Supporting Information

Deciphering Structural Elements of Mucin Glycoprotein Recognition

Authors: Andrew Borgert,[†] Jamie Heimburg-Molinaro[‡] Xuezheng Song,[‡] Yi Lasanajak,[‡] Tongzhong Ju,[‡] Mian Liu,[#] Pamela Thompson,[#] Govind Ragupathi,[&] George Barany,[^] David F. Smith, [‡] Richard D. Cummings,[‡,*] and David Live[#,*]

[†]Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, MN 55455
[‡]Department of Biochemistry, Emory University, Atlanta, GA 30322
[#]Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602
[&]Memorial Sloan-Kettering Cancer Center, New York, NY 10065
[^]Department of Chemistry, University of Minnesota, Minneapolis, MN 55455

Supporting Information: Results and Discussion

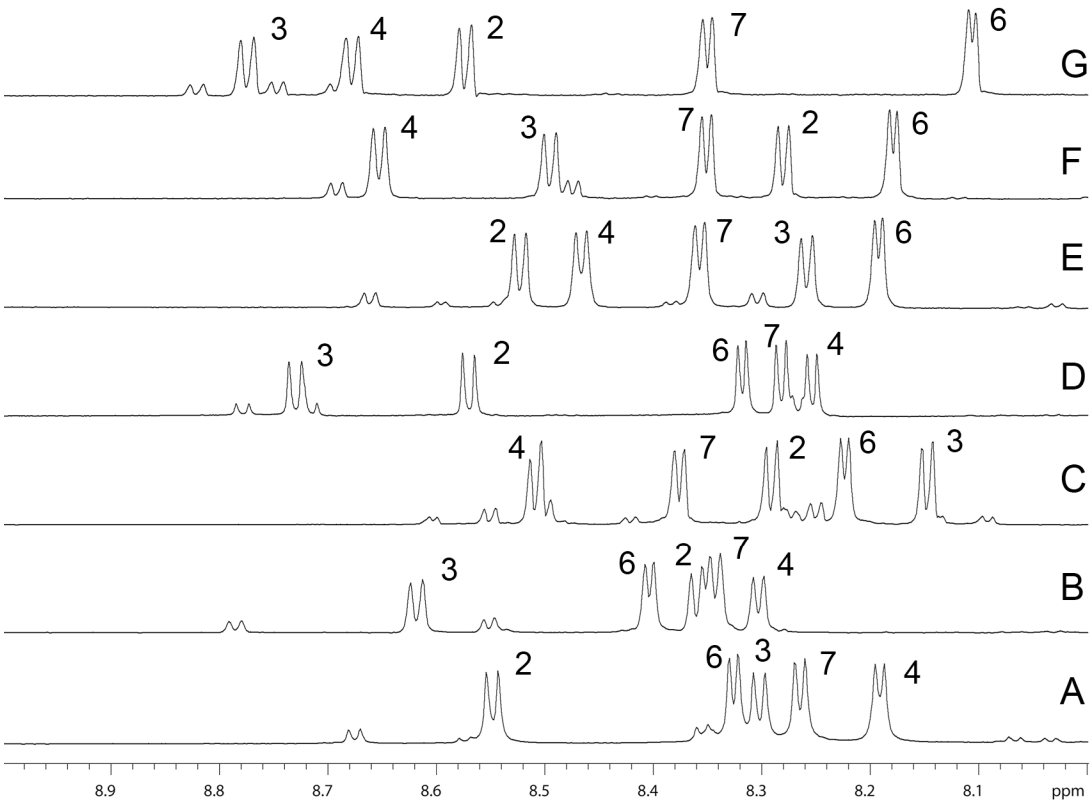


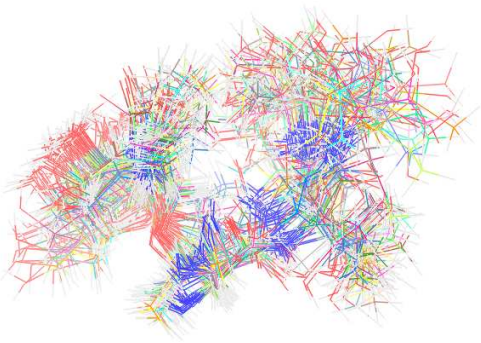
Figure S1. 800 MHz ^{1}H NMR of the peptide amide region of the glycopeptide constructs A through G in 90% H_2O/10% D_2O. Watergate solvent suppression[1] was used. Small peaks arise from the minor proline cis amide bond forms.

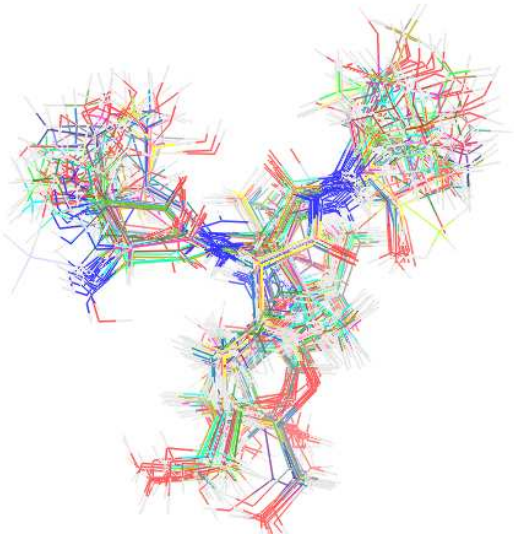
Table S1 Restraint Information, Deviations from Ideal Geometry and Structural Statistics related to Constructs A-G in Figure 2*

A Structural Information PDB ID 2LHV		
Distance Restraints		
Peptide-Peptide		
Intra-residue:		49
Inter-residue:		14
Peptide-Sugar		
Proximal:		5
Non-Proximal:		1
Sugar-Sugar		
GalNAc-GalNAc		8
Total:		77
^3J Coupling Restraints:		5
Torsion Restraints:		7
Restraint Violations		
NOE Violations (>.25 Å)		0
J-Coup Violations (>.5 Hz)		0
Torsion Violations (>5°)		0
Deviations From Ideal Geometry		
Bond (Å):	0.00205 +/- 0.00006	
Angle (°):	0.472 +/- 0.019	
Improper (°)	0.355 +/- 0.009	
Average RMSD Over 78 Structures		
Backbone Heavy Atoms		
+Thr+GalNAc Heavy atoms		
	2.000 +/- 0.662	
All Heavy Atoms		
	2.545 +/- 0.616	

B Structural Information PDB ID 2LI2		
Distance Restraints		
Peptide-Peptide		
Intra-residue:		52
Inter-residue:		20
Peptide-Sugar		
Proximal:		9
Non-Proximal:		4
Sugar-Sugar		
GalNAc-GalNAc		9
Total:		94
^3J Coupling Restraints:		5
Torsion Restraints:		7
Restraint Violations		
NOE Violations (>.25 Å)		0
J-Coup Violations (>.5 Hz)		0
Torsion Violations (>5°)		0
Deviations From Ideal Geometry		
Bond (Å):	0.00208 +/- 0.00012	
Angle (°):	0.472 +/- 0.011	
Improper (°)	0.363 +/- 0.011	
Average RMSD Over 36 Structures		
Backbone Heavy Atoms		
+Thr+GalNAc Heavy atoms		
	2.049 +/- 0.617	
All Heavy Atoms		
	2.530 +/- 0.656	

Figure S2 Overlay of all heavy atoms of T3, T4, and T5 core with associated GalNAcs for all accepted structures A-G. T3 to the right.

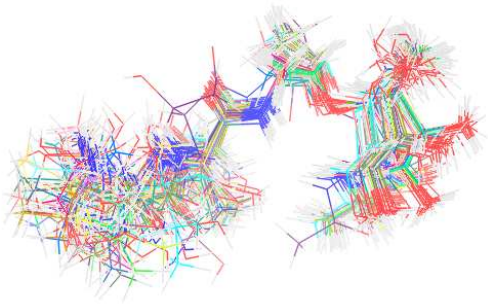




C Structural Information PDB ID 2LI1

Distance Restraints

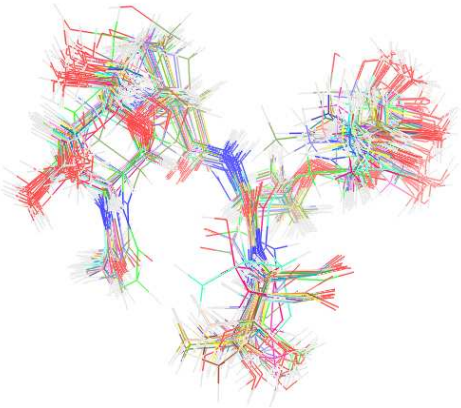
Peptide-Peptide		
	Intra-residue:	56
	Inter-residue:	20
Peptide-Sugar		
	Proximal:	8
	Non-Proximal:	6
Sugar-Sugar		
	GalNAc-GalNAc	9
Total:		99
^3J Coupling Restraints:		5
Torsion Restraints:		7
Restraint Violations		
NOE Violations (>.25 Å)		0
J-Coup Violations (>.5 Hz)		0
Torsion Violations (>5°)		0
Deviations From Ideal Geometry		
Bond (Å):	0.00206 +/- 0.00013	
Angle (°):	0.472 +/- 0.012	
Improper (°)	0.377 +/- 0.018	
Average RMSD Over 74 Structures		
Backbone Heavy Atoms		
+Thr+GalNAc Heavy atoms		
	1.632 +/- 0.220	
All Heavy Atoms		
	2.272 +/- 0.281	



D Structural Information PDB ID 2LI0

Distance Restraints

Peptide-Peptide		
	Intra-residue:	50
	Inter-residue:	21
Peptide-Sugar		
	Proximal:	19
	Non-Proximal:	8
Sugar-Sugar		
	GalNAc-GalNAc	10
Total:		108
^3J Coupling Restraints:		5
Torsion Restraints:		9
Restraint Violations		
NOE Violations (>.25 Å)		0
J-Coup Violations (>.5 Hz)		0
Torsion Violations (>5°)		0
Deviations From Ideal Geometry		
Bond (Å):	0.00192 +/- 0.00013	
Angle (°):	0.443 +/- 0.012	
Improper (°)	0.358 +/- 0.009	
Average RMSD Over 49 Structures		
Backbone Heavy Atoms		
+Thr+GalNAc Heavy atoms		
	1.774 +/- 0.322	
All Heavy Atoms		
	2.461 +/- 0.313	



E Structural Information PDB ID 2LHZ

Distance Restraints

Peptide-Peptide		
	Intra-residue:	44
	Inter-residue:	22
Peptide-Sugar		
	Proximal:	22
	Non-Proximal:	17
Sugar-Sugar		
	GalNAc-GalNAc	25
Total:		130
^3J Coupling Restraints:		5
Torsion Restraints:		9

Restraint Violations

NOE Violations (>.25 Å)	0
J-Coup Violations (>.5 Hz)	0
Torsion Violations (>5°)	0

Deviations From Ideal Geometry

Bond (Å):	0.00193 +/- 0.00016
Angle (°):	0.447 +/- 0.016
Improper (°)	0.365 +/- 0.020

Average RMSD Over 29 Structures

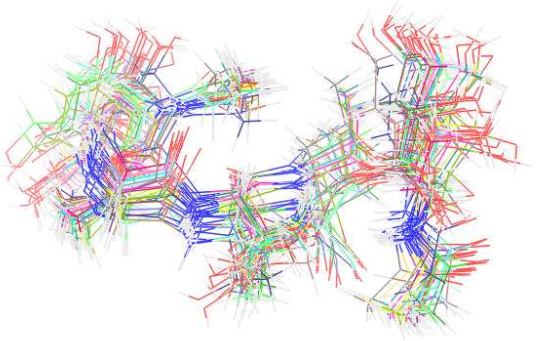
Backbone Heavy Atoms

+Thr+GalNAc Heavy atoms

1.262 +/- 0.296

All Heavy Atoms

1.794 +/- 0.379



F Structural Information PDB ID 2LHY

Distance Restraints

Peptide-Peptide		
	Intra-residue:	51
	Inter-residue:	22
Peptide-Sugar		
	Proximal:	21
	Non-Proximal:	13
Sugar-Sugar		
	GalNAc-GalNAc	15
Total:		
122		
^3J Coupling Restraints:		5
Torsion Restraints:		9

Restraint Violations

NOE Violations (>.25 Å)	0
J-Coup Violations (>.5 Hz)	0
Torsion Violations (>5°)	0

Deviations From Ideal Geometry

Bond (Å):	0.00187 +/- 0.00006
Angle (°):	0.442 +/- 0.011
Improper (°)	0.665 +/- 0.014

Average RMSD Over 24 Structures

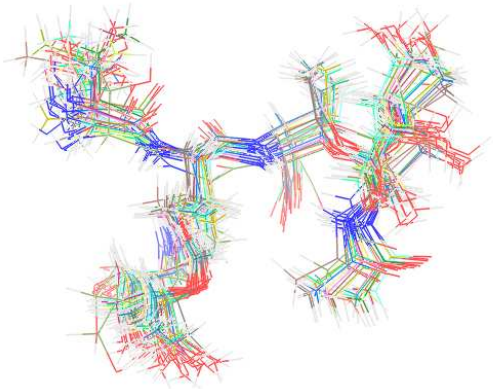
Backbone Heavy Atoms

+Thr+GalNAc Heavy atoms

1.173 +/- 0.237

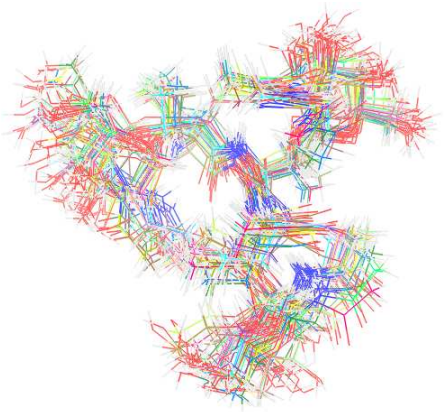
All Heavy Atoms

1.787 +/- 0.408



G Structural Information PDB 2LHX

Distance Restraints		
	Peptide-Peptide	
		Intra-residue: 36
		Inter-residue: 14
	Peptide-Sugar	
		Proximal: 27
		Non-Proximal: 14
	Sugar-Sugar	
		GalNAc-GalNAc 25
	Total:	116
^3J Coupling Restraints:		5
Torsion Restraints:		11
Restraint Violations		
	NOE Violations (>.25 Å)	0
	J-Coup Violations (>.5 Hz)	0
	Torsion Violations (>5°)	0
Deviations From Ideal Geometry		
	Bond (Å):	0.00179 +/- 0.00010
	Angle (°):	0.423 +/- 0.014
	Improper (°)	0.363 +/- 0.016
Average RMSD Over 54 Structures		
	Backbone Heavy Atoms	
	+Thr+GalNAc Heavy atoms	
		1.250 +/- 0.262
	All Heavy Atoms	
		1.664 +/- 0.417



*Some of the sugar-sugar NOEs are between protons attached to members of the sugar 6 atom skeleton that was held fixed in the calculations

Table S2 RDC Values, accuracy +/- 0.2 Hz				
construct	D	E	F	G
P1 HαCα	-7.8	-4.9	-4.6	-5.3
T2 HαCα	-10.8*	-7.7*	-14.5	-10.0*
T3 HαCα	-10.6*	-9.0	-15.4*	-10.5*
T4 HαCα	-7.2	-8.0*	-9.5*	-6.6*
P5 HαCα	-5.8	-0.7	-2.5	-3.2
L6 HαCα	-8.5	-7.2	-11.4	-7.9
K7 HαCα	-6.7	-4.9	-10.7	-7.2
T2 HβCβ	8.6*	5.2*	< -0.5	7.4*
T2 γ2ME	2.9*	2.2*	8.9	4.1*
T3 HβCβ	8.8*	4.3	15.2*	11.1*
T3 γ2ME	3.9*	3.6	4.9*	4.3*
T4 HβCβ	1.0	11.1*	13 9*	8.4*
T4 γ2ME	5.6	1.0*	2.6*	0.9*
G9 H1C1	6.4	5.3	N/A	7.1
GN10 H1C1	2.9	N/A	9.0	6.7
GN11 H1C1	N/A	4.4	0.5	2.0
T2 HN	-6.4*	-5.9*	-7.3	-5.2*
T3 HN	-8.8*	-7.3	-12.1*	-8.3*
T4 HN	-7.0	-8.5*	-13.1*	-9.2*
L6 HN	-4.2	-5.0	-7.0	-6.7
K7 HN	-6.1	-5.5	-7.8	-5.7
GN9 HN	6.0	3.9	N/A	4.4
GN10 HN	7.6	N/A	5.7	7.6
GN11 HN	N/A	3.9	7.8	7.8
*=GalNAc Glycosylation				

Supporting Information: Experimental Procedures

Glycopeptide Immobilization

The MUC2 and alpha-dystroglycan related constructs were immobilized through coupling of side chain amine of the C-terminal lysine residue to the NHS-functionalized slide. The Ac-T-(α-*O*-GalNAc)-NH-$(CH_2)_3$-NH_2 and Ac-[T-(α-*O*-GalNAc)]$_3$-NH-$(CH_2)_3$-NH_2 (Chart IDs 19 and 20) were attached through the propylamine linker, and the other structures were coupled through the peptide *N*-terminal amino group.

Vaccination Protocol

Patients were vaccinated subcutaneously three times at one week intervals, once four weeks later and then once or twice more at three month intervals as described earlier study.[2] The vaccines were well tolerated. Serum samples were drawn at regular intervals and tested for reactivity against MUC1 and Tn. Peak titer sera were utilized in the analyses.

Synthesis of Ac-PTTDSTTPAPTTK-NH_2, Ac-PTTDSTT(α-D-GalNAc)PAPTTK-NH_2, (EA2) TSAPDTRPAP-NH_2, and TSAPDT(α-D-GalNAc)RPAP-NH_2 (MUC1-1)

Starting with Fmoc-PAL-PEG-PS resin (250 mg, 0.20 mmol/g), the peptides were assembled on bench-top facility in a glass vessel (5 ml) containing porous polypropylene frits. Fmoc-amino acids (4 equiv) were coupled for 2 h in the presence of HCTU (4 equiv), HOBt (4 equiv) and DIEA (6 equiv) in DMF at 25ºC. Fmoc-Thr(Ac_3-α-D-GalNAc)-OH (1.5 equiv) was double coupled for 4 h in the presence of HCTU (1.5 equiv), HOBt (1.5 equiv) and DIEA (2 equiv) in NMP at 25ºC. The following Fmoc-Thr(^{t}Bu)-OH (4 equiv) was double coupled. Fmoc removal was achieved with piperidine–DMF (1:4) for 20 min. Washings between reactions were carried out with DMF and CH_2Cl_2, and no intermediate capping steps were done. After complete chain assembly, *N*-acetylation was achieved by treatment with Ac_2O–DMF (1:4) for 20 min. Cleavage of the peptides was achieved by treatment with TFA–H_2O (19:1) and precipitated in cold ether (100 ml). Analytical RP-HPLC was performed on an Agilent system with detection at 220 nm. Samples were chromatographed at 1.0 mL/min using linear gradients of 0.1% aqueous TFA (buffer A) and 0.1% TFA in CH_3CN (buffer B), from 0 to 40% buffer B over 40 min. Crude peptides and glycopeptides were purified by semi-preparative RP-HPLC on a Agilent system, using gradients which varied according to the properties of the particular sequence, and detection at 220 nm. After purification, the glycopeptides were treated with NaOMe in MeOH (pH ~9, as detected by wet litmus paper) for 6 h as monitored by analytical RP-HPLC, followed by semi-preparative HPLC of the fully deprotected material. Fractions with the desired peptides or glycopeptides were combined and lyophilized. Yields of peptides: ~60%. Yields of glycopeptides: ~ 40%. Yields were calculated based on the loading of the resin used. ESI-MS: Ac-PTTDSTTPAPTTK-NH_2, 679.8 $[M+2H]^{2+}$, 690.8 $[M+H+Na]^{2+}$, t_R = 16.0 min; Ac-PTTDSTT(α-D-GalNAc)PAPTTK-NH_2, 800.3 $[M+H+K]^{2+}$, 803.3 $[M+2Na]^{2+}$, t_R = 13.9 min. MALDI-TOF: H-TSAPDTRPAP-NH_2, 1011.6 $[M+H]^+$, t_R = 12.4 min; H-TSAPDT(α-D-GalNAc)RPAP-NH_2, 1340.5 $[M+H]^+$. t_R = 19.3 min.

Synthesis of APGSTAPP-NH_2 and APGS(α-D-GalNAc)T(α-D-GalNAc)APP-NH_2 (MUC1-2)

Starting with Rink Amide AM (180 mg, 0.36 mmol/g), the first three amino-acid derivatives were assembled on an ABI 433A peptide synthesizer, by sequential couplings of Fmoc-amino acid derivatives (4 equiv) using standard protocol in the presence of HCTU (4 equiv), HOBt (4 equiv) and DIEA (8 equiv). Then, the peptide APP-resin (50 mg) was elongated by double couplings of Fmoc-Thr(Ac_3-α-D-GalNAc)-OH (1.5 equiv) and Fmoc-Ser(Ac_3-α-D-GalNAc)-OH (1.5 equiv) manually for 12 h in the presence of HATU (1.5 equiv), HOBt (1.5 equiv) and DIEA (2 equiv) in DMF at 25ºC respectively. The remaining three amino-acid derivatives (4 equiv) were coupled manually. The peptide was cleaved from resin by TFA–H2O (19:1, 2 ml) and precipitated in cold ether (100 ml). After purification, the peptide APGS(Ac_4-α-D-GalNAc)T(Ac_4-α-D-GalNAc)APP-NH_2 was treated with NaOMe in MeOH (pH ~9, as detected by wet litmus paper) for 6 h as monitored by analytical RP-HPLC. The target product was obtained by purification on semi-preparative RP-HPLC. The structure was confirmed by MALDI-TOF. After synthesis on ABI 433A peptide synthesizer, peptide APGSTAPP was cleaved from the resin by TFA/H_2O/triisopropylsilane (38:1:1), followed by purification on semi-preparative RP-HPLC.

Structure Calculation Protocol

Starting from an extended structure, peptide backbone and glycosidic linkage torsion angles were randomized and the structure was then subjected to an initial minimization consisting of four iterations of 2000 steps each. The first iteration included only bond length and angle terms, the second iteration added torsion terms, the third added improper torsion terms and the fourth iteration added a repulsive van der Waals term. After initial minimization, the van der Waals term was turned off, and NMR derived distance and torsion restraints were activated. The system was then warmed to 50,000 K over 2 ps, which was followed by 30ps of dynamics and 32ps of dynamics with ramped repulsive van der Waals. This was followed by 16ps of slow cooling to 0 K and 2000 steps of final Cartesian minimization. Electrostatic terms were turned off for the duration of the

calculations. All simulated annealing dynamics stages utilized a time-step of 2 fs. An ensemble of 100 structures was generated for each variant, which were then evaluated for restraint violations and deviations from ideal geometry based on the following criteria: no distance restraint violations greater than 0.25Å, no backbone torsion restraint violations greater than 5.0°, no bond length constraint violations greater than 0.05Å and no angle or improper torsion constraint violations greater than 5.0º. All structure determination calculations utilized Torsion Angle Dynamics[3] along with the Internal Variable Module[4]. Throughout the calculations, proline and GalNAc rings were held rigid, with the GalNAc rings held in the chair conformation.

Supporting Information References

(1) Piotto M, Saudek V, Sklenar V. (1992) Gradient-Tailored Excitation for Single-Quantum NMR-Spectroscopy of Aqueous-Solutions. *J Biomol NMR 2*, 661-665.

(2) T., Ragapathi, G., Bhuta, S., Clausen, H., Norton, L., Houghton, A. N., and Livingston, P. O. (2001) Preliminary data: Vaccination with glycosylated mucin1-keyhole limpet hemocyanin conjugate plus the immunological adjuvant qs-21 in breast cancer patients, *Proceedings of the American Society of Clinical Oncology 20*, 271a.

(3) Stein, E.G., Rice, L.M. and Brunger, A.T. (1997) Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *J. Magn.. Reson*, *124*, 154-164.

(4) Schwieters, C.D. and Clore, G.M. (2001) Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *J. Magn. Reson*., *152*, 288-302.