Supporting Information

Regulatory Role of One Critical Catalytic Loop of Polypeptide N-Acetyl-Galactosaminyltransferase-2 in Substrate Binding and Catalysis during Mucin-Type O-Glycosylation

Jiaqi Tian‡, Feng Liu‡, Zhijue Xu, Jingjing Shi, Tao Liang, Yan Zhang, Lin-Tai Da**

Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China.

Corresponding Author

*E-mail for Lin-Tai Da: darlt@sjtu.edu.cn

*E-mail for Yan Zhang: yanzhang2006@sjtu.edu.cn

‡These authors contributed equally.

Contents

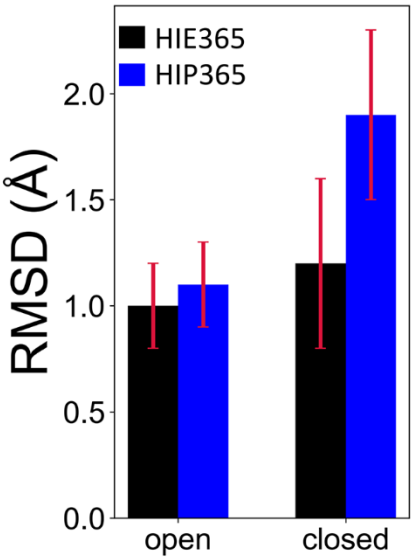

Figure S1. The RMSD of the C_α atoms of the catalytic loop (from R362 to A378) and W331 during the 10-ns MD simulations for the T2 complex with two different protonated forms of His365: HIE365 (black) and HIP365 (blue). The mean and standard errors were calculated based on the corresponding 10-ns MD simulations starting from the open and closed-loop T2 complex.

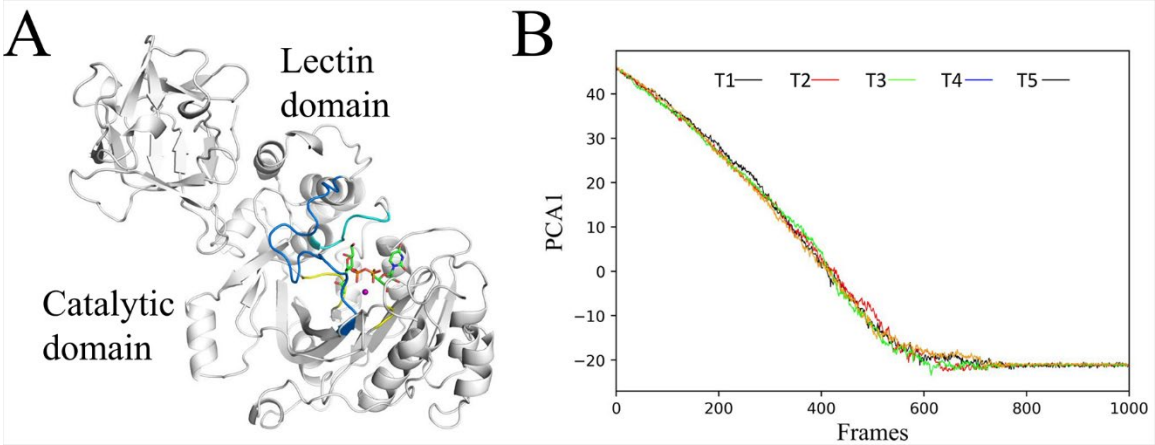

Figure S2. (A) We selected several discontinuous motifs as the targeting regions, including the residues H359-A378 (blue), V330-N335 (cyan), A307-L310 (yellow), S225-H226 (yellow), Mn^{2+} (purple), and UDP-GalNAc (green). (B) The PCA analysis based on the 5 independent TMD simulations with different pulling force constants: 1 (denoted as T1), 2 (T2), 3 (T3), 4 (T4), and 5 (T5) kcal/mol/$Å^2$, respectively.

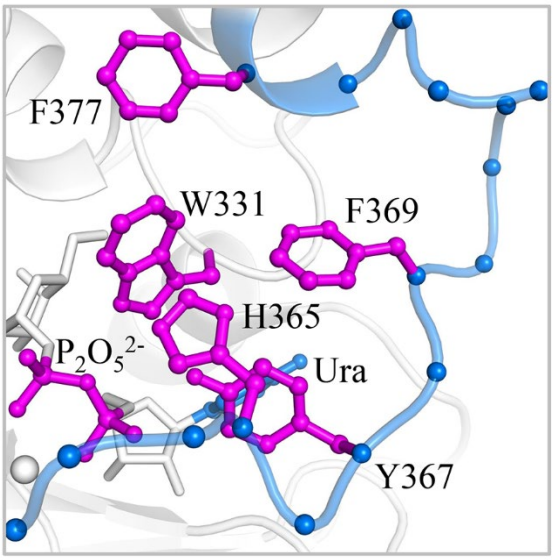

Figure S3. Selected atoms used for tICA, including the C_α atoms of residues R362-F377; sidechain heavy atoms of W331, H365, Y367, F369, and F377; the heavy atoms of pyrophosphate moiety ($P_2O_5^{2-}$) and the uracil moiety (Ura) of UDP-GalNAc. The selected C_α atoms and uracil moiety are shown in blue spheres; other selected heavy atoms are shown in magenta spheres.

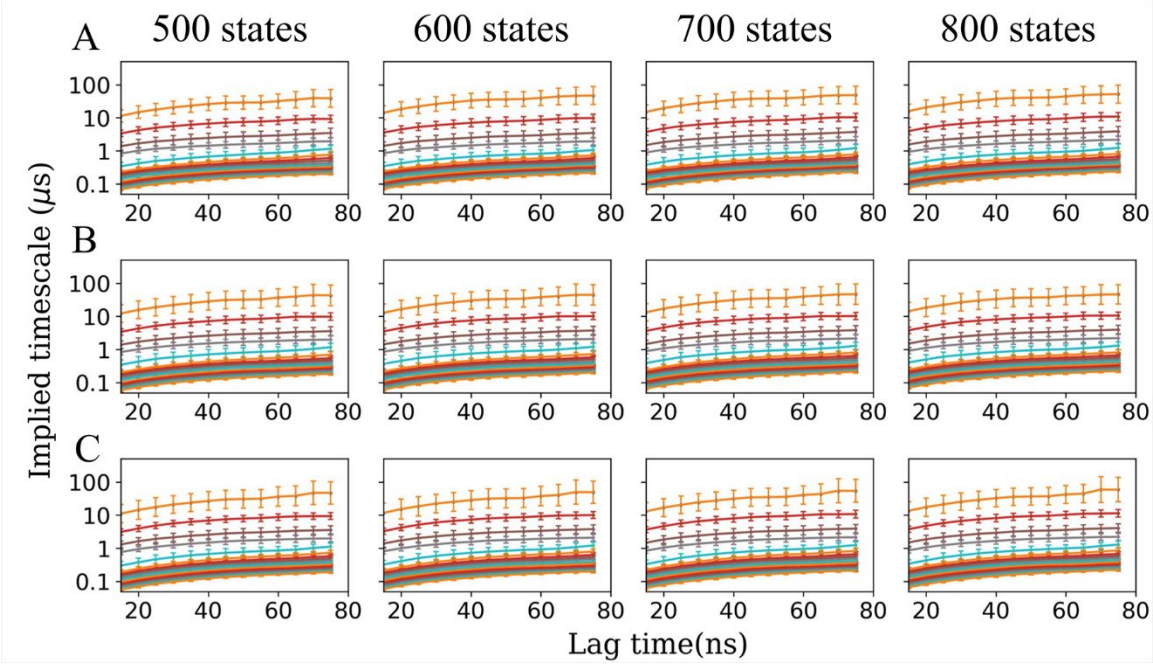

Figure S4. The implied timescale plots as a function of various lag time using different number of microstates and tICA lag time. Under different tICA correlation lag time, 20 ns (A), 30 ns (B), 40 ns (C), the implied timescales were calculated based on the MSM constructed using different number of microstates: 500, 600, 700, and 800, respectively.

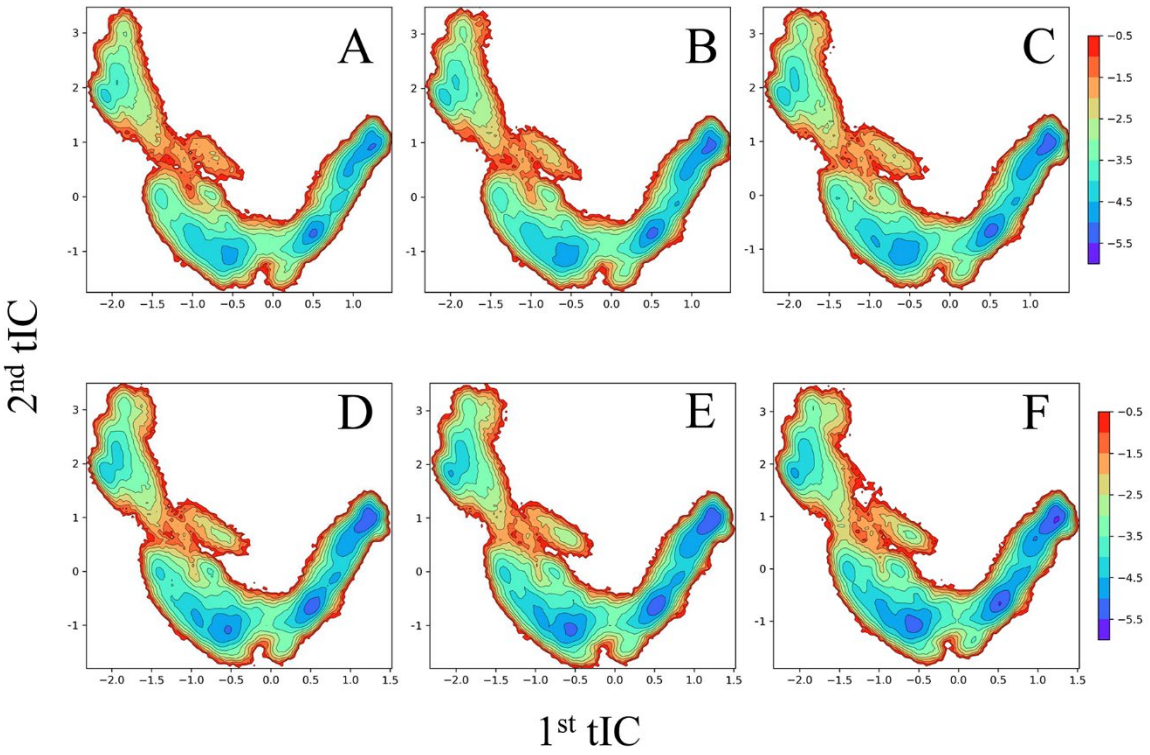

Figure S5. Free energy projections of the MD conformations onto the top two tICs for different datasets with varying aggregated simulation times: (A) 10 μs (50 ns × 199); (B) 12 μs (60 ns × 199); (C) 14 μs (70 ns × 199); (D) 16 μs (80ns × 199); (E) 18 μs (90 ns × 199), and (F) 20 μs (100 ns × 199), respectively.

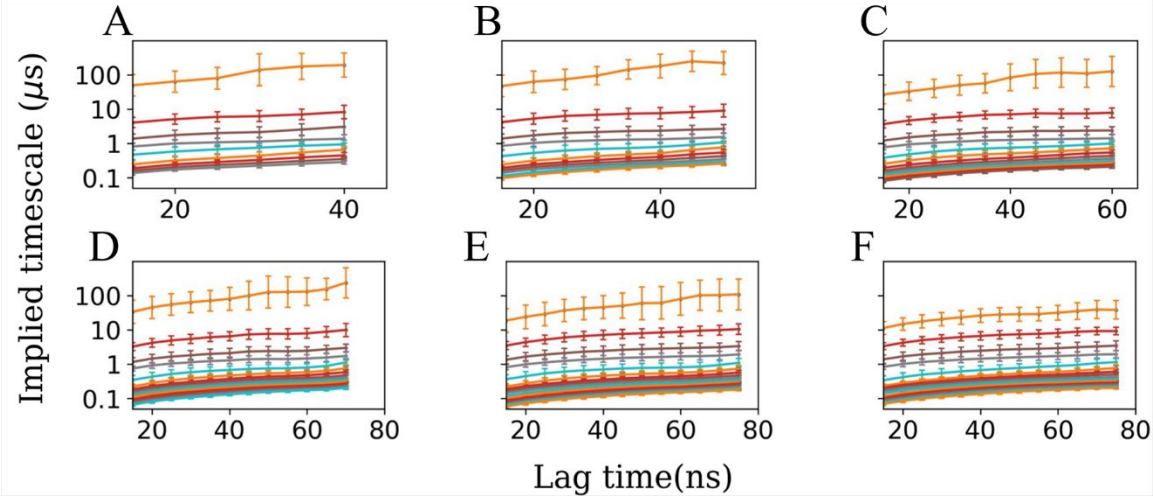

Figure S6. The implied timescale plots for different datasets with varying aggregated simulation times: (A) 10 μs (50 ns × 199); (B) 12 μs (60 ns × 199); (C) 14 μs (70 ns × 199); (D) 16 μs (80 ns × 199); (E) 18 μs (90 ns × 199); (F) 20 μs (100 ns × 199), respectively. We used the microstate number of 500 and tICA correlation lag time of 20 ns to construct each MSM.

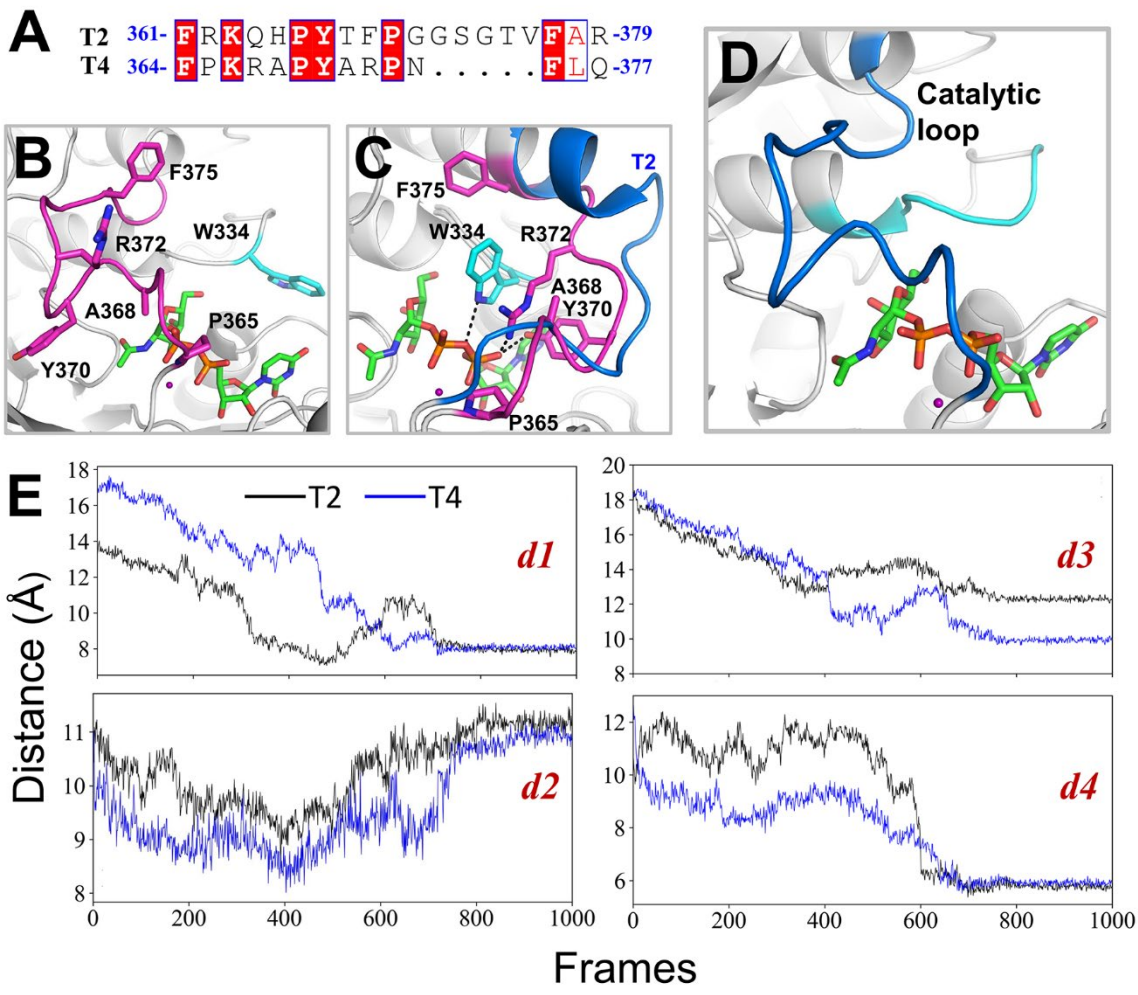

Figure S7. (A) Sequence alignment of the catalytic-loop region between T2 and T4. (B) The modeled UDP-GalNAc-T4 complex with the catalytic loop (residues P365-F375) in an open conformation (in magenta), in which several catalytic-loop residues are shown in magenta sticks; W334 is shown in cyan sticks; UDP-GalNAc and Mn^{2+} are represented in green sticks and magenta sphere, respectively. (C) The constructed UDP-GalNAc-T4 complex where the catalytic-loop adopts a closed form (in magenta), refer to **B** for other representations. The closed-loop conformation in T2 is also overlaid (in blue). (D) The selected regions in T4 used for the TMD simulations, including the residues P365-T379 (in blue), V333-N338 (in cyan), Mn^{2+} (in purple), and UDP-GalNAc (in green). (E) The changes of four distance pairs along the

simulation time for T2 and T4, namely ***d1*** (between the center of mass (COM) of UDP-GalNAc and Y367 in T2); ***d2*** (between the COM of UDP-GalNAc and C_α atom of R362); ***d3*** (between the COM of UDP-GalNAc and C_α atom of F369 in T2); and ***d4*** (between the COM of UDP-GalNAc and W331 in T2). In comparison, the counterpart distances in T4 were also measured.

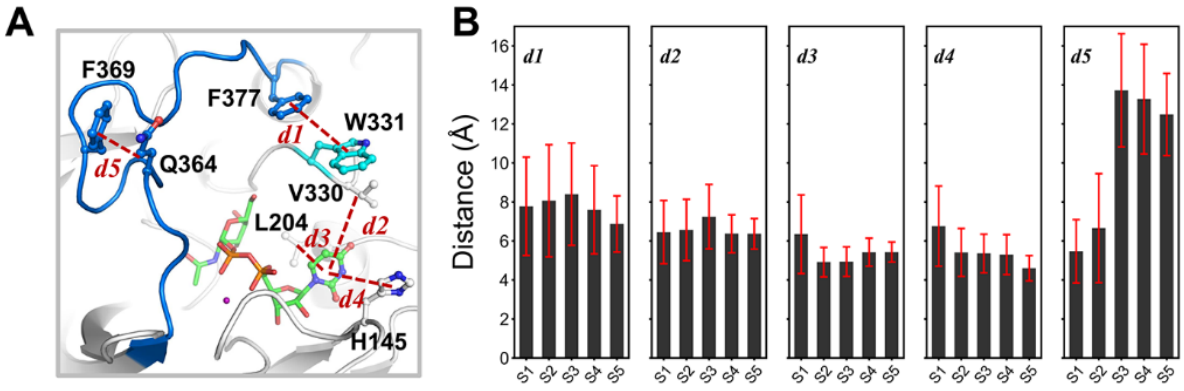

Figure S8. (A) Several residue pairs are defined for the distance calculations, including the pairs between the center of mass (COM) of F377 sidechain and COM of W331 sidechain (*d1*); COM of V330 sidechain and COM of the uracil moiety of UDP-GalNAc (*d2*); COM of L204 sidechain and COM of the uracil moiety of UDP-GalNAc (*d3*); COM of H145 sidechain and COM of the uracil moiety of UDP-GalNAc (*d4*); COM of F369 sidechain and COM of Q364 sidechain (*d5*). (B) The measured distances for *d1, d2, d3, d4,* and *d5* in each state. For each calculation, the mean value (in black) was averaged over all the microstates that belong to the same macrostate, and the corresponding standard error (in red) was calculated.

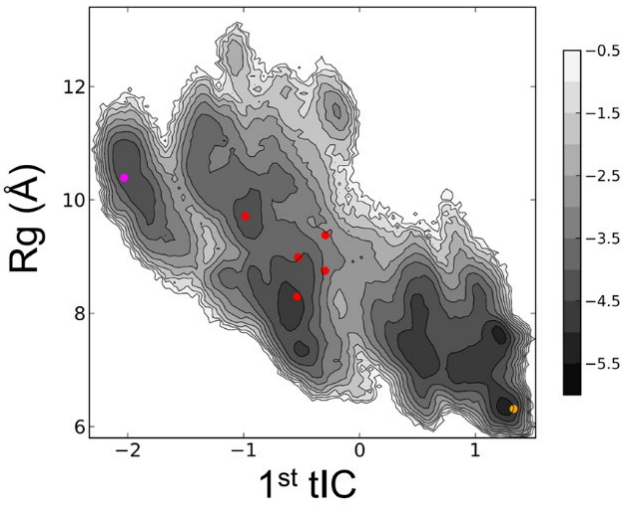

Figure S9. Free energy projections of all the MD conformations onto the first tIC and Rg. The five red points represent the starting structures selected from the S2 state for additional MD simulations. The magenta and orange points represent the crystal structures of T2 complex with the catalytic-loop in open (PDB ID: 2ffv) and closed (PDB ID: 4d0t) states, respectively.

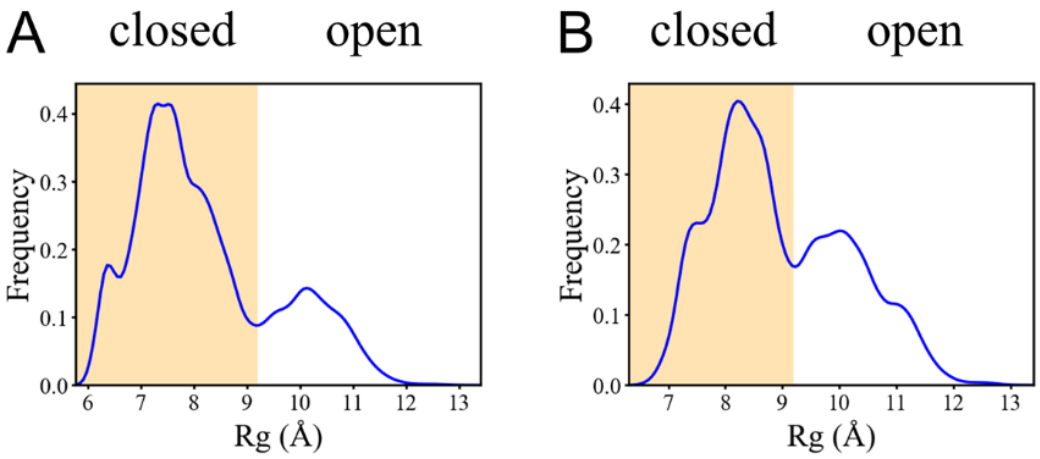

Figure S10. The Rg distributions calculated for H145, L204, V330, W331, H365, Y367, F369, F377, and the uracil moiety (Ura) of UDP-GalNAc based on the complete simulation dataset (A) and only for the S2 conformations (B). Based on the Rg curves, a region with the Rg ~9 Å is defined as a transition state connecting the closed (shadowed in orange background) and open states.

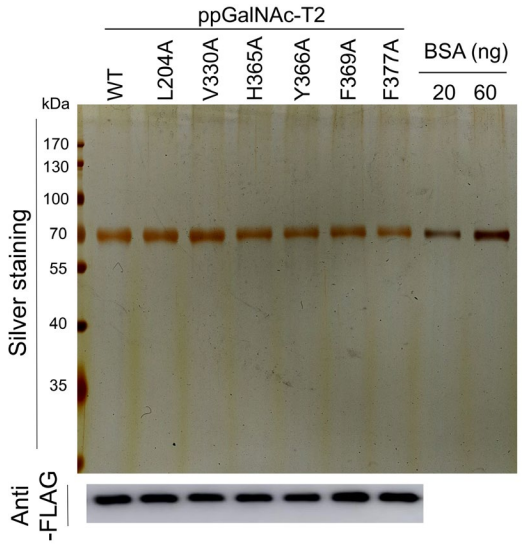

Figure S11. Silver staining and western blot (anti-FLAG) of purified wt T2 and its mutants.

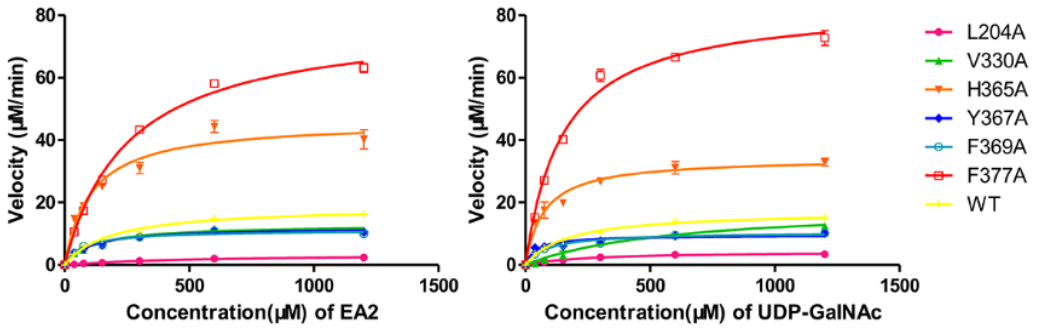

Figure S12. The kinetic curves of wt T2 and its mutants against EA2 and UDP-GalNAc. Each data point represents as mean ± SD (n = 3). The data were fitted using the Michaelis-Menten equation with GraphPad Prism version 5.

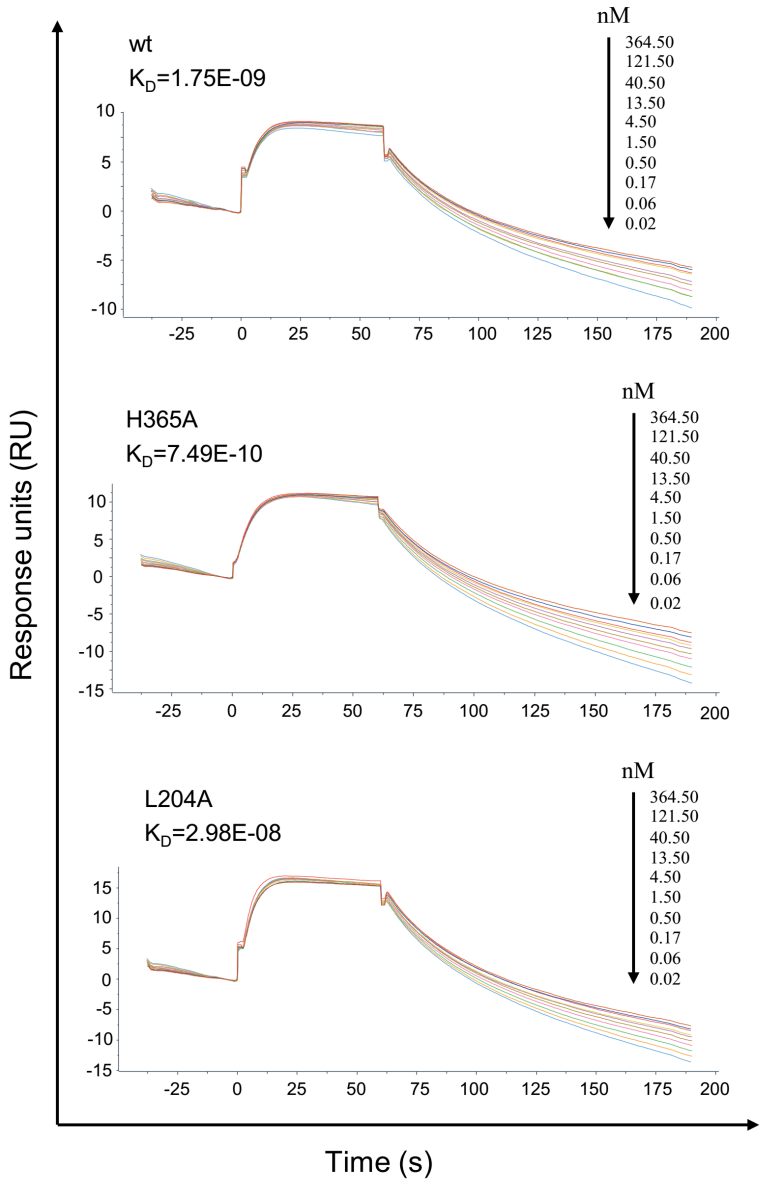

Figure S13. SPR assay (Biacore) of the binding between wt T2 and its mutants with UDP-GalNAc. The change of response units over time is shown.

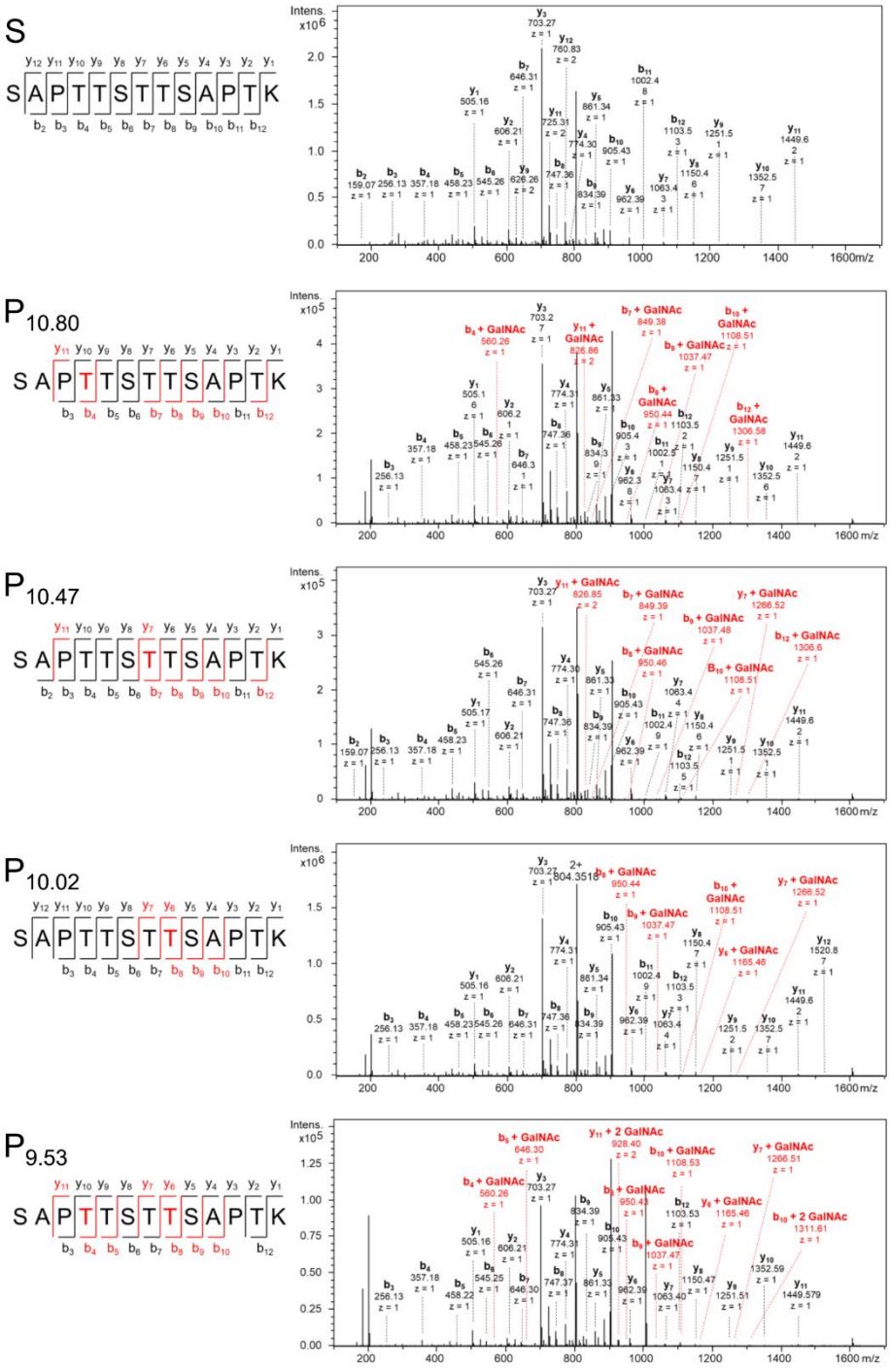

Figure S14. MS/MS spectra of the substrate and glycosylated products of MUC5AC catalyzed by wt T2. T residues in red indicate the glycosylated sites. Glycopeptide b- and y-ions (peptide backbone ions carrying GalNAc residues) were colored in red.

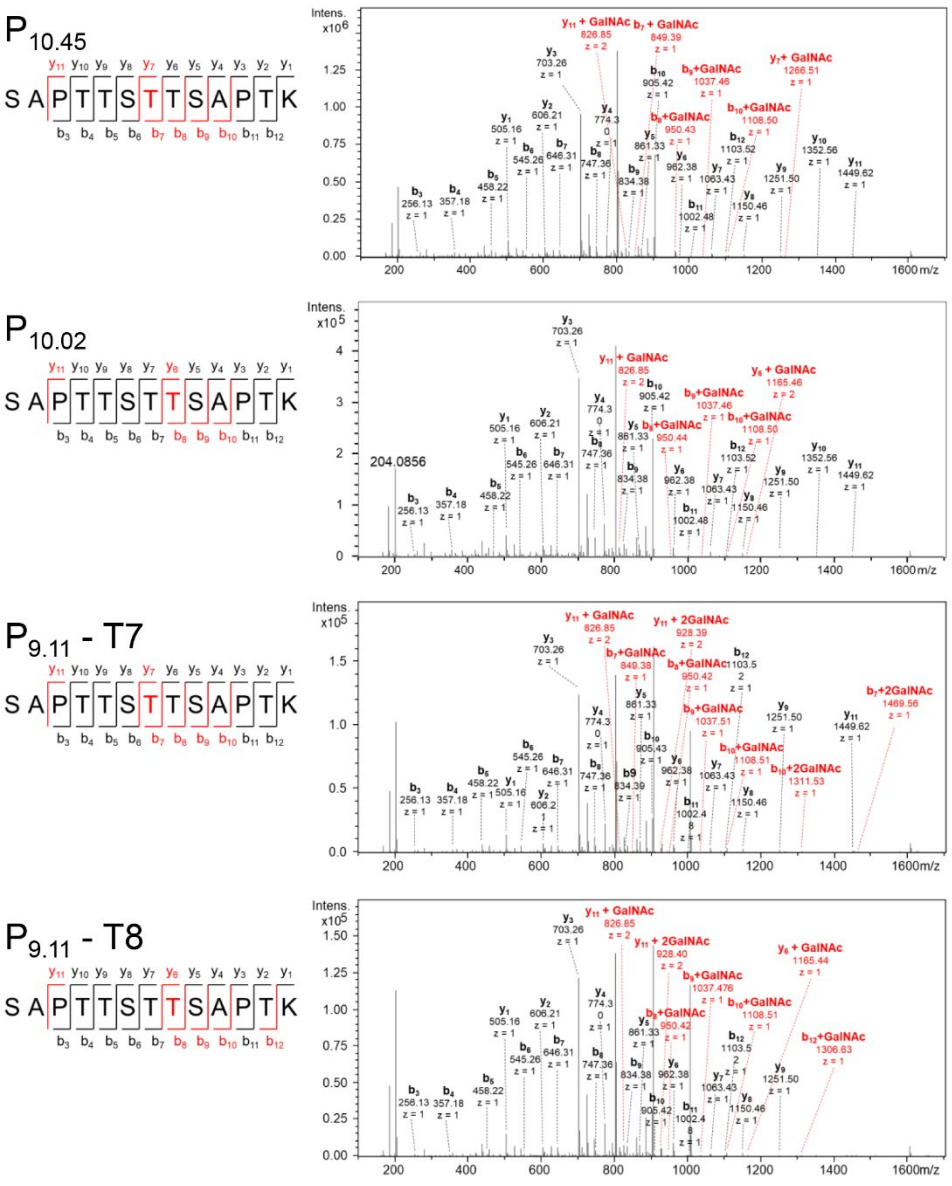

Figure S15. MS/MS spectra of the glycosylated products of MUC5AC catalyzed by T2 H365A mutant. T residues in red indicate the glycosylated sites. Glycopeptide b- and y-ions (peptide backbone ions carrying GalNAc residues) were colored in red. $P_{9.11}$ - T7 indicates T7 is one of the two glycosylated sites on $P_{9.11}$. $P_{9.11}$ - T8 indicates T8 is one of the two glycosylated sites on $P_{9.11}$.

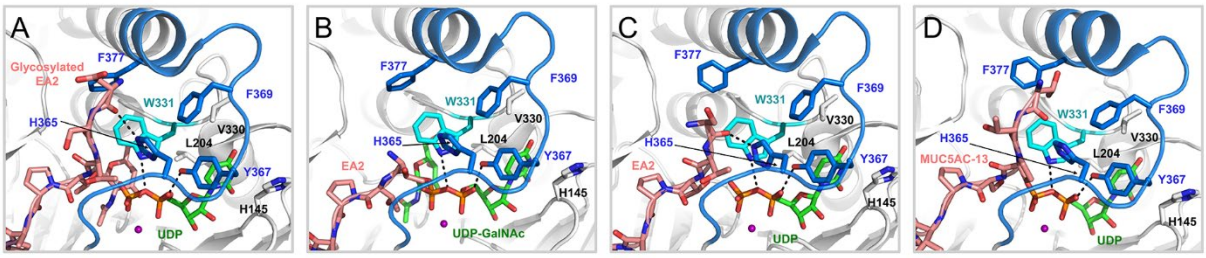

Figure S16. X-ray structures of several ternary T2 complexes. (A) T2 in complex with UDP and glycosylated EA2 (PDB ID: 4d0t); (B) T2 in complex with UDP-GalNAc and EA2 (PDB ID: 4d0t); (C) T2 in complex with UDP and EA2 (PDB ID: 2ffu); (D) T2 in complex with UDP and MUC5AC-13 (PDB ID: 5ajp, and MUC5AC-13 sequence is: GTTPSPVPTTSTT*SA, the T* represents the T is O-linked with a GalNAc moiety). UDP-GalNAc and Mn^{2+} are shown in green sticks and purple sphere, respectively; the catalytic loop is highlighted in blue and the key residues are shown in blue sticks; W331 is shown in cyan sticks; other key residues in T2 are show in gray sticks; the EA2 and MUC5AC-13 peptides are shown in pink sticks.

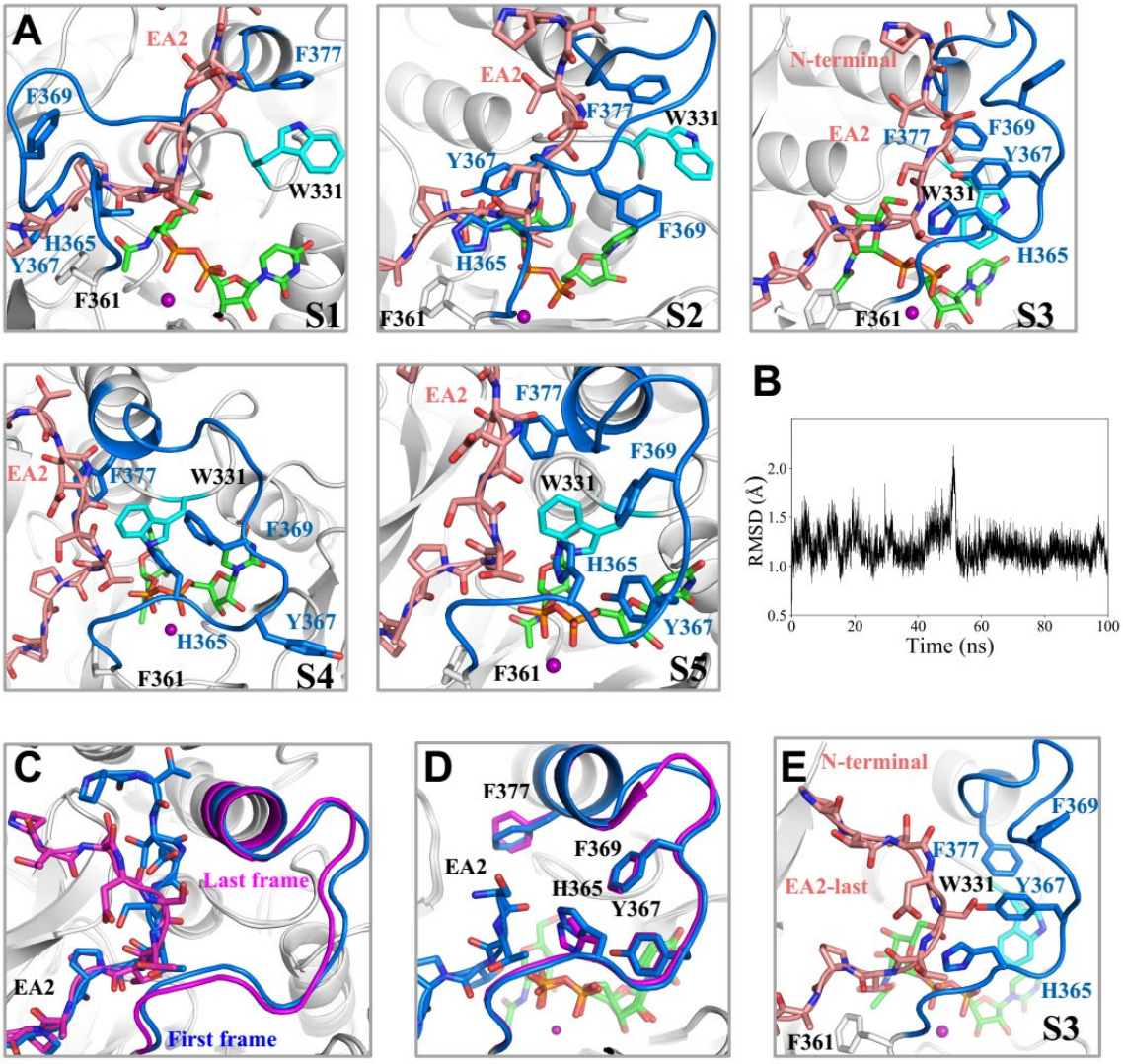

Figure S17. (A) The modeled EA2-UDP-GalNAc-T2 complexes for each macrostate (S1-S5), in which the EA2 peptide is shown in pink sticks. Refer to **Figure 2B** for other representations. (B) The RMSD of the catalytic loop during the 100-ns MD simulations for EA2-UDP-GalNAc-T2 complex with respect to the first frame. (C) The alignment of two EA2-UDP-GalNAc-T2 complexes obtained from the first (in blue) and last frame (in magenta) of the 100-ns simulations starting from the closed UDP-GalNAc-T2-EA2 complex. (D) Structural alignment of the crystal structure of the EA2-UDP-T2 complex shown in blue (PDB ID: 2ffu) and UDP-GalNAc-T2

complex shown in magenta (PDB ID: 4d0t). (E) The modeled EA2-UDP-GalNAc-T2 complexes for S3, in which the EA2 is obtained from the last snapshot of the MD simulations.

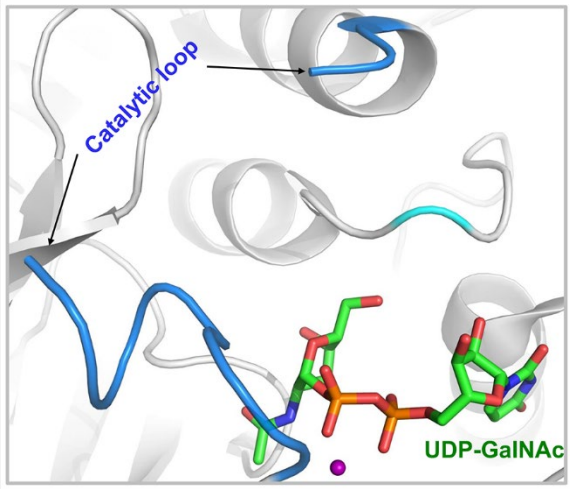

Figure S18. X-ray structure of the UDP-GalNAc-T2 complex where the UDP-GalNAc is in an "inverted" conformation (PDB ID: 6egs).

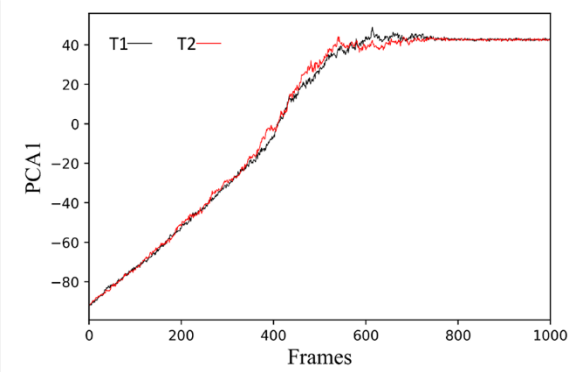

Figure S19. The PCA analysis based on the 2 TMD trajectories starting from the open-loop conformation with two different binding modes of UDP-GalNAc: the catalytic form (T1) and an invert form (T2) of UDP-GalNAc.

Table S1 Products of MUC5AC peptide glycosylated by wt and H365A mutant T2.

	T2-WT	T2-H365A	T2-WT	T2-H365A	Theoretical	T2-WT	T2-H365A	GalNAc No.	Glycosylated Peptide[b]
	RT (min)[a]		Peak area (%)		m/z (z = 2)				
S	11.47	11.46	33.6	23.7	804.341	804.351	804.352	0	SAPTTSTTSAPTK
$P_{10.80}$	10.80		15.2		905.881	905.890		1	SAPT*TSTTSAPTK
$P_{10.47}/P_{10.45}$	10.47	10.45	6.4	42.6	905.881	905.892	905.892	1	SAPTTST*TSAPTK
$P_{10.02}$	10.02	10.02	29.5	28.1	905.881	905.893	905.887	1	SAPTTSTT*SAPTK
$P_{9.53}$	9.53		15.4		1007.421	1007.932		2	SAPT*TSTT*SAPTK
$P_{9.11}$		9.11		5.5	1007.421		1007.433	2	SAPTTST*T*SAPTK

a, RT indicates retention time.

b, * indicates glycosylated site.