGYSRVMYEALKN-----TDCQQILFMDDDIRIEPDSILRVLALHRFAK-TPMLI	277
Map	232	GGYSRVMYEALKN-----TDCQQILFMDDDIRIEPDSILRVLALHRFAK-TPMLI	280
Mjl	245	GGYSRVMYEALKN-----TDCEQILFMDDDIRVEPDSVLRALALNRFK-SPMLV	293
Mmc	245	GGYSRVMYEALKN-----TDCEQILFMDDDIRVEPDSVLRALALNRFK-SPMLV	293
Mva	254	GGYSRVMYEALKN-----TDCEQILFMDDDIRIEPDSILRALALNRFK-VPTLV	302
Mgi	247	GGYSRVMYEALKN-----TDCEQILFMDDDIRVEPDSILRALALNRFK-VPTLV	295
Msm	245	GGYSRVMYEALKN-----TDCEQILFMDDDIRIEPDSILRALAMNRFK-SPILV	293
Mab	245	GGYSRVMYEALKN-----TDCEQILFMDDDIRIEPDSILRALALNRFK-SPMLV	293
Rr2	246	GGYSRIMYEALKT-----TDCQHILFMDDIEIEPDSILRALAMSRFAK-TPMLV	294
Nf2	245	GGYSRVMYEALKT-----TDAEYIVFMDDIEIEPDSILRALAFARFAK-SPVLV	293
Cur	269	GGYSRIMYEARHAERGT-----TSPFILYMDDDIQIEPDSVLRSLAAARYAR-SPMLI	320
Cje	276	GGYSRIMYEALHEERGT-----TSPYILYMDDIIAIEPDSVLRALAAARYAK-SPMLI	327
Cg1	250	GGYSRIMFEALGGVDGKGEAGAAKSPYILYMDDIIAIEPDSVLRALQVARYAK-SPILV	307
Cg2	253	GGYSRIMFEALGGVDGKGEAGAAKSPYILYMDDIIAIEPDSVLRALQVARYAK-SPILV	310
Cg3	243	GGYSRIMFEALGGVDGKGEAGAAKSPYILYMDDIIAIEPDSVLRALQVARYAK-SPILV	300
Cgr	243	GGYSRIMFEALGGVDGKGEAGAAKSPYILYMDDIIAIEPDSVLRALQVARYAK-SPILV	300
Cef	285	GGYSRIMFEALGGVDGTGEAGASKSPYILYMDDIIAIEPDSVLRALQAARYAK-SPMLI	342
Cdi	242	GGYSRIMYEALGGVDGTKPAGAAQSPYILYMDDIIAIEPDSVLRALQVARYAK-SPILV	299
Se1	263	GGYARIMYEALTT-----TDCEQILFMDDDVQLEPDTILRALAFSRFAR-QPMLV	311
Se2	260	GGFGRVMYEGVYN-----SDAEVMLIDDDIALEPDGVLNANAFARASS-QPVIV	308
Rsa	265	GGFSRGMLEAVDNG-----SDYVLLDDDDVMVEPESINRLLTFaelCK-KPTIV	312
Afb	271	GGFARGMFEAVENG-----SDYVLLMDDDIVVEPESIIIRLLTFADRCK-TPTIV	318
Aau	271	GGFARNMYEMVSDK-----SDYVMLLDDDIETEgVMRAVAFADLCR-KPTIV	319
Kra	268	GGFARGMFEAVRGER-----SKYVLLLDDDVNIEPEGIVRAVQFGDACR-KPTLV	316
Krh	236	GGFARGMYEASREGG-----SRYVMLLDDDVVEVHPESILRAVRFGDFAR-HNTIV	284
Njs	264	GGYARGQLESVRKGT-----ATYTMMDDDIVCEPEGVIRAITFADLAR-RPTIV	312
Cms	271	GGFSRGMYEETLKEGA-----SDYVLVMDDDITLPEsIRRavKFADYAR-TPTIV	319
Cmm	271	GGFSRGMYEETLKEGA-----SDYVLVMDDDITLPEsIRRavKFADYAR-TPTIV	319
Bl2	248	GGFARGMYEVEHHGE-----SGYALLLDDDTVLEPEsVSRAIAFANHCE-KPTLV	296
Sen	213	GGFMRGLIEAGKIN-----DIKHVIFMDDDGsCEIEsICRTHAFLLMAKDKNtVV	262
Seh	213	GGFMRGLIEAGKIN-----DIKHVIFMDDDGsCEIEsICRTHAFLLMAKDKNtVV	262
Sew	213	GGFMRGLIEAGKIN-----DVKHVIFMDDDGsCEIEsICRTHAFLLMAKDKNtVV	262
Sep	213	GGFMRGLIEAGKIN-----DVKHVIFMDDDGsCEIEsICRTHAFLLMAKDKNtVV	262
Sej	213	GGFMRGLIEAGKIN-----DVKHVIFMDDDGsCEIEsICRTHAFLLMAKDKNtVV	262
Rba	208	GGFTRTMVEATSVATP-----ATHHLLMDDDIILDPRVNRALGFLAYVE-GELAV	257
Gox	169	AGFTRGIIeALSdPN-----ITHVVLMDDDVeVDAGLLCRIRsALAYIS-POICI	217

**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.

Nbr 383 WDDVEYGIRAREAGFVTVTLPNAAVWHADFYWKDYD-DWARYFSMRNSLIVGALHTDLD 440  
 Nf1 381 WDDVEYGLRAREHGFTVTVTLPNAAVWHADFYWKDYD-DWARYFSTRNSLIVGAMHTDLD 438  
 Rr1 377 WDDIEYGIRARAAGFVTVTLPNAGVWHADFWKDRD-DWAKYFSIRNSLIAAALHSDFD 434  
 Ser 414 WDDVEFGLRARAAGYPTIALPGAGVWHGDSRWRSQD-DHAGYFHLRNLITAGLHGGFG 471  
 Bde 430 FDDTEYAVRALEHGHTVCLPGVAVWHQAWHDKDPSRTWEEYFFQRNRWICGLLHCPKP 488  
 Ban 428 FDDIEYGLRAKEHGFTVSLPGVAVWHQAWHDKDPPGRTWEEYFNRRNRWVCALLHAEKP 486  
 Bl1 391 FDDIEYGLRAKEHGVTISLPGVAVWHMGWHDKDPARGWEEYFTQRNRWICALLHFPNA 449  
**Mth** 370 WDDADYGLRAAEHGYPTVTLPGAIIWHMAWSKDDDAIDWQAYFHLRNLVVAAMHWDGP 428  
 Mtc 370 WDDADYGLRAAEHGYPTVTLPGAIIWHMAWSKDDDAIDWQAYFHLRNLVVAAMHWDGP 428  
 Mma 368 WDDADYGLRAAEHGYPTVTLPGAIVWHMAWSKDDDAIDWQAYFHLRNLVVAAMHWDG- 425  
 Mul 368 WDDADYGLRAAEHGYPTVTLPGAIVWHMAWSKDDDAIDWQAYFHLRNLVVAAMHWDG- 425  
 Mle 371 WDDADYGLRAAEHGYPTVTLPGAIIWHMAWSKDDDAIDWQAYFHLRNLVVAAMHWDG- 428  
 Mav 365 WDDAEYGLRAGEHGYPVTLPGAIIWHMAWSKDDDAIDWQAYFHLRNLVVAALHWDG- 422  
 Map 368 WDDAEYGLRAGEHGYPVTLPGAIIWHMAWSKDDDAIDWQAYFHLRNLVVAALHWDG- 425  
 Mj1 381 WDDAEYGLRAAEQGYPTATMPGTAIWHMAWSKDDDAIDWQAYFHLRNLVVAALHWDG- 438  
 Mmc 381 WDDAEYGLRAAEQGYPTATMPGTAIWHMAWSKDDDAIDWQAYFHLRNLVVAALHWDG- 438  
 Mva 391 WDDADYGLRAGEHGYPVTLPGAIIWHMAWSKDDDAIDWQAYFHLRNLVVAALHWDG- 448  
 Mgi 384 WDDAEYGLRAGEHGYPVTLPGAIIWHMAWSKDDDAIDWQAYFHLRNLVVAALHWDG- 441  
 Msm 382 WDDAEYGLRAGEHGYPVTLPGAIIWHMAWSKDDDAIDWQAYFHLRNLVVAALHWDG- 439  
 Mab 382 WDDAEYGLRANEHGYPVTLPGAIIWHMAWSKDDDAIDWQAYFHLRNLVVAALHWDN- 439  
 Rr2 381 WDDAEYGLRARDAGYPTVTPGAAIWHMAWSKDDDAIDWQAYFHLRNLVVAALHMPG- 438  
 Nf2 380 WDDVEYGLRAREAGFPTVSLPGAAVWHMAWSKDDDAIDWQAYFHLRNLVVAALHLPG- 437  
 Cur 418 WDDGEFGLRAKDAGFPTASWPGIAIWHMAWSKDDDAIDWQAYFHLRNLIVAAIQHDG- 475  
 Cje 445 WDDGEFGLRAKDAGFPTASWPGIAIWHMAWSKDDDAIDWQAYFHLRNLIVAAIQHDG- 502  
 Cg1 403 WDDAEYGLRARKAGFPTATWPGIAIWHMAWSKDDDAIDWQAYFHLRNLVVAAMYHQG- 460  
 Cg2 406 WDDAEYGLRARKAGFPTATWPGIAIWHMAWSKDDDAIDWQAYFHLRNLVVAAMYHQG- 463  
 Cg3 396 WDDAEYGLRARKAGFPTATWPGIAIWHMAWSKDDDAIDWQAYFHLRNLVVAAMYHQG- 453  
 Cgr 396 WDDAEYGLRARKAGFPTATWPGIAIWHMAWSKDDDAIDWQAYFHLRNLVVAAMYHQG- 453  
 Cef 438 WDDAEYGLRAGAAGFPTATWPGIAIWHMAWSKDDDAIDWQAYFHLRNLIVAAAMNHDG- 495  
 Cdi 395 WDDAEYGLRAGNAGFATATWPGVAIWHLAWSDKDDAIDWQAYFHLRNLIVGAIE-NG- 451  
 Sel 395 WDDVEYGLRAGAAGFPTVTVPGIAVWHMSFADKHDLSLWQYFHLTRNQLVVAALHGPP- 452  
 Se2 392 WDDAEYSLRAGGHGYPTVTLPGAIIWHMPWTDKNDATDWTAYFHTRNRLILAALHSPDE 450  
 Rsa 397 WDDAEYGLRAKQHGATVSLPGAIVHVSVIDKDDLGVGQAYFHTRNRLITALLHSPYE 455  
 Afb 403 WDDSEYGLRAKAGFPTVSLPGSAVHVSVIDKDDLGVGQAYFHARNRVIAALLHSPYE 461  
 Aau 402 WDDAEYSLRGAAGFPTVTLPGAIVHVSVIDKDDSDVDWQAYYHERNRLIATLLHSPFP 460  
 Kra 400 WDDAEYGLRAMEHGYSTVSLPGACVHVSVIDKDDTIDWQAYYHERNRLVALLHSPYE 458  
 Krh 368 WDDAEYGLRAKQAGYSTVSLPGVAVWHMPWTEKDDTIDWQAYYHARNRWLAALLHSPYS 426  
 Njs 396 WDDSEYGLRAKEAGYPTVTFPGAIVHVPWTDKNDGLDWQAYFHQRNRFVAALLHSPYP 454  
 Cms 401 WDDAEYALRAKEGVPTVTLPGAIVHVSVIDKDDSDQDWQAFFHARNRLIAALLHSPYE 459  
 Cmm 401 WDDAEYALRAKEGVPTVTLPGAIVHVSVIDKDDSDQDWQAFFHARNRLIAALLHSPYE 459  
 Bl2 380 NDDVEYGVRAQRAGYRTVTVPGVCLWHQSFVDDKDDQLDWQAYYHIRNRTIMGLLYANQQ 438  
 Sen 338 GDDLLFGYMHKKH--NIVTLNGVASWQMDFER--KISVLNSYLNFRVAVPALISKRK- 391  
 Seh 338 GDDLLFGYMHKKH--NIVTLNGVASWQMDFER--KISVLNSYLNFRVAVPALISKRK- 391  
 Sew 338 GDDLLFGYMHKKH--NIVTLNGVASWQMDFER--KISVLNSYLNFRVAVPALISKRK- 391  
 Sep 338 GDDLLFGYMHKKH--NIVTLNGVASWQMDFER--KISVLNSYLNFRVAVPALISKRK- 391  
 Ses 338 GDDLLFGYMHKKH--NIVTLNGVASWQMDFER--KISVLNSYLNFRVAVPALISKRK- 391  
 Sev 338 GDDLLFGYMHKKH--NIVTLNGVASWQMDFER--KISVLNSYLNFRVAVPALISKRK- 391  
 Sej 338 GDDLLFGYMHKKH--NIVTLNGVASWQMDFER--KISVLNSYLNFRVAVPALISKRK- 391  
 Rba 335 GDDIEYGCRMAANGVETICLPGVAVWHESFHKK--TSDWLTYYDMRNRLFVASLYPQLV 391  
 Gox 298 GDDAEYGLRLKRAGFPTVMWPGVYVAHPNLQN--QTRPWHHYDRRNALICALLERGAV 354

**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 -----
2z86A 111 PLDWPSDLTL PPLPESTNDY VWAGKRKELL II--DGLSIV IPTYNRAKIL

Glft2 176 VNALRELTA -PLVDQVIGA VIVPDQGERK VRDHPDFPAA AARLGSRLSI
2z86A 166 AITLACL CNQ KTIYD---YE VIVADDGSK- --EN--IEEI VREFESLLNI

Glft2 225 HDQPNLGGSG GYSRVMYEAL KNTDCQQILF MDDDIRLEPD SILRVLAMHR
2z86A 208 KYVRQKDYGY QLCAVRNLGL RAAKYNVVAI LDCDMPNPL WVQSYMELLA

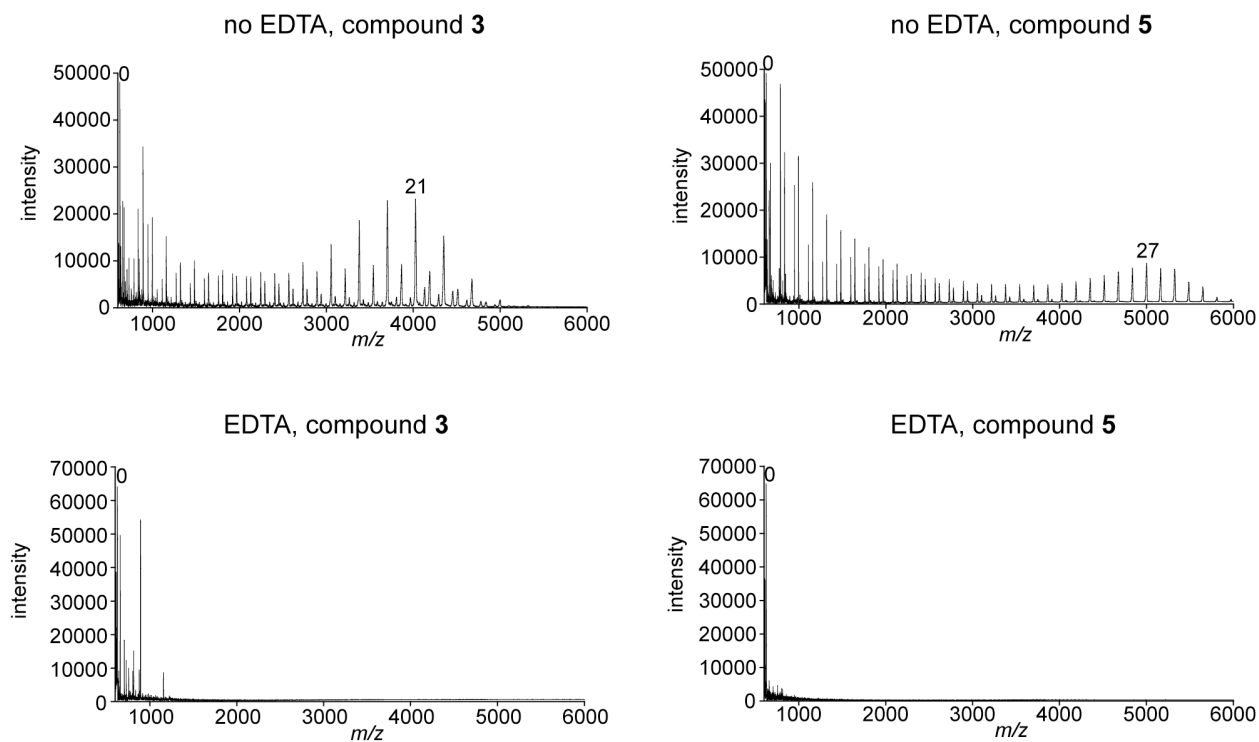
Glft2 275 FAKAPMLVGG QMLNLQE-PS HLHIMGEVVD RSIFMWTAAP HAEYDHDFAE
2z86A 258 VDDNVALIGP RKYIDTSKHT Y----LDFLS QKSLINEIPE SV--DWRIEH

Glft2 324 YPLNDNNSRS KLLHRRIDVD YNGWWTMIP RQVAEELGQP LPLFIKWDDA
2z86A 317 FKN-TDN--- -LRLCNTPFR FFSGGNVAFA KKWLFRAGWF DEEFTHWGGE

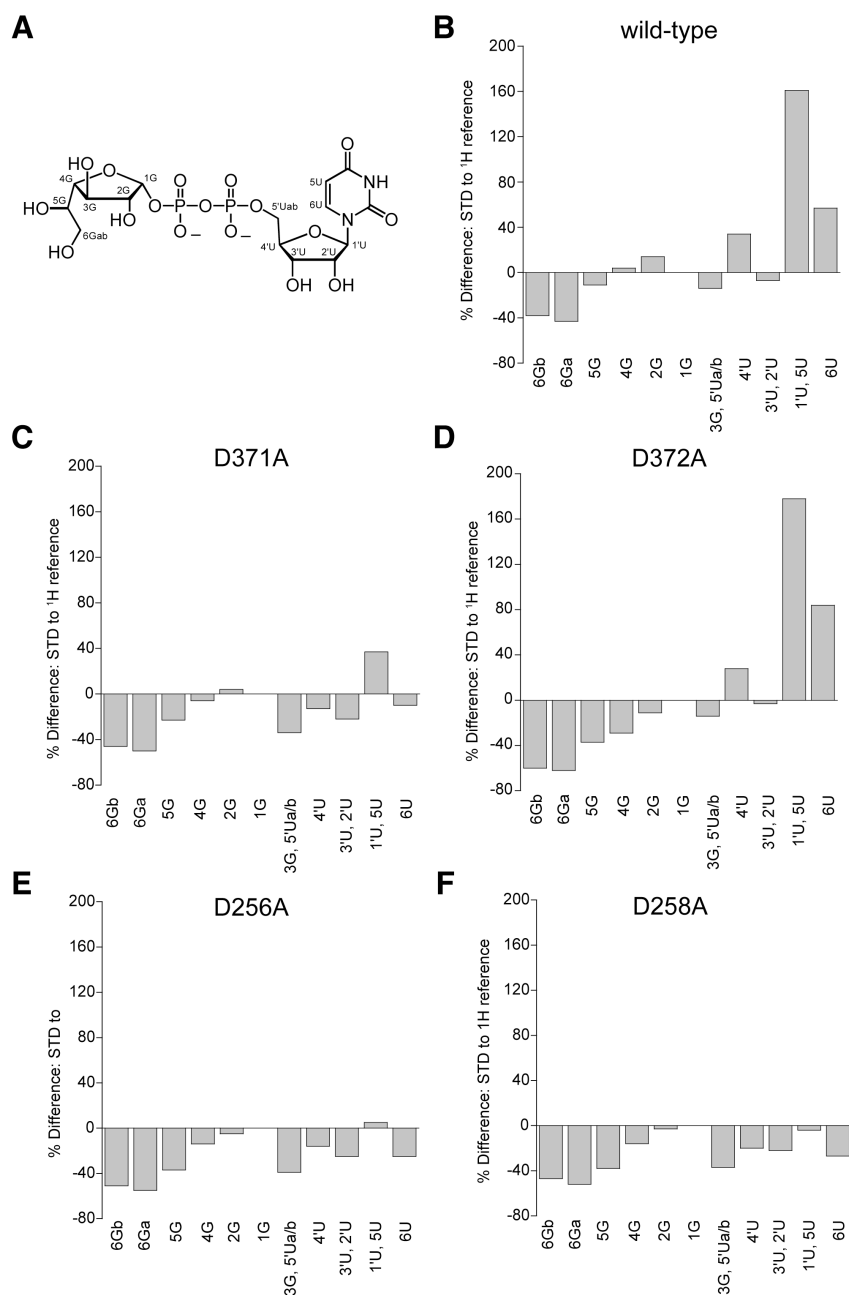
Glft2 374 --DYGLRAAE HGYPTVTLPG AAIWHMA--- -----
2z86A 362 DNEFGYRLYR EGCYFRSVEG AMAYHQEPPQ LLQQKVPYFY RKKEKIESAT

```

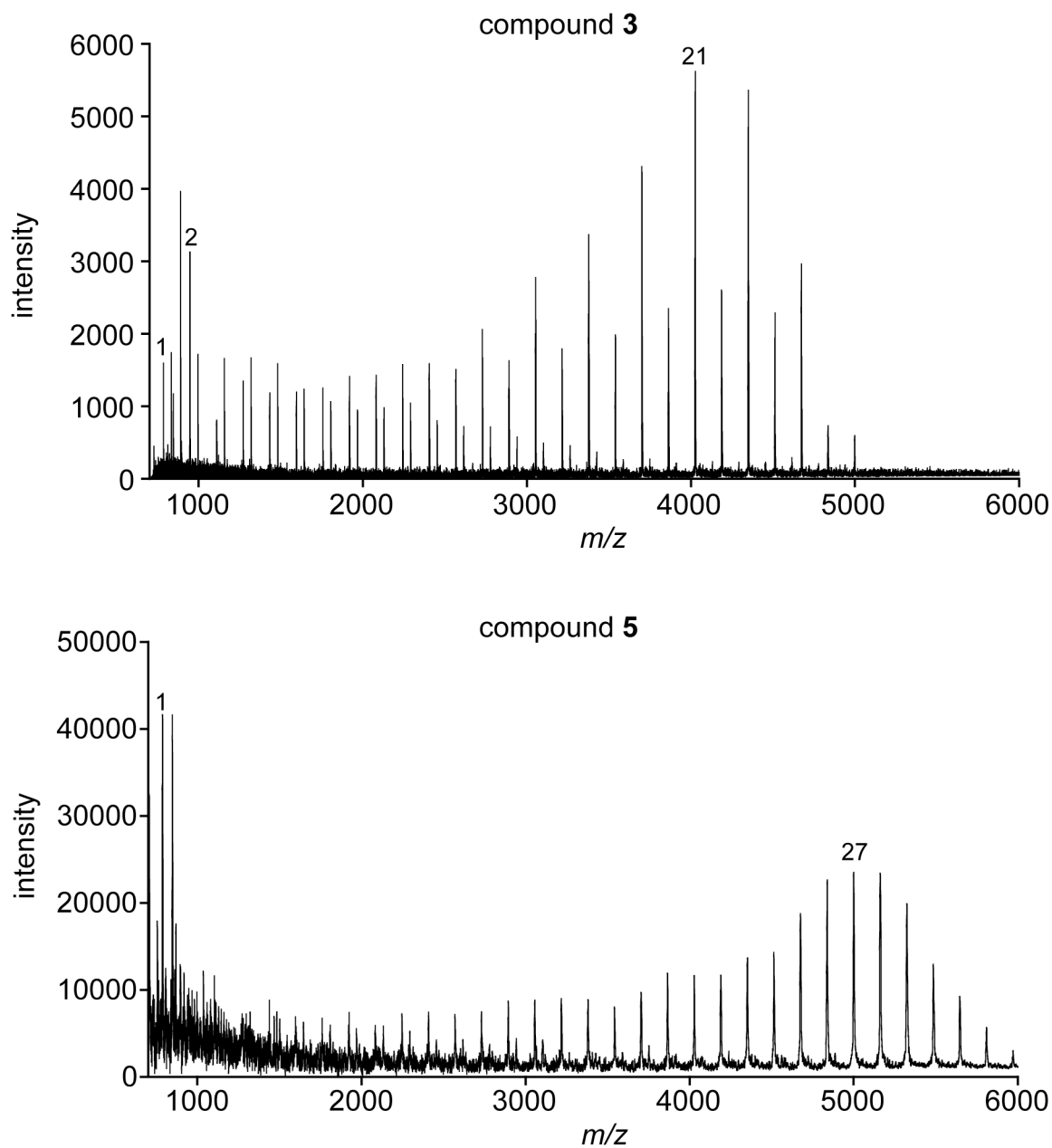
**Figure S3.** Alignment used for homology model of Glft2. This alignment was generated by the SWISS-MODEL server and used as the basis for the homology model shown in Figure 3B. This region of the Glft2 sequence corresponds to its GT2 domain. The DDD and DDA motifs in Glft2 are highlighted in red. This alignment showed 14.8% identity between the two sequences with an e-value of  $3.6 \times 10^{-9}$ .



**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).

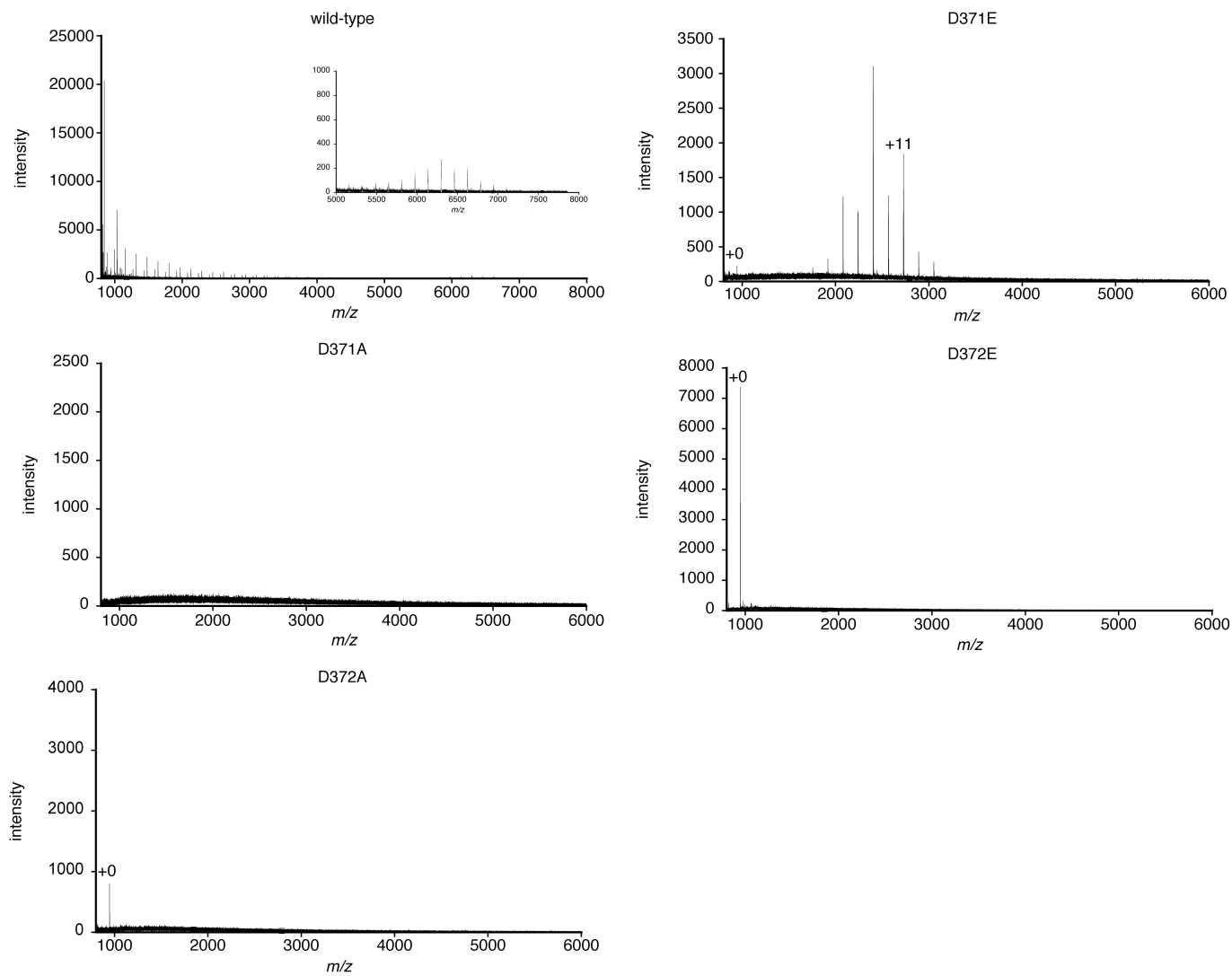


**Figure S5.** Analysis of STD NMR spectra for GlfT2 variants binding to UDP-Galf. Addition of wild-type GlfT2 (B), D371A GlfT2 (C), D372A GlfT2 (D), D256A GlfT2 (E), or D258A GlfT2 (F) to a solution containing UDP-Galf results in saturation transfer difference NMR signals at the indicated protons in the UDP-Galf 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-Galf in a similar manner, D256A and D258A bind UDP-Galf similarly to each other but distinctly from wild-type GlfT2, and D371A binds UDP-Galf 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  $n$  GalF 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 $\beta$  residues to a tetrasaccharide acceptor that is elongated by wild-type GlfT2 (up to +37 Gal $\beta$  residues) or D371E GlfT2 (up to +17 Gal $\beta$  residues).

**A**

```

SpHasA  41  -----KVA AV  IPSYNEDAES  LLET LKSVLA  QTYPLSEIYI  VDDGSSNTDA
3bcvB   3  LIP--KVSVI  VPIYNVE-KY  LDQC VQALLA  QTLSDIEIIL  IDDESP-DNC

SpHasA  86  IQLIEEYVNR  EVDICRNVIV  HRSLV NKGKR  HAQAWAFERS  DADVFLTVDS
3bcvB   49  PKICDDYAAQ  YPN-----IK  VIHKKNAGLG  XACNSGLDVA  TGEYVAFCD S

SpHasA  136  DTYYIPNALE  ELLKSFNDET  VYAATGHLNA  RNRQTNLLTR  LTDIRYDNAF
3bcvB   94  DDYVDSXYYX  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  NRSLPDVRL E  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  LKGA AVSRGQ  VITFLDAHCE  CTAGWLEPLL  ARIKHDRRTV  VCP II-DVIS

CesA    271  PDPFQRNLSA  GYR-----  -----TPPE  GNLFYGVVQD  GND FWDATFF
1xhbA  243  DDTFEYMAGS  DMTYGGFNWK  LNFRWYPVPQ  REMDRR-KGD  RTL PVRTPTM

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  $1.3 \times 10^{-14}$ . (B) Alignment of cellulose synthase from *Gluconacetobacter xylinus* (CesA; GenBank accession number P21877.3) with the glycosyltransferase UDP-GalNAc:polypeptide  $\alpha$ -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  $1.6 \times 10^{-11}$ .

**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.

Gift2 D256A	5' – CAGATCCTATTCATGG <u>CCG</u> GACGACATCCGCCTC – 3'
Gift2 D256E	5' – CAGATCCTATTCATGG <u>GAG</u> GACGACATCCGCCTC – 3'
Gift2 D258A	5' – CTATTCATGGACGACG <u>CC</u> ATCCGCCTCGAGCCG – 3'
Gift2 D258E	5' – CTATTCATGGACGACG <u>AG</u> ATCCGCCTCGAGCCG – 3'
Gift2 D371A	5' – GCCGTTGTTTCATCAAATGGG <u>CCG</u> GACGCCGATTACGGCC – 3'
Gift2 D371E	5' – GCCGTTGTTTCATCAAATGGG <u>AAG</u> GACGCCGATTACGGCC – 3'
Gift2 D372A	5' – GCCGTTGTTTCATCAAATGGGACG <u>CCG</u> CCGATTACGGCC – 3'
Gift2 D372E	5' – GCCGTTGTTTCATCAAATGGGACG <u>AAG</u> CCGATTACGGCC – 3'