nature chemical biology

SUPPLEMENTARY INFORMATION

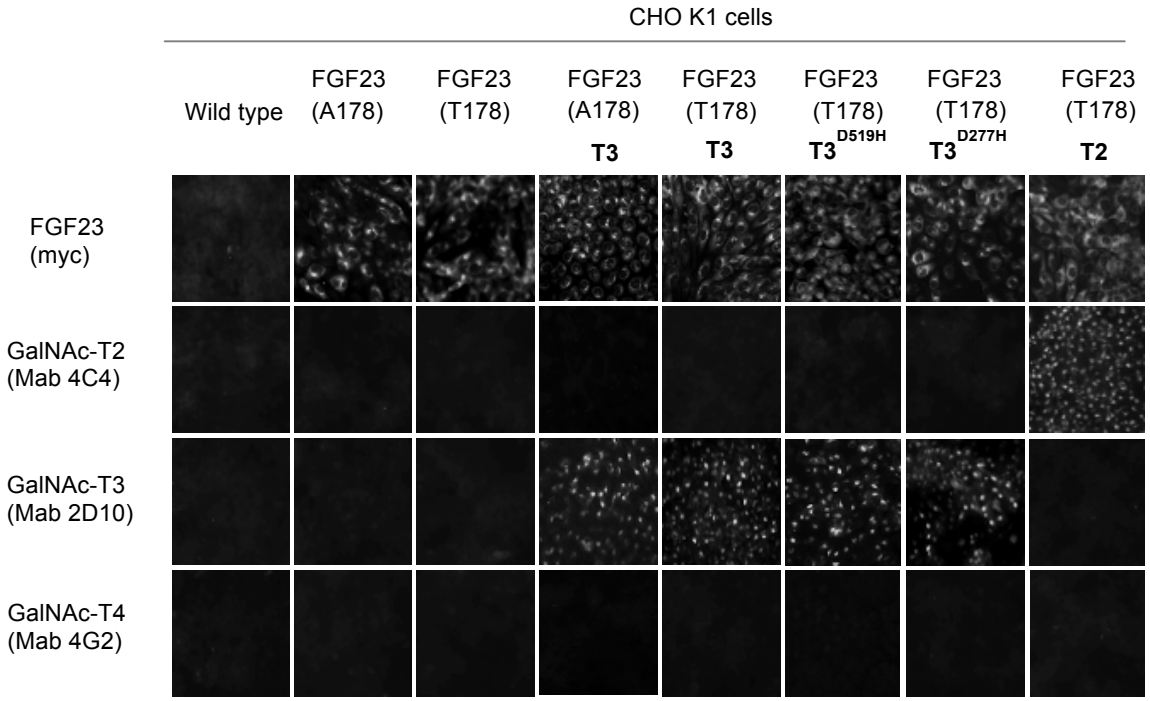
https://doi.org/10.1038/s41589-019-0444-x

In the format provided by the authors and unedited.

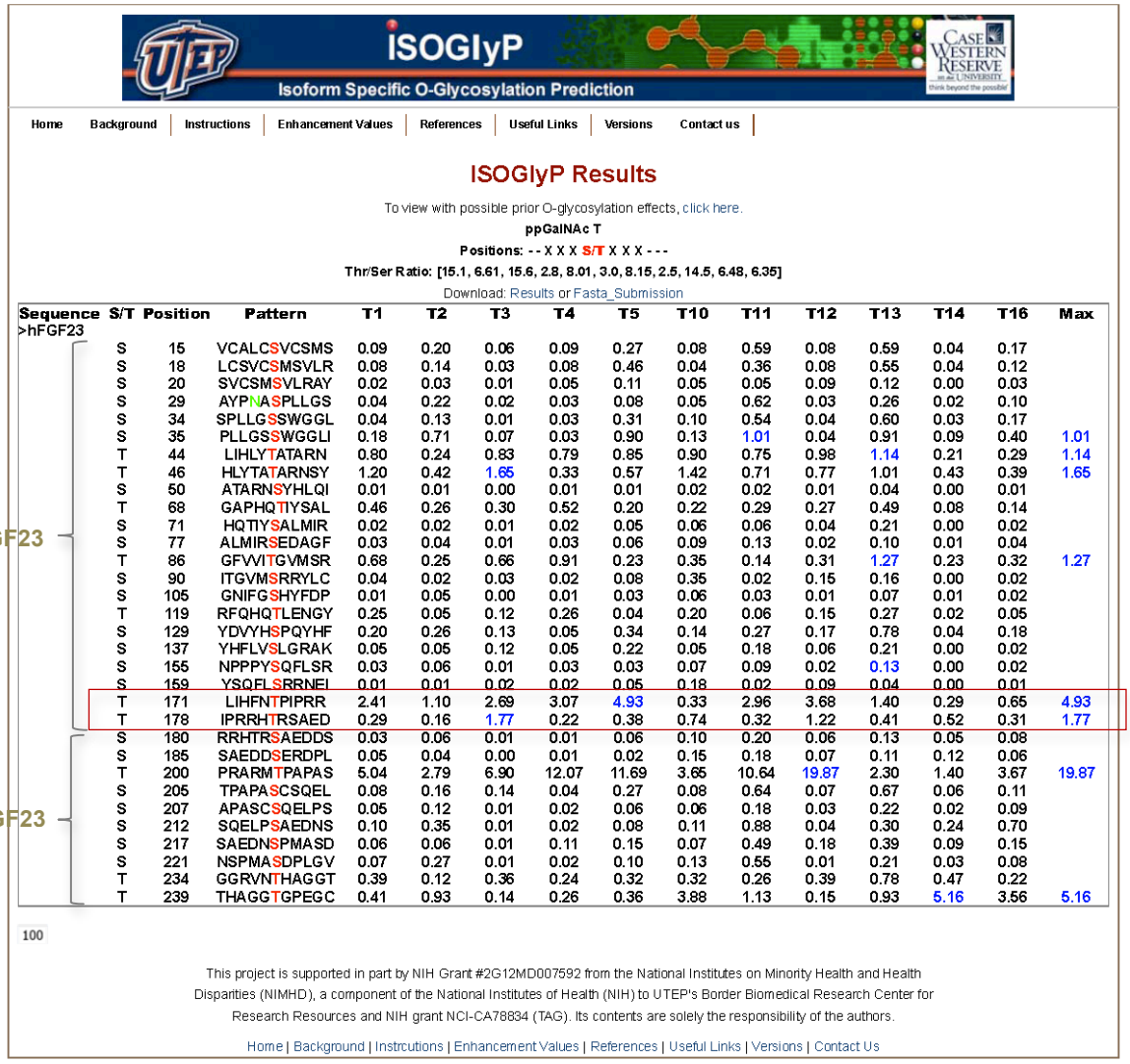
Molecular basis for fibroblast growth factor 23 *O*-glycosylation by GalNAc-T3

Matilde de las Rivas[1,14], Earnest James Paul Daniel[2,14], Yoshiki Narimatsu[3,14], Ismael Compañón[4,14], Kentaro Kato[3,5], Pablo Hermosilla[6], Aurélien Thureau[7], Laura Ceballos-Laita[1], Helena Coelho[8,9], Pau Bernadó[10], Filipa Marcelo[8], Lars Hansen[3], Ryota Maeda[11], Anabel Lostao[6,12,13], Francisco Corzana[4], Henrik Clausen[3], Thomas A. Gerken[2] and Ramon Hurtado-Guerrero[1,3,12]*

[1]BIFI, University of Zaragoza, Mariano Esquillor s/n, Campus Rio Ebro, Edificio I+D, Zaragoza, Spain. [2]Department of Biochemistry, Case Western Reserve University, Cleveland, OH, USA. [3]Copenhagen Center for Glycomics, Department of Cellular and Molecular Medicine, School of Dentistry, University of Copenhagen, Copenhagen, Denmark. [4]Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química, Logroño, Spain. [5]Department of Eco-epidemiology, Institute of Tropical Medicine Nagasaki University, Nagasaki, Japan. [6]Laboratorio de Microscopías Avanzadas, Instituto de Nanociencia de Aragón, Universidad de Zaragoza, Zaragoza, Spain. [7]Swing Beamline, Synchrotron SOLEIL, Gif sur Yvette, France. [8]UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade de Nova de Lisboa, Caparica, Portugal. [9]CIC bioGUNE, Bizkaia Technology Park, Derio, Spain. [10]Centre de Biochimie Structurale. INSERM, CNRS, Université de Montpellier, Montpellier, France. [11]Department of Hematology, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [12]Fundación ARAID, Zaragoza, Spain. [13]Instituto de Ciencia de Materiales de Aragón, Universidad de Zaragoza-CSIC, Zaragoza, Spain. [14]These authors contributed equally: Matilde de las Rivas, Earnest James Paul Daniel, Yoshiki Narimatsu, Ismael Compañón. *e-mail: rhurtado@bifi.es

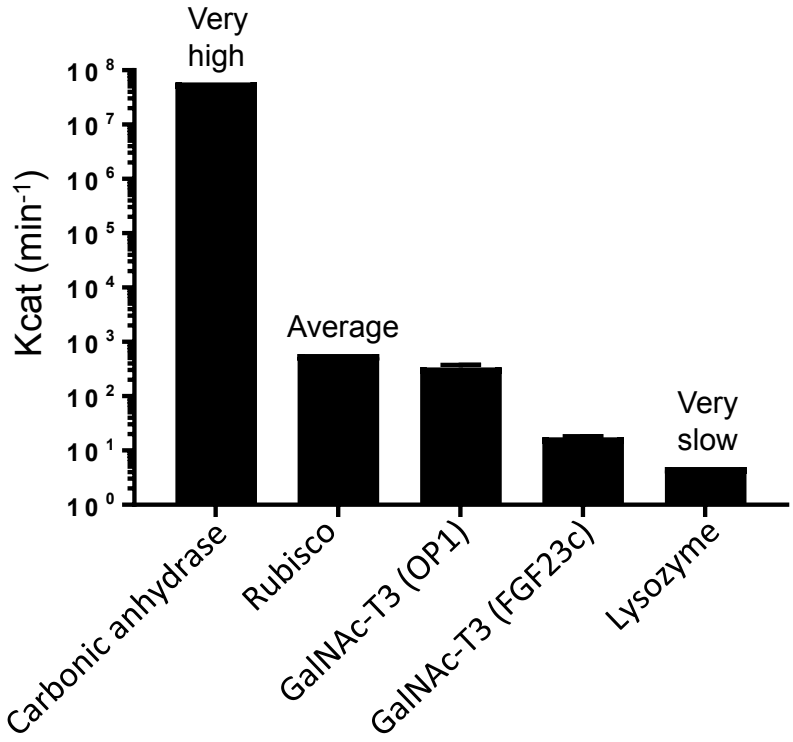


Supplementary Figure 1. Characterization of stably transfected CHO K1 cell lines. Cells immunofluorescence stained with anti-myc (FGF23), Mab 4C4 (GalNAc-T2), Mab 2D10 (GalNAc-T3), and Mab 4G2 (GalNAc-T4) (magnification x200). All experiments were done in sextuplicate.

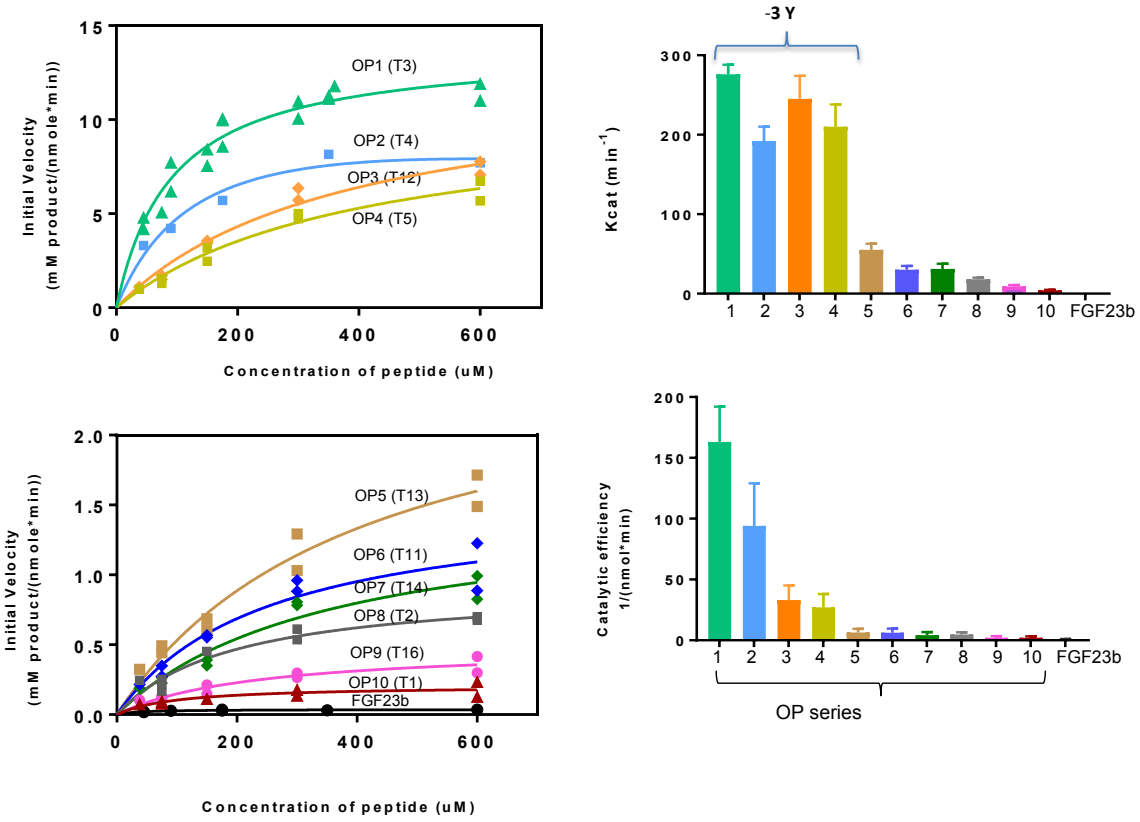


Sequence >hFGF23	S/T	Position	Pattern	T1	T2	T3	T4	T5	T10	T11	T12	T13	T14	T16	Max
	S	15	VCALCSVCSMS	0.09	0.20	0.06	0.09	0.27	0.08	0.59	0.08	0.59	0.04	0.17	
	S	18	LCSVCSMSVLR	0.08	0.14	0.03	0.08	0.46	0.04	0.36	0.08	0.55	0.04	0.12	
	S	20	SVCSMSVLRAY	0.02	0.03	0.01	0.05	0.11	0.05	0.05	0.09	0.12	0.00	0.03	
	S	29	AYPNASPLLGS	0.04	0.22	0.02	0.03	0.08	0.05	0.62	0.03	0.26	0.02	0.10	
	S	34	SPLLGSSWGGL	0.04	0.13	0.01	0.03	0.31	0.10	0.54	0.04	0.60	0.03	0.17	
	S	35	PLLGSSWGGLI	0.18	0.71	0.07	0.03	0.90	0.13	1.01	0.04	0.91	0.09	0.40	1.01
	T	44	LIHLYTATARN	0.80	0.24	0.83	0.79	0.85	0.90	0.75	0.98	1.14	0.21	0.29	1.14
	T	46	HLYTATARNSY	1.20	0.42	1.65	0.33	0.57	1.42	0.71	0.77	1.01	0.43	0.39	1.65
	S	50	ATARNSYHLQI	0.01	0.01	0.00	0.01	0.01	0.02	0.02	0.01	0.04	0.00	0.01	
	T	68	GAPHQTIYSAL	0.46	0.26	0.30	0.52	0.20	0.22	0.29	0.27	0.49	0.08	0.14	
	S	71	HQTIYSALMIR	0.02	0.02	0.01	0.02	0.05	0.06	0.06	0.04	0.21	0.00	0.02	
	S	77	ALMIRSEDAGF	0.03	0.04	0.01	0.03	0.06	0.09	0.13	0.02	0.10	0.01	0.04	
	T	86	GFVVITGVMSR	0.68	0.25	0.66	0.91	0.23	0.35	0.14	0.31	1.27	0.23	0.32	1.27
	S	90	ITGVMSRRYLC	0.04	0.02	0.03	0.02	0.08	0.35	0.02	0.15	0.16	0.00	0.02	
	S	105	GNIFGSHYFDP	0.01	0.05	0.00	0.01	0.03	0.06	0.03	0.01	0.07	0.01	0.02	
	T	119	RFQHQTLENGY	0.25	0.05	0.12	0.26	0.04	0.20	0.06	0.15	0.27	0.02	0.05	
	S	129	YDVYHSPQYHF	0.20	0.26	0.13	0.05	0.34	0.14	0.27	0.17	0.78	0.04	0.18	
	S	137	YHFLVSLGRAK	0.05	0.05	0.12	0.05	0.22	0.05	0.18	0.06	0.21	0.00	0.02	
	S	155	NPPPYSQFLSR	0.03	0.06	0.01	0.03	0.03	0.07	0.09	0.02	0.13	0.00	0.02	
	S	159	YSQFLSRRNEI	0.01	0.01	0.02	0.02	0.05	0.18	0.02	0.09	0.04	0.00	0.01	
	T	171	LIHFNTPIPRR	2.41	1.10	2.69	3.07	4.93	0.33	2.96	3.68	1.40	0.29	0.65	4.93
	T	178	IPRRHTRSAED	0.29	0.16	1.77	0.22	0.38	0.74	0.32	1.22	0.41	0.52	0.31	1.77
	S	180	RRHTRSAEDDS	0.03	0.06	0.01	0.01	0.06	0.10	0.20	0.06	0.13	0.05	0.08	
	S	185	SAEDDSERDPL	0.05	0.04	0.00	0.01	0.02	0.15	0.18	0.07	0.11	0.12	0.06	
	T	200	PRARMTPAPAS	5.04	2.79	6.90	12.07	11.69	3.65	10.64	19.87	2.30	1.40	3.67	19.87
	S	205	TPAPASCSQEL	0.08	0.16	0.14	0.04	0.27	0.08	0.64	0.07	0.67	0.06	0.11	
	S	207	APASCSQELPS	0.05	0.12	0.01	0.02	0.06	0.06	0.18	0.03	0.22	0.02	0.09	
	S	212	SQELPSAEDNS	0.10	0.35	0.01	0.02	0.08	0.11	0.88	0.04	0.30	0.24	0.70	
	S	217	SAEDNSPMASD	0.06	0.06	0.01	0.11	0.15	0.07	0.49	0.18	0.39	0.09	0.15	
	S	221	NSPMASDPLGV	0.07	0.27	0.01	0.02	0.10	0.13	0.55	0.01	0.21	0.03	0.08	
	T	234	GGRVNTHAGGT	0.39	0.12	0.36	0.24	0.32	0.32	0.26	0.39	0.78	0.47	0.22	
	T	239	THAGGTGPEGC	0.41	0.93	0.14	0.26	0.36	3.88	1.13	0.15	0.93	5.16	3.56	5.16

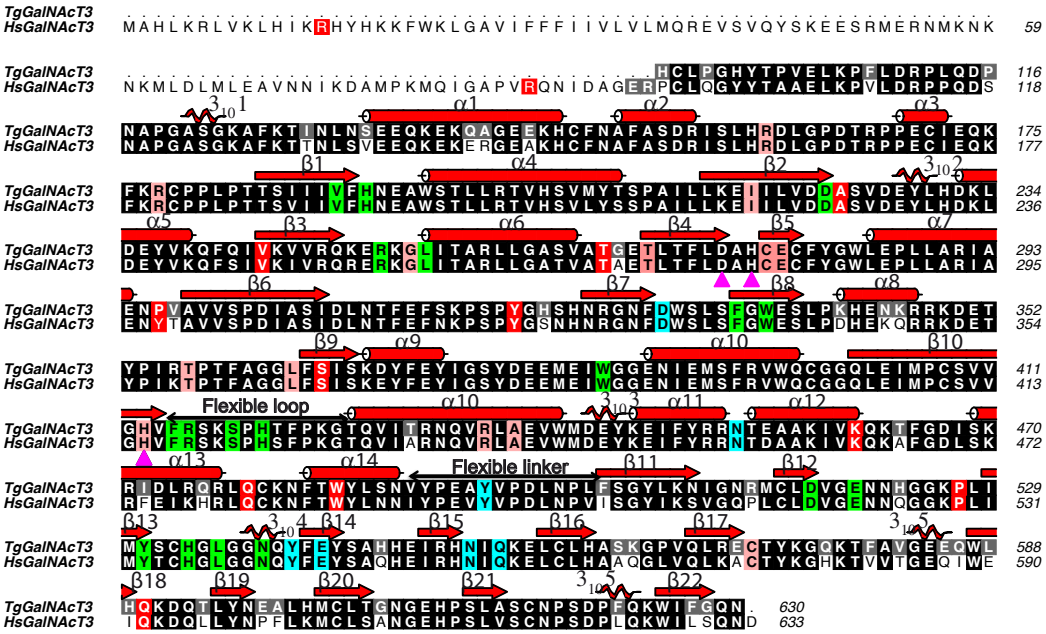
Supplementary Figure 2. Prediction of potential FGF23 *O*-glycosylation sites by the isoform specific *O*-glycosylation prediction program, ISOGlyP. Enhancement value products (EVP) are given for the 11 GalNAc-T isoforms available in ISOGlyP. EVP values >1 suggest potential glycosylation sites, values <1 suggest the site would not be glycosylated by the listed GalNAc-T. ISOGlyP predicts Thr171 could be glycosylated by GalNAc-T1, T2, T3, T4, T5, T11, T12 & T13 while Thr178 could be glycosylated by only GalNAc-T3 and -T12 (highlighted in red box). Overall, there are 6 sites predicted to be glycosylated in the N-terminal fragment (n-FGF23), 5 of which are very weak and 2 strong sites (T200 and T239) in the C-terminal fragment (c-FGF23). ISOGlyP beta 2.1, http://isoglyp.utep.edu/index.php was used for these calculations.



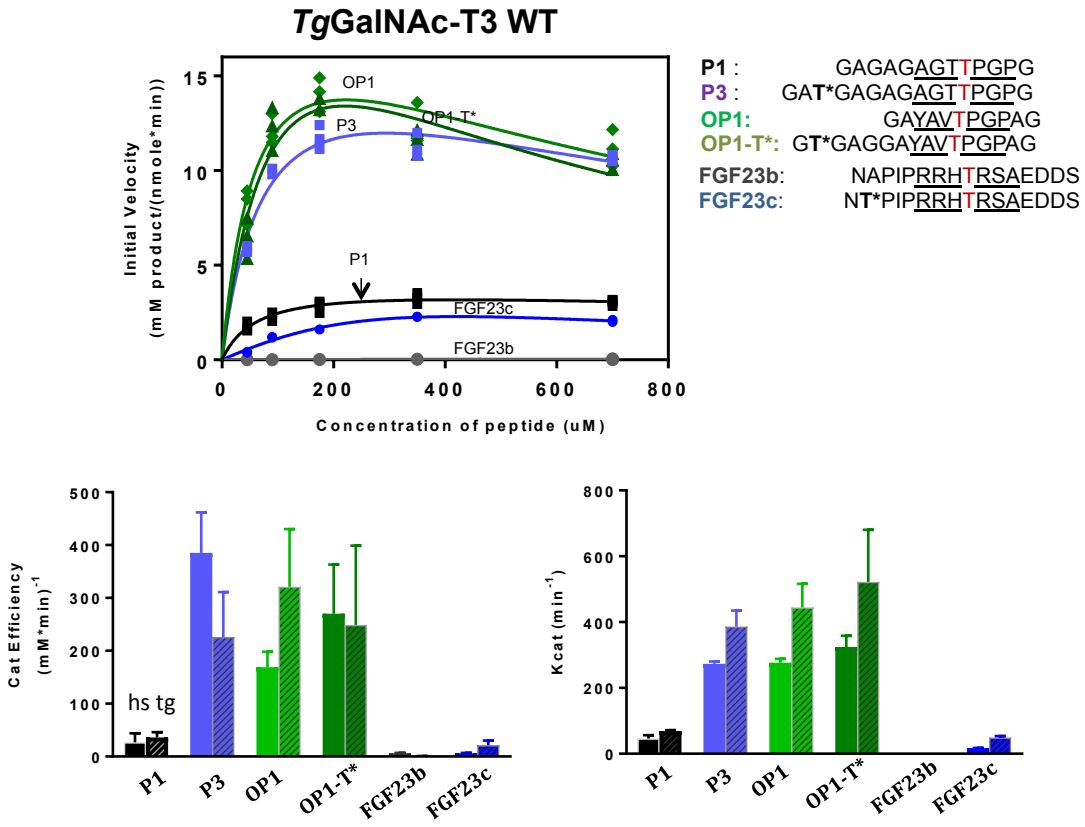
Supplementary Figure 3. Enzymes can display a wide range of Kcat values. Plot shows the range of possible enzyme Kcat values with carbonic anhydrase as one of the fastest enzymes in Nature and lysozyme as one of the slowest. GalNAc-T3 is a fairly slow enzyme against the **FGF23c** (T*171,T178) substrate but an average enzyme against its optimal substrate **OP1**. Note Kcat values are plotted on a logarithmic scale. Kcat values for GalNAc-T3 against **OP1** and **FGF23c** were obtained from the GraphPad Prism Michaelis-Menten non-linear fit of the data shown in Figure 3 and Supplementary Table 2. Error bars represent the standard deviation calculated by the GraphPad Prism fit of the data. A total of 21 and 24 independent initial velocity determinations were used for the calculations for **OP1** and **FGF23c** respectively.



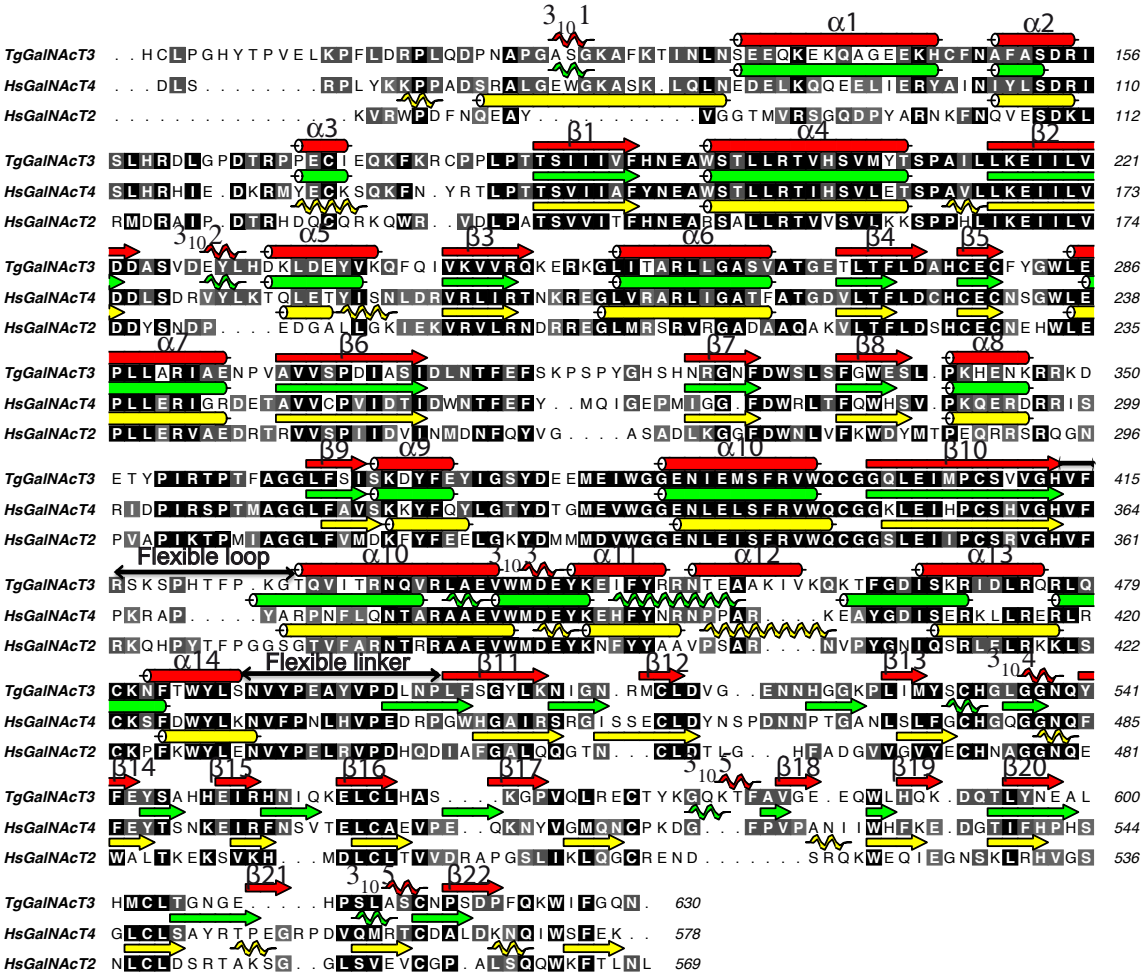
Supplementary Figure 4. Kinetics studies of GalNAc-T3 against a series of transferase specific optimal peptide sequences. Optimal peptide sequences were designed based on the ISOGlyP predictions (ISOGlyP beta 2.1, http://isoglyp.utep.edu/index.php) for the indicated GalNAc-T isoforms (given in brackets). Substrates, designated **OP1** to **OP10**, consist of the GAGA-XXXTPZP-AGAG sequence where the XXX residues varied based on the isoform and Z is Gly except in **OP2** (for GalNAc-T12) where it is Arg. See Supplementary Table 1 for full sequences. Top left panel shows the four most active substrates which contain a Tyr residue at the -3 position of the acceptor (i.e YAV-, -YPI-, -YYI- & -YYP-, for **OP1**, **OP2**, **OP3** & **OP4** respectively). Bottom left panel displays plots for the remaining less active substrates plotted on an expanded scale. Plotted lines represent the Michaelis–Menten non-linear fit (using GraphPad Prism 7.03) of the initial velocity data using the kinetic values given in Supplementary Table 2**.** Initial velocities (individual points) were obtained in duplicate or higher for each peptide concentration giving a total of 9, independent determinations for OP1 to OP10 and 15 determinations for FGF23b. Graphs to the right summarize Kcat and catalytic efficiency values obtained from the plots on the left. Error bars represent the standard deviation calculated by the GraphPad Prism fit of each data set as described in the Online Methods. Note that data for the **OP1** plot also included partial data from the shorter **OP1** peptide (GA-YAVTPGP-AG) that gave indistinguishable results from the longer **OP1** peptide.



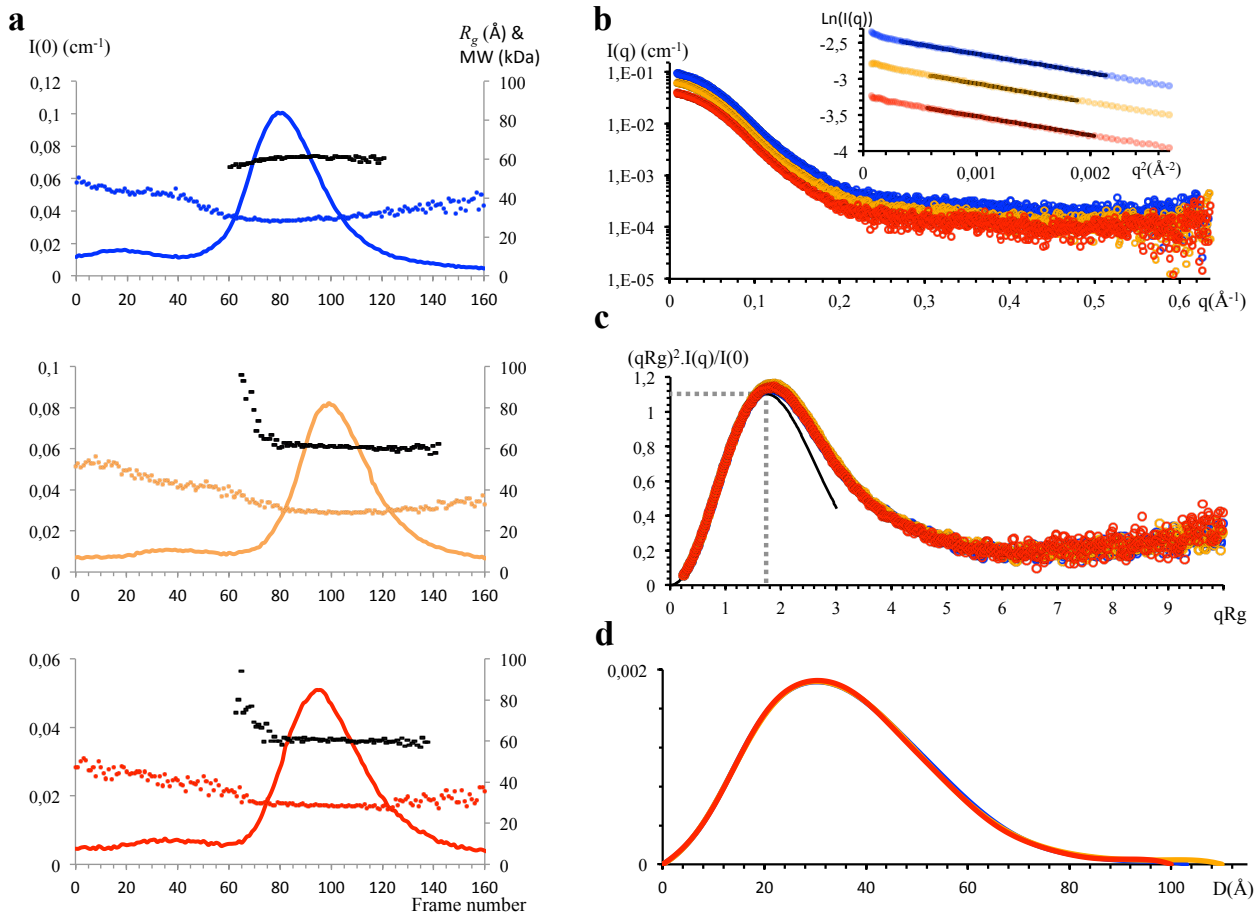
Supplementary Figure 5. Sequence alignment of *Tg*GalNAc-T3 and human GalNAc-T3 (*Hs*GalNAc-T3). Residues are color-coded by degree of sequence conservation where black and grey colours denote identity/high similarity and dissimilarity, respectively. Note that for *Tg*GalNAc-T3, only the residues present in the crystal structure are shown in the alignment while for the human enzyme, the entire sequence is shown. Shown above the sequence, in red, are the secondary structure elements (α-helices, 3_{10}-helices and β-strands) based on the *Tg*GalNAc-T3 structure. The double-sided arrows indicate regions encompassing the catalytic domain flexible loop and the linker connecting the catalytic and lectin. The DxH motif and the His coordinating the Mn^{+2} ion are indicated with magenta diamonds. Residues interacting with peptide ligands are highlighted in green. Residues of the catalytic domain (spanning residues 91-491) contacting residues of the lectin domain (505-631) are highlighted in blue. Missense and frameshift/point-nonsense mutations found in disease are highlighted in salmon and red, respectively (see also Fig. 6 and Supplementary Fig 16).



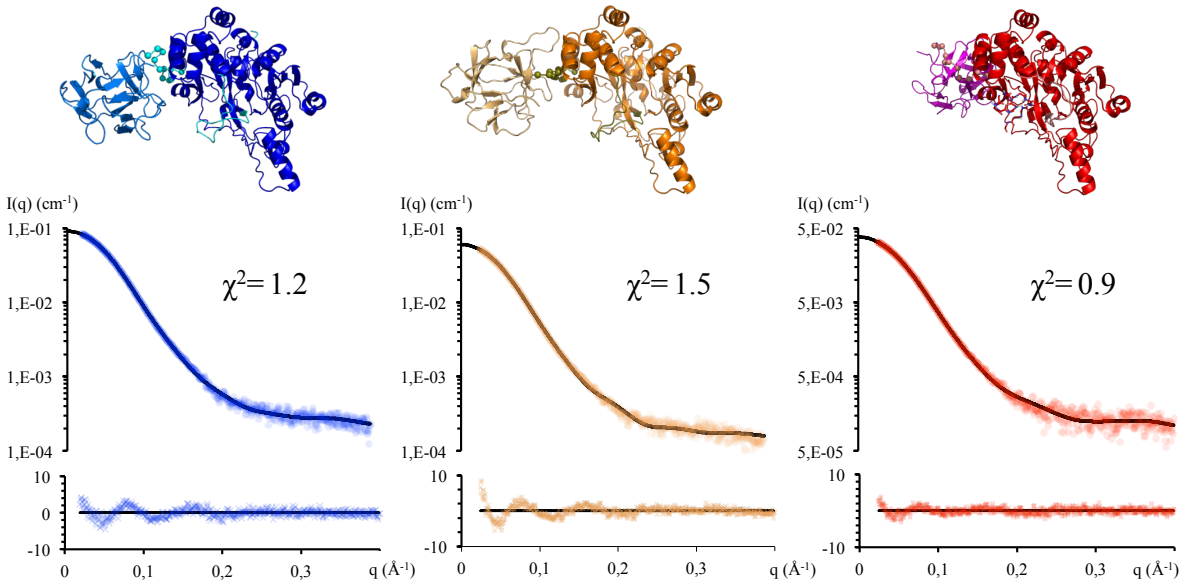
Supplementary Figure 6. *Tg*GalNAc-T3 shows nearly identical kinetics as *Hs*GalNAc-T3 against a series of (glyco)peptide substrates including FGF23c. (Upper panel) Glycosylation kinetics of *Tg*GalNAc-T3 against a series of (glyco)peptides also characterized by the *Hs*GalNAc-T3 **(**Fig. 1d and 3a). Plotted lines represent the Michaelis–Menten (or Michaelis–Menten with substrate inhibition) non-linear fits (using GraphPad Prism 7.03) of the initial velocity data using the kinetic values given in Supplementary Table 2**.** Initial velocities (individual points) were obtained in duplicate or higher for each peptide concentration giving a total of 15, 15, 15, 14, 15 and 10, independent determinations for P1, P3, OP1, OP1-T*, FGF23b and FGF23c against *Tg*GalNAc-T3 respectively. (Lower panels) Plots comparing the catalytic efficiency (V_{max}/K_m) and Kcat values of the human GalNAc-T3 (solid bars) and *Tg*GalNAc-T3 (striped bars) derived from the plots in the upper panels and from Fig. 1d and 3a. Error bars represent the standard deviation calculated by the GraphPad Prism fit of each data set as described in the Online Methods.



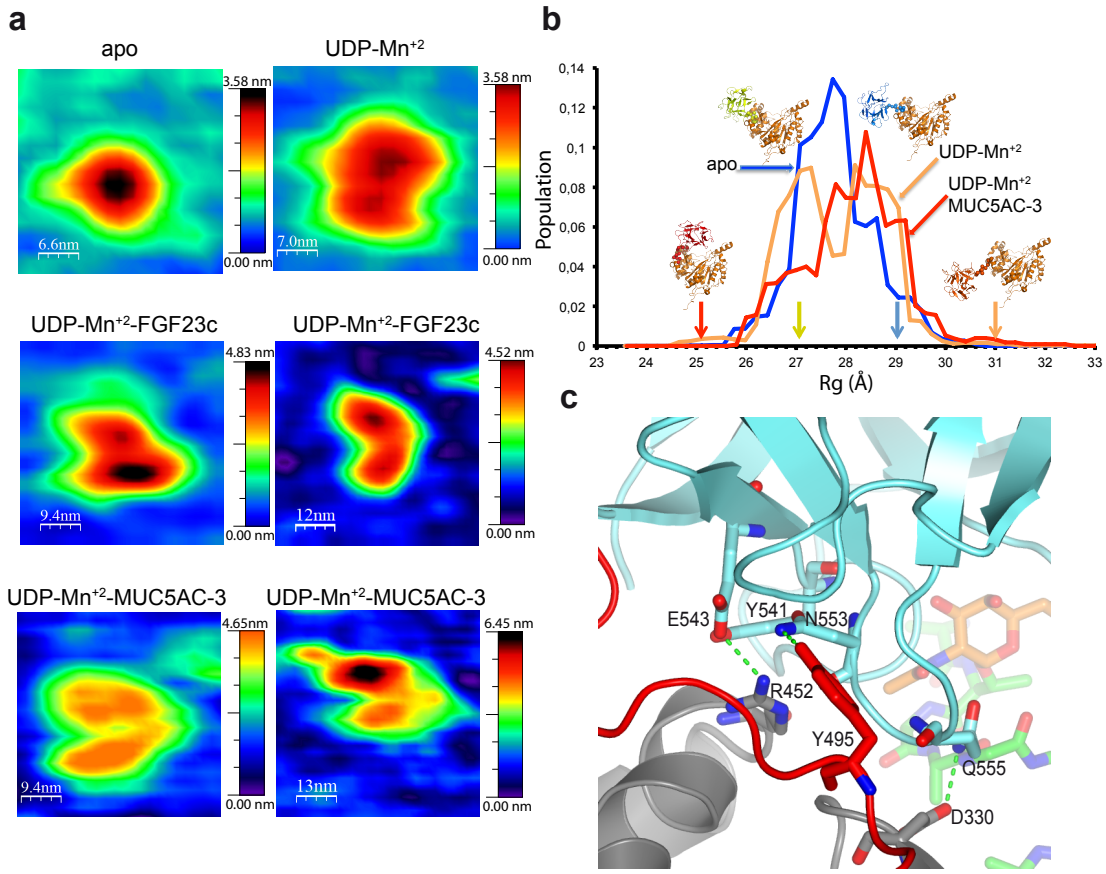
Supplementary Figure 7. Multiple sequence alignment of *Tg*GalNAc-T3 and the *Hs*GalNAc-T2 and T4. Secondary structure elements (α-helices/3_{10}-helices/β-strands) are shown for the *Tg*GalNAc-T3 (in red), ***Hs***GalNAc-T2 (in yellow) and T4 (in green). The double sided arrows indicate regions encompassing the catalytic domain flexible loop and the linker connecting the catalytic and lectin. Note that all the three enzymes share nearly the same secondary structure elements. The major structural difference between GalNAc-T3/T4 and GalNAc-T2 is the placement of their lectin domains relative to their catalytic domains (see **Fig. 4c**)



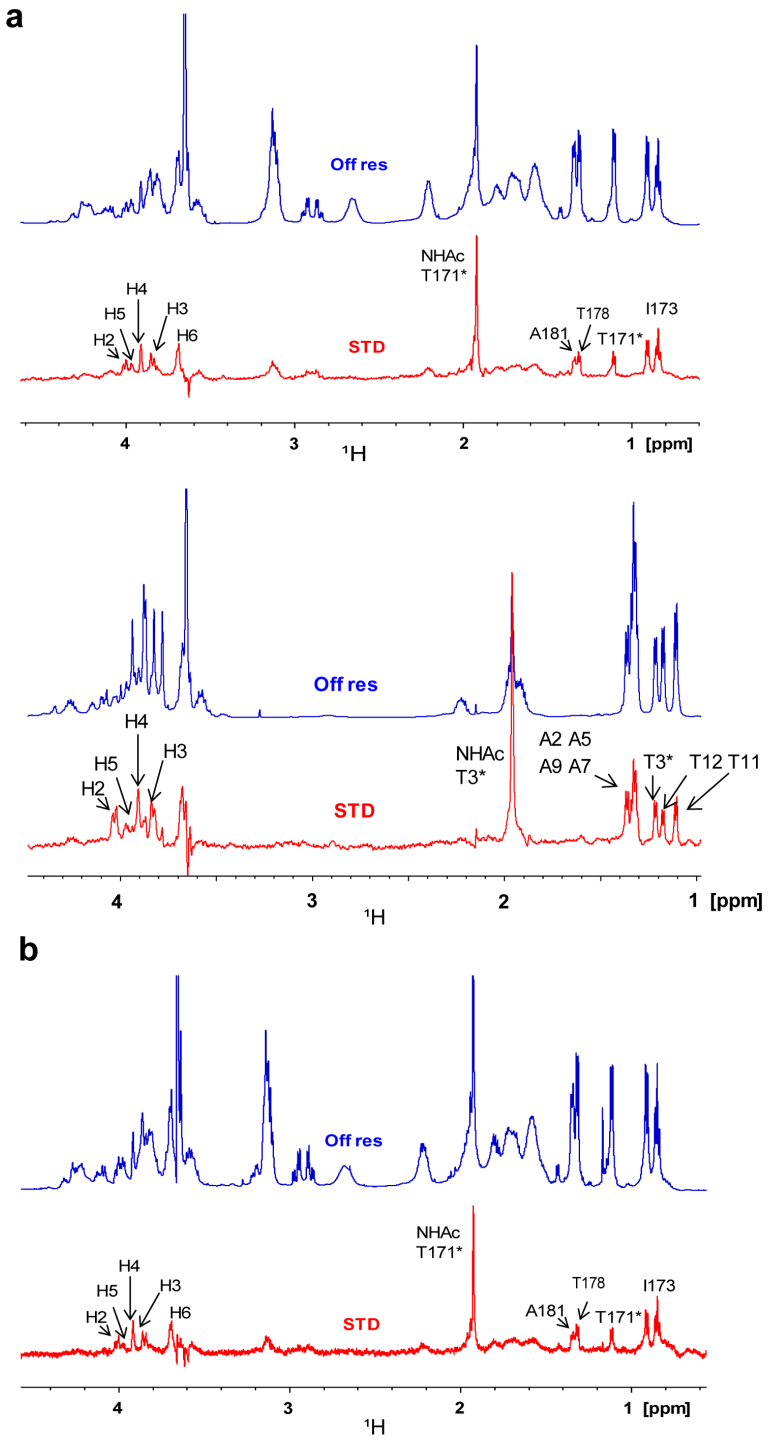
Supplementary Figure 8. **Small angle X-ray scattering (SAXS) data obtained for *Hs*GalNAc-T3**. (**a**) Size-exclusion SEC-SAXS profiles of *Hs*GalNAc3 without ligands (blue), with UDP/$MnCl_2$ (yellow) and UDP/$MnCl_2$ and **MUC5AC-3** (red). Lines represent the profile of scattering at zero angle *I(0)* (cm^{-1}) in absolute scale. Plots also display the R_g (Å) with the same color code and the MW (estimated using Vc, in kDa) in black derived from all the SAXS curves measured. The flat R_g and MW profiles indicate that the elution pic in all samples is monodisperse. (**b**) Scattering intensity, *I(q)*, in absolute scale as a function of the momentum transfer *q*, (**c**) dimensionless Kratky plot [$(qR_g)^2I(q)$ *vs.* qR_g], and (**d**) Pair-wise distance distribution, *P(r)*, for *Hs*GalNAc-T3 without ligands (blue), with UDP/$MnCl_2$ (yellow) and UDP/$MnCl_2$ and **MUC5AC-3** (red). The inset in panel (**b**) displays the Guinier fits for the SAXS curves in the *AutoRg* selected areas with *Rg* values of 28.1, 28.4 and 28.0 Å for *Hs*GalNAc-T3 without ligands, with UDP/$MnCl_2$ and UDP/$MnCl_2$ and **MUC5AC-3**, respectively. For panels (**c**) and (**d**), the 3 curves are overlapped indicating that the protein adopt a similar conformation in the three conditions although some differences can be observed. The dimentionless Kratky plots in (**c**) present maxima close to the reference black curve (1.1 at qR_g=1.75), confirming that the protein in the three conditions is mainly globular with reduced inherent flexibility. The *P(r)* functions in (**d**), which represent the distribution of interatomic distances in the particle, indicate maximum intramolecular distances, *Dmax*, of 102 Å, 110 Å and 100 Å for *Hs*GalNAc-T3 without ligands, with UDP/$MnCl_2$ and UDP/$MnCl_2$ and **MUC5AC-3**, respectively.



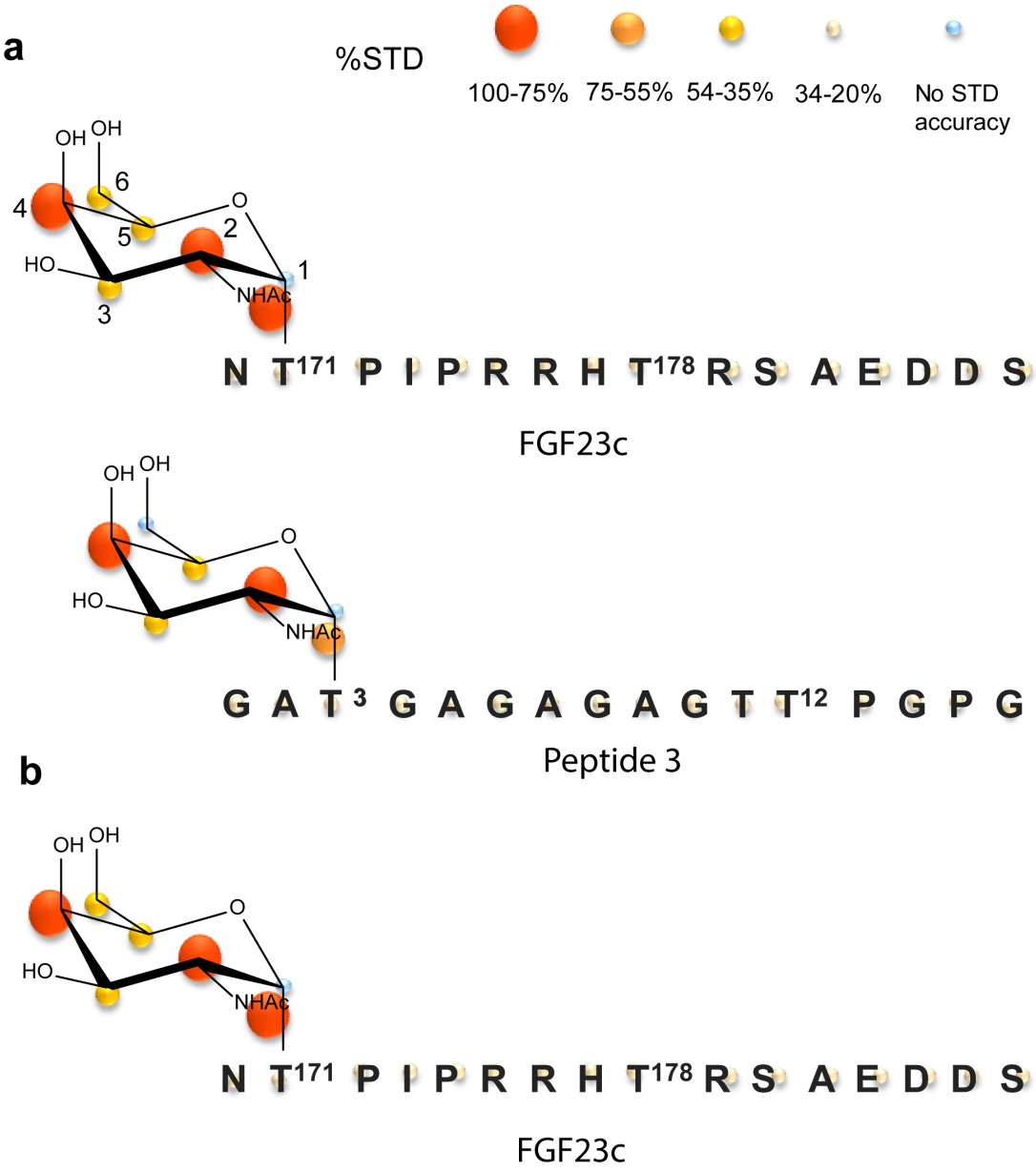
Supplementary Figure 9. **Analysis of the small angle X-ray scattering (SAXS) data from Supplementary Fig. 8.** Best rigid-body Dadimodo models (upper panels) and curve fits and residuals (lower panels) for *Hs*GalNAc-T3 without ligands (blue), with UDP/$MnCl_2$ (yellow) and UDP/$MnCl_2$ and **MUC5AC-3** (red). Residuals were computed with the equation $[(I^{exp}(s)-I^{Dad}(s))/\sigma(s)]$ where $I^{exp}(s)$ is the experimental curve, $I^{Dad}(s)$ is the theoretical curve of the Dadimodo optimized structure, and $\sigma(s)$ are the experimental error associated to all points in the curve. Excellent fits indicate that the derived model is a good representation of the conformations present in solution.



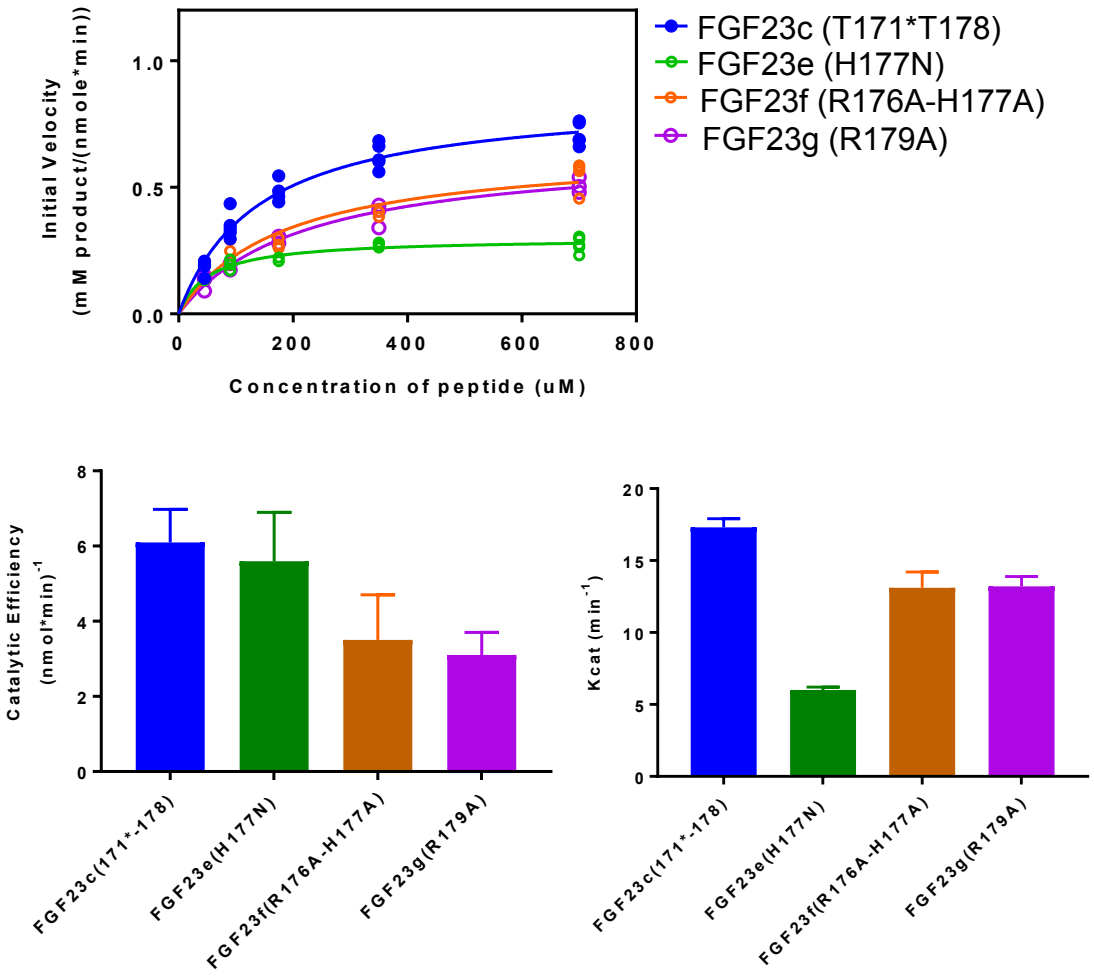
Supplementary Figure 10. **Topography AFM images of single *Hs*GalNAc-T3 molecules showing different conformations and R_g analysis of the SAXS data for *Hs*GalNAc-T3.** (**a**) (Top-left panel) AFM image of a very compact apo form showing that the lectin and catalytic domains overlap each other. The average Z-height of the feature is ≈3.5 nm. (Top-right panel) AFM image showing a relatively compact form of the enzyme bound to UDP/Mn^{+2}. (Remaining panels) AFM images showing less-compact structures of the enzyme bound to UDP/Mn^{+2}/**FGF23c** (middle panels) or UDP/Mn^{+2}/MUC5AC-3 (lower panels). Note that in the presence of the monoglycopeptides both the catalytic and lectin domains are clearly visualized. (**b**) Radius of gyration (R_g) distributions derived from the ensemble optimization method (EOM) analysis of the GalNAc-T3 apo form (blue), the GalNAc-T3 complexed with UDP/Mn^{+2} (orange), and the GalNAc-T3 complexed with UDP/Mn^{+2}/**MUC5AC-3** (red). Representative GalNAc-T3 conformations with different R_gs are displayed with the EOM R_g profiles. The EOM analysis suggests that, without ligand, GalNAcT3 is compact and displays a narrow R_g peak. In the presence of UDP/Mn^{+2}, GalNAcT3 has a broader R_g distribution suggesting an enhanced interdomain flexibility. The addition of UDP/Mn^{+2} and **MUC5AC-3** to the enzyme leads to a peak that is displaced towards less-compact structures (R_g≈28.5Å) indicating that the distance between both domains is increased, in agreement with the AFM data. Both AFM and SAXS data indicate that the less-compact structures are more prevalent in the presence of monoglycopeptides. (**c**) The absence of predominantly extended conformations in GalNAc-T3 can be explained by specific interactions between the catalytic and lectin domains. These interactions include an electrostatic interaction between $Arg452^{cat\ dom}$ & $Glu543^{lec\ dom}$, hydrogen bond interactions between $Asp330^{cat\ dom}$ & $Gln555^{lec\ dom}$ and $Tyr495^{flexible\ linker}$ & $Asn553^{lec\ dom}$. Note that all of the above interactions are established between the side chains of the indicated amino acids and are likely not present in GalNAc-T2.



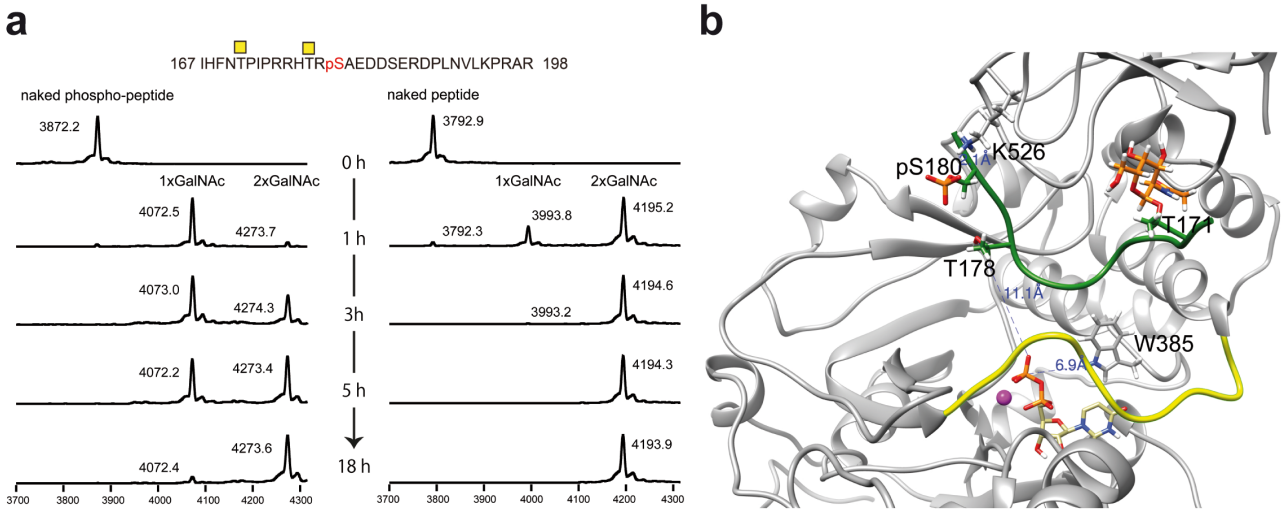
Supplementary Figure 11. Saturation transfer difference (STD) NMR spectra showing glycopeptide binding to *Tg*- and *Hs*GalNAc-T3. (a) (upper panel) STD-NMR experiment of **FGF23c** (1690 µM), in the presence of *Tg*GalNAc-T3 (26 µM) and UDP (75 µM) and $MnCl_2$ (75 µM) at 298 K. (lower panel) STD-NMR experiments with glycopeptide **P3** (1365 µM) in the presence of *Tg*GalNAc-T3 (21 µM), UDP (75 µM), $MnCl_2$ (75 µM) at 298 K. (**b**) STD-NMR experiments of **FGF23c** (750 µM) in the presence of *Hs*GalNAc-T3 (11.5 µM), UDP (75 µM), $MnCl_2$ (75 µM) at 298 K. Key proton resonances are marked in each STD spectrum. Proton resonance assignments are given in Supplementary Tables 5 and 6. All experiments were done in triplicate giving similar results.



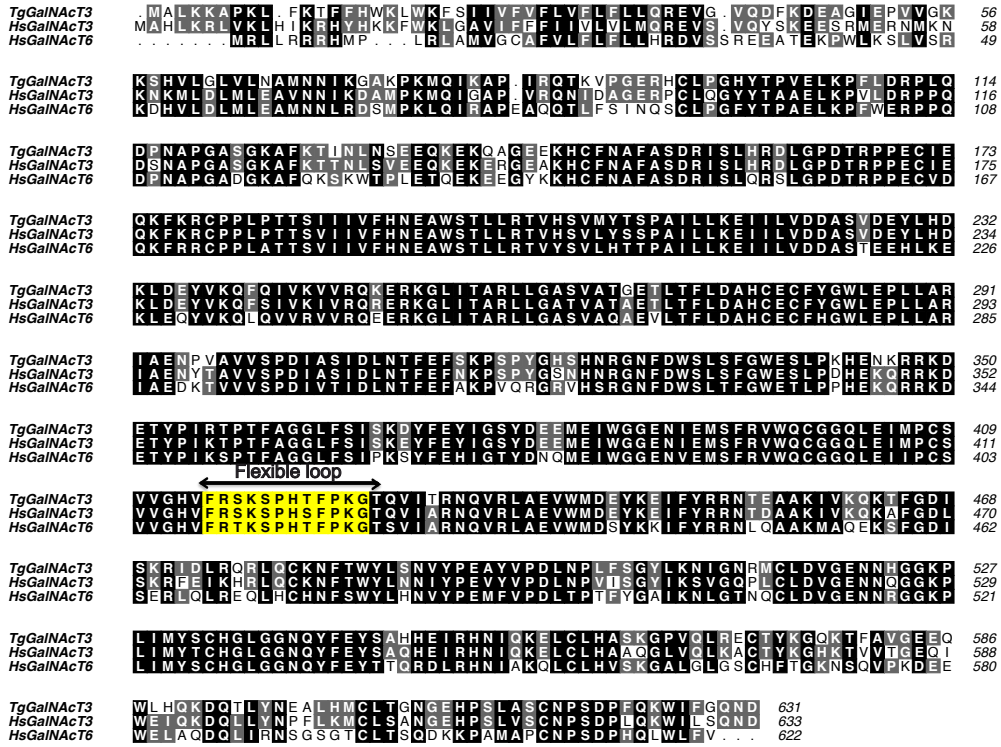
Supplementary Figure 12. Saturation transfer difference (STD) NMR epitope mapping of FGF23c and peptide P3 bound to *Tg*GalNAc-T3 and *Hs*GalNAc-T3. (**a**) STD-NMR epitope mapping in the presence of *Tg*GalNAc-T3. (Upper panel) Glycopeptide **FGF23c** presents a GalNAc STD-NMR derived binding epitope where H2, H4, and the N-acetyl group of the GalNAc moiety have the highest STD-NMR responses, closely followed by the H3/H5/H6 protons. (Lower panel) Monoglycopeptide **P3** presents a STD-NMR derived binding etpitope where H2 and H4 have the highest STD-NMR response followed by the N-acetyl group of the GalNAc moiety and the H3/H5 protons. In the case of **P3**, no accurate STD data was obtained for H6s protons. (**b**) STD-NMR mapping showing that *Hs*GalNAc-T3 binds **FGF23c** identically to *Tg*GalNAc-T3. Note that the STD-NMR binding paterns of **FGF23c** are also similar to those previously reported for GalNAc-T4 binding to remotely glycosylated glycopeptides.



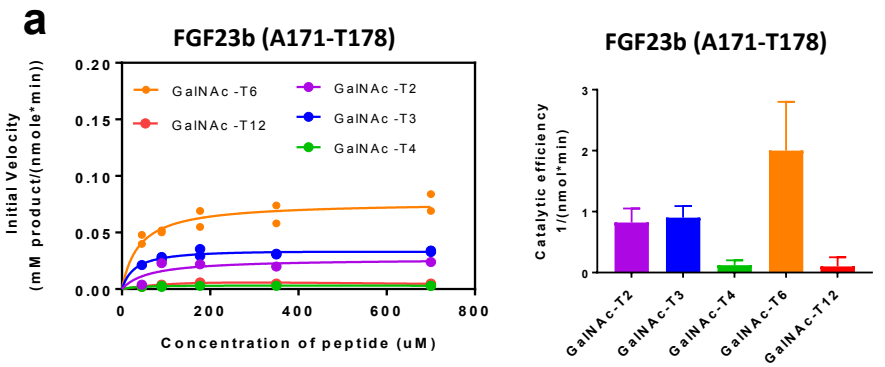
Supplementary Figure 13. Kinetics of *Hs*GalNAc-T3 against a series of FGF23c mutants and FGF23c altered charge derivatives. (upper panel) Comparison kinetic plots. Plotted lines represent the Michaelis–Menten non-linear fits (using GraphPad Prism 7.03) of the initial velocity data using the kinetic values given in Supplementary Table 2**.** Initial velocities (individual points) were obtained in duplicate or higher for each peptide concentration giving a total of 15 independent determinations for **FGF23e**, **FGF23f** and **FGF23g** and 24 determinations for **FGF23c**. (Lower panels) Plots of the catalytic efficiency (V_{max}/K_m) and Kcat of *Hs*GalNAc-T3 derived from the plots in the upper panels. Error bars represent the standard deviation calculated by the GraphPad Prism fit of each data set as described in the Online Methods.



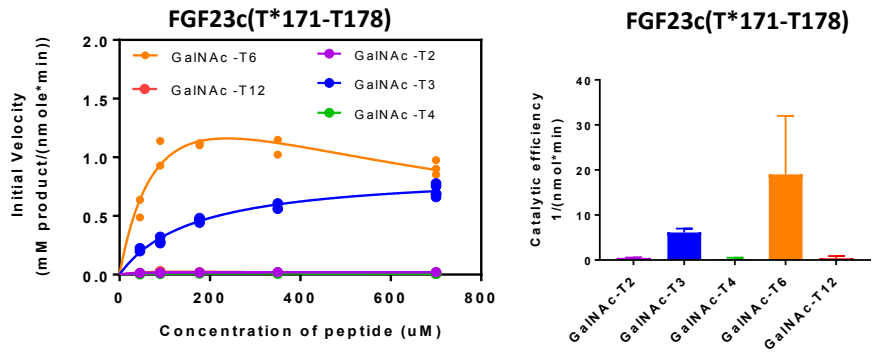
Supplementary Figure 14. Impact of FGF23 Ser180 phosphorylation on glycosylation under *in vitro* conditions. (**a**) Comparison of the glycosylation of the FGF23 pS180 phosphopeptide, **FGF23i**, with the nonphosphorylated peptide, **FGF23h**, by *Hs*GalNAc-T3 monitored by MALDI-TOF MS. The estimated number of GalNAc residues incorporated are indicated next to each mass value. All experiments were done in triplicate yielding similar results. (**b**) Close-up view of the *Tg*GalNAc-T3 **FGF23c**-pSer180 structure derived from 500 µs MD simulation.



Supplementary Figure 15. Multiple sequence alignment of *Tg*GalNAc-T3 and the *Hs*GalNAc-T3 and T6. The region encompassing the catalytic domain flexible loop is indicated with double sided arrows and highlighted in yellow. The sequence identity between *Hs*GalNAc-T3 and *Hs*GalNAc-T6 is 65%.



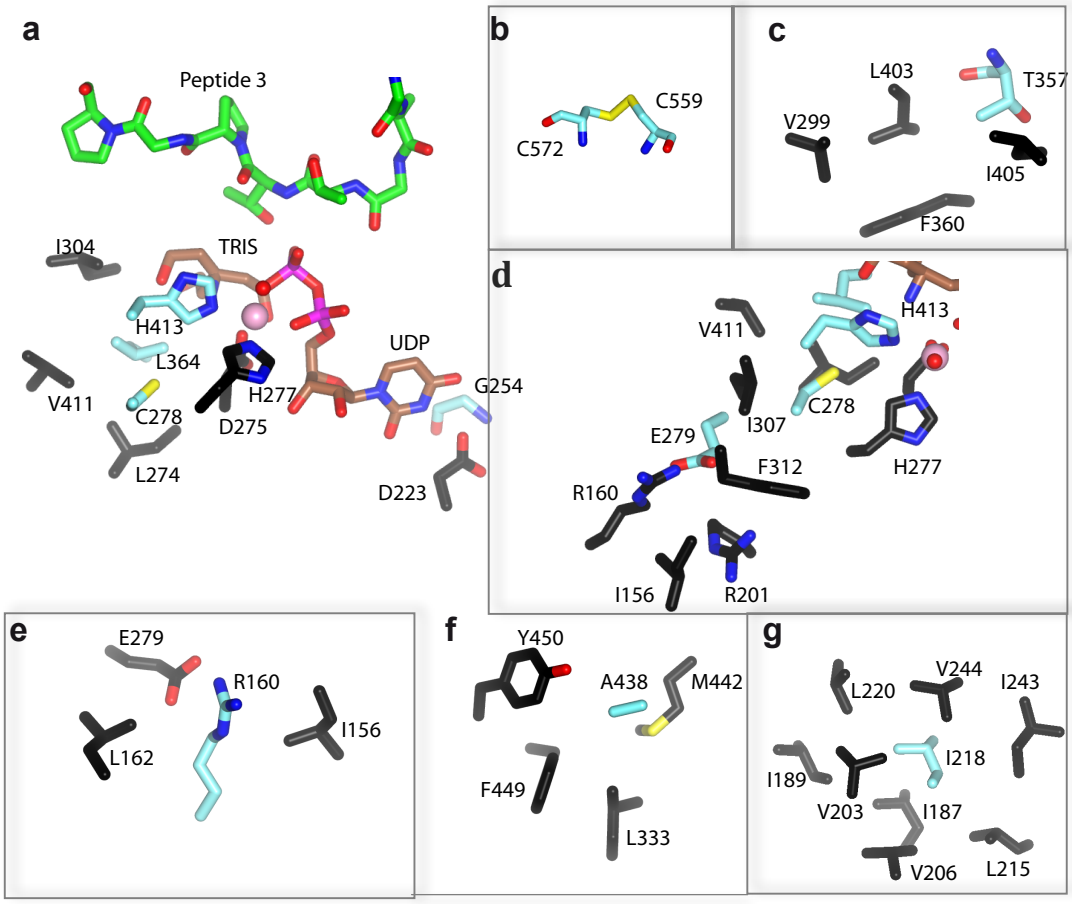
FGF23b	Vmax (mM/nmol*min)	Km (μM)	Vmax/km (1/(nmol*min))	R^2	Hydro -lysis %
GalNAc-T2	0.04 ± 0.04	115 ± 260	0.82 ± 0.23	0.61	75%
GalNAc-T3	0.03 ± 0.001	36 ± 6.9	0.9 ± 0.19	0.81	70%
GalNAc-T4	0.003 ± 0.001	26 ± 8	0.12 ± 0.08	0.72	90%
GalNAc-T6	0.077± 0.005	39 ± 12	2.0 ± 0.81	0.73	70%
GalNAc-T12	0.005 ± 0.001	50 ± 35	0.1 ± 0.15	0.69	85%



FGF23c	Vmax (mM/nmol*min)	Km (μM)	Vmax/km (1/(nmol*min))	R^2	Hydro-lysis %
GalNAc-T2	0.02 ± 0.001	42 ± 5	0.48 ± 0.08	0.98	50%
GalNAc-T3	0.86 ± 0.03	140 ± 15	6.1 ± 0.88	0.95	10%
GalNAc-T4	0.008 ± 0.001	38 ± 26	0.21 ± 0.32	0.68	40%
GalNAc-T6	2.1 ± 0.58	106 ± 40	19.0 ± 13.7	0.81	5%
GalNAc-T12	0.04 ± 0.07	109 ± 250	0.37 ± 0.5	0.52	30%

Supplementary Figure 16. Peptide glycosylation kinetics of the *Hs*GalNAc-T2, T3, T4, T6 & T12 against FGF23b (A171,T178) and FGF23c (T*171,T178). (a) (Left panel) Kinetic plots of *Hs*GalNAc-T2, T3, T4, T6 & T12 against **FGF23b**. Plots represent the Michaelis–Menten non-linear fits (using GraphPad Prism 7.03) of the initial velocity data using the kinetic values given in the table below**.** Initial velocities (individual points) were obtained for each peptide concentration giving a total of 5, 15, 10, 10 and 5 independent determinations for GalNAc-T2, T3, T4, T6 and T12. **(Right panel)** Plots summarizing the catalytic efficiency (V_{max}/K_m) derived from the plots in the left panel. Error bars represent the standard deviation calculated by the GraphPad Prism fit of each data set as described in the Online Methods. (b) **(Left panel)** Kinetic plots of *Hs*GalNAc-T2, T3, T4, T6 & T12 against **FGF23c**. Plots represent the Michaelis–Menten (or Michaelis–Menten with substrate inhibition) non-linear fits (using GraphPad Prism 7.03) of the initial velocity data using the kinetic values given in

the lower table**.** Initial velocities (individual points) were obtained for each peptide concentration giving a total of 5, 15, 11, 11 and 5 independent determinations for GalNAc-T2, T3, T4, T6 and T12. (**Right panel**) Plots summarizing the catalytic efficiency (V_{max}/K_m) derived from the plots in the left panel. Error bars represent the standard deviation calculated by the GraphPad Prism fit of each data set as described in the Online Methods. Percent UDP-GalNAc hydrolysis was determined by G10 gel filtration chromatography of the reaction products as described in the Online Methods.



Supplementary Figure 17. Structural environments of the *Hs*GalNAc-T3 missense mutations mapped onto the *Tg*GalNAc-T3 structure. *Tg*GalNAc-T3 residues shown in aquamarine correspond to the human residues that are mutated in FTC (see numbering of the human residues in Fig. 6 and Supplementary Fig. 5) while in some cases, additional residues in black interact with residues in aquamarine. Note that residue numbers in the figure and legend represent the *Tg*GalNAc-T3 numbering. **(a)** His413 and Gly254 are in the UDP-GalNAc binding site while Leu364 and Cys278 are very close to it. His413 is one of the residues coordinating the Mn^{+2} atom and its mutation to Gln will likely cause the loss of this interaction and in turn lead to steric hindrance with UDP-GalNAc pyrophosphate. **(b)** Cys572 forms a disulfide bridge with Cys559, both are located in the lectin domain. (**c**) The methyl group of Thr357 faces a hydrophobic pocket formed by Leu403, Val299, Phe360 and Ile405. (**d**) Glu279 interacts by a salt bridge with Arg160. (**e**) Arg160 establishes a salt bridge interaction with Glu279 and cation-CH interactions with Ile156 and Leu162. (**f**) Ala438 is surrounded by hydrophobic residues such as Tyr450, Phe449, Leu333 and Met442. (**g**) Ile218 is surrounded by hydrophobic residues such as Val244, Leu220, Ile243, Ile189, Val203, Ile187, Val206 and Leu215.

Supplementary Table 1. Names and sequences of the peptides used in this study.

Peptide	Sequence[a,b]
FGF23a (T171-T178)	IHFNT^{171}PIPRRHT^{178}RSAEDD
FGF23b (A171-T178)	NA^{171}PIPRRHT^{178}RSAEDDS
FGF23c (T*171-T178)	NT*^{171}PIPRRHT^{178}RSAEDDS
FGF23d (T171)	IHFNT^{171}PIPRRH
FGF23e (H177N)	NT*^{171}PIPRRNT^{178}RSAEDDS
FGF23f (R176A-H177A)	NT*^{171}PIPRAAT^{178}RSAEDDS
FGF23g (R179A)	NT*^{171}PIPRRHT^{178}ASAEDDS
FGF23h	IHFNT^{171}PIPRRHT^{178}RSAEDDSERDPLNVLKPRAR
FGF23i	IHFNT^{171}PIPRRHT^{178}RpSAEDDSERDPLNVLKPRAR
MUC5AC-3	GTT*PSPVPTTSTTSAP[c]
P1 (-TPGP-)	GAGAGAGTT^{9}PGPG
P2 (-TPGP—T*-)	AGAGTT^{6}PGPGAGAT*GA
P3 (-T*--TPGP-)	GAT*GAGAGAGTT^{12}PGPG
OP1 (GalNAc-T3)	(GA)GAYAVT^{6}PGPAG(GA)[d]
OP1-T* (GalNAc-T3)	GT*GAGGAYAVT^{11}PGPAG
OP2 (GalNAc-T4)	GAGAYPIT^{8}PAPAGAG
OP3 (GalNAc-T12)	GAGAYYIT^{8}PRPAGAG
OP4 (GalNAc-T15)	GAGAYYPT^{8}PGPAGAG
OP5 (GalNAc-T13)	GAGAPPVT^{8}PGPAGAG
OP6 (GalNAc-T11)	GAGAPPPT^{8}PGPAGAG
OP7 (GalNAc-T14)	GAGAVGPT^{8}PGPAGAG
OP8 (GalNAc-T2)	GAGAPGPT^{8}PGPAGAG
OP9 (GalNAc-T16)	GAGAIGPT^{8}PGPAGAG
OP10 (GalNAc-T1)	GAGADAPT^{8}PGPAGAG

[a] Acceptor sites are highlighted in blue.
[b] Mutated residues are highlighted in red.
[c] Preferred acceptor site is in bold.
[d] Short and long **OP1** peptides were used, each giving identical kinetic values.

Supplementary Table 2. Kinetic parameters[#] of the peptides used in this study using the *Hs*GalNAc-T3 and the *Tg*GalNAc-T3.

***Hs*GalNAc-T3**	**Vmax** (mM/nmol* min)	**Km** (uM)	**Catalytic efficiency (Vmax/Km)** (1/(nmol*min))	**Kcat** (min^{-1})	**MM curve fitness r^2**	**Hydro-lysis* (%)**	**Individual determinants** */no. of substrate concentration*
P1 (-TPGP-)	2.2 ± 0.5	81 ± 30	27 ± 18	46 ± 10.0	0.69	20%	19/5
P2 (-TPGP--T*-)	2.0 ± 0.1	77 ± 13	26 ± 6	40 ± 1.9	0.89	20%	16/5
P3 (-T*--TPGP-)	13.7 ± 0.4	35 ± 5	389 ± 68	273 ± 7.2	0.84	2%	21/5
OP1 (T3)	13.9 ± 0.6	85 ± 11	163 ± 29	276 ± 12	0.91	2%	21/5
OP1-T* (T3)	16.2 ± 1.7	61 ± 14	265 ± 93	325 ± 34	0.90	2%	15/5
OP2 (T4)	9.5 ± 0.9	101 ± 28	94 ± 37	192 ± 18	0.92	2%	9/5
OP3 (T12)	12.4 ± 1.4	376 ± 88	33 ± 12	245 ± 29	0.97	2%	5/5
OP4 (T5)	10.5 ± 1.4	393 ± 103	27 ± 11	210 ± 28	0.96	2%	9/5
OP5 (T13)	2.6 ± 0.40	404 ± 112	6.4 ± 3.0	55 ± 7.8	0.95	20%	9/5
OP6 (T11)	1.60 ± 0.22	252 ± 83	6.3 ± 3.3	30 ± 4.9	0.92	25%	9/5
OP7 (T14)	1.50 ± 0.26	350 ± 121	4.3 ± 2.5	31 ± 6.5	0.92	25%	9/5
OP8 (T2)	0.92 ± 0.11	194 ± 52	4.7 ± 1.9	18 ± 2.0	0.92	25%	9/5
OP9 (T16)	0.50 ± 0.08	233 ± 85	2.1 ± 1.3	9.2 ± 1.6	0.83	45%	9/5
OP10 (T1)	0.20 ± 0.03	100 ± 41	2.0 ± 1.3	4.2 ± 0.6	0.70	55%	9/5
FGF23b (A171-T178)	0.03 ± 0.001	36 ± 6.9	0.9 ± 0.19	0.7 ± 0.02	0.81	70%	15/5
FGF23c (T*171-T178)	0.86 ±0.03	140 ± 15	6.1 ± 0.88	17.3 ± 0.6	0.95	10%	24/5
FGF23d (T171)	0.25 ± 0.01	423 ± 71	0.6 ± 0.13	5.0 ± 0.4	0.95	70%	20/5
FGF23e (H177N)	0.29 ± 0.01	52 ± 10	5.6 ± 1.3	6.0 ± 0.2	0.83	15%	15/5
FGF23f (R176A-H177A)	0.65 ±0.05	185 ± 45	3.5 ± 1.2	13.1 ± 1.1	0.85	15%	15/5
FGF23g (R179A)	0.66 ±0.03	216 ± 31	3.1 ± 0.6	13.2 ± 0.7	0.92	15%	15/5
***Tg*GalNAc-T3**	**Vmax** (mM/nmol* min)	**Km** (mM)	**Catalytic efficiency (Vmax/Km)** (1/(nmol*min))	**Kcat** (min^{-1})	**MM curve fitness r^2**	**Hydro-lysis* (%)**	**Individual determinants** */no. of substrate concentration*
P1 (-TPGP-)	3.0 ± 0.19	83 ± 17	36.1 ± 10	69 ± 2.9	0.86	2%	15/5
P3 (-T*--TPGP-)	18.6 ± 2.2	82 ± 20	226 ± 87	386 ± 52	0.89	2%	15/5
OP1 (T3)	22.2 ± 3.4	69 ± 22	321 ± 119	445 ± 79	0.77	2%	15/5
OP1-T* (T3)	26.6 ± 7.9	112 ± 40	241 ± 181	522 ± 158	0.78	2%	14/5
FGF23b (A171-T178)	0.04 ± 0.006	170 ± 62	0.29 ± 0.16	0.3 ± 0.05	0.72	70%	15/5
FGF23c (T*171-T178)	2.7 ± 0.26	127 ± 37	21.3 ± 9.0	49 ± 4.8	0.89	12%	10/5

Kinetic parameters and standard deviations were calculated using the Michaelis–Menten (MM) (or Michaelis–Menten with substrate inhibition) non-linear fit in GraphPad Prism 7.03. Initial velocities were obtained typically in triplicate or greater for each substrate concentration as indicated in the right most column. See Online Methods for further details.

* Percent UDP-GalNAc hydrolysis was determined by G10 gel filtration chromatography of the reaction products as described in the Online Methods. Note that UDP-GalNAc hydrolysis tends to be highest against poor substrates.

Supplementary Table 3. Data collection and refinement statistics.

	*Tg*GalNAc-T3 in complex with UDP-Mn^{+2} and **P3**	*Tg*GalNAc-T3 in complex with UDP-Mn^{+2} and **FGF23c**
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, *b*, *c* (Å)	50.65, 104.66, 143.19	50.96, 104.82, 143.47
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	20-2.12 (2.23-2.12)*	50.96-1.96 (2.07-1.96)
R_{merge}	0.110 (0.791)	0.131 (1.031)
$I / \sigma I$	7.7 (1.2)	8.9 (2.3)
Completeness (%)	94.0 (71.9)	100 (100)
Redundancy	5.3 (3.1)	7.0 (7.1)
Refinement		
Resolution (Å)	2.12	1.96
No. reflections	218424	394278
R_{work} / R_{free}	0.191/0.237	0.179/0.216
No. atoms		
Protein	4335	4269
UDP	25	25
Peptide	75	98
GalNAc	14	14
Waters	201	304
Ethylenglycol	-	56
Sulphate	-	5
TRIS	11	-
Phosphate	80	-
B-factors (Å^2)		
Protein	43.99	36.38
UDP	45.20	58.50
Peptide	75.40	86.99
GalNAc	45.93	34.10
Waters	44.15	42.46
Ethylenglycol	-	53.55
TRIS	63.14	-
Sulphate	-	82.79
Phosphate	81.90	-
R.m.s. deviations		
Bond lengths (Å)	0.0091	0.0111
Bond angles (°)	1.688	1.729

1 crystal was used to determine each structure. *Values in parentheses are for highest-resolution shell.

Supplementary Table 4. SAXS structural parameters.

	*Hs*GalNAc-T3	*Hs*GalNAc-T3 + $MnCl_2$ + UDP	*Hs*GalNAc-T3 + $MnCl_2$ + UDP + **MUC5AC-3**
Guinier Analysis			
I(0) (1/cm)	0.0915	0.0610	0.0385
Rg (Å)	28.09	28.36	28.04
qmin (1/Å)	0.0185	0.0243	0.0235
qRg max	1.3	1.23	1.26
Coefficient of correlation, R^2	0.9999	0.9998	1.0000
*M from I0 or from Atsas Prediction**	64 (top of elution pic) [54-60] (0.87-0.97)	[54-60] (0.86-0.95)	[51-58] (0.80-0.92)
P(r) analysis			
I(0) (1/cm)	0.0918	0.0613	0.0384
Rg (Å)	28.48	28.89	28.30
Dmax (Å)	103	110	100
q range (1/Å)	0.4	0.4	0.4

Atomistic modelling			
Homology model	*Hs*GalNAc-T3 (obtained from *Tg*GalNAc-T3 crystal structure)	*Hs*GalNAc-T3 (obtained from *Tg*GalNAc-T3 crystal structure)	*Hs*GalNAc-T3 (obtained from *Tg*GalNAc-T3 crystal structure)
q range for all modelling	0.0186-0.4	0.0242-0.4	0.0242-0.4
EOM			
χ2	3.5**	3.6**	5.9**
Constant subtraction	No	No	No
No. of representative structures	4	5	5
Dadimodo			
χ2	1.2	1.5	0.9
Number of runs	10	10	10
Rigid body definition	body1 = A: 5 - 321, A: 341 - 400 body2 = A: 412 - 541	body1 = A: 5 - 321, A: 341 - 400 body2 = A: 412 - 541	body1 = A: 5-321, A: 341-400, B: 548-563 body2 = A: 412 - 541

* The prediction of the Molar Mass (M) from *I(0)* was only possible for the free form. The undetermined quantity of UDP around the protein during the column elution influences the absorbance at 280 nm and protein concentrations could not be determined accurately for the two others forms.
** χ2 values from EOM were high due to the presence of a flexible loop that was considered in a fixed conformation in EOM analysis.

Supplementary Table 5. ^{1}H-NMR assignments of the **FGF23c** peptide (ppm) recorded at Bruker Avance 600 MHz spectrometer equipped with a triple channel cryoprobe head with H_2O/D_2O (90%/10%) as solvent at 278K.

Residue	Proton	ppm
GalNAc-T171*	H1	4.74
	H2	3.78
	H3	3.62
	H4	3.67
	H5	3.74
	H6	3.45
	NHAc	1.69
N170	H	8.73
	Hα	4.27
	$H\beta_3$	2.71
	$H\beta_2$	2.65
T171*	H	8.06
	Hα	4.04
	Hβ	3.89
	Hγ	0.89
P172	Hα	4.20
	$H\delta_3$	3.47
	$H\delta_2$	3.34
	Hβ	1.98
	Hγ	1.67
I173	H	8.00
	Hα	3.83
	Hβ	1.42
	$H\gamma_1$	0.88
	$H\gamma_2$	0.68
	Hδ	0.61
P174	Hα	4.06
	Hδ	3.62
	Hβ	1.97
	Hγ	1.72
R175	H	8.22
	Hα	3.97
	Hδ	2.91
	Hβ	1.43
	Hγ	1.32
R176	H	8.23
	Hα	3.99
	Hδ	2.88
	Hβ	1.47
	Hγ	1.30
H177	H	8.55
	Hα	4.47
	Hβ	2.92
T178	H	8.17
	Hα	4.01
	Hγ	1.08
R179	H	8.35
	Hα	4.09
	Hδ	2.92
	Hβ	1.53
	Hγ	1.35
S180	H	8.27
	Hα	4.13
	Hβ	3.61
A181	H	8.31
	Hα	4.00
	Hβ	1.09
E182	H	8.05
	Hα	3.99
	Hγ	2.12
	$H\beta_3$	1.77
	$H\beta_2$	1.66
D183	H	8.19
	Hα	4.45
	Hβ	2.58
D184	H	8.07
	Hα	4.37
	Hβ	2.54
S185	H	8.03
	Hα	4.04
	Hβ	3.59

Supplementary Table 6. ^{1}H-NMR assignments of the **P3** (ppm) recorded at Bruker Avance 600 MHz spectrometer equipped with a triple channel cryoprobe head with H_2O/D_2O (90%/10%) as solvent at 278K.

Residue	Proton	δ ppm
T3* GalNAc	H1	4.89
	H2	4.02
	H3	3.83
	H4	3.90
	H5	3.96
	H6	3.68
	NHAc	7.76
G1	NH	8.43
	Hα	3.59
A2	NH	8.43
	Hα	4.23
	Hβ	1.13
T3*	NH	8.49
	Hα	4.28
	Hβ	4.04
	Hγ	0.97
G4	NH	8.24
	Hα	3.67
A5	NH	8.24
	Hα	3.98
	Hβ	1.09
G6	NH	8.32
	Hα	3.61
A7	NH	8.04
	Hα	3.99
	Hβ	1.07
G8	NH	8.26
	Hα	3.65
A9	NH	7.95
	Hα	4.03
	Hβ	1.11
G10	NH	8.20
	Hα	3.75
T11	NH	7.97
	Hα	4.33
	Hβ	4.14
	Hγ	1.09
T12	NH	8.23
	Hα	4.56
	Hβ	4.09
	Hγ	1.16
P13	Hα	4.13
	$H\varepsilon_3$	3.50
	Hβ	2.01
	Hγ	1.71
	$H\varepsilon_2$	3.45
G14	NH	8.19
	Hα	3.85
P15	Hα	4.10
	Hγ	1.72
	Hβ	1.98
	Hε	3.37
	Hε	3.34
G16	NH	8.42
	Hα	3.59

Supplementary Video 1. 500 ns molecular Dynamics simulation of *Tg*GalNAc-T3 complexed to UDP-Mn^{+2} and FGF23c in explicit water.

Supplementary Video 2. 500 ns molecular Dynamics simulation of *Hs*GalNAc-T4 complexed to UDP-Mn^{+2} and FGF23c in explicit water.

Supplementary Video 3. 500 ns molecular Dynamics simulation of *Hs*GalNAc-T6 complexed to UDP-Mn^{+2} and FGF23c in explicit water.

Supplementary Video 4. 500 ns molecular Dynamics simulation of *Hs*GalNAc-T12 complexed to UDP-Mn^{+2} and FGF23c in explicit water.