Structural and mechanistic insights into the catalytic domain-mediated short-range glycosylation preferences of GalNAc-T4

Matilde de las Rivas[1,12], Earnest James Paul Daniel[2,12], Helena Coelho[3,4], Erandi Lira-Navarrete[5], Lluis Raich[6], Ismael Compañón[7], Ana Diniz[3], Laura Lagartera[8], Jesús Jiménez-Barbero[4,9], Henrik Clausen[3], Carme Rovira[6,10], Filipa Marcelo[3], Francisco Corzana[7], Thomas A. Gerken[2]*, and Ramon Hurtado-Guerrero[1,11]*

[1]. BIFI, University of Zaragoza, BIFI-IQFR (CSIC) Joint Unit, Mariano Esquillor s/n, Campus Rio Ebro, Edificio I+D, Zaragoza, Spain.

[2]. Departments of Biochemistry, Pediatrics and Chemistry, Case Western Reserve University, Cleveland, OHIO, USA

[3]. UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade de Nova de Lisboa, Caparica, Portugal.

[4]. CIC bioGUNE, Bizkaia Technology Park, Building 801A, 48170 Derio, Spain; Departament of Organic Chemistry II, Faculty of Science & Technology, University of the Basque Country, 48940 Leioa, Bizkaia, Spain

[5]. Copenhagen Center for Glycomics, Department of Cellular and Molecular Medicine, School of Dentistry, University of Copenhagen, Copenhagen, Denmark.

[6]. Departament de Química Inorgànica i Orgànica (secció de Química Orgànica) & Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain.

[7]. Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química, E-26006 Logroño, Spain.

[8] Instituto de Química Médica, IQM-CSIC, 28006 Madrid, Spain.

[9]. Ikerbasque, Basque Foundation for Science, Maria Diaz de Haro 13, 48009 Bilbao, Spain

[10]. Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys, 23, 08010 Barcelona, Spain.

[11]. Fundación ARAID, 50018, Zaragoza, Spain.

[12]. These authors contributed equally to this work.

* To whom correspondence should be addressed.

E-mail: rhurtado@bifi.es and txg2@cwru.edu

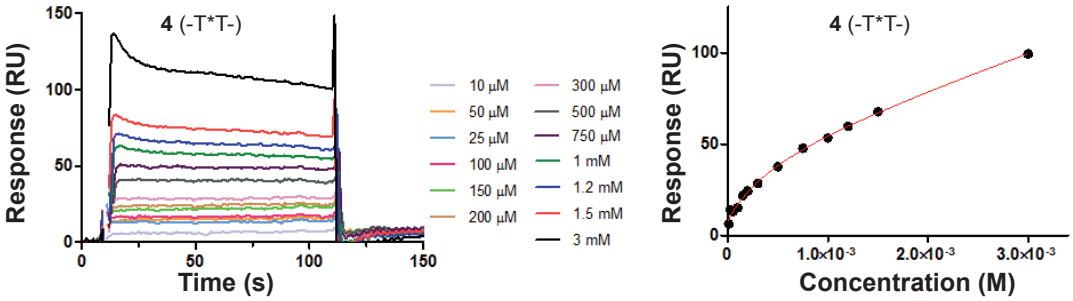


Figure S1. SPR sensogram of glycopeptide 4 binding to GalNAc-T4. Sensogram (left) and fitting (right) of SPR data for binding of glycopeptide **4** to GalNAc-T4**.** Glycopeptide concentrations are reported in the inset legend of the sensogram (left). The end-point values for each injection were plotted against glycopeptide concentration (right). Note that the Kd could not be determined for peptide **4** because binding saturation could not be achieved. However, the Kd was accurately determined for glycopeptide **6** as 70 ± 15 µM (see Figure 1b).

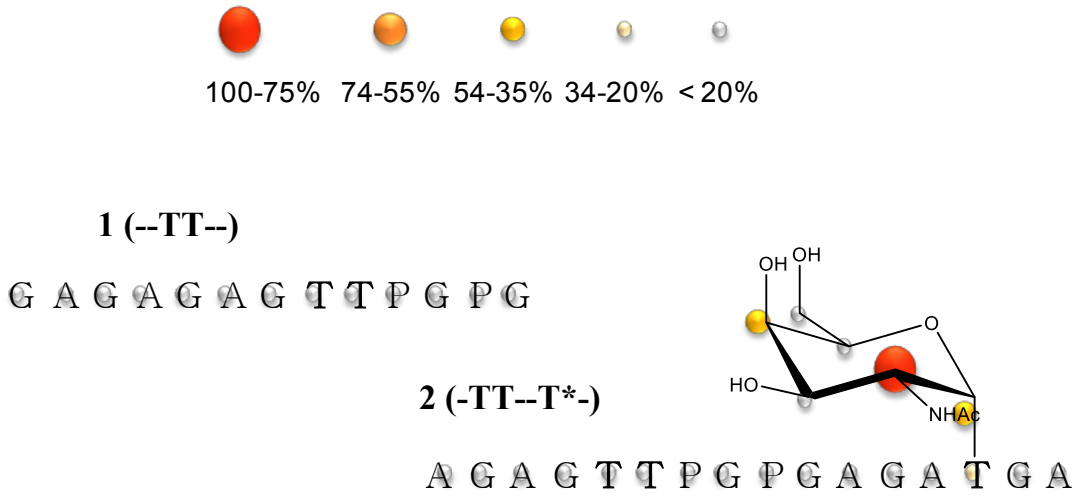


Figure S2. STD-NMR-derived epitope mapping of compounds 1 and 2 with GalANc-T4. Data taken from previous studies[1] for comparison to the present work.

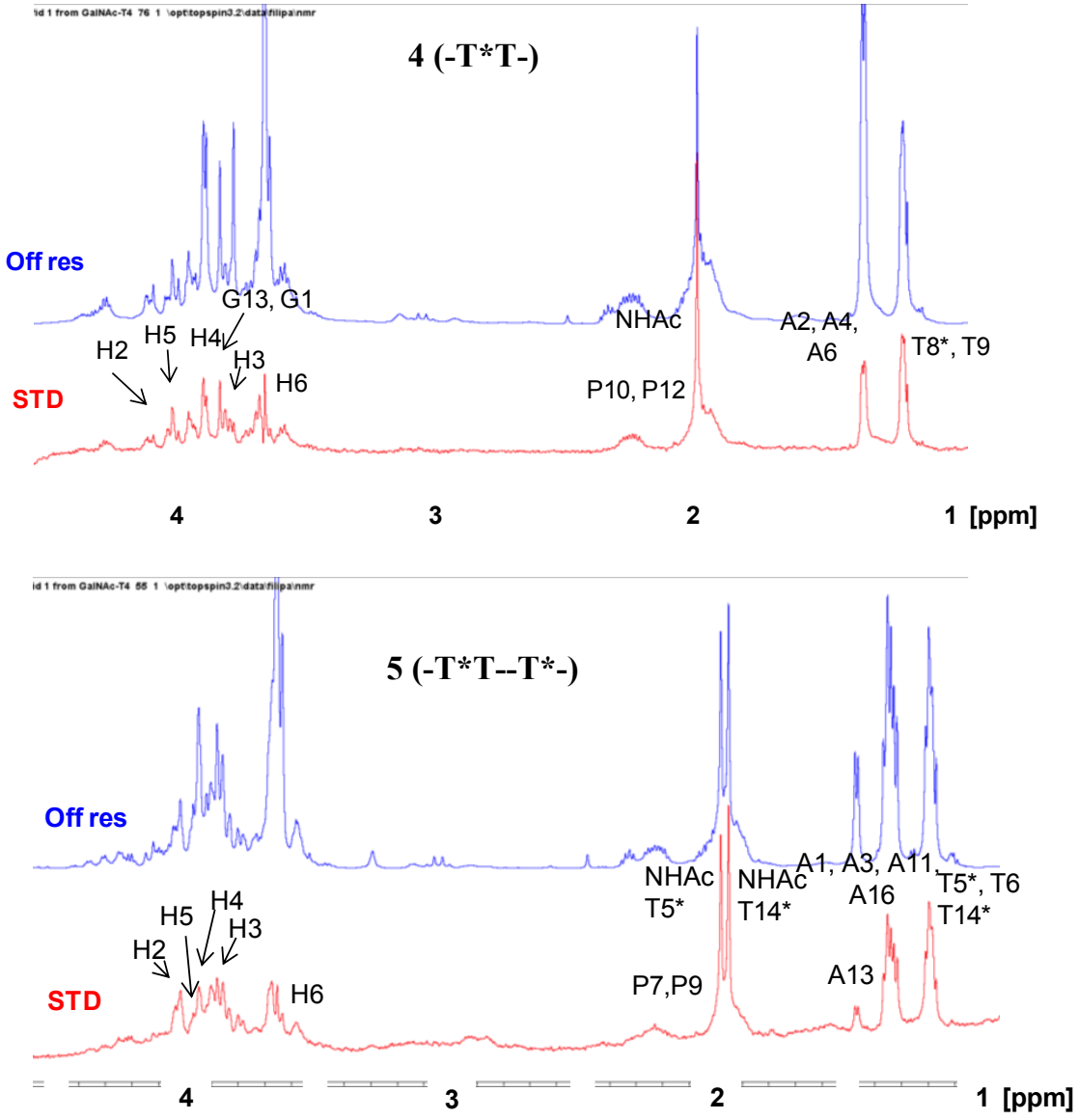
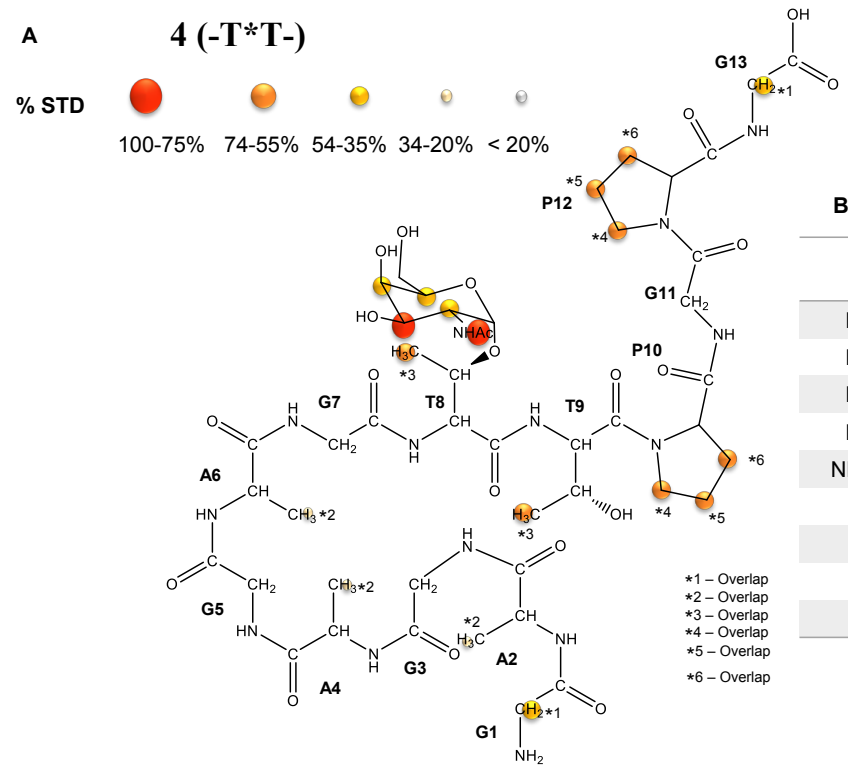


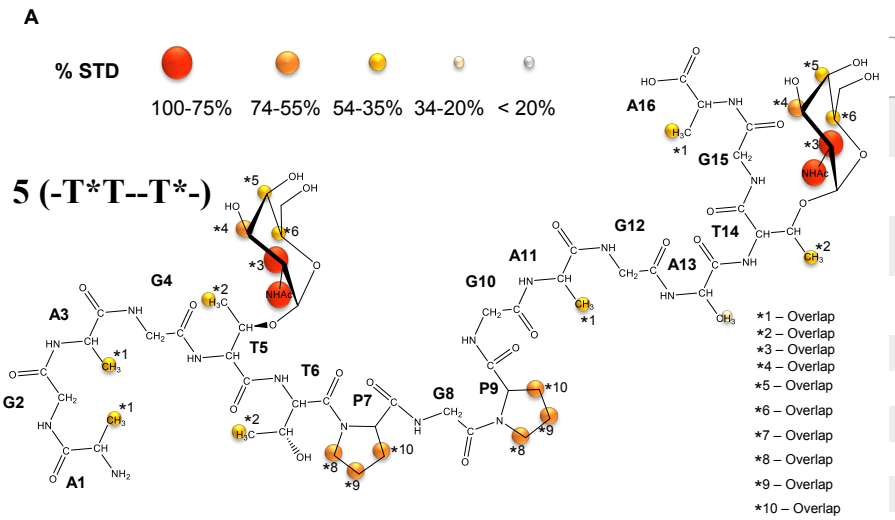
Figure S3. STD-NMR experiments for glycopeptides 4 and 5. Spectra obtained at 877.5 µM in the presence of GalNAc-T4 (13.5 µM), UDP (75 µM), MnCl2 (75 µM) at 298 K. The reference spectrum (Off res) is displayed in blue color while the STD spectrum (STD) is displayed in red. The key proton resonances are marked in each STD spectrum.



B

Proton	% STD relative
H2 – GalNAc-T8	53
H3 – GalNAc-T8	87
H4 – GalNAc-T8	35
H5 – GalNAc-T8	42
NHAc – GalNAc-T8	100
A2, A4, A6	26
G1, G13	48
T8, T9	60
P10, P12	62

Figure S4. **(a) STD-NMR-derived epitope mapping obtained for 4 (-T*T-) with GalNAc-T4. Protons** that could not be accurately analyzed in the STD spectrum were not mapped. In particular, the signal of the anomeric proton H1 of GalNAc could not be analyzed in the STD spectra due to their close distance to the HDO resonance, H6 protons could not be analysed in the STD spectra due to Tris buffer interference and the Hα of Gly amino acids protons due to their overlap with the sugar protons. The proton resonances that overlap in the spectrum are identified in the figure and displayed with *. **(b)** Relative STD percentages for proton resonances relative to the NHAc methyl.



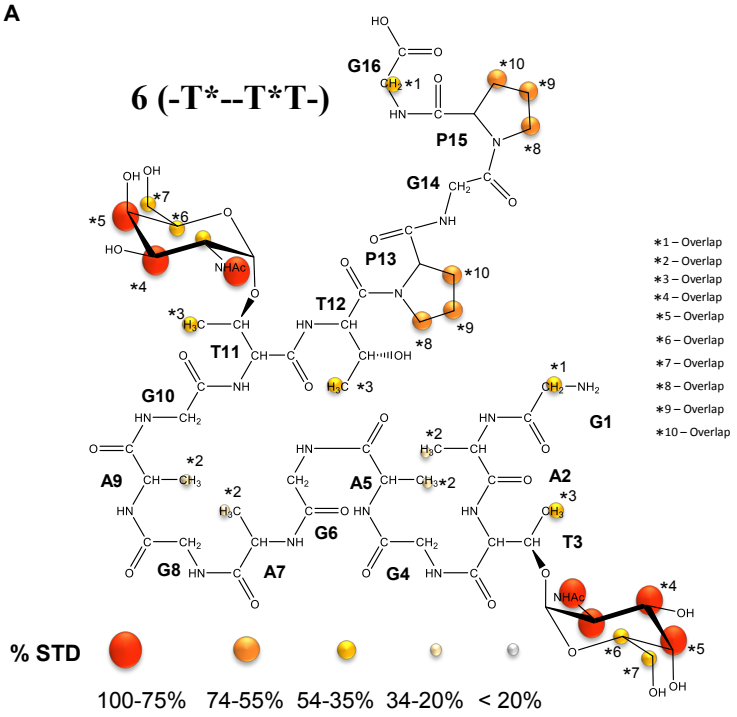
B

Proton	% STD relative
H2 – GalNAc-T5 &GalNAc-T14	100
H3 – GalNAcT5 &GalNAc-T14	70
H4 – GalNAc-T5 &GalNAc-T14	38
H5 – GalNAc-T5& GalNAc-T14	43
NHAc – GalNAc-T5	85
NHAc – GalNAc-T14	85
A1, A3, A11, A16	41
A13	20
T5, T6, T14	45
P7, P9	65

Figure S5. **(a) STD-NMR-derived epitope mapping obtained for 5 (-T*T—T*-) with GalNAc-T4.** Protons that could not be accurately analyzed in the STD spectrum were not mapped. In particular, the signal of the anomeric proton H1 of GalNAc could not be analyzed in the STD spectra due to their close distance to the HDO resonance, H6 protons could not be analysed in the STD spectra due to Tris buffer interference and the Hα of Gly amino acids protons due to their overlap

with the sugar protons The proton resonances that overlap in the spectrum are identified in the figure and displayed with *. **(b)** Relative STD percentages for proton resonances relative to the overlapping H2 GalNAc protons.

A



B

Proton	% STD relative
H2 – GalNAc-T3	88
H2 – GalNAc-T11	54
H3 – GalNAc-T3 &GalNAc-T11	78
H4 – GalNAc-T3 &GalNAcT11	88
H5 – GalNAcT3 &GalNAcT11	54
H6 – GalNAc-T3& GalNAc-T11	50
NHAc – GalNAc-T3	100
NHAc – GalNAc-T11	100
A2, A5, A7, A9	30
G1,G16	46
T3, T11, T12	45
P13, P15	64

Figure S6. **(a) STD-NMR-derived epitope mapping obtained for 6 (-T*--T*T-) with GalNAc-T4**. Protons that could not be accurately analyzed in the STD spectrum were not mapped. In particular, the signal of the anomeric proton H1 of GalNAc could not be analyzed in the STD spectra due to their close distance to the HDO resonance, H6 protons could not be analysed in the STD spectra due to Tris buffer interference and the Hα of Gly amino acids protons due to their overlap with the sugar protons. The proton resonances that overlap in the spectrum are identified in the figure and display with *. **(b)** Relative STD percentages for proton resonances relative to the NHAc methyl.

Table S1. ^{1}H-NMR assignments of glycopeptide 4 (-T*T-) in H_2O/D_2O (90:10) at 278K.

Residue	Proton	δ ppm
T8* GalNAc	H1	4.473
	H2	3.859
	H3	3.564
	H4	3.648
	H5	3.705
	H6	3.436
	NHAC	1.742
G1	H	8.426
	QA	3.581
A2	H	7.977
	QB	1.086
	HA	4.010
G3	H	8.221
	QA	3.687
A4	H	8.125
	QB	1.114
	HA	4.026
G5	H	8.302
	QA	3.631
A6	H	7.985
	HA	4.031
	QB	1.103
G7	H	8.196
	QA	3.737
T8*	H	8.269
	HA	4.343
	HB	4.026
	QG2	0.919
T9	H	8.199
	HA	4.293
	QG2	0.953
P10	HD3	3.358
	HA	4.104
	HD2	3.335
	QB	1.977
	QG	1.689
G11	H	8.186
	QA	3.765
P12	HD3	3.497
	HA	4.116
	HD2	3.458
	HB3	2.033
	HB2	2.020
	HG3	1.763
	HG2	1.753
G13	H	8.411
	QA	3.593

Table S2. ^{1}H-NMR assignments of glycopeptide 5 (-T*T—T*-) in H_2O/D_2O (90:10) at 278K.

Residue	Proton	δ ppm
T5* GalNAc	H1	4.651
	H2	3.924
	H3	3.666
	H4	3.714
	H5	3.765
	H6	3.435
	NHAc	17.474
T14* GalNAc	H1	4.481
	H2	3.924
	H3	3.666
	H4	3.714
	H5	3.765
	H6	3.435
	NHAc	17.197
A1	H	8.170
	QB	1.095
	HA	3.947
G2	H	8.440
	QA	3.715
A3	H	8.262
	QB	1.112
	HA	4.049
G4	H	8.279
	QA	3.710
T5*	H	8.335
	QG2	0.950
	HB	4.022
	HA	4.339
T6	H	8.193
	QG2	0.972
	HB	3.877
	HA	4.291
P7	HD3	3.513
	HA	4.114
	HD2	3.449
	QB	2.014
	QD	1.712
G8	H	8.199
	QA	3.688
P9	HA	4.137
	QD	3.345
	QB	1.979
	QG	1.696

Residue	Proton	δ ppm
G10	H	8.364
	QA	3.635
A11	H	7.939
	HA	4.173
	QB	1.122
G12	H	8.248
	QA	3.629
A13	H	7.899
	QB	1.138
	HA	4.183
T14*	H	8.287
	HA	4.302
	HB	4.070
	QG2	0.977
G15	H	8.150
	QA	3.602
A16	H	8.027
	QB	1.133
	HA	3.995

Table S3. ^{1}H-NMR assignments of glycopeptide 6 (-T*--T*T-) in H_2O/D_2O (90:10) at 278K.

Residue	Proton	δ ppm
T3* GalNAc	H1	4.469
	H2	3.853
	H3	3.569
	H4	3.641
	H5	3.708
	H6	3.442
	NHAC	17.675
T11* GalNAc	H1	4.682
	H2	3.762
	H3	3.569
	H4	3.641
	H5	3.708
	H6	3.442
	NHAC	1.742
G1	H	8.430
	QA	3.586
A2	H	8.430
	HA	4.226
	QB	1.130
T3*	H	8.494
	HA	4.282
	HB	4.035
	QG2	0.969
G4	H	8.244
	QA	3.668
A5	H	8.236
	HA	3.979
	QB	1.087
G6	H	8.320
	QA	3.612
A7	H	8.040
	HA	3.991
	QB	1.074
G8	H	8.264
	QA	3.651
A9	H	7.945
	HA	4.028
	QB	1.105
G10	H	8.203
	QA	3.746
T11*	H	8.260
	HA	4.343
	HB	4.041
	QG2	0.952
T12	H	8.195
	HA	4.294
	QG2	0.964
P13	HA	4.128
	HD3	3.495
	QB	2.013
	QG	1.709
	HD2	3.453
G14	H	8.189
	QA	3.846
P15	HA	4.101
	QG	1.718
	QB	1.982
	HD3	3.366
	HD2	3.335
G16	H	8.415
	QA	3.591

Table S4. Data collection and refinement statistics. Values in parentheses refer to the highest resolution shell. Ramachandran plot statistics were determined with PROCHECK.

	GalNAc-T4 in complex with the peptide **6** and UDP/Mn^{+2}
Space group	P1
Wavelength (Å)	0.97
Resolution (Å)	20-1.80 (1.90-1.80)
Cell dimensions a, b, c (Å) α, β, γ (°)	*a = 66.09* *b* = 81.33 *c* = 84.86 *65.55, 68.66, 74.31*
Mn(I) half-set correlation CC(1/2)	0.998 (0.529)
Unique reflections	133764
Completeness	97.3 (95.6)
R_{pim}	0.044 (0.742)
$I/\sigma(I)$	9.2 (0.9)
Redundancy	3.7 (3.8)
R_{work} / R_{free}	0.189/0.238
RMSD from ideal geometry, bonds (Å)	0.014
RMSD from ideal geometry, angles (º)	1.823
<*B*> *GalNAc-T4* (Å^2)	54.31
<*B*> UDP (Å^2)	38.48
<*B*> Peptide **6** (Å^2)	66.49
<*B*> solvent (Å^2)	52.86
<*B*> Ethylenglycol(Å^2)	68.01
<*B*> Glycerol (Å^2)	79.02
Ramachandran plot: Most favoured (%) Additionally allowed (%) Disallowed (%)	 95.14 3.69 1.17
PDB ID	6H0B

Table S5. HPLC and MS Characterization Glycopeptides 4, 5 and 6

Peptide 4 GAGAGAGT*TPGPG: HPLC: Rt = 10.73 min (Grad: water 0.1% TFA/acetonitrile (95:5) → (85.6:14.4), 15 min, λ = 212 nm). HRMS ESI+ (m/z) calcd. for $C_{47}H_{78}N_{15}O_{20}$ $[M+H]^+$ 1172.5542, found 1172.5494.

Peptide 5: AGAGT*TPGPGAGAT*GA: HPLC: Rt = 11.62 min (Grad: water 0.1% TFA/acetonitrile (95:5) → (86.3:13.7), 14 min, λ = 212 nm). HRMS ESI+ (m/z) calcd. for $C_{65}H_{109}N_{19}O_{29}$ $[M+2H]^{2+}$ 809.8814, found 809.8829.

Peptide 6 GAT*GAGAGAGT*TPGPG: HPLC: Rt = 11.20 min (Grad: water 0.1% TFA/acetonitrile (95:5) → (86.3:13.7), 14 min, λ = 212 nm). HRMS ESI+ (m/z) calcd. for $C_{64}H_{107}N_{19}O_{29}$ $[M+2H]^{2+}$ 802.8736, found 802.8743.

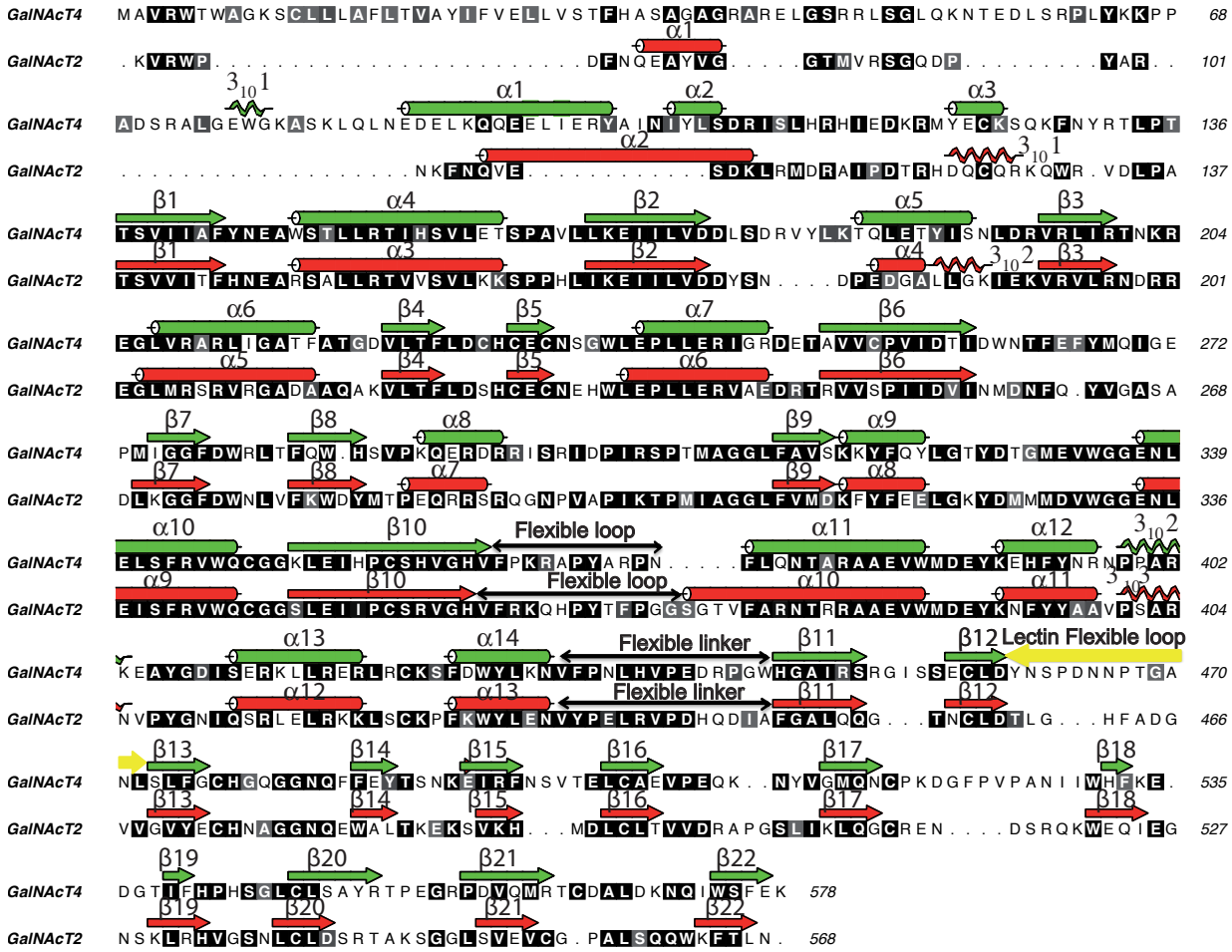
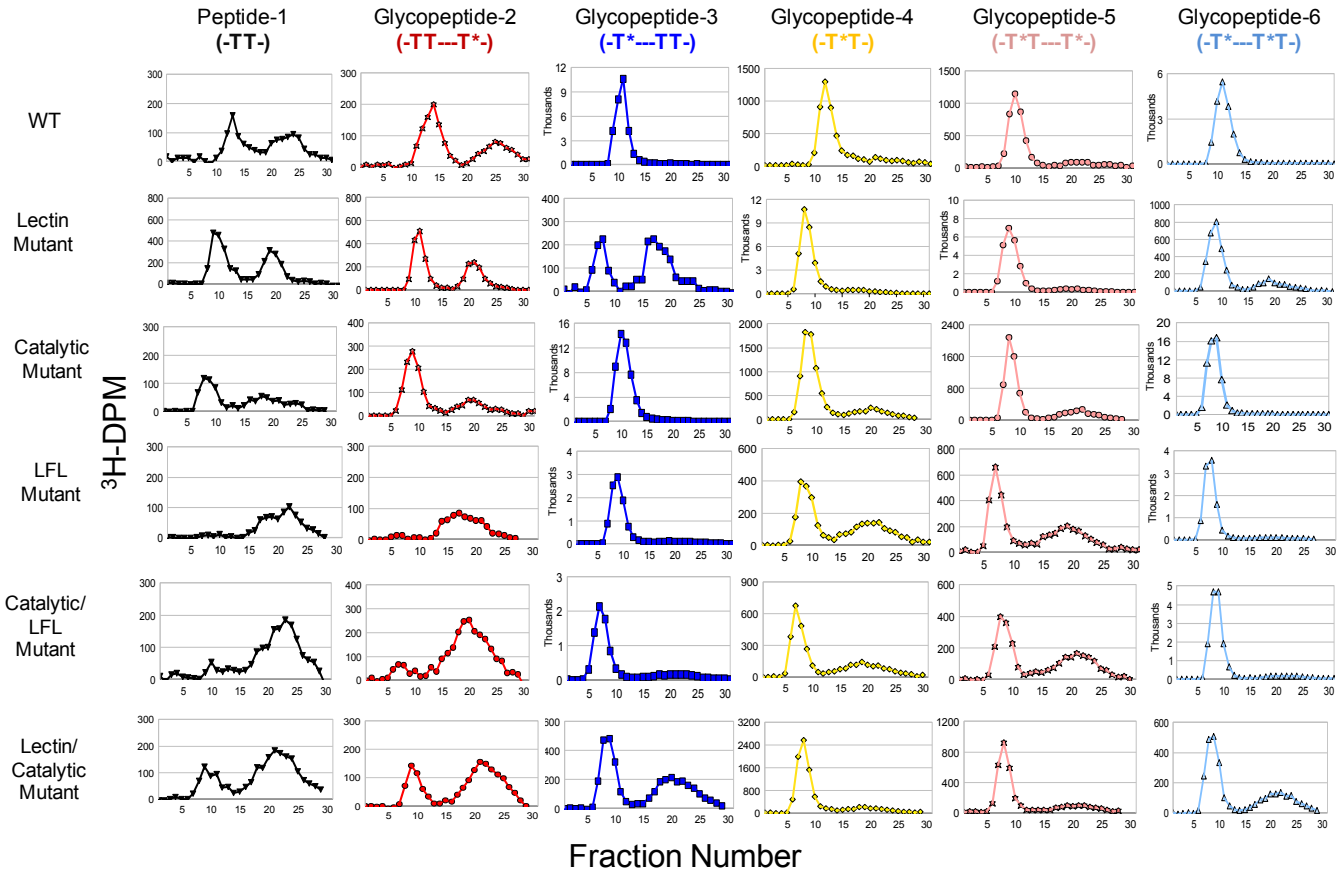


Figure S7. Sequence alignment of human GalNAc-T2 and GalNAc-T4. Secondary structure elements from the GalNAc-T2 and T4 structures are shown, with α-helices, 3_{10}-helices and β-strands in red and green, respectively. The regions encompassing the flexible loop and linker are indicated with double sided arrows. The yellow arrow indicates the residues encompassing the lectin flexible loop.



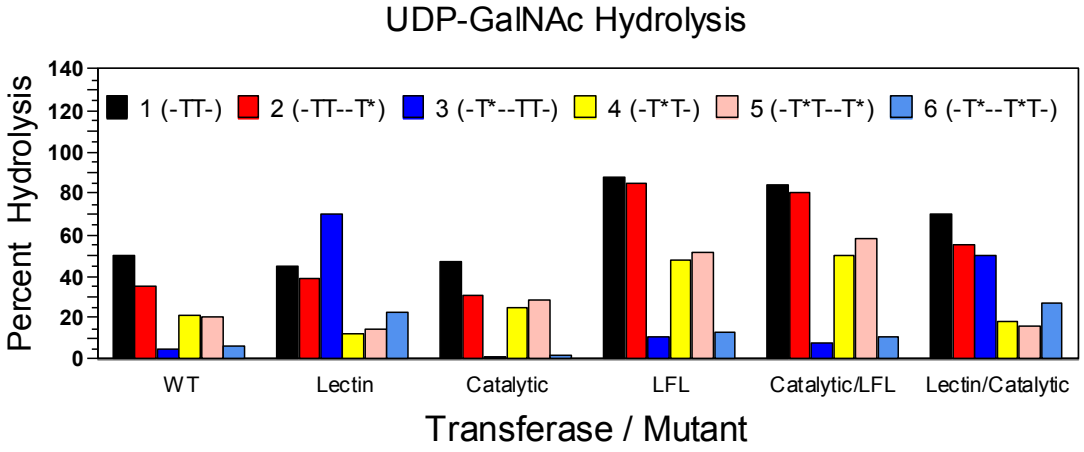


Figure S8. Determination of UDP-^{3}H-GalNAc hydrolysis. Top Panels: Representative Sephadex G10 chromatograms showing extent of UDP-GalNAc hydrolysis. Columns represent each (glyco)peptide substrate while rows represent GalNAc-T4 transferase construct. Transferase reaction products (after Dowex-1X8) were applied to Sephadex G10 chromatography and the eluted fractions counted for ^{3}H-activity. Note the first eluting peak represents transfer to (glyco)peptide while the second eluting peak represents transfer to water (i.e. free ^{3}H-GalNAc). **Bottom Panel: Summary of UDP-^{3}H-GalNAc hydrolysis.** Average hydrolysis for each substrate and transferase combination obtained from 2 to 5 individual Sephadex G10 determinations.

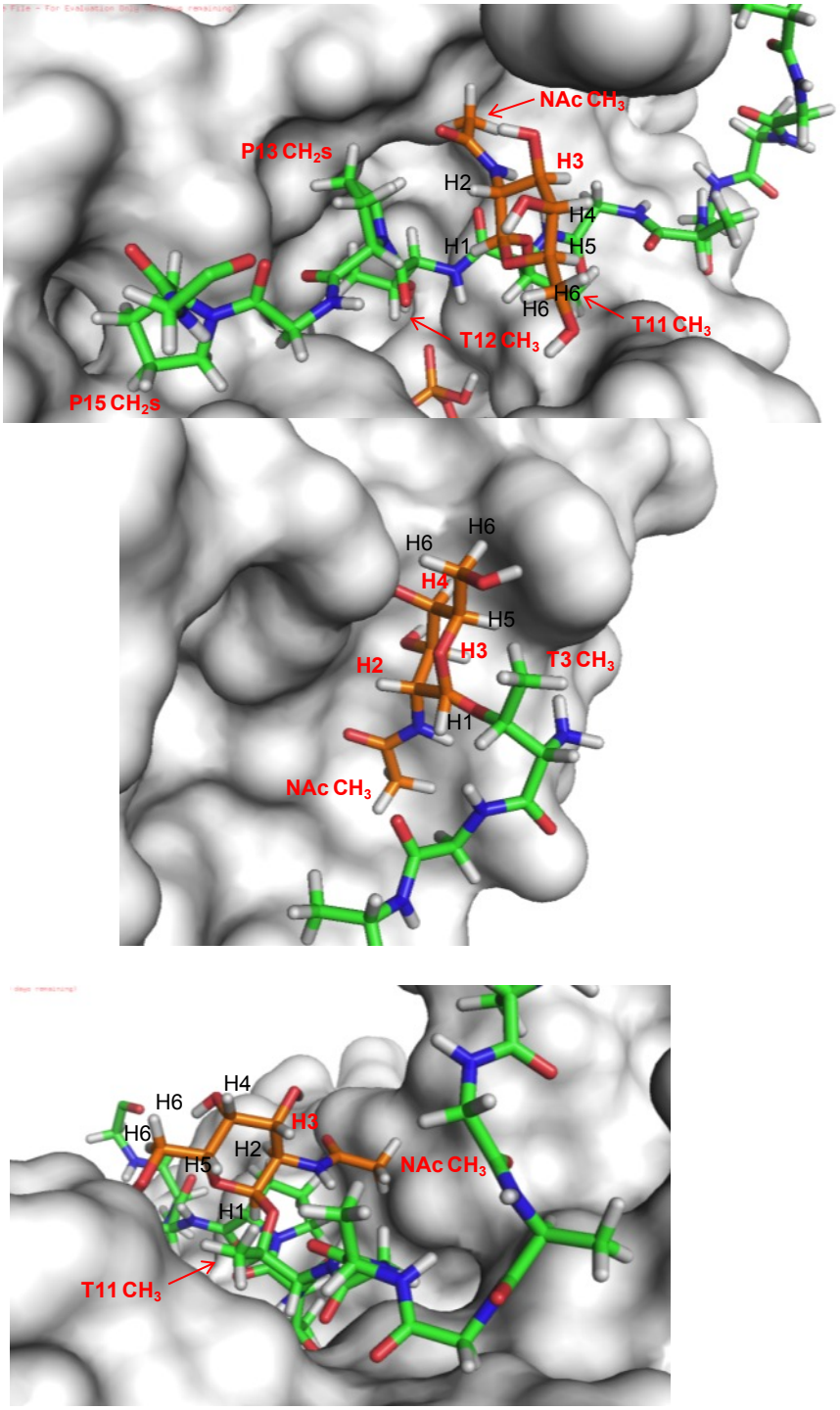


Figure S9. **STD-NMR enhancements for glycopeptides 3 and 4 mapped onto the structure of glycopetide 6 bound to GalNAc-T4**. Top and center figures show the catalytic domain bound -T*TPGP- sequence with the protons labeled in red exhibiting STD-NMR enhancements in glycoptide 4 (Figure S4). Bottom figure displays the lectin domain bound GalNAc STD-NMR enhancements, labeled in red, observed for glycopeptide $\mathbf{3}^{2}$. The STD-NMR enhancements and structure of bound glycopeptide **6** are consistent with each other demonstrating the different catalytic and lectin domain binding modes for the GalNAc residue.

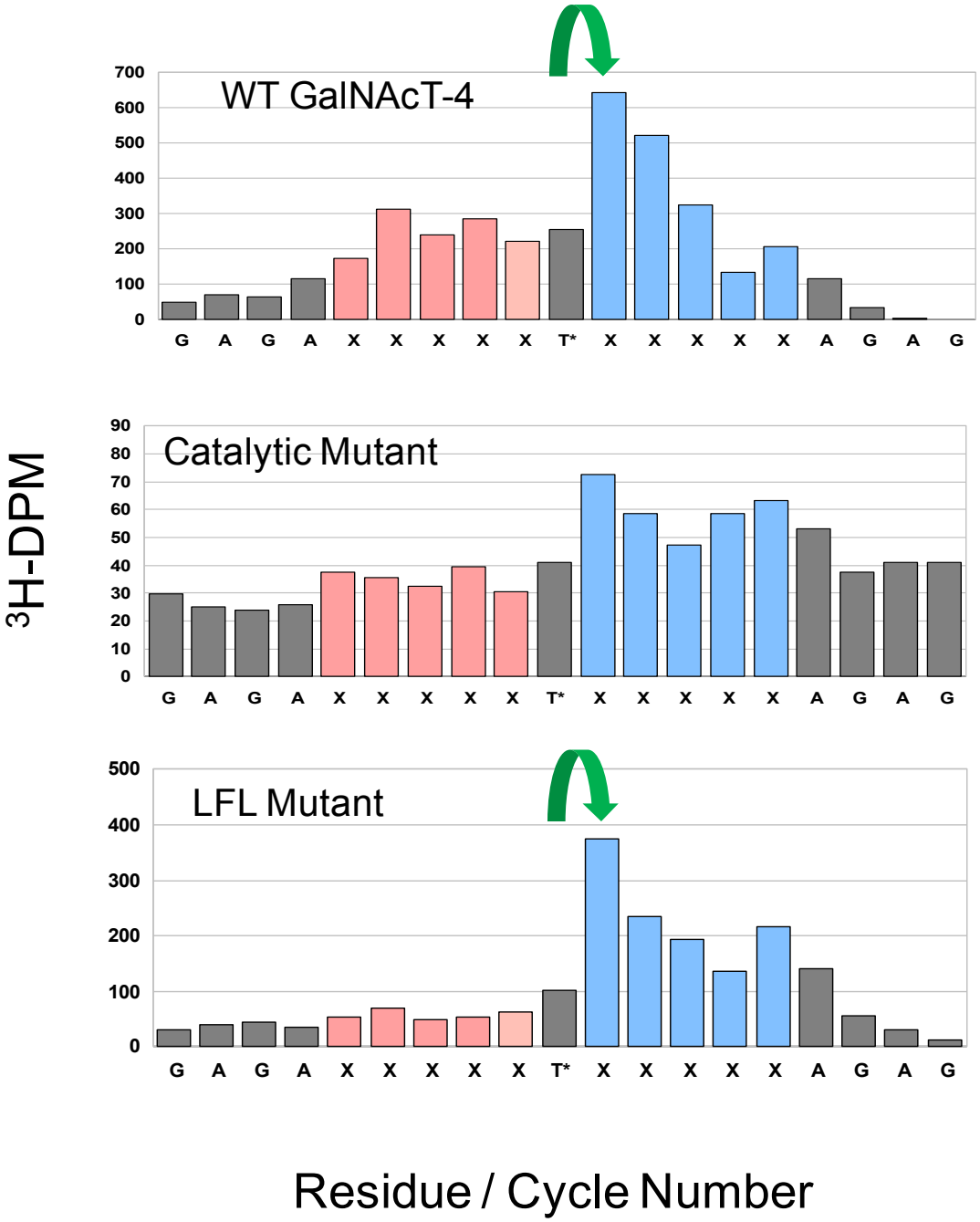


Figure S10. Neighboring glycosylation preferences of GalNAc-T4 and its catalytic and LFL mutants. Random glycopeptide GPIID GAGAXXXXXT*XXXXXAGAG (X= GARPNEYV & T) was glycosylated by the indicated transferases and Edman amino acid sequenced (after Sephadex G10 gel filtration) to determine the sites of incorporation of ^{3}H-GalNAc. Pink and light blue bars represent the possible sites of Thr residue glycosylation. Note that Thr acceptor residues are present in all of the random X sites. The green arrow represents the observed prior neighboring glycosylation activity of GalNAc-T4.

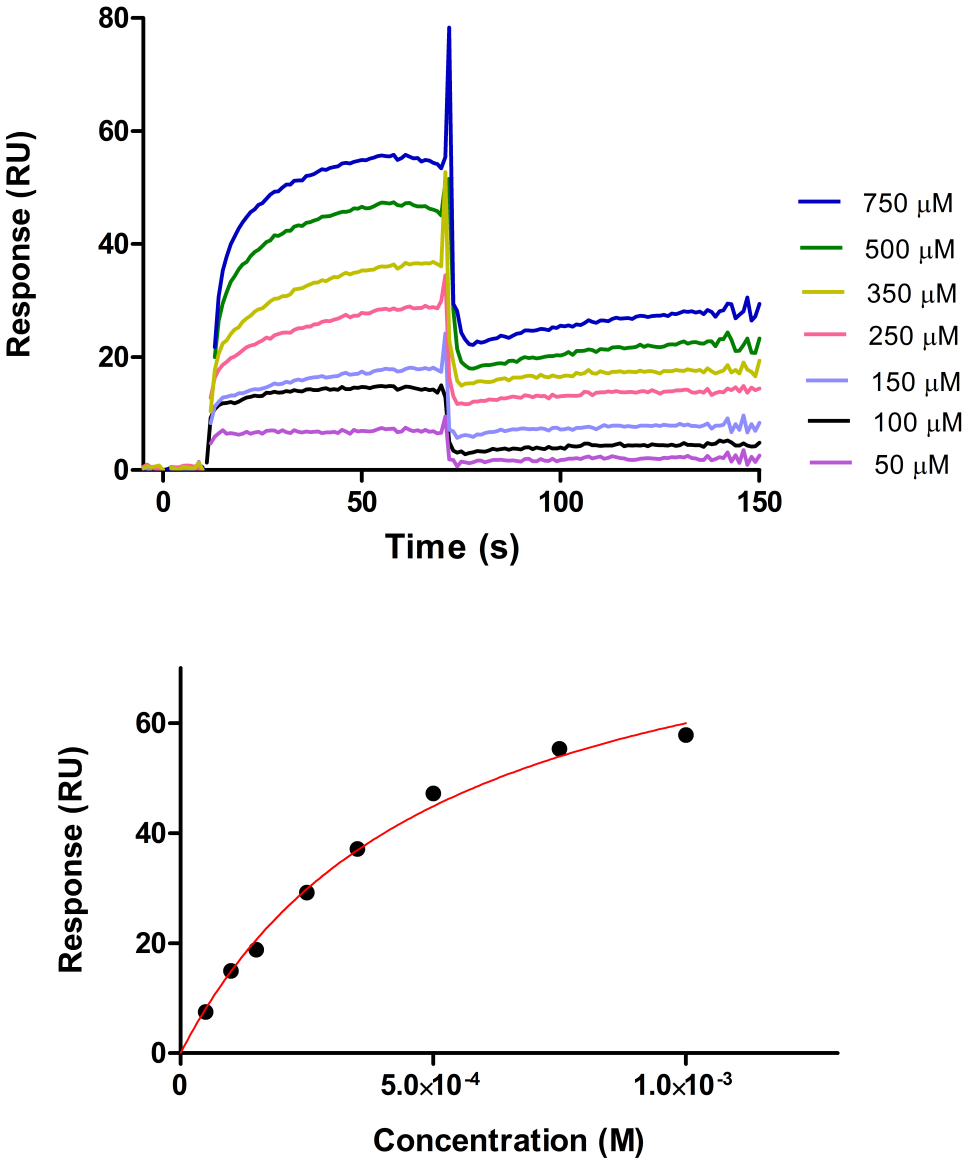


Figure S11. SPR sensogram of glycopeptide 6 binding to the GalNAc-T4 catalytic mutant. Sensogram (top) and fitting (bottom) of SPR data for the binding of glycopeptide **6** to the catalytic domain mutant of GalNAc-T4. Glycopeptide ligand concentrations are reported in the inset legend of the sensogram (top). End-point values after each injection were plotted against glycopeptide concentration which were fit to a Kd of 435 μM ± 87 μM (solid line, bottom plot).

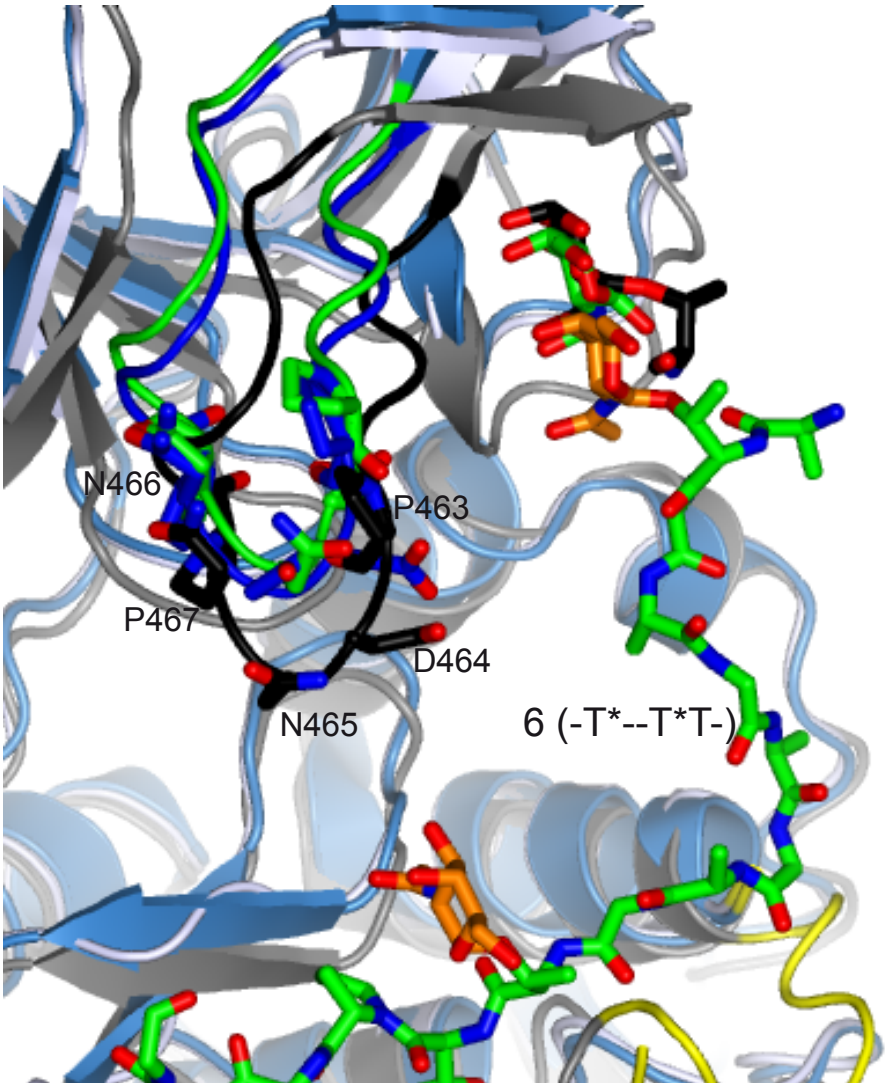


Figure S12. Conformational changes of the LFL. The crystal structure of GalNAc-T4 in complex with the peptide **3** was superimposed with the crystal structure of GalNAc-T4 in complex with the peptide **6.** Note that the crystals for both crystal structures have two molecules in the asymmetric unit. For this superposition analysis, we used the two monomers of GalNAc-T4 that were soaked with peptide **3** (LFL in green and blue) whereas we have only used the monomer of GalNAc-T4 in complex with UDP and peptide **6** (LFL in black). Note that for one of the monomers of the GalNAc-T4-peptide **3** complex asymmetric unit is there only electron density visualized for the GalNAc moiety, for the other monomer we visualized Thr3 and its covalently bound GalNAc residue of peptide **3**. The GalNAc moieties of peptide **3** were shown as green (free GalNAc) and black (sugar bound to Thr3) carbon atoms. The GalNAc moieties for peptid**e 6** were shown as orange carbon atoms. The amino acids of peptide **3** (only Thr3 is visualized) and **6** were shown as black and green carbon atoms, respectively. The residues that are listed in the figure correspond to the LFL of the structure of GalNAc-T4-UDP-peptide **6**. Note that only the residues eliminated and replaced by GGG to generate the LFL mutant are labelled.

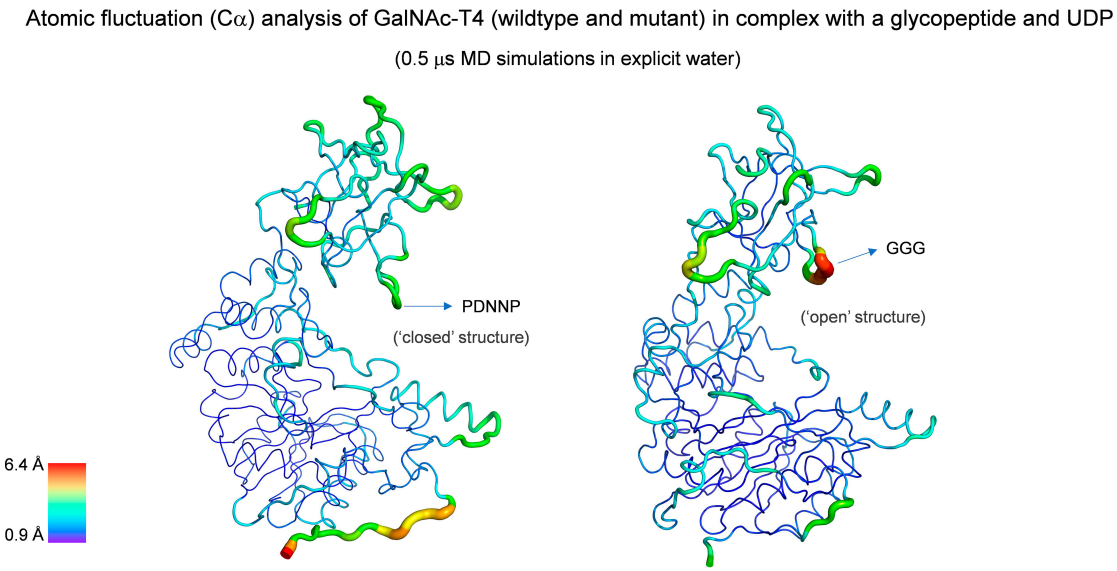


Figure S13. Atomic fluctuation (Cα) analysis of wild-type (left) and LFL mutant (right) obtained from 0.5 μs MD simulations.

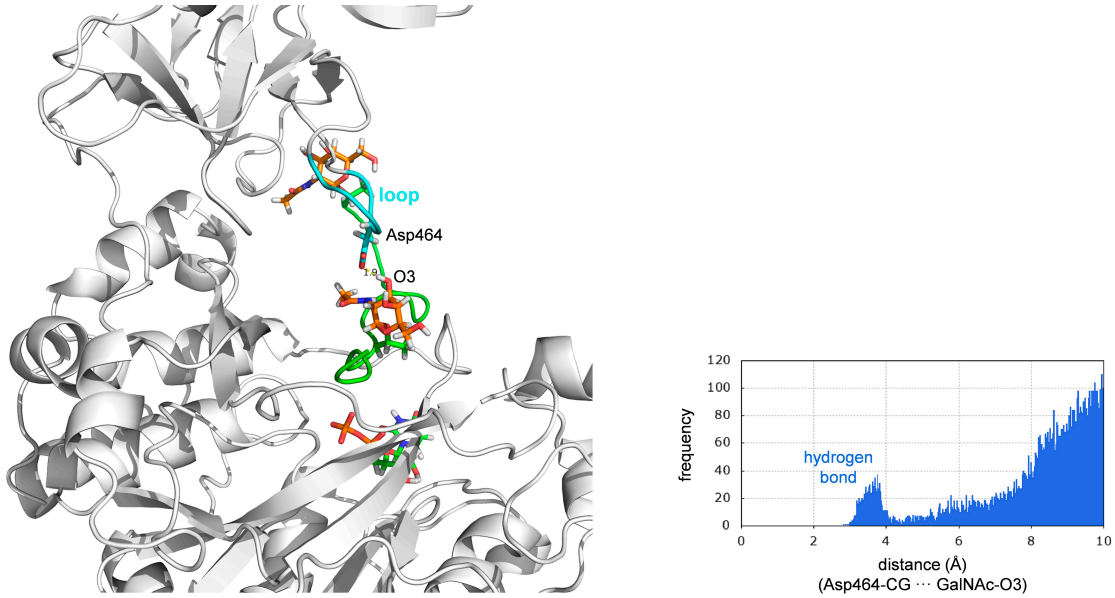


Figure S14. (left) An example of a transient hydrogen bond between the peptide **6** GalNAc moiety located in the catalytic domain and Asp464, (right) together with the distribution of the distance between the GalNAc O3 and the carboxylate group of the side chain of Asp464 obtained from 0.5 μs MD simulations.

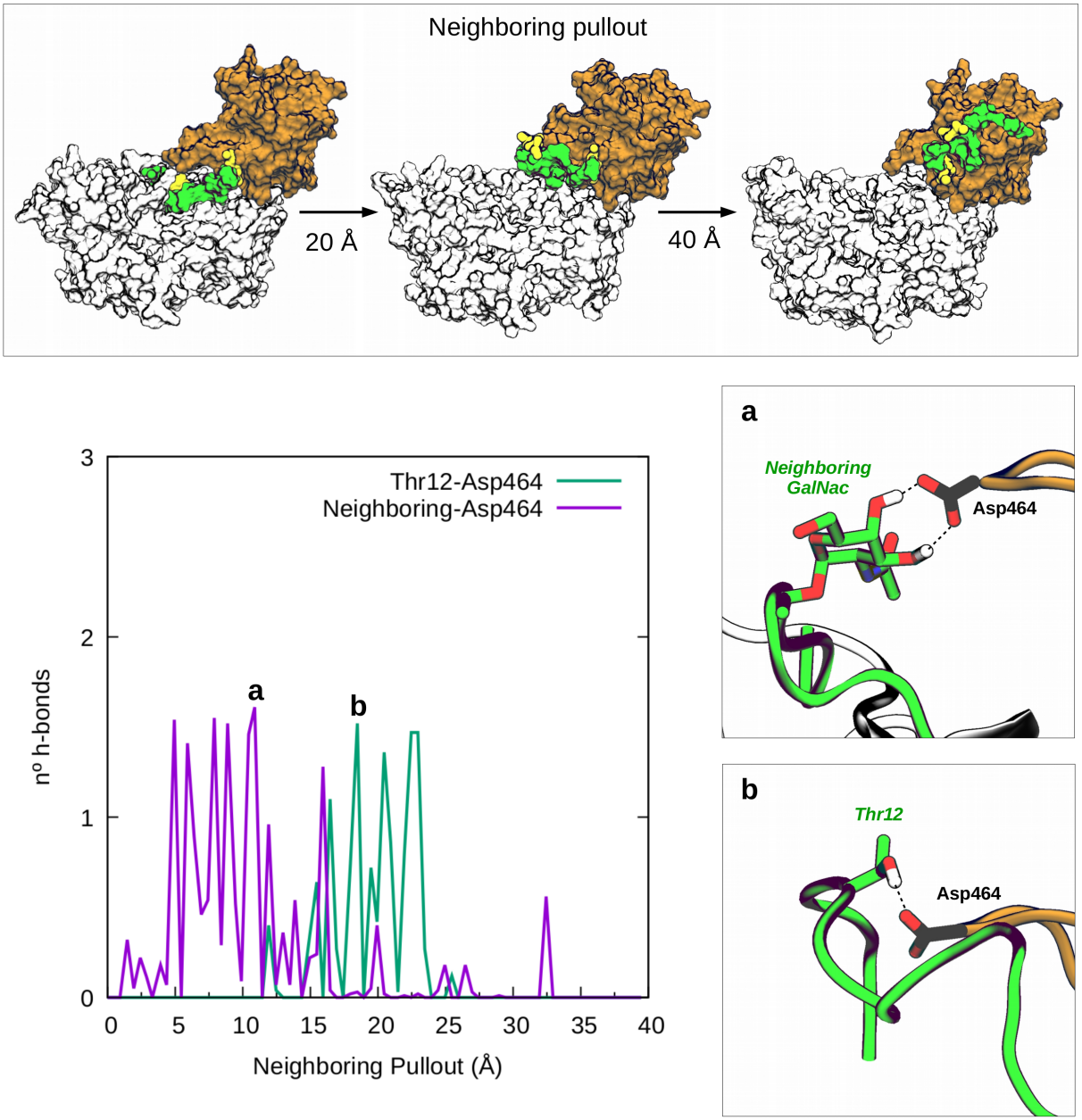


Figure S15. Pulling out of bound glycopeptide 6 from the catalytic domain of GalNAc-T4. (Top panel) Molecular motion of the peptide along the pathway. The catalytic domain is represented in white color, the lectin domain in brown, and the peptide and the GalNAc glycans in green and yellow, respectively. (Bottom panel) Mean number of H-bonds between the lectin flexible loop Asp464 and (a) the neighboring Thr11-GalNAc and (b) the Thr12 residue.

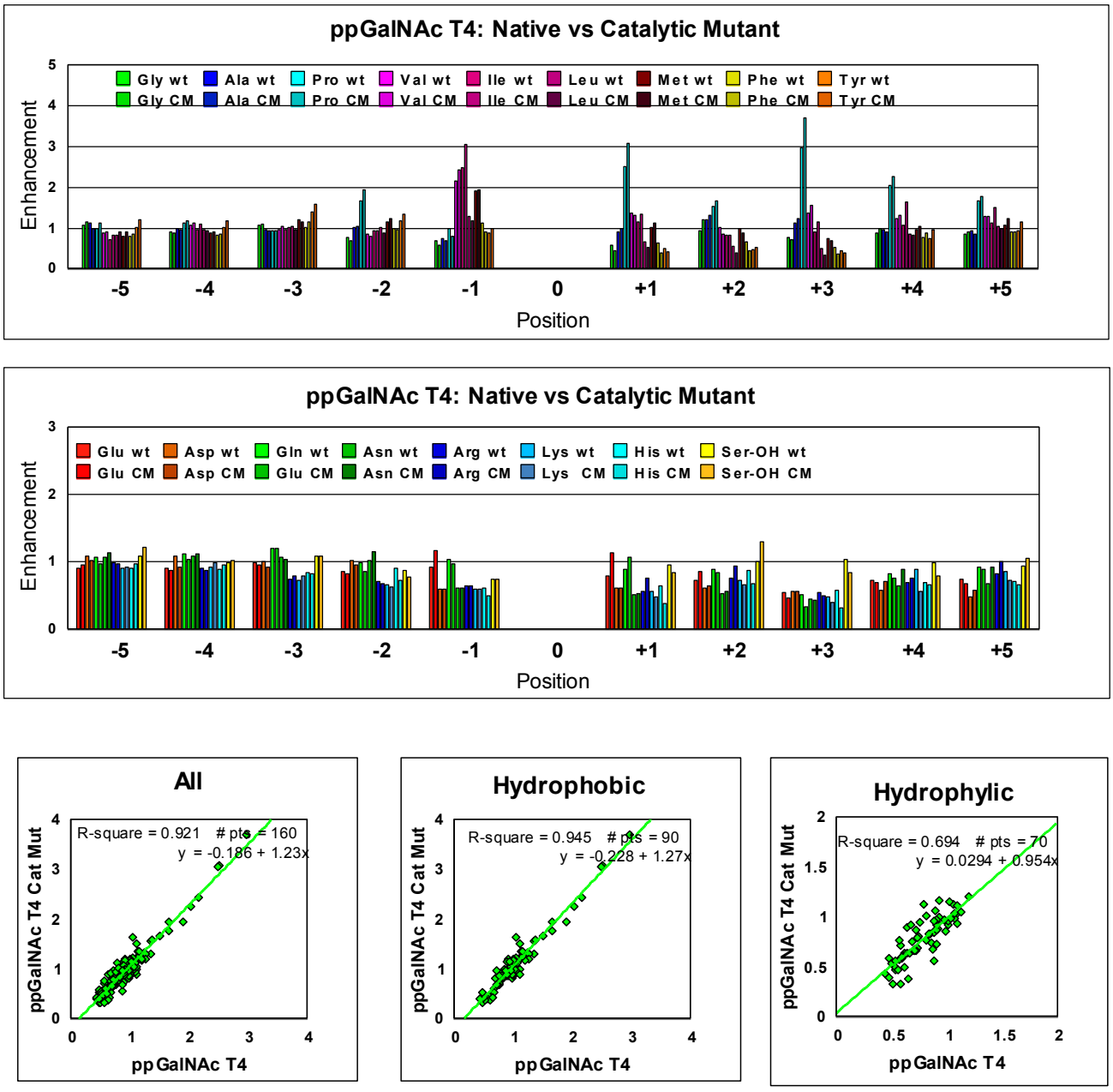


Figure S16. **Comparison of wt and catalytic mutant GalNAc-T4 enhancement values (EV) derived from random peptide substrates**. Comparison plots of hydrophobic residues (top panel) and hydrophylic residues (center panel) are shown. Correlation plots between wt and catalytic mutant transferases are shown at the bottom. Note there are no significant differences between the wt and catalytic mutant preferences. See Methods section for details.

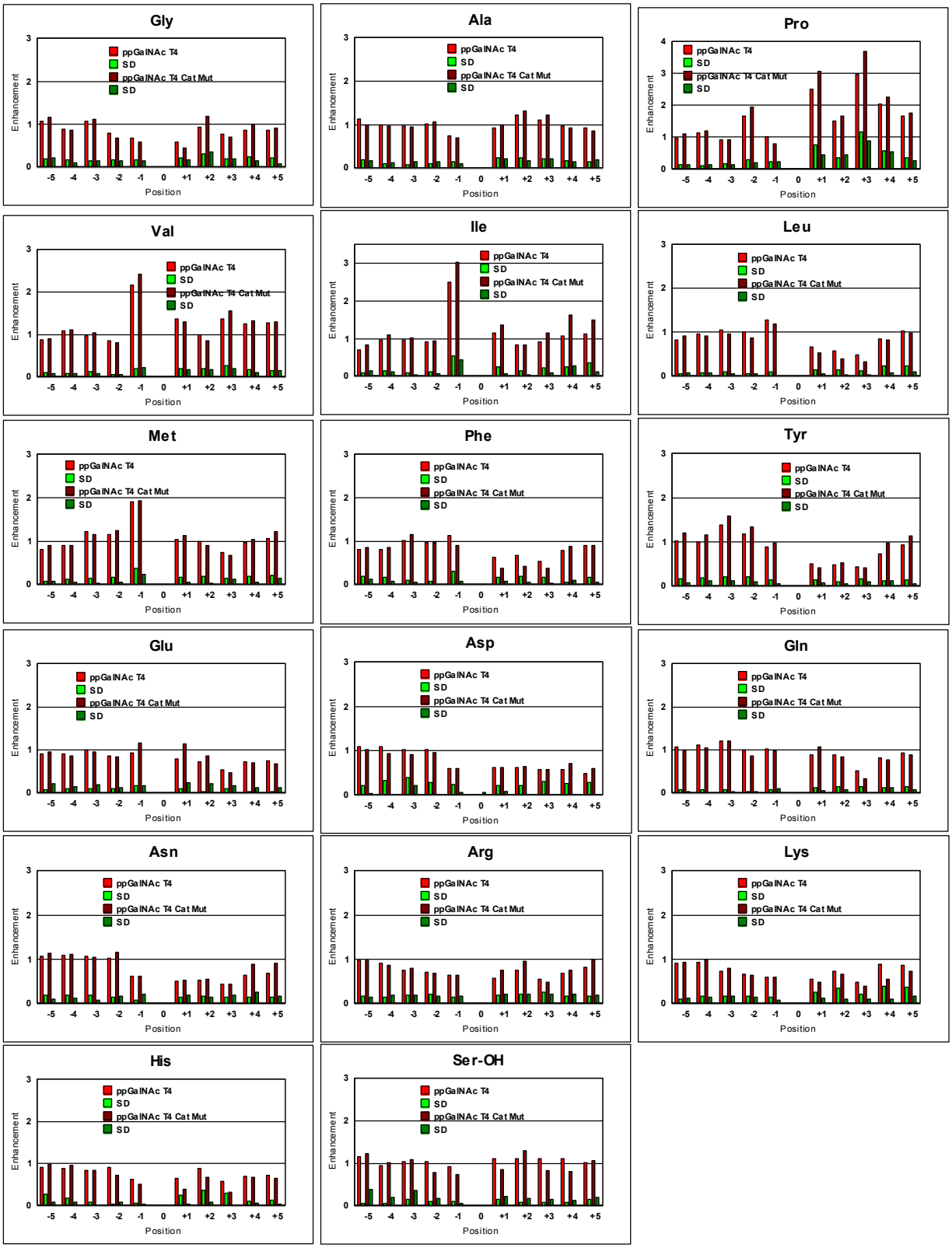


Figure S17. **Comparison of wt and catalytic mutant GalNAc-T4 enhancement values (EV) for each amino acid residue.** Red and brown bars represent wt and catalytic mutant averaged EV's while the the green bars represent their standard deviation (SD). the calculated standard deviation. Note there are no significant differences between the wt and catalytic mutant preferences. See **Methods** section for details.

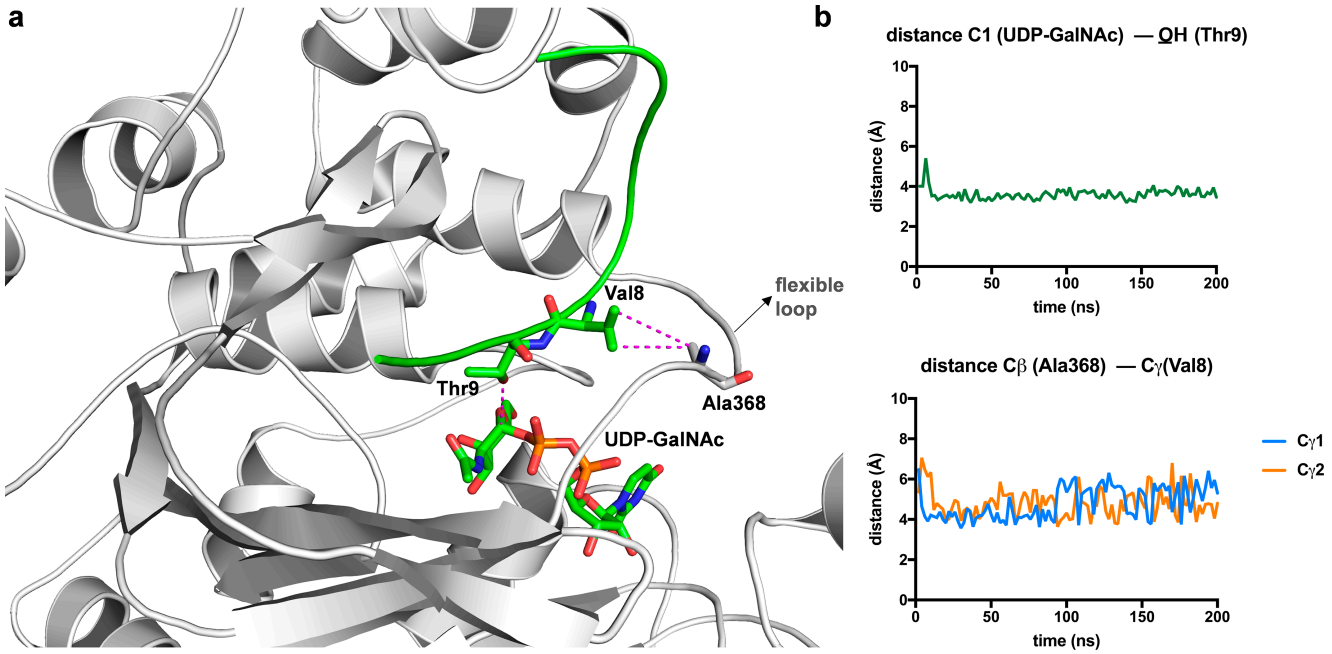


Figure S18. (**a**) A representative frame derived from 200 ns MD simulations illustrating the hydrophobic contacts between the side-chains of Val8 and Ala368. (**b**) Evolution of the distances between C1 of GalNAc and oxygen of the hydroxyl group of Thr9 (upper panel), together with Cβ of Ala368 and the methyl groups of Val8 (lower panel) along the 200 ns MD simulations.

REFERENCES

1. Rivas, M. L.; Lira-Navarrete, E.; Daniel, E. J. P.; Companon, I.; Coelho, H.; Diniz, A.; Jimenez-Barbero, J.; Peregrina, J. M.; Clausen, H.; Corzana, F.; Marcelo, F.; Jimenez-Oses, G.; Gerken, T. A.; Hurtado-Guerrero, R., The interdomain flexible linker of the polypeptide GalNAc transferases dictates their long-range glycosylation preferences. *Nature communications* **2017,** *8* (1), 1959.
2. de Las Rivas, M.; Coelho, H.; Diniz, A.; Lira-Navarrete, E.; Companon, I.; Jimenez-Barbero, J.; Schjoldager, K. T.; Bennett, E. P.; Vakhrushev, S. Y.; Clausen, H.; Corzana, F.; Marcelo, F.; Hurtado-Guerrero, R., Structural Analysis of a GalNAc-T2 Mutant Reveals an Induced-Fit Catalytic Mechanism for GalNAc-Ts. *Chemistry* **2018,** *24* (33), 8382-8392.