



HOKKAIDO UNIVERSITY

Title	Significance of the Posttranslational O-glycosylations in the Protein Conformation and Function
Author(s)	早川, 瞬
Degree Grantor	北海道大学
Degree Name	博士(生命科学)
Dissertation Number	甲第12268号
Issue Date	2016-03-24
DOI	https://doi.org/10.14943/doctoral.k12268
Doc URL	https://hdl.handle.net/2115/64870
Type	doctoral thesis
File Information	Shun_Hayakawa.pdf



Significance of the Posttranslational O-glycosylations in the Protein Conformation and Function

(タンパク質立体構造および機能に関する O-結合型糖鎖修飾の意義)

Doctoral Thesis

2016

Shun HAYAKAWA

Laboratory of Advanced Chemical Biology

Graduate School of Life Science, Hokkaido University

Table of Contents

Table of Contents.....	0
Abbreviations	4
Chapter 1 General Introduction	8
1-1. Protein Glycosylation.....	9
1-2. Organic Chemistry for Glycobiology.....	14
1-3. Reference.....	16
Chapter 2 Consecutive O-GalNAc Modifications on Mucin Glycoproteins.....	20
2-1. Introduction	21
2-2. Result and Discussion	24
2-2-1. Sequence analysis of mucin tandem repeat and design of mucin-model glycopeptides containing TTX motif.....	24
2-2-2. Synthesis of 9-mer model glycopeptides and 19-mer MUC2 glycopeptides	28
2-2-3. NMR spectral analysis of synthetic glycopeptides	31
2-2-4. NMR structural analyses of 9-mer mucin model glycopeptides.....	39
2-2-5. NMR structure analyses of MUC2 glycopeptides (5-7).	50
2-2-6. Auto-antibody analysis against MUC2 glycopeptides in cancer patients	64
2-3. Conclusion.....	69
2-4. Experimental Section	71
General methods and materials.....	71
Solid-phase synthesis of glycopeptides (1-6).	73
Enzymatic synthesis of sialyl glycopeptide (7).	74
Nuclear magnetic resonance (NMR) spectroscopic experiments.....	75
Structure Calculation	76
Microarray analyses of synthetic MUC2 glycopeptides	76
2-5. Reference.....	78
Chapter 3 O-Fucosylation and O-Glucosylation on Human Notch1 EGF12	84
3-1. Introduction	85
3-2. Result and Discussion	88

3-2-1. Design of human Notch1 EGF12 glycopeptide analogues	88
3-2-2. Synthesis of Fmoc-Ser[Ac ₃ Xyl(α 1 $\rightarrow$ 3)Ac ₂ Xyl(α 1 $\rightarrow$ 3)Ac ₃ Glc β 1 $\rightarrow$]-OH.	90
3-2-3. Synthesis of human Notch1 EGF12 glycopeptide analogues	93
3-2-4. Glutathione-mediated folding reaction	94
3-2-5. Enzymatic sugar elongation toward total synthesis of fully modified EGF12	106
3-2-6. NMR-based conformational analysis of EGF12 analogues.....	108
3-2-7. Cellular assay of Synthetic Notch EGF domain	131
3-3. Conclusion.....	136
3-4. Experimental Section	137
Reagents and General Methods	137
Solid Phase Peptide Synthesis of EGF12 analogues	138
Disulfide bond formation of EGF analog.....	140
Thermolysin digestion	141
Enzymatic carbohydrate elongation	141
Nuclear magnetic resonance (NMR) spectroscopic experiment	142
Structure Calculation	143
Synthesis of Fmoc-Ser[Ac ₃ Xyl α (1 $\rightarrow$ 3)Ac ₂ Xyl α (1 $\rightarrow$ 3)Ac ₃ Glc β 1 $\rightarrow$]-OH (21)..	144
3-5. Reference.....	159
Chapter 4 Concluding Remarks.....	166
Acknowledgement	169

Abbreviations

Ac	acetyl
ACN	acetonitrile
Bn	benzyl
Bu	butyl
CMP	cytidine monophosphate
COSY	correlation spectroscopy
DBU	1,8-diazabicyclo[5.4.0]-7-undecene
DCM	dichloromethane
DHB	dihydrobenzoic acid
DMAP	<i>N,N</i> -dimethyl-4aminopyridine
DMF	<i>N,N</i> -dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDT	ethanedithiol
EDTA	ethylenediaminetetraacetic acid
ESI	electro spray ionization
Et	ethyl
EtOAc	ethyl acetate
Fmoc	9-fluorenylmethyloxycarbonyl
Fuc	fucose
Gal	galactose
GalNAc	<i>N</i> -acetyl galactosamine

Glc	glucose
GlcNAc	<i>N</i> -acetyl glucosamine
GnF	GlcNAc β 1 $\rightarrow$ 3Fuc α 1 $\rightarrow$
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
HBTU	2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOAt	1-hydroxy-7-azabenzotriazole
HOBt	1-hydroxybenzotriazole
HPLC	high performance liquid chromatography
HSQC	hetero-nuclear multiple quantum coherence
m/z	mass/charge ratio
MALDI	matrix assisted laser desorption ionization
Me	methyl
MS	mass
MS4A	molecular sieve 4A
Neu5Ac	<i>N</i> -acetyl neuraminic acid
NIS	<i>N</i> -iodosuccinimide
NMR	nuclear magnetic resonance
NOE	nuclear overhauser effect
NOESY	NOE correlated spectroscopy
PyBOP	(Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate

Sia	sialic acid
SPPS	solid phase peptide synthesis
TEA	triethyl amine
TFA	trifluoroacetic acid
TfOH	trifluoromethanesulfonic acid
THF	tetrahydrofuran
TIS	triisopropylsilane
TLC	thin-layer chromatography
TMSOTf	Trimethylsilyl trifluoromethanesulfonate
TOCSY	totally correlated spectroscopy
TOF	time of flight
Tris	<i>tris</i> (hydroxymethyl)aminomethane
XXG	Xyl α 1 $\rightarrow$ 3Xyl α 1 $\rightarrow$ 3Glc β 1 $\rightarrow$
Xyl	xylose

Chapter 1

General Introduction

1-1. Protein Glycosylation

Post-translational modification is a chemical modification of protein, which is DNA-independent manner. Protein glycosylation is known as one of the most important post-translational modifications because over 50% of proteins are found to be glycosylated.¹ These glycosylations are performed by specific glycosyltransferase and sugar nucleotides as glycosyl donor in ER, Golgi apparatus or cytoplasm. These DNA-independent carbohydrate modifications produce protein structural diversity in terms of heterogeneous glycosylation sites as well as glycan structure, which generates protein functional diversity (**Figure 1-1**).²

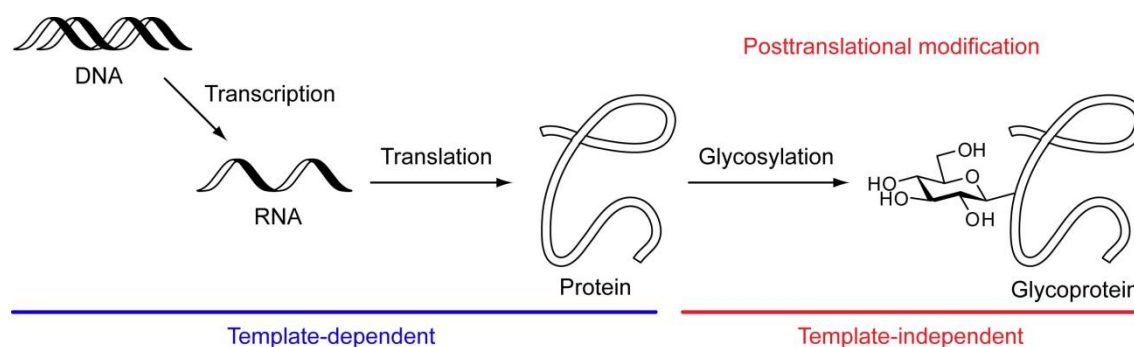
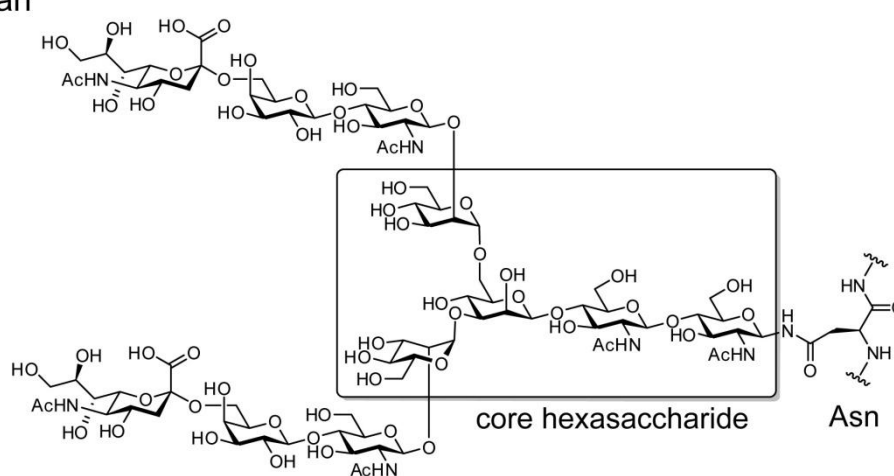


Figure 1-1. Protein expression and glycosylation

There are two major types of glycan structure, *N*-glycan and *O*-glycan. *N*-glycans are covalently attached to nitrogen atom on asparagines residue to form GlcNAc β 1 $\rightarrow$ Asn, where Asn locates in Asn-X-Thr consensus sequence (X: any amino acid residue other than Pro).³ *N*-glycans have pentasaccharide core structure, and it is majorly classified into complex, hybrid and high mannose type according to their branching and terminal structure. On the other hand, *O*-glycan is linked to Ser or Thr residue via oxygen atom, and various sugar residues are found to be attached. GalNAc α 1 $\rightarrow$ Thr/Ser modification is one of the most abundant *O*-glycosylations, which is mainly found in mucin protein.⁴

In addition, GlcNAc β 1 $\rightarrow$, Fuc β 1 $\rightarrow$, Man α 1 $\rightarrow$, Glc β 1 $\rightarrow$ and Xyl β 1 $\rightarrow$ modifications at Thr/Ser have been found in several proteins (**Figure 1-2**).⁵⁻⁸ These O-glycans are further elongated by specific sugar residues, producing heterogeneous glycan structure. For example, GalNAc α 1 $\rightarrow$ Thr/Ser can be further glycosylated to form Gal β 1 $\rightarrow$ 3GalNAc α 1 $\rightarrow$, which is called as core1 O-glycan.⁹ In this context, O-glycans would have more complex structure than N-glycan in terms of structural diversity.

N-glycan



O-glycan

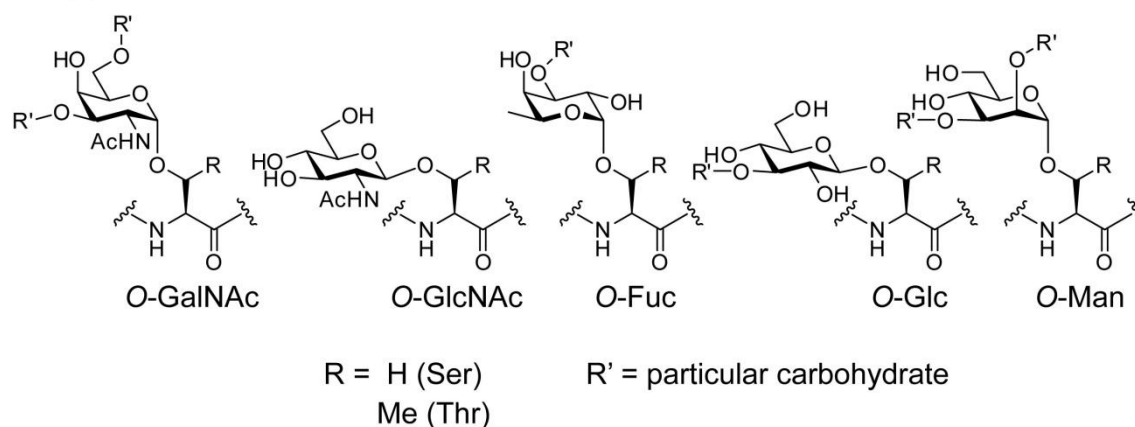


Figure 1-2. Chemical structure of typical N- and O-glycans

These *N*- and *O*-glycans are related to various biological phenomena involving cell-cell/protein-protein interaction, infection, immunity, quality control of protein expression, enhancement of protein stability, conformational alteration of protein backbone, and so on.¹⁰ Aberrant glycosylation on proteins as well as aberrant expression of glycoproteins relate to several diseases, such as cancer, leukemia, autoimmune disease etc.¹¹⁻¹³ Especially, hyper-branched *N*-glycans and truncated *O*-glycans have been found to be expressed in several cancer cells. These tumor-associated carbohydrate antigens have been applied as bio-markers of cancer diagnosis and therapy. Therefore, glycoproteins have great potential to therapeutic applications to several diseases.¹⁴

The protein biological functions can be modulated by glycan structure, but “how does glycosylation affect protein functions?” Investigation of molecular mechanism of glycosylation effects on protein functions could be important for application of glycopeptide towards diagnosis or therapy for several diseases including cancer.

Anti-freeze glycoprotein (AFGP) acquires anti-freeze property after disaccharide modifications, Gal β 1 $\rightarrow$ 3GalNAc α 1 $\rightarrow$, at threonine residue on repeating triplet (Thr-Ala-Ala). In prior research, chemical synthesis and NMR conformational analysis of AFGP demonstrated that synthetic *O*-glycosylated AFGP adopted polyproline type II helix structure in which disaccharide moieties are on the same side of the molecule.^{15,16} This glycosylation-induced amphiphilic structure is presumed to be crucial for the AFGP specific activity (**Figure 1-3A**). In addition, NMR study of MUC1 glycopeptide revealed that glycan structure affected not only antibody recognition but also peptide backbone conformation.¹⁷ Pro-Asp-Thr-Arg-Pro-Ala-Pro region of MUC1 tandem repeat carrying core 1 glycan (Gal β 1 $\rightarrow$ 3GalNAc α 1 $\rightarrow$) at Thr residue has found to be minimal epitope against anti-KL6 monoclonal antibody which is applied for diagnosis

of interstitial pneumonia, lung adenocarcinoma, breast cancer, colorectal adenocarcinoma, and hepatocellular carcinoma etc. Chemical synthesis and NMR conformational analysis of MUC1 glycopeptide derivatives revealed that disialylated core 2 [Neu5Ac α 2 $\rightarrow$ 3Gal β 1 $\rightarrow$ 3(Neu5Ac α 2 $\rightarrow$ 3Gal β 1 $\rightarrow$ 4GlcNAc β 1 $\rightarrow$ 6)GalNAc α 1 $\rightarrow$] gave conformational flexibility between two conformers of γ -turn and β -turn on PDTR region while naked and core 2-modified PDTR regions adopt γ -turn and β -turn, respectively. Thus, it is suggested that anti-KL6 mAb recognize carbohydrate structure and conformational alteration of peptide backbone induced by *O*-glycosylation at PDTR region (**Figure 1-3B**).

These results clearly indicate that regulation of protein function as well as antigenicity is demonstrated by glycosylation-induced conformational alteration. Therefore, investigation of glycosylation impacts on peptide backbone would be important for understanding of molecular mechanism of glycobiology. Furthermore, investigation of conformational motif which represents specific glycan-induced 3D-structure can become good probe to develop therapeutic as well as diagnostic methods.

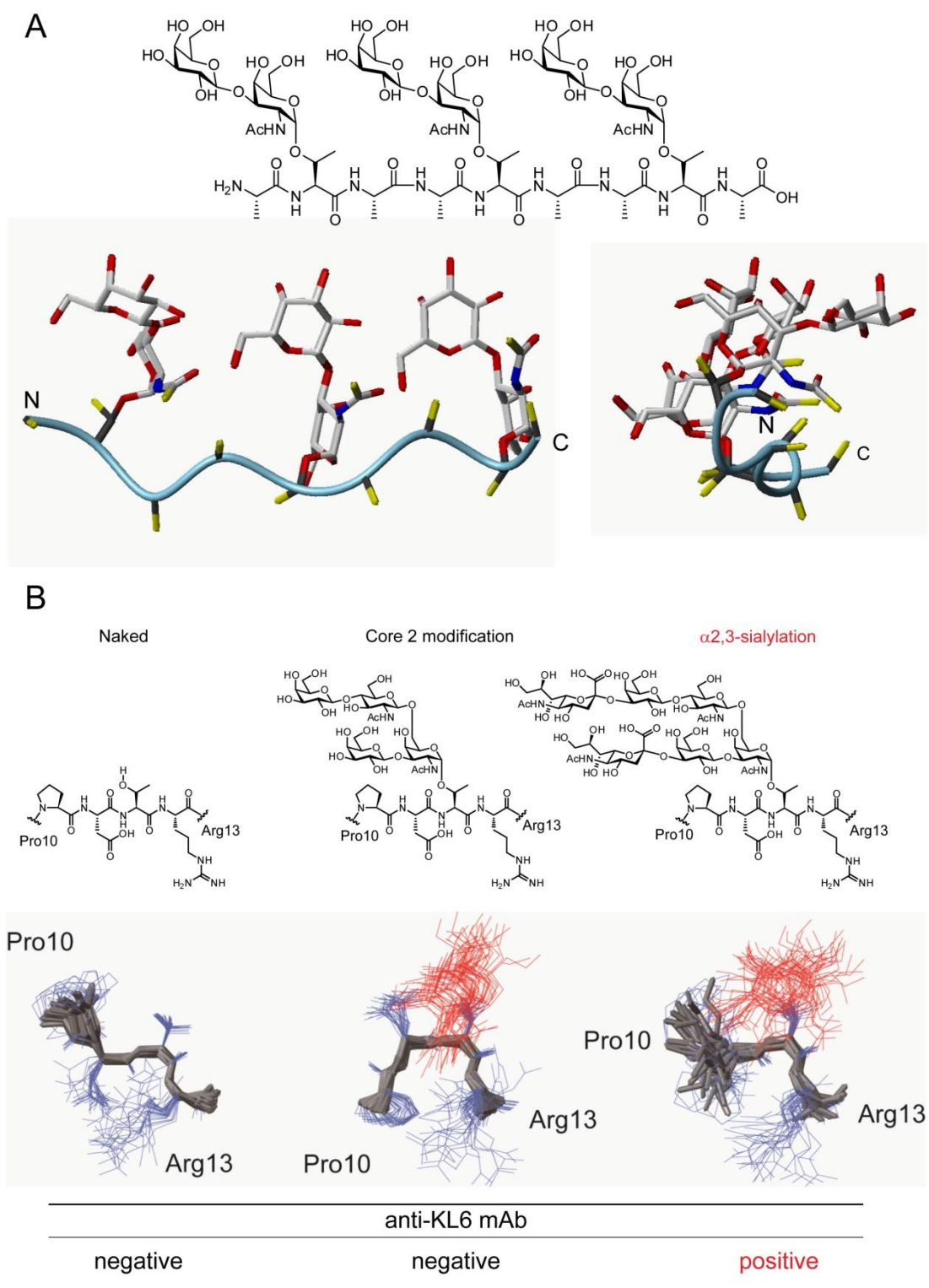


Figure 1-3. Glycan-induced conformational alteration. (A) synthetic AFGP adopts poly proline type-II helix. (B) PDTR sequence of MUC1 showed γ - and β -turn during glycosylation.

1-2. Organic Chemistry for Glycobiology

Organic chemistry of glycopeptide has been developed for powerful tool to understand the molecular mechanism of *O*-glycosylation impacts on conformational alteration as well as functional alteration because homogeneous glycopeptide is required to investigate functions of individual *O*-glycans.¹⁸⁻²² Unlike the protein expressed with DNA-dependent manner, the difficulties in the production of structurally-well-defined glycopeptide often limit the proceeding of glycobiology. Since protein glycosylation is DNA-independent modification, biological approach will give heterogeneous glycopeptides in terms of glycan structure as well as glycosylation site.

To overcome this barrier, organic chemical approach to efficiently afford structurally-defined glycopeptide should be necessary.^{15-17,23,24} In addition, this approach can modify some specific functional group on glycopeptide such as keton-modification required for microarray immobilization which facilitate functional study of glycopeptides. In present thesis, I demonstrated organic chemical approach of glycobiology focusing on (1) establishment of efficient synthetic methodology to obtain structurally-well-defined glycopeptides, (2) NMR-based conformational analysis of synthetic glycopeptides to find conformational motif/modules induced by glycosylation on peptide backbone, and (3) investigation of biological function of glycan-induced conformational motifs with microarray analysis as well as cellular assay. These strategies will provide important relationship between glycan-induced conformational alteration and biological functions of glycopeptides, which can be applicable to therapeutic target of glycosylation-associated diseases.

In this thesis, general *O*-GalNAc-modified glycopeptide and unusual *O*-fucose- as well as *O*-glucose-modified glycopeptide are focused as target molecule to investigate

conformational and functional impacts.

In chapter 2, mucin glycopeptide carrying consecutive *O*-GalNAc modification is focused as conformational-defined motif. Consecutive two *O*-GalNAc modifications are abundantly found in mucin glycoprotein. NMR-based structural analysis revealed that synthetic glycopeptide *O*-glycosylated T*T*X triplets (which * indicates glycosylated residues) is key motif to represent important conformational features which strongly restrict the peptide conformation. Microarray analysis of synthetic MUC2 glycopeptide demonstrated that cancer patient sera showed significant level of auto-antibody recognition and glycosylation-dependent affinity alteration.

In chapter 3, impacts of unusual *O*-glycans such as *O*-fucosylation and *O*-glucosylation on human Notch1 EGF12 domain are discussed. These glycosylations are known to regulate Notch signaling pathway which has crucial roles for cellular development. Chemical synthesis of 38-mer human Notch1 EGF12 glycopeptide carrying *O*-fucose and *O*-glucose glycan with three disulfide bonds were carried out by microwave-assisted solid-phase peptide synthesis and Ca^{2+} -assisted folding reaction. Three-dimensional NMR structure of *O*-glycosylated human Notch1 EGF12 demonstrated that *O*-fucose and *O*-glucose have different effects on peptide conformational stabilization. Preliminary cellular assay with synthetic EGF modules were demonstrated that these conformational modules affect cancer cell viability.

1-3. Reference

- (1) Varki, a. Biological Roles of Oligosaccharides: All of the Theories Are Correct. *Glycobiology* **1993**, 3, 97–130.
- (2) Lis, H.; Sharon, N. Protein Glycosylation. Structural and Functional Aspects. *Eur. J. Biochem.* **1993**, 218, 1–27.
- (3) Kornfeld, R.; Kornfeld, S. Assembly of Asparagine-Linked Oligosaccharides. *Annu. Rev. Biochem.* **1985**, 54, 631–664.
- (4) Ten Hagen, K. G.; Fritz, T. T. A.; Tabak, L. A; Hagen, K. Ten. All in the Family: The UDP-GalNAc: Polypeptide N-Acetylgalactosaminyltransferases. *Glycobiology* **2003**, 13, 1–16.
- (5) Taylor, M. E.; Drickamer, K. Introduction to Glycobiology, Oxford University Press, **2011**
- (6) Endo, T. O-Mannosyl Glycans in Mammals. *Biochim. Biophys. Acta* **1999**, 1473, 237–246.
- (7) Harris, R. J.; Spellman, M. W. O-Linked Fucose and Other Post-Translational Modifications Unique to EGF Modules. *Glycobiology* **1993**, 3, 219–224.
- (8) Wells, L.; Hart, G. W. O-GlcNAc Turns Twenty: Functional Implications for Post-Translational Modification of Nuclear and Cytosolic Proteins with a Sugar. *FEBS Lett.* **2003**, 546, 154–158.
- (9) Tarp, M. A.; Clausen, H. Mucin-Type O-Glycosylation and Its Potential Use in Drug and Vaccine Development. *Biochim. Biophys. Acta.* **2008**, 1780, 546–563.
- (10) Varki, a. Biological Roles of Oligosaccharides: All of the Theories Are Correct. *Glycobiology* **1993**, 3, 97–130.
- (11) Brockhausen, I. Pathways of O-Glycan Biosynthesis in Cancer Cells. *Biochim. Biophys. Acta.* **1999**, 1473, 67–95.
- (12) Kannagi, R.; Izawa, M.; Koike, T.; Miyazaki, K.; Kimura, N. Carbohydrate-Mediated Cell Adhesion in Cancer Metastasis and Angiogenesis. *Cancer Sci.* **2004**, 95, 377–384.
- (13) Hollingsworth, M. A.; Swanson, B. J. Mucins in Cancer: Protection and Control of the Cell Surface. *Nat. Rev. Cancer* **2004**, 4, 45–60.

- (14) Yarema, K. J.; Bertozzi, C. R. Chemical Approaches to Glycobiology and Emerging Carbohydrate-Based Therapeutic Agents. *Curr. Opin. Chem. Biol.* **1998**, *2*, 49–61.
- (15) Tachibana, Y.; Fletcher, G. L.; Fujitani, N.; Tsuda, S.; Monde, K.; Nishimura, S.-I. Antifreeze Glycoproteins: Elucidation of the Structural Motifs That Are Essential for Antifreeze Activity. *Angew. Chemie Int. Ed.* **2004**, *43*, 856–862.
- (16) Izumi, R.; Matsushita, T.; Fujitani, N.; Naruchi, K.; Shimizu, H.; Tsuda, S.; Hinou, H.; Nishimura, S.-I. Microwave-Assisted Solid-Phase Synthesis of Antifreeze Glycopeptides. *Chem. Eur. J.* **2013**, *19*, 3913–3920.
- (17) Matsushita, T.; Ohyabu, N.; Fujitani, N.; Naruchi, K.; Shimizu, H.; Hinou, H.; Nishimura, S.-I. Site-Specific Conformational Alteration Induced by Sialylation of MUC1 Tandem Repeating Glycopeptides at an Epitope Region for the Anti-KL-6 Monoclonal Antibody. *Biochemistry* **2013**, *52*, 402–414.
- (18) Ohyabu, N.; Hinou, H.; Matsushita, T.; Izumi, R.; Shimizu, H.; Kawamoto, K.; Numata, Y.; Togame, H.; Takemoto, H.; Kondo, H.; Nishimura, S.-I. An Essential Epitope of Anti-MUC1 Monoclonal Antibody KL-6 Revealed by Focused Glycopeptide Library. *J. Am. Chem. Soc.* **2009**, *131*, 17102–17109.
- (19) Buskas, T.; Thompson, P.; Boons, G.-J. Immunotherapy for Cancer: Synthetic Carbohydrate-Based Vaccines. *Chem. Commun.* **2009**, *36*, 5335–5349.
- (20) Vergati, M.; Intrivici, C.; Huen, N.-Y.; Schlom, J.; Tsang, K. Y. Strategies for Cancer Vaccine Development. *J. Biomed. Biotechnol.* **2010**, *2010*, 596432.
- (21) Matsushita, T.; Takada, W.; Igarashi, K.; Naruchi, K.; Miyoshi, R.; Garcia-Martin, F.; Amano, M.; Hinou, H.; Nishimura, S.-I. A Straightforward Protocol for the Preparation of High Performance Microarray Displaying Synthetic MUC1 Glycopeptides. *Biochim. Biophys. Acta.* **2014**, *1840*, 1105–1116.
- (22) Martínez-Sáez, N.; Castro-López, J.; Valero-González, J.; Madariaga, D.; Compañón, I.; Somovilla, V. J.; Salvadó, M.; Asensio, J. L.; Jiménez-Barbero, J.; Avenoz, A.; Busto, J. H.; Bernardes, G. L.; Peregrina, J. M. Hurtado-Guerrero, R.; Corzana, F. Deciphering the Non-Equivalence of Serine and Threonine O -Glycosylation Points: Implications for Molecular

- Recognition of the Tn Antigen by an Anti-MUC1 Antibody. *Angew. Chemie Int. Ed.* **2015**, *54*, 9830–9834.
- (23) Hashimoto, R.; Fujitani, N.; Takegawa, Y.; Kurogochi, M.; Matsushita, T.; Naruchi, K.; Ohyabu, N.; Hinou, H.; Gao, X. D.; Manri, N.; Satake, H.; Kaneko, A.; Sakamoto, T.; Nishimura, S.-I.. An Efficient Approach for the Characterization of Mucin-Type Glycopeptides: The Effect of O-Glycosylation on the Conformation of Synthetic Mucin Peptides. *Chem. Eur. J.* **2011**, *17*, 2393–2404.
- (24) Hiruma-Shimizu, K.; Hosoguchi, K.; Liu, Y.; Fujitani, N.; Ohta, T.; Hinou, H.; Matsushita, T.; Shimizu, H.; Feizi, T.; Nishimura, S.-I. Chemical Synthesis, Folding, and Structural Insights into O-Fucosylated Epidermal Growth Factor-like Repeat 12 of Mouse Notch-1 Receptor. *J. Am. Chem. Soc.* **2010**, *132*, 14857–14865.

Chapter 2

***Consecutive O-GalNAc Modifications
on Mucin Glycoproteins***

2-1. Introduction

O-GalNAc modification is one of the most abundant O-glycosylation found mainly in mucin glycoprotein, which is heavily glycosylated protein produced by epithelial cells.¹ For this reason, O-GalNAc modification is called as mucin-type glycosylation. The construction of matured O-GalNAc modification is achieved by multiple enzymatic glycosylation. First, GalNAc residue is transferred from UDP-GalNAc to oxygen atom of Thr or Ser side chain by polypeptides *N*-acetylgalactosaminyltransferase (ppGalNAc-Ts) *via* α -linkage. This GalNAc α 1 $\rightarrow$ Ser/Thr is the major basal structure, which abundantly exist in the mucin-type glycoproteins. Then, Gal, GlcNAc and Neu5Ac residues are transferred by specific glycosyltransferases in Golgi apparatus and ER (endoplasmic reticulum) to form various core subtypes. In particular, the core 1 subtype (Gal β 1 $\rightarrow$ 3GalNAc α 1 $\rightarrow$ Ser/Thr) can be recognized as a key structure to organize core mucin scaffolds including *N*-acetylglucosamine repeats *via* formation of core 2 branched subtype (GlcNAc β 1 $\rightarrow$ 6(Gal β 1 $\rightarrow$ 3)GalNAc α 1 $\rightarrow$ Ser/Thr, see **Figure 2-1**).² These core structures provide a lot of biological aspects involving differentiation, cell-cell recognition, inflammation, immune response, and tumor metastasis through further modification processes.³

In tumor cells, mucin glycoproteins are notably over expressed, and higher expression levels of unusual glycosylation with aberrantly truncated core-type glycans are found in mucin O-glycosylation. Especially, O-GalNAc, core 1, and sialyl core 1 type glycans (Neu5Ac α 2 $\rightarrow$ 3Gal β 1 $\rightarrow$ 3GalNAc α 1 $\rightarrow$ Ser/Thr) are known as Tn-, T-, and sialyl T (ST) antigen, and these truncated carbohydrates are considered as tumor-associated carbohydrate antigen (**Figure 2-1**).⁴ These tumor-associated truncated glycan might

expose core peptide backbone, inducing enhancement of antibody reactivity toward human tumor cells.

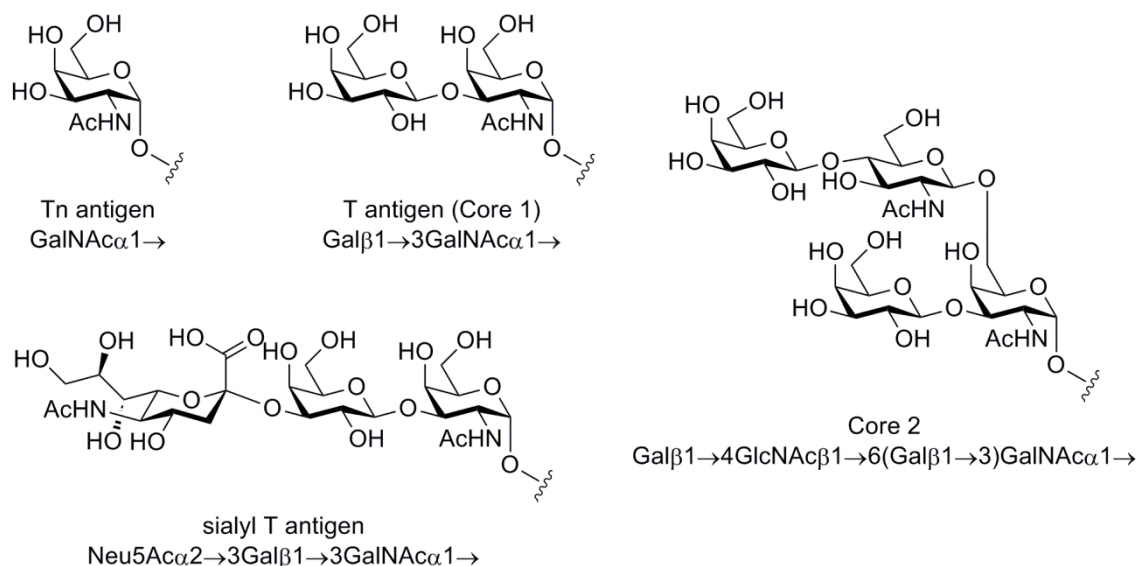


Figure 2-1. Typical O-GalNAc isoform and tumor associated carbohydrate antigen

The reactivity of monoclonal antibodies against human tumor cells inferred the presence of not only dispersed but also clustered O-glycans antigens on the cancer-related mucins, which are appeared on colonic mucosa as clustered dimmers and trimmers with transformation to malignancy.⁵⁻⁷ Some monoclonal antibodies raised against a human colorectal cancer line LS 180 cells especially indicated the pronounced reactivity toward the glycopeptide epitopes including three or four consecutive residues of Ser/Thr carrying carbohydrate antigens such as Tn and sialyl Tn.⁸⁻¹¹ These high immunogenic motifs could be synthesized as the conjugates with protein carriers, which have proven to be useful for exploiting vaccine candidates in monomer or cluster approach.¹²⁻¹⁶

Despite of the growing importance of mucin glycopeptide structural information, there are a few attempts to elucidate conformational features of mucin glycopeptides. In this study, we focused on TTX motif as a key representing typical mucin structural features. To elucidate the mucin conformational features, we demonstrated the NMR-based structural analyses by using two series of the synthetic mucin-type glycopeptides focusing on the T*T*X motifs where * indicates the position of attachment of specific *O*-glycan. Artificial mucin-like model glycopeptides composed of the amino acids abundantly found in native mucin families allow to highlight the conformational influences from hydrophobic X residue located on the T*T*X motifs. This is because we hypothesized X plays an important role in building the triplet T*T*X structural motifs in mucin context. Hydrophobic amino acids are commonly occupied at X in the glycopeptides from CD43 (X = A),¹⁷ MUC2 (X = P),¹⁸ and MUC5AC (X = V)¹⁹ that had already been solved their NMR structures, in which the rigid and extended natures on the T*T* backbones are prone to propagate to the adjacent X. Another synthetic glycopeptides from MUC2 have two T*T*X motifs carrying different tumor associated *O*-glycans on their sequence: Tn, T, and α 2,3-sialyl T antigens. These could be nice probes for elucidating carbohydrate-induced impacts on the MUC2 glycopeptide conformations. Furthermore, these synthetic MUC2 glycopeptides were applied to microarray analysis to evaluate antigenicity against serum auto-antibody.

2-2. Result and Discussion

2-2-1. Sequence analysis of mucin tandem repeat and design of mucin-model glycopeptides containing TTX motif

22 members of human mucins (MUC) have been identified, in which 19 of them had tandem repeat domains (MUC1, MUC2, MUC3A, MUC3B, MUC4, MUC5AC, MUC5B, MUC6, MUC7, MUC8, MUC9, MUC11, MUC12, MUC13, MUC16, MUC17, MUC19, MUC20, and MUC21).^{1,20} In tandem repeats of these mucins, consecutive Thr-Thr segments were found in 14 members including transmembrane mucins (MUC3A, MUC3B, MUC11, MUC12, MUC16, MUC17, and MUC21) and secretory mucins (MUC2, MUC5AC, MUC5B, MUC6, MUC7, MUC8, and MUC19; **Table 2-1**). Composition ratio of X amino acid residues and proximal residues on TTX segment were summarized in **Figure 2-2**. Threonine, serine, and proline residues were found as predominant amino acid residues in any position around consecutive Thr-Thr segments. Particularly in X6 position corresponding to X of TTX motif, approximately 50% of X6 residues were occupied by hydrophobic amino acid residues, in which 80% are proline, alanine, leucine and valine.

		TR	X = Hydrophobic							X = Hydrophilic						SUM
			TTV	TTP	TTA	TTL	TTI	TTF	TTG	TTM	TTY	TTH	TTE	TTN	TTD	
MUC2	secreted	23	1	1												2
		16		1		1										2
MUC3A	membrane	17											1			1
		375	1	3	4	2	1	2	1	2		1	1	2	1	21
MUC3B	membrane	17											1			1
		375	1	3	4	2	1	2	1	2		1	1	2	1	21
MUC5AC	secreted	5	1													1
MUC5B	secreted	29		1	1							1				3
MUC6	secreted	169				2					1	1				4
MUC7	secreted	23		1	1											2
MUC11	membrane	28				1										1
MUC12	membrane	28				1		1								2
MUC16	membrane	156											1			1
MUC17	membrane	59		2	1											3
MUC19	secreted	7	1													1
		7	1													1
		16	1													1
		8	1													1
Total			8	12	11	9	2	5	2	4	1	4	5	4	2	

Table 2-1. The number of TTX motif in mucin tandem repeat sequence.

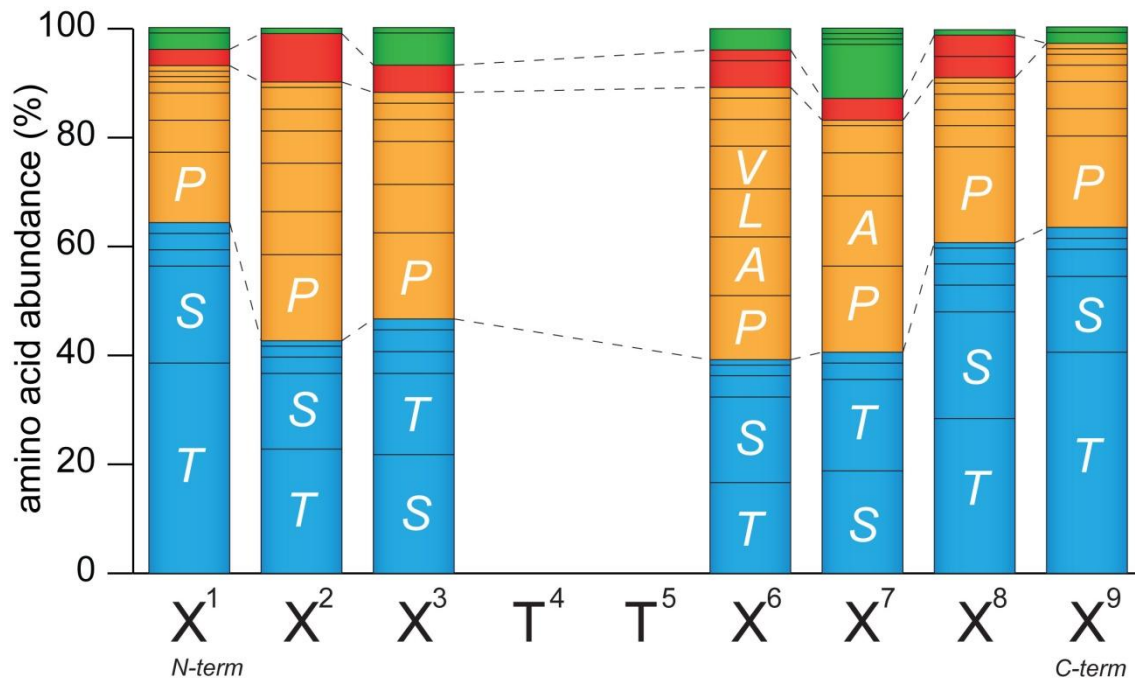


Figure 2-2. The frequency rates of amino acids at each flanking position of consecutive double threonines (Thr-Thr) among the tandem repeat regions of a human mucin family of glycoproteins (MUCs). Lite blue: hydrophilic residues, orange: hydrophilic residues, red: acidic residue, green: basic residue.

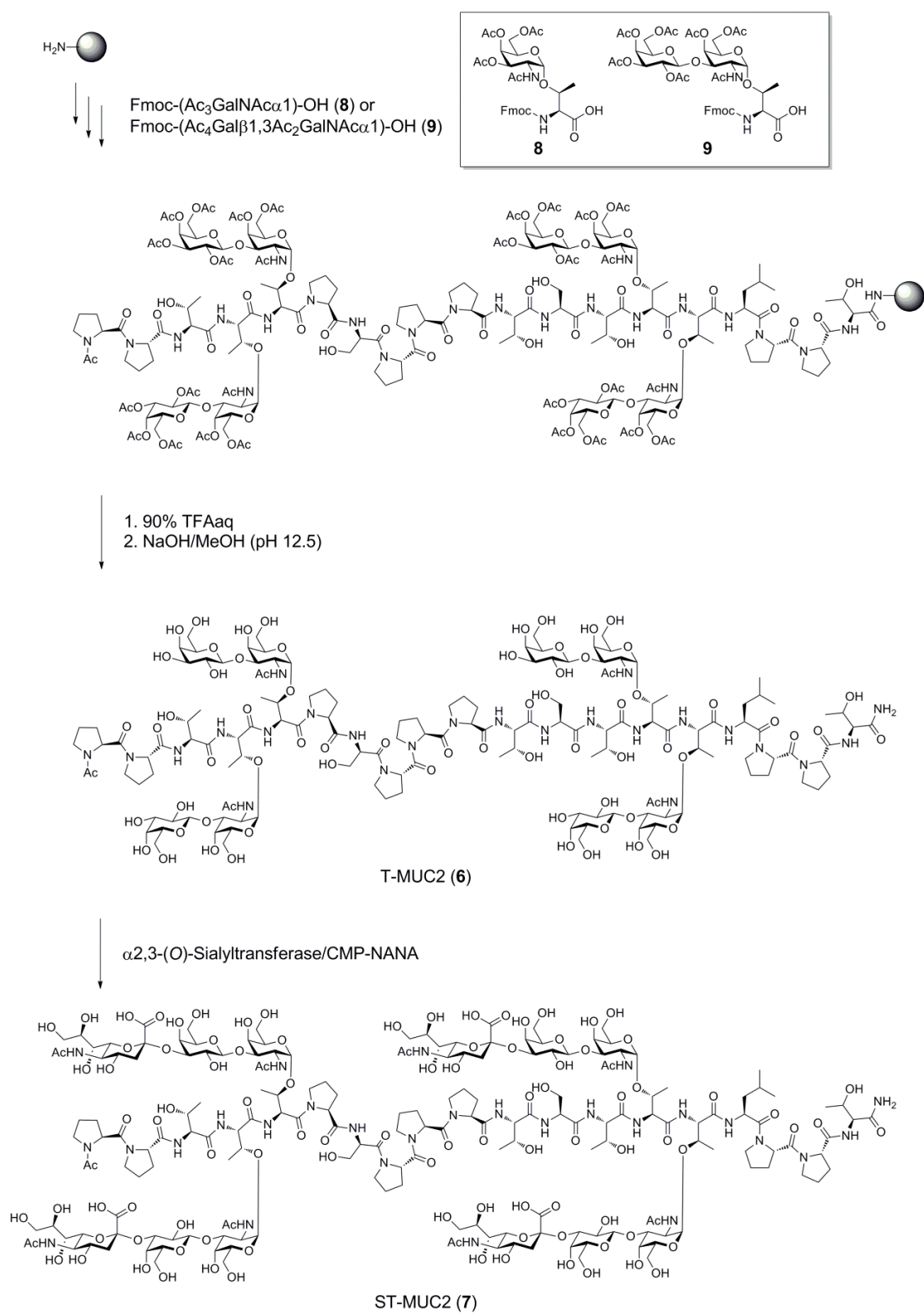
To evaluate *O*-GalNAc modification impacts on TTX motif conformation, 9-mer mucin-model glycopeptides containing TTX motif were designed using most abundant proximal residues, Ac-¹TTST*T*XSTT⁹-NH₂ where X is Ala for **1**, Val for **2**, Leu for **3**, or Pro for **4**. Uniform T-antigen modifications were designed at two consecutive threonine residues in TTX motif indicated by *. In addition, 19-mer MUC2 tandem repeat containing TTP and TTL motif, Ac-¹PPTT*T*PSPPPTSTT*T*LPPT⁹-NH₂ where * indicates uniform attachment of specific tumor associated *O*-glycan (Tn for **5**, T for **6**, or ST for **7**), were designed focusing on the influences derived from *O*-glycan type on the structures in and around the individual TTX motifs (**Table 2-2**).

Entry	Sequence											O-glycan								
		1		5		9														
TTA (1)	Ac-	T	T	S	T	T	A	S	T	T	-NH ₂	T								
TTV (2)	Ac-	T	T	S	T	T	V	S	T	T	-NH ₂	T								
TTL (3)	Ac-	T	T	S	T	T	L	S	T	T	-NH ₂	T								
TTP (4)	Ac-	T	T	S	T	T	P	S	T	T	-NH ₂	T								
		1		5		10		15		19										
Tn-MUC2 (5)	Ac-	P	P	T	T	P	S	P	P	P	T	S	T	T	L	P	P	T	-NH ₂	Tn
T-MUC2 (6)	Ac-	P	P	T	T	P	S	P	P	P	T	S	T	T	L	P	P	T	-NH ₂	T
ST-MUC2 (7)	Ac-	P	P	T	T	P	S	P	P	P	T	S	T	T	L	P	P	T	-NH ₂	sialyl T

Table 2-2. Mucin-model and MUC2 glycopeptides synthesized in this thesis

2-2-2. Synthesis of 9-mer model glycopeptides and 19-mer MUC2 glycopeptides

Chemoenzymatic approach with microwave-assisted solid-phase peptide synthesis (Fmoc SPPS) and enzymatic sugar elongation is performed to prepare 9-mer mucin model glycopeptides **1-4** and 19-mer MUC2 glycopeptides **5-7** with synthetic Fmoc-protected glycoamino acid derivatives **8, 9** (**Scheme 2-1**).²¹ Scheme 2-1 represented the chemoenzymatic procedure to synthesize 19-mer ST-MUC2 glycopeptide (**6**). In Fmoc SPPS, incorporation of glycoamino acid residue is one of the most challenging steps because of bulky pendant and difficulty of glycoamino acid synthesis. To overcome this synthetic barrier, double activation method was demonstrated, which enables us to perform efficient glycoamino acid incorporation with small equivalents of Fmoc-glycoamino acid derivatives (1.2 equiv.).²² After coupling reaction with glycoamino acid under microwave irradiation, extra coupling reagents were added into the reaction mixture to re-activate unreacted Fmoc glycoamino acid derivatives, and stirred under microwave irradiation. In this method, PyBOP/HOAt activation system was employed instead of our conventional HBTU/HOBt system. The advantage of this activation system is that PyBOP can avoid guanydilation of unreacted amino group on resin (**Figure 2-3**). HPLC comparison between conventional and double activation method were shown in **Figure 2-4**. This HPLC results clearly indicated efficiency of double activation method.



Scheme 2-1. Chemoenzymatic synthesis of MUC2 glycopeptides

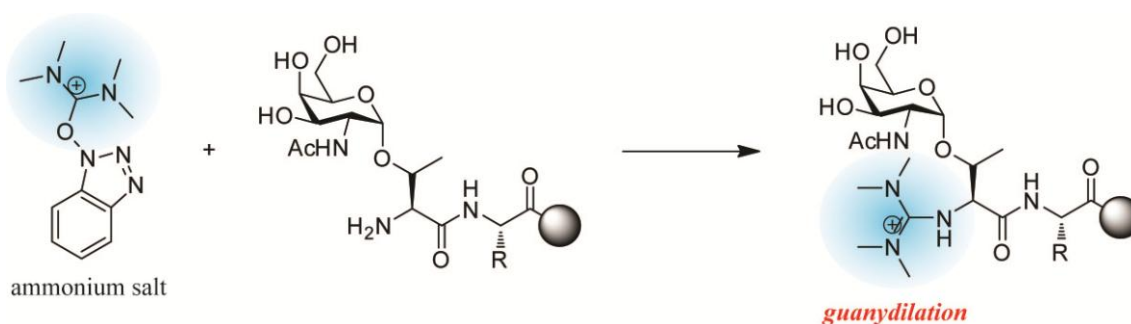


Figure 2-3. Undesired guanydilation of amino group by extra HBTU reagents

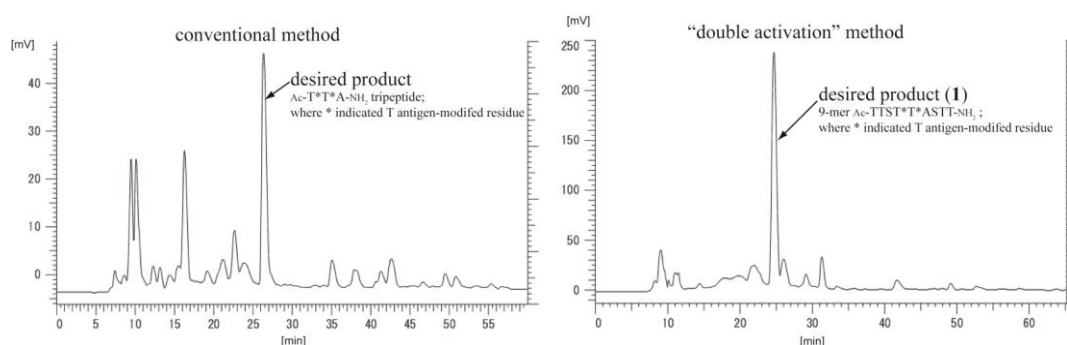


Figure 2-4. HPLC profile of conventional method and double activation method

9-mer mucin model glycopeptides **1-4** and 19-mer MUC2 tandem repeats **5-6** were prepared with Tentagel S RAM and ChemMatrix resin support, which gave CONH₂ at C-terminus, respectively. After peptide elongation and N-terminus acetylation on the resin, peptide cleavage from resin and deprotection of peptide side chain were performed. Then, acetyl groups were removed by basic condition and crude material was purified by RP-HPLC. Asialo MUC2 glycopeptide **6** carrying uniform T antigens was subjected to enzymatic reaction to obtain ST MUC2 glycopeptide **7** with α 2,3-(O)-sialyltransferase in the presence of glycosyl donor CMP-NANA.

2-2-3. NMR spectral analysis of synthetic glycopeptides

Purified glycopeptides **1-7** were subjected to solution NMR analyses. These glycopeptides **1-7** were dissolved in 300 μ l of 90% H₂O/10% D₂O or 99.9% D₂O at the concentration of 1.5~2.0 mM. The pH of NMR solution was adjusted at 6.0 by 0.1 N NaOH_{aq}. The two-dimensional NMR spectra of TOCSY, NOESY, DQF-COSY, ¹³C-edited HSQC, and HSQC-TOCSY of **1-7** were successfully obtained using 600 MHz NMR at proton frequency equipped with Cryo-probe. The cross-peaks on these spectra were fully identified with respect to both peptide and carbohydrate moiety.

2-2-3-1. ³J coupling constant

The vicinal couplings were analyzed from the assignment of the cross peaks on DQF-COSY spectra of **1-7** (**Table 2-3**). In 9-mer mucin-like model glycopeptides **1-4**, the coupling values of ³J_{HN,H α} on O-glycosylated Thr4 and Thr5 were in a range of 8.77-10.08 Hz, which are higher than those of unglycosylated amino acid residues (**Table 2-3A**). These higher ³J coupling values are indicative of local extended conformations.²³ The highest ³J_{HN,H α} value was observed on Thr5 in **3**, suggesting the backbone dihedral angle $\phi \sim -120^\circ$ with limitation in possible torsion angular averaging.²⁴ In addition, the ³J coupling constants between the N-acetyl amide HN2 protons and H2 protons in T4-GalNAc and T5-GalNAc residues, ³J_{HN2,H2}, also fall in a high value range of 8.77-10.39 Hz. Close to or more than 10 Hz of the ³J_{HN2,H2} values indicate the corresponding torsion angle is $\sim 180^\circ$.²⁵

In 19-mer MUC2 glycopeptides **5-7**, as similar to 9-mer mucin-like models **1-4**, all of the O-glycosylated threonines (except for Thr14 of **5**, not detected because of peak overlap) and the attached GalNAc residues showed high values on the backbone ³J_{HN,H α}

(8.33-10.09 Hz) and the GalNAc $^3J_{\text{HN},\text{H}_2}$ (8.33-10.39 Hz) couplings, respectively (**Table 2-3B**). Interestingly, the conversion from **5** to **6** leads to increases in the $^3J_{\text{HN},\text{H}\alpha}$ coupling values of the inside Thr5 and Thr15 (**bold**) as well as in the $^3J_{\text{HN},\text{H}_2}$ coupling values of the GalNAc residues attached to the outside threonine residues Thr4 and Thr14 (*italic*) on the $^4T^*\mathbf{T}^*P^6$ and $^{14}T^*\mathbf{T}^*L^{16}$ motifs, while the conversion from **6** to **7** results in substantial decreases in them, even though the values are yet high. The synchronized alteration patterns appear to be associated with systematic organization of the core scaffold with increasing in the size of sugar chains.

A

residue	1	2	3	4
T1	4.82	7.90	N.D.	7.45
T2	7.89	7.89	N.D.	7.89
S3	7.01	7.02	7.47	7.02
T4	9.21	8.77	9.20	9.20
T5	9.21	9.64	10.08	8.77
X6	7.58	8.77	7.89	—
S7	N.D.	N.D.	N.D.	6.14
T8	7.45	N.D.	8.33	7.45
T9	8.34	8.33	8.77	7.84
T4-GalNAc	10.09	10.39	9.64	9.65
T5-GalNAc	10.08	10.08	8.77	10.09

B

residue	5	6	7
T3	7.46	N.D.	N.D.
T4	9.00	9.65	9.65
T5	9.21	9.65	8.77
S7	6.58	N.D.	7.45
T11	N.D.	7.45	N.D.
S12	7.89	N.D.	7.45
T13	8.33	8.77	8.33
T14	N.D.	9.21	9.65
T15	9.21	10.09	8.33
L16	6.14	7.86	7.90
T19	8.33	7.46	N.D.
T4-GalNAc	10.09	10.52	8.33
T5-GalNAc	9.65	9.21	9.21
T14-GalNAc	10.08	10.08	9.21
T15-GalNAc	10.09	10.52	8.33

Table 2-3. ^3J coupling constants of synthetic glycopeptides **1-7**

2-2-3-2. NOE contacts

NOE study will provide essential clues to determine the structure of our probes as well as NOE assay gives information about spatial contacts. The 9-mer mucin-like model glycopeptides **1-4** share the consecutive strong to medium $d_{\alpha N}(i,i+1)$ connectivities coupled with absent $d_{NN}(i,i+1)$ NOEs spanning on the regions of ${}^4T^*T^*XS^7$, suggesting the dominance of extended backbone conformations.²⁶ In the case of **4**, a strong NOE between the $H\alpha$ of Thr5 and the $H\delta$ of Pro6 is observed, which is indicative of restriction of Pro6 backbone to the *trans* conformation. NOEs between the *N*-acetyl amide protons (HN2) of the GalNAc residues (bound on Thr*i*) and the backbone HN protons of the underlying threonine residue (*i*) are detected both on Thr4 and Thr5 with weak intensities. Similar NOE or ROE connectivities have been found in the other mucin-like glycopeptide systems.^{17,19,27-31} NOEs between the HN2 protons of T5-GalNAc and the $H\alpha$ protons of Thr5 are also observed in all cases, although such interactions are lack between T4-GalNAc and Thr4. On contrary, NOEs between the HN2 protons of T4-GalNAc and the HN protons of the following Thr5 (the *i*+1 residue) were consistently found between **1-4**, although such contacts were not detected on T5-GalNAc and the following X6 (the *i*+1 residue). Additionally, **1** and **2** have extra NOE contacts between the HN2 protons of T4-GalNAc and the $H\alpha$ protons of Thr5. These NOEs suggest that *N*-acetyl groups of T4-GalNAc and T5-GalNAc adopt different orientations to the peptide chain, in which the difference in bulkiness of the side chain of X6 potentially gives rise to further difference. Nonetheless, these NOE patterns on the *O*-glycosylated threonine dimer segment infer direct interactions between the *N*-acetyl HN2 protons of the GalNAc residues (bound on Thr*i*) and the carbonyl oxygen atoms of the underlying threonine residues (the *i* residue), possibly

through hydrogen bonds.

A lot of medium-weak NOEs between the methyl protons of the GalNAc residues (bound on Thr*i*) and the peptide backbone HN plus the side chain protons of proximal residues (the *i*+1 and/or *i*+2 residue) were observed. With respect to the NOEs between Ti-GalNAc and the *i*+1 residue, the methyl protons of T4-GalNAc interact with the H α proton of Thr5 in all compounds, while those of T5-GalNAc contact with the H α proton of X6 only in **2** (Val6) and **3** (Leu6). With regard to the NOEs between Ti-GalNAc and the *i*+2 residue, the methyl protons of T4-GalNAc interact with the side chain methyl protons of X6 in **1-3** (medium-weak) and with the H δ methylene protons of P6 in **4** (medium), while those of T5-GalNAc contact with the H α (strong-medium) and H β (weak) protons of S7 in all cases (**Figure 2-5**).

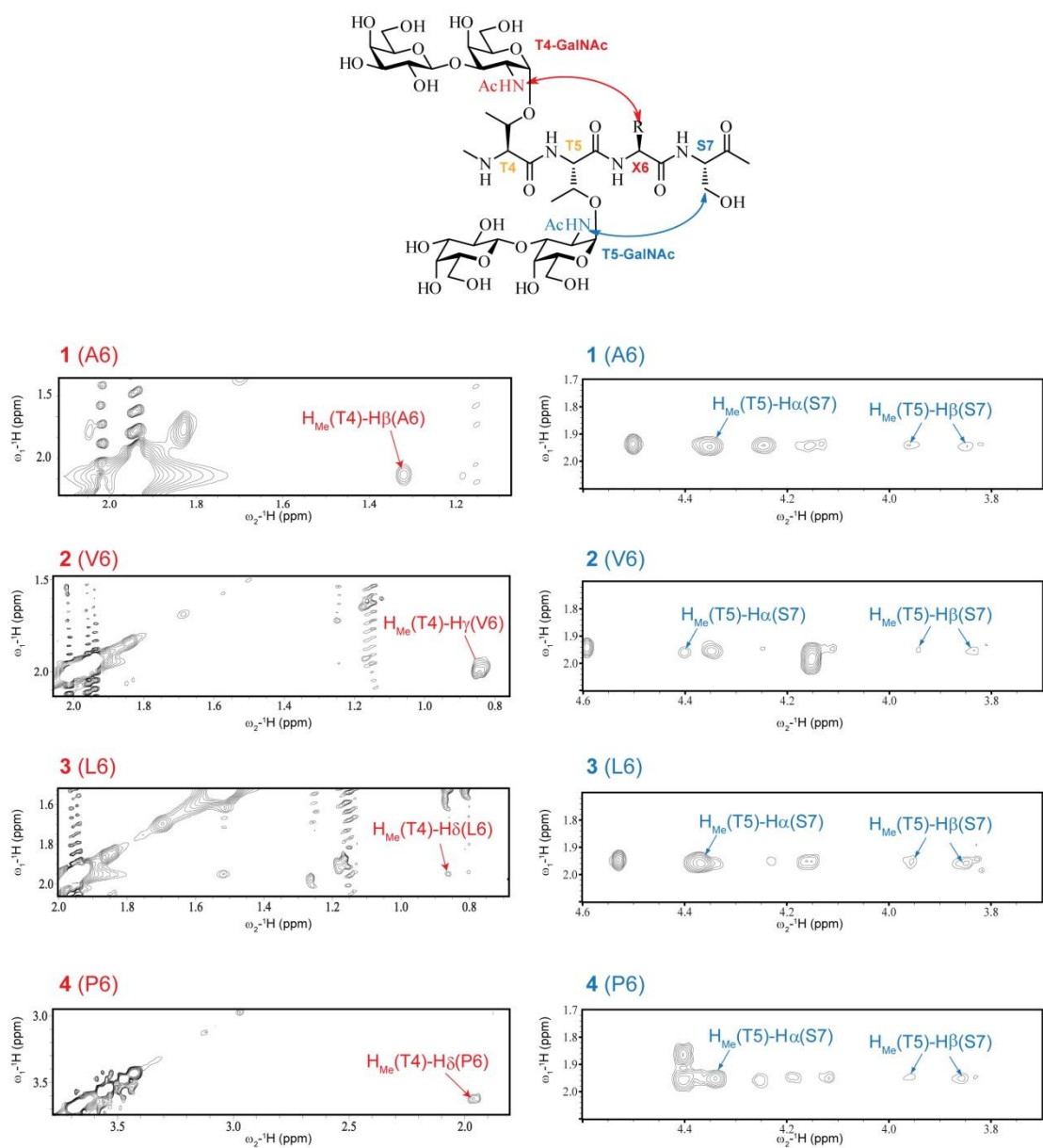


Figure 2-5. NOE interaction between sugar and peptide in 1-4

These NOE interactions are indicative of the limited orientations of the *N*-acetyl groups of GalNAc residues to the peptide chain. Especially the NOEs involving the *i*+2 residue also infer the existence of hydrophobic interactions. The putative hydrogen bonds and hydrophobic interactions may reflect the restriction of conformation and solvation on local constructs.

In addition, NOEs between the H1 proton of the Gal residue and the H3 proton of the adjacent GalNAc residue are detected on all of the sugar chains in **1-4**. There are no NOE interactions between the distal Gal residues and core peptide.

Further, 19-mer MUC2 glycopeptides **5-7** also showed the consecutive strong $d_{\alpha N}(i, i+1)$ connectivities coupled with absent $d_{NN}(i, i+1)$ NOEs over the regions of ${}^4\text{T}^*\text{T}^*\text{PS}^7$ and ${}^{14}\text{T}^*\text{T}^*\text{LP}^{17}$ in their NOESY spectra. As similar to mucin-like models, NOEs between the HN2 protons of the GalNAc residues (bound on Thr*i*) and the HN plus H α protons of the underlying threonine residues (*i*) are observed in **5-7**. Exceptionally, there are the NOE contacts of the HN2 proton of T15-GalNAc only in **7** with the H α and H β protons of Thr15, although such NOEs are absent in **5** and **6**.

NOE contacts between the HN2 protons of the GalNAc residues (bound on Thr*i*) and the backbone and side chain protons of the *i*+1 residues were altered with increase in the size of glycans. The NOE contact of the HN2 proton of T4-GalNAc with the H α proton was found in **5**. Partner protons of the NOEs with the same HN2 proton of T4-GalNAc was changed to the H α and HN protons of Thr5 in **6** and to the HN proton of Thr5 in **7**. Similarly, the HN2 proton of T14-GalNAc has NOE interaction with the HN of the Thr14 in **5**, which was shifted to the HN proton of the Thr15 in **6**, and the H α proton of Thr15 in **7**. On the other hand, the HN2 proton of T5-GalNAc connect with the H α and H δ protons of Pro6 in **5** and with the H α proton of Pro6 in **6**, although such connection

is absent in **7**. Similarly, the NOE contact of the HN2 proton of T15-GalNAc with the HN proton of Leu16 was observed only in **6**, although such contact was not detected in **5** and **7**. Considering these analyses, β 1,3-galactosylation and subsequent α 2,3-sialylation to clustered Tn glycans potentially make changes in the backbone rigidity on the underlying T*T*X motifs and the residues adjacent to them.

As can be seen in **1-4**, NOEs between the methyl protons of the GalNAc residues (bound on Thr*i*) and the side chain protons of the *i*+2 residues are consistently observed in **5-7**, suggesting the existence of hydrophobic interactions between T(*i*)-GalNAc and the *i*+2 residues. In addition, NOEs between the H1 proton of the Gal residue and the H3 proton of the adjacent GalNAc residue were detected on all of the sugar chains in **6** and **7**. Similarly, the axial and equatorial H3 protons of the peripheral sialic acid residues have NOE interactions with the H3 proton of the adjacent Gal residues on all of the sugar chains in **7**. As similar to **1-4**, there are no NOE interactions between the distal carbohydrate residues and core peptide. These may be attributed to the spatial arrangement of sugar chains sticking out from the peptide chain in all compounds.

2-2-4. NMR structural analyses of 9-mer mucin model glycopeptides

The NMR solution structures of **1-4** were computed from distance restraints from NOESY spectra and torsion angle constraints from DQF-COSY spectra. We obtained well-defined structures with heavy atom root-mean-square deviation (rmsd) values of 0.56 ± 0.25 Å (**1**), 0.50 ± 0.19 Å (**2**), 0.40 ± 0.10 Å (**3**), and 0.47 ± 0.17 Å (**4**) for residues 4-6 with two GalNAc moieties. The thirty lowest-energy structures of all mucin-like models have extended and rigid structures on the backbone of the ${}^4\text{T}^*\text{T}^*\text{X}^6$ motifs (**Figure 2-6**). The sugar chains bound on Thr4 and Thr5 also exhibit restricted orientations permitting to distinguish their global shapes, even though not so clear as compared to the ${}^4\text{T}^*\text{T}^*\text{X}^6$ backbones. Commonly used definitions of the geometric parameters and the atom labels in this study are illustrated in **Figure 2-7**.

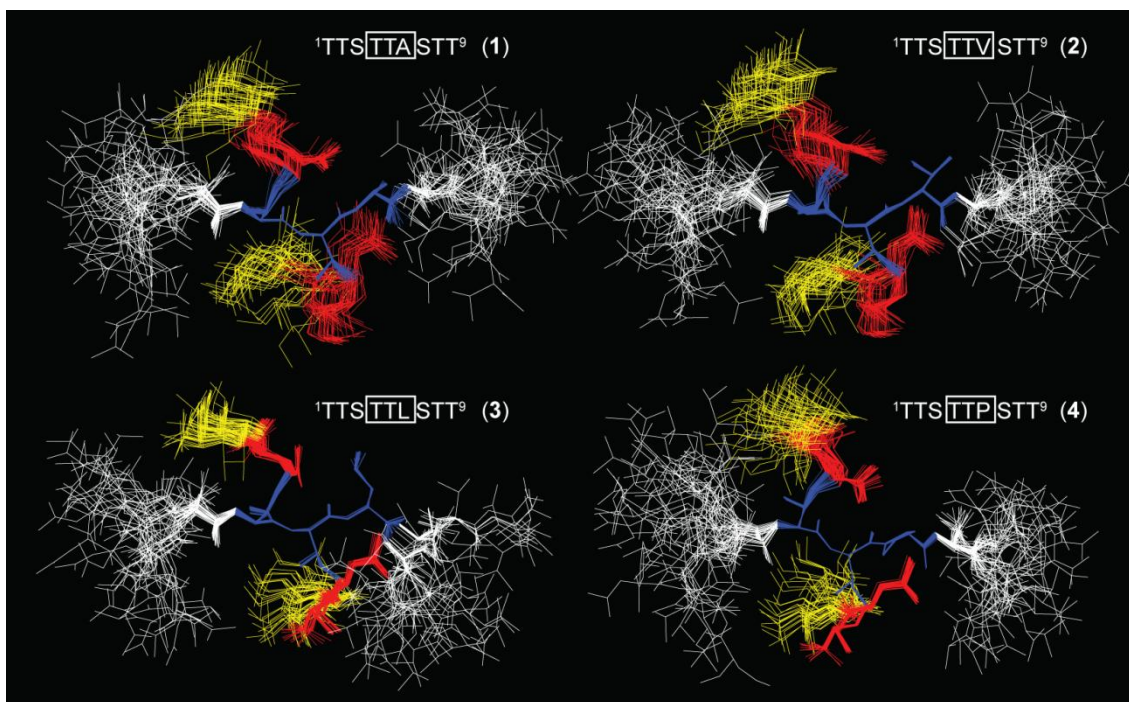


Figure 2-6. NMR structures of mucin model glycopeptides **1-4**

Blue: focused TTX sequence, Red: GalNAc, Yellow: Gal, White: other region.

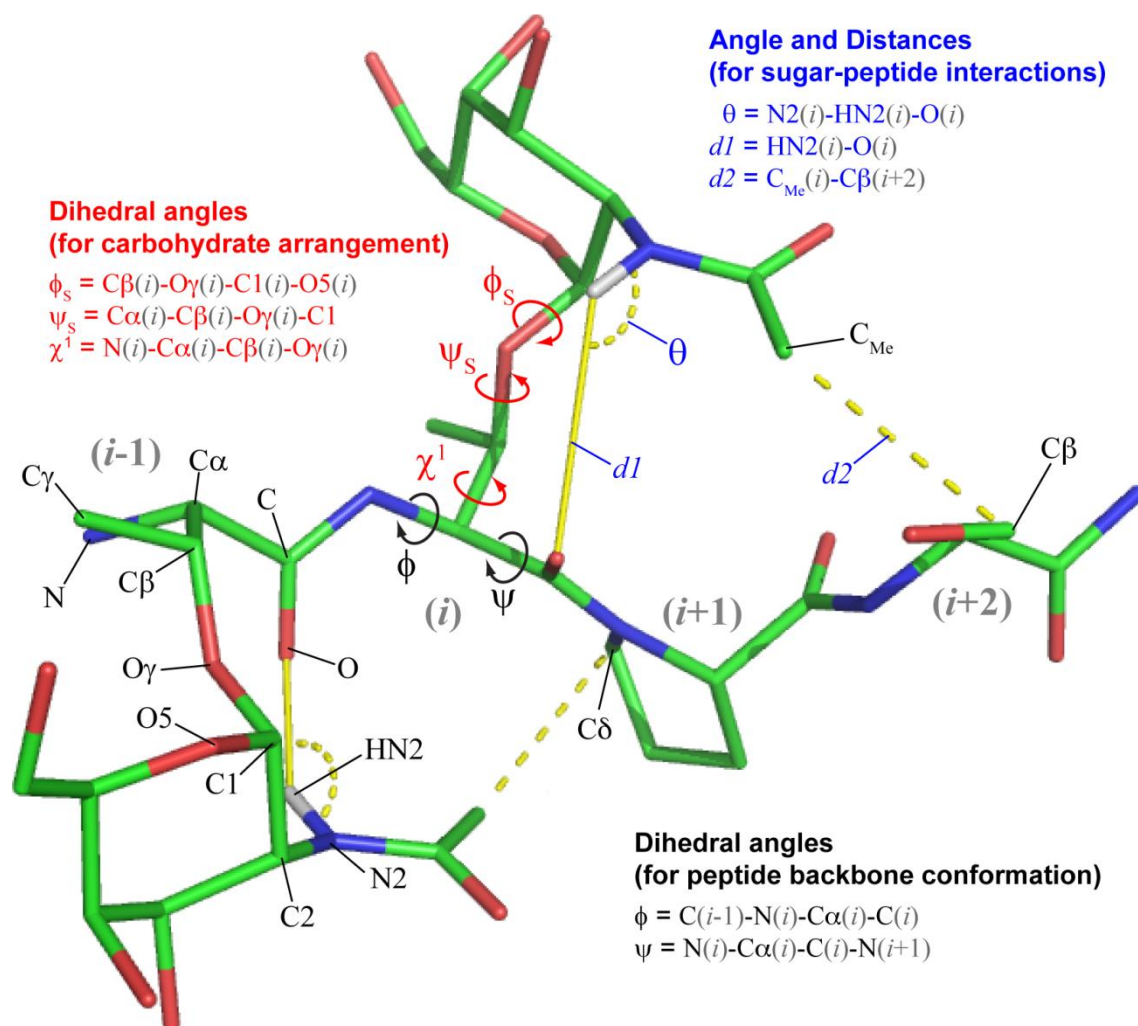


Figure 2-7. Definitions of geometric parameters and atom labels in this study

2-2-4-1. Peptide backbone conformations.

The torsion angles of peptide main chains were obtained from the thirty lowest-energy structures of **1-4**. The deviations of the individual ϕ or ψ torsion angles reflect the rigidity/flexibility of peptide backbone conformation in calculated structure (**Table 2-4**). Extended conformation can be assessed from the allowable extended region of mean torsion angles. In order to compare the conformational features between mucin-like models, we used the four-level rating for rigidity from the standard deviations of the mean torsion angles. The mean ϕ or ψ with high rigidity from first to third rates were additionally used for evaluating extended conformations. The arbitrary definitions of these criteria and the conformational propensity of **1-4** are shown in **Table 2-5**. Highly restricted angular variations are conserved on $\phi(\text{X6})$ and $\psi(\text{Thr5})$ in all cases. The rigidity on $\phi(\text{Thr5})$, $\psi(\text{Thr4})$, and $\phi(\text{Thr4})$ progressively decrease with distance from them, although the patterns in rigidity differ depending on X6 residue. The high rigidity was uniquely found on $\psi(\text{X6})$ and $\phi(\text{Ser7})$ in **4**, which may be attributed to conformational restriction of a distinctive cyclic structure of the proline residue at X6. On the other hand, consistent extended natures with allowable rigidity lie on $\psi(\text{Thr4})$ and $\psi(\text{Thr5})$, while $\phi(\text{Thr4})$ and $\phi(\text{X6})$ are inconsistently extended between the mucin-like models. These assessments suggest that the threonine dimer carrying T antigen cluster and an additional hydrophobic X residue with high abundance in native mucins could form a consistent structural motif with high rigidity and extended nature on mucin glycopeptides. In addition, X residue plays a role in the control of the structural preferences on the peptide backbone.

residue		1 (A6)	2 (V6)	3 (L6)	4 (P6)
Thr1	ϕ	-93.1 $\pm$ 80.3	-122.3 $\pm$ 79.0	-124.5 $\pm$ 89.2	-114.0 $\pm$ 62.7
	ψ	107.9 $\pm$ 66.1	82.8 $\pm$ 79.4	115.2 $\pm$ 79.0	117.9 $\pm$ 83.3
	χ^1	13.4 $\pm$ 81.5	-110.3 $\pm$ 96.6	-13.2 $\pm$ 96.5	11.3 $\pm$ 91.3
Thr2	ϕ	-97.2 $\pm$ 81.5	-129.5 $\pm$ 74.6	-110.7 $\pm$ 81.7	-116.8 $\pm$ 72.0
	ψ	61.2 $\pm$ 83.4	112.0 $\pm$ 72.7	124.9 $\pm$ 39.0	86.7 $\pm$ 81.8
	χ^1	1.5 $\pm$ 101.4	7.7 $\pm$ 76.0	6.4 $\pm$ 82.8	15.1 $\pm$ 90.6
Ser3	ϕ	-111.2 $\pm$ 86.7	-106.3 $\pm$ 92.2	-62.8 $\pm$ 92.7	-124.4 $\pm$ 95.6
	ψ	162.9 $\pm$ 25.8	156.9 $\pm$ 27.4	161.5 $\pm$ 17.5	131.5 $\pm$ 33.3
	χ^1	-94.7 $\pm$ 85.3	-101.9 $\pm$ 98.7	-3.2 $\pm$ 97.5	-109.9 $\pm$ 68.1
Thr4	ϕ	-141.4 $\pm$ 9.7	-129.7 $\pm$ 14.0	-116.6 $\pm$ 17.4	-126.4 $\pm$ 29.9
	ψ	177.7 $\pm$ 7.5	172.7 $\pm$ 9.2	-179.1 $\pm$ 3.2	132.5 $\pm$ 5.9
	χ^1	66 $\pm$ 4.3	57.5 $\pm$ 4.8	76.2 $\pm$ 1.0	81.5 $\pm$ 4.0
Thr5	ϕ	-90.8 $\pm$ 1.5	-96.9 $\pm$ 7.0	-113.1 $\pm$ 0.9	-89.6 $\pm$ 0.1
	ψ	153.8 $\pm$ 0.3	160.1 $\pm$ 2.4	-168.7 $\pm$ 0.6	165.1 $\pm$ 0.8
	χ^1	57.7 $\pm$ 3.6	69.0 $\pm$ 4.0	46.2 $\pm$ 2.0	43.5 $\pm$ 0.6
Xaa6	ϕ	-68.2 $\pm$ 4	-79.9 $\pm$ 0.9	-129.5 $\pm$ 4.3	-81.0 $\pm$ 1.6
	ψ	178 $\pm$ 36.6	159.7 $\pm$ 61.5	69.4 $\pm$ 42.5	134.5 $\pm$ 0.4
	χ^1	-5.1 $\pm$ 83.1	-169.6 $\pm$ 2.0	-69.9 $\pm$ 2.4	-
Ser7	ϕ	-137.5 $\pm$ 50.7	-124.3 $\pm$ 69.2	43.0 $\pm$ 65.5	-75.4 $\pm$ 7.7
	ψ	117.7 $\pm$ 37.3	105.1 $\pm$ 34.5	-136.1 $\pm$ 96.6	61.9 $\pm$ 85.9
	χ^1	-154.6 $\pm$ 79.9	-171.2 $\pm$ 74.0	39.4 $\pm$ 100.5	-83.5 $\pm$ 98.4
Thr8	ϕ	12.4 $\pm$ 77.4	25.1 $\pm$ 78.0	58.0 $\pm$ 71.5	-127.3 $\pm$ 62.4
	ψ	93.2 $\pm$ 81.3	117.4 $\pm$ 83.4	92.6 $\pm$ 63.4	90.6 $\pm$ 79.8
	χ^1	-8.8 $\pm$ 89.4	-2.2 $\pm$ 103.0	-26.3 $\pm$ 86.6	14.7 $\pm$ 97.3
Thr9	ϕ	-116.8 $\pm$ 47.3	-111.2 $\pm$ 72.0	-107.0 $\pm$ 61.4	-117.1 $\pm$ 54.9
	χ^1	31 $\pm$ 81.9	16.2 $\pm$ 65.1	-8.5 $\pm$ 88.9	31.2 $\pm$ 77.2

Table 2-4. Average dihedral angles for peptide backbone in mucin model 1-4

							*		*										* O-glycosylation site E $150^\circ < \phi \text{ or } \psi \leq 180^\circ$
	T1		T2		S3		T4		T5		X6		S7		T8		T9		
	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	
1 (A6)						E		E		E		E							 $\sigma \geq 15$ $10 \leq \sigma < 15$ $5 \leq \sigma < 10$ $\sigma < 5$
2 (V6)						E		E				E							
3 (L6)						E		E		E									
4 (P6)								E		E									

Table 2-5. Relationships between rigidity and extended nature of the backbone torsion angles obtained from the thirty lowest-energy structures of mucin model glycopeptides **1-4**

2-2-4-2. Carbohydrate arrangements.

Orientation of initial GalNAc residue to the peptide chain can be represented by the χ^1 , ϕ_S , and ψ_S torsion angles, which are associated with the tethering linkage composed of the three covalent bonds (C α -C β -O γ -C1) (**Figure 2-7**). These torsion angles are obtained from the thirty lowest-energy structures of **1-4**. The relationships between O-glycosylated Thr4 and Thr5 with respect to individual types of torsion angles are illustrated in **Figure 2-8**.

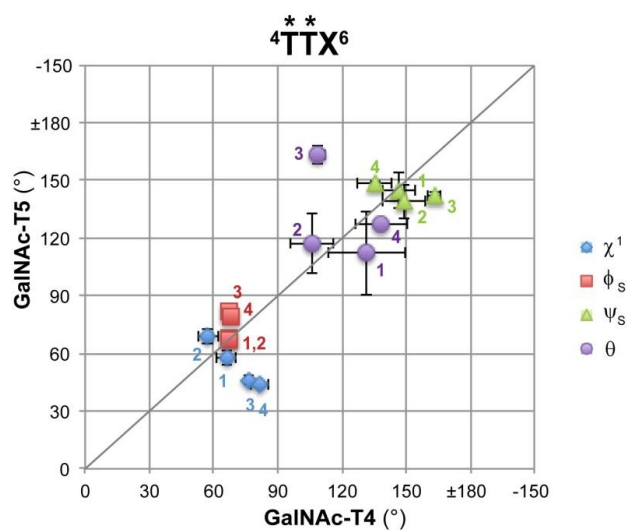


Figure 2-8. Distributions of the χ^1 , ϕ_s , and ψ_s torsion angles and the θ angle for the O-glycosylated threonines (T4 and T5) obtained from the thirty lowest-energy structures of mucin model glycopeptides **1-4**.

2-2-4-2-1. χ^1 torsion angles.

All mean χ^1 torsion angles of Thr4 and Thr5 fall into a range of $62.5 \pm 19.0^\circ$. The $\chi^1(\text{Thr4}, \text{Thr5})$ coordinate of **1** ($66.0^\circ, 57.7^\circ$) and **2** ($57.5^\circ, 69.0^\circ$) are nearly symmetric about a diagonal line, while those of **3** ($76.2^\circ, 46.2^\circ$) and **4** ($81.5^\circ, 43.5^\circ$) are far from equal. The $\chi^1(\text{Thr4}, \text{Thr5})$ coordinates of **1**, **3**, and **4** are located at the same area (Thr4>Thr5) and that of **2** is located at the opposite area (Thr4<Thr5). Interestingly, all $\chi^1(\text{Thr4}, \text{Thr5})$ coordinates are aligned on a roughly orthogonal line [$\chi^1(\text{Thr5}) = -1.0857 \times \chi^1(\text{Thr4}) + 130.4$, R^2 value is approximately 0.98] about a diagonal with the point of intersection at ($62.5^\circ, 62.5^\circ$).

These data suggest that X6 residue has potential to shift not only proximal $\chi^1(\text{Thr5})$ but also distal $\chi^1(\text{Thr4})$ on the ${}^4\text{T}^*\text{T}^*\text{X}^6$ motif. Moreover, their motions appear to vary in inverse proportion each other. Sterically hindered Leu6 and Pro6 tend to generate larger differences in χ^1 between Thr4 and Thr5 than Ala6 and Val6 in our models.

2-2-4-2-2. ϕ_S and ψ_S torsion angles.

The ϕ_S distributions between Thr4 and Thr5 are more heavily constrained than those of the χ^1 in all mucin models. The $\phi_S(\text{Thr4})$ in **1-4** and the $\phi_S(\text{Thr5})$ in **1** and **2** are close to 68° , although the $\phi_S(\text{Thr5})$ in **3** and **4** are both shifted to approximately 80° . Little influences from steric bulk of X6 are suggested on the ϕ_S of immediately adjacent Thr5. The $\psi_S(\text{Thr4,Thr5})$ distributions are relatively fluctuated on the $\psi_S(\text{Thr4})$ in a range of $149.2 \pm 13.8^\circ$ than on the $\psi_S(\text{Thr5})$ in a range of $143.7 \pm 4.6^\circ$. The minimal mean value at 135.4° for **3** and the maximal mean value at 163.0° for **4** owe to the diversity in $\psi_S(\text{Thr4})$. The mean values of the $\psi_S(\text{Thr4})$ in **1** and **2** are both around 148° . The backbone ϕ,ψ and GalNAc ϕ_S,ψ_S dihedral spaces on the Thr4 and Thr5 in **1-4** are both in agreement with those of the corresponding Thr1 and Thr2 obtained from MD-tar simulations for the small models Ac-¹T(Tn)²T(Tn)-NHMe.³²

2-2-4-2-3. Intramolecular non-covalent interactions between GalNAc and peptide.

Hydrogen bonds and hydrophobic interactions between GalNAc and peptide chains have been proposed as key elements for stabilizing structures of threonine-glycosylated peptides so far. In the next section, the geometric parameters indicative of these existences and/or differences were determined from the ensembles of **1-4** (Figure 2-9).

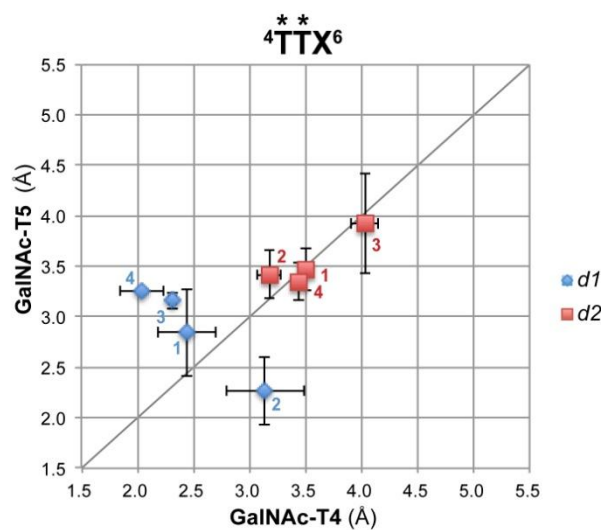


Figure 2-9. Distributions of the *d1* and *d2* distances for T4 and T5 obtained from the thirty lowest-energy structures of mucin model glycopeptides **1-4**.

2-2-4-2-4. d1 distances and θ angles for hydrogen bonding geometries.

The formation of direct^{17,18,29-31,33-35} or water-mediated^{32,36,37} hydrogen bonds between the peptide backbone and proximal GalNAc have been observed in many mucin-like glycopeptide structures. For instance, the calculated structures of IgA-hinge glycopeptides with separately glycosylated with Tn antigens mostly show the probable hydrogen bonds between GalNAc amide N2-HN2 and the underlying threonine backbone O.³⁵ This leads to large restriction of the ψ torsion angle of the peptide backbone to $\sim 155^\circ$, which is accompanied by inducing significant dominance of the *trans* conformation on the following proline residue.

In our case, the ψ torsion angle of Thr5 ahead of Pro6 is constrained to 165° in average and the Pro6 backbone are restricted to the *trans* conformation in all of the thirty lowest-energy structures of **4**, even though our compound is *O*-glycosylated with successive T antigens. The good agreement between mucin-like model **4** and IgA-hinge glycopeptide strongly inferred the existence of hydrogen bond between T5-GalNAc

amide and Thr5 carbonyl oxygen in **4**. The favored or allowed geometries for hydrogen bonding might be reflected by the arrangement of these groups. The related parameters of the $d1$ distances (HN2-O) and the θ angles (N2-HN2 $\cdots$ O) for Thr4 and Thr5 in **1-4** are shown in **Figure 2-9**. With respect to Thr5 in **4**, the mean values of $d1$ and θ obtained from the thirty lowest-energy structures are well constrained to 3.3 Å and 127.5°, respectively. These arrangements are consistent with weak stabilizing hydrogen bonding geometries.¹⁷ With regard to Thr4 in **4**, the mean values of $d1$ and θ are 2.0 Å and 138.4°. Thus, Thr4 has shorter $d1$ and higher θ angle than Thr5, suggesting the stronger interaction in Thr4. The other mucin models (**1-3**) also demonstrate the uneven arrangements between T4-GalNAc and Thr4 and between T5-GalNAc and Thr5 moieties. All mean values of $d1$ and θ in **1-4** fall into a range of 2.8 ± 0.5 Å and $134.6 \pm 28.7^\circ$, respectively. These distances and angles are in consistent with the reported geometries for potential hydrogen bonds.¹⁷

Interestingly, as similar to the cases of $\chi^1(\text{Thr4, Thr5})$, the all $d1(\text{Thr4, Thr5})$ coordinates in **1-4** are aligned on a roughly orthogonal line [$d1(\text{Thr5}) = -0.9329 \times d1(\text{Thr4}) + 5.2009$, R^2 value is approximately 0.96] about a diagonal with the point of intersection at (2.7 Å, 2.7 Å) (**Figure 2-9**). These coordinates seem to correlate inversely with the $\chi^1(\text{Thr4, Thr5})$ in **Figure 2-8**. The $d1(\text{Thr4, Thr5})$ in **2** is located in the Thr4 dominant space (Thr4 > Thr5 in $d1$), while the $\chi^1(\text{Thr4, Thr5})$ in **2** is located in the Thr5 dominant space (Thr4 < Thr5 in χ^1). On contrary, the $d1(\text{Thr4, Thr5})$ in **1, 3, and 4** are located in the Thr5 dominant space (Thr4 < Thr5 in $d1$), while the $\chi^1(\text{Thr4, Thr5})$ in **1, 3, and 4** are located in the Thr4 dominant space (Thr4 > Thr5 in χ^1) with maintenance of the aligned order from a diagonal in each case. This systematic correlation between $\chi^1(\text{Thr4, Thr5})$ and $d1(\text{Thr4, Thr5})$ in mucin-like models **1-4** is considered as one of

conformational insights into stabilization for the T*T*X motifs on mucin constructs.

2-2-4-2-5. d2 distances for hydrophobic interaction geometries.

The hydrophobic interaction between *N*-acetyl methyl group of Thr(*i*)-GalNAc and the side chain of the (*i*+2) amino acid residue is thought to be another key element for increasing stability on mucin-type glycopeptide systems.^{17,18,30,38} The geometries associated with hydrophobic interaction could be assessed by the *d2* distance parameter (**Figure 2-7**). As the partner atom of *N*-acetyl methyl carbon C_{Me} of T4-GalNAc residue for *d2* in each compound, methyl carbon of C β (Ala6) for **1**, C γ (Val6) for **2**, C δ (Leu6) for **3**, and methylene carbon of C δ (Pro6) for **4** were selected on the basis of the NOE contacts. In the cases of **2** and **3**, the one of two methyl groups on the side chain closer to the C_{Me} of T4-GalNAc were used for determination. Similarly, C β (S7) was selected as the common partner atom for C_{Me} of Thr5-GalNAc in **1-4**. As shown in **Figure 2-9**, average coordinates of the *d2*(Thr4) and *d2*(Thr5) in all compounds fall into a range of 3.6 ± 0.4 Å, which are consistent with the pattern of medium-range interactions.¹⁷ The *d2*(Thr4) and *d2*(Thr5) are almost uniform in the all cases. The *d2*(Thr4,Thr5) coordinate in **3** is located far from the others (**1**, **2**, and **4**), which seem to be associated with the differences in the θ (Thr4,Thr5).

2-2-4-3. Differences in shapes on the TTX motifs between mucin model glycopeptides.

The lowest-energy structure of each compound is regarded as the most stable conformer avoiding steric crowding of structural components as possible. The superimposed backbones of residue 4-6 in the lowest-energy structures of **1-4** show high

similarity with the rmsd value of 0.391 Å (**Figure 2-10**). The sugar chains bound on Thr4 and Thr5 are directed away from each other. The spatial arrangements of T antigens bound on Thr4 and Thr5 are both different between all models. As can be seen in **Figure 2-8**, those considerations highlight the concerted motions of the χ^1 , ψ_s , and ϕ_s torsion angles.

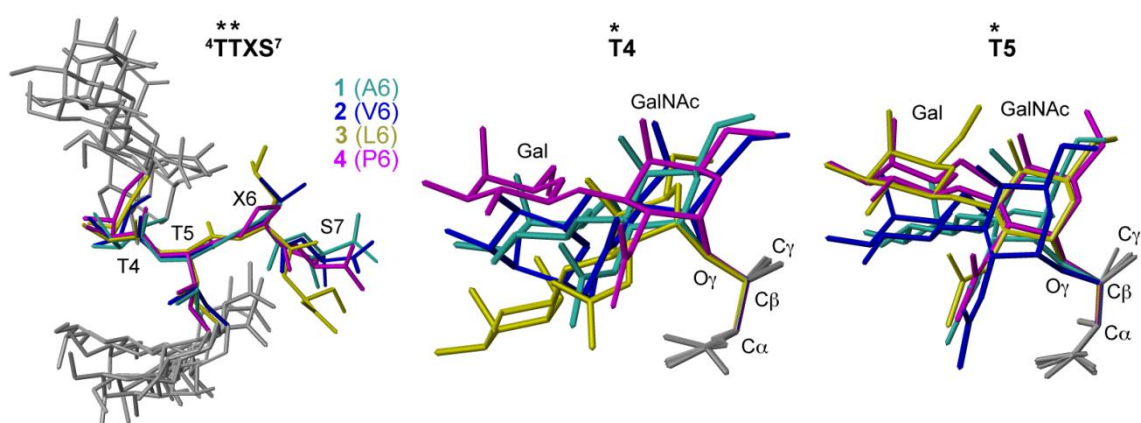


Figure 2-10 Comparison in the lowest-energy structure between mucin model glycopeptides **1-4**. The peptide main chains are partially shown in the ${}^4\text{TTXS}^7$ regions (left), which are aligned with the ${}^4\text{TTX}^6$ backbone atoms. The glycan chains attached to T4 (middle) and T5 (right) are illustrated with the underlying threonine residues. The focused moieties in all structures are colored in cyan for **1**, navy blue for **2**, yellow for **3**, and magenta for **4**. The other moieties are masked in grey.

2-2-5. NMR structure analyses of MUC2 glycopeptides (5-7).

As similar to mucin-like models **1-4**, the NMR solution structures of MUC2 glycopeptides (**5-7**) were computed from distance and dihedral angle constraints from NMR spectra. We obtained the well-defined structures with heavy atom rmsd values of 0.33 ± 0.11 Å for **5**, 0.69 ± 0.22 Å for **6**, and 0.86 ± 0.27 Å for **7** in the ${}^4\text{T}^*\text{T}^*\text{P}^6$ regions including two GalNAc residues as well as those of 0.42 ± 0.13 Å for **5**, 0.52 ± 0.25 Å for **6**, and 0.45 ± 0.11 Å for **7** in the ${}^{14}\text{TTL}^{16}$ region including GalNAc residues. The thirty lowest-energy structures of individual compounds indicate the well-ordered and extended conformations on the both ${}^4\text{T}^*\text{T}^*\text{P}^6$ and ${}^{14}\text{T}^*\text{T}^*\text{L}^{16}$ motifs (**Figure 2-11**). We validated the NMR structure of MUC2 glycopeptide **5** carrying clustered Tn antigens by comparison with the NMR structure of Tn-bound MUC2 glycopeptide Ac-PTT * T * PLK-NH₂, where * represents the position of O-glycosylated site with Tn antigen, deposited in the Protein Data Bank (PDB ID: 2LHX).¹⁸ Both structures show similar extended backbone conformations and they can be well superimposed on the backbone atoms on the common T * T * P motifs with an rmsd value of 0.362 Å (**Figure 2-12**).

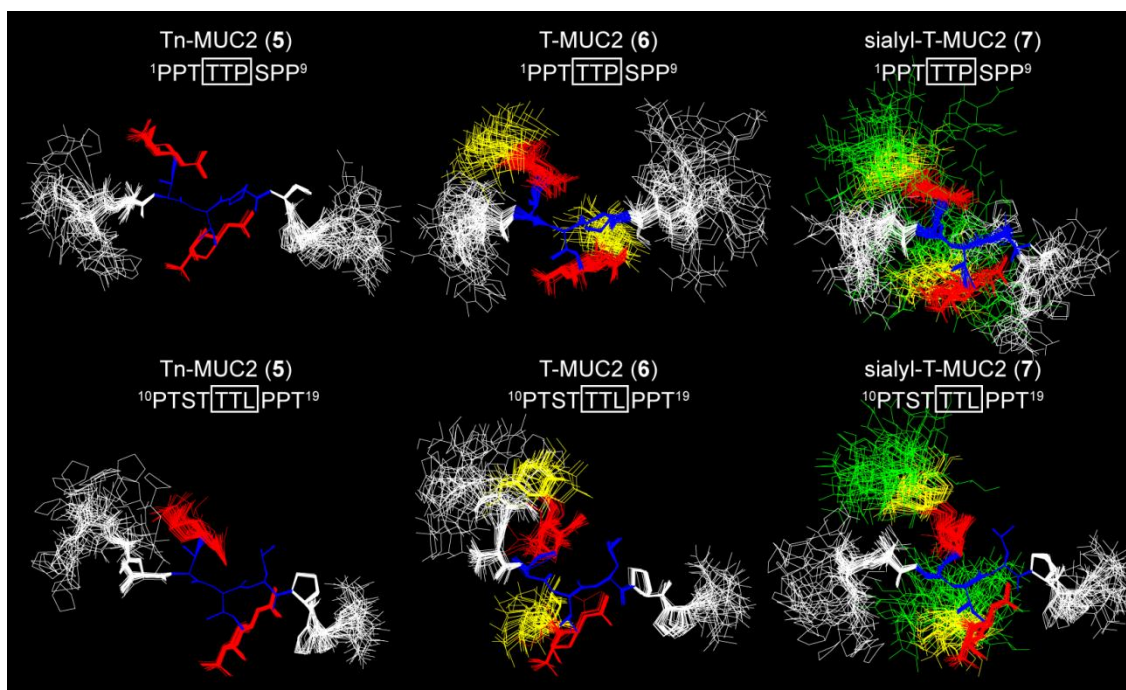


Figure 2-11. Overlay of the thirty lowest-energy structures of MUC2 glycopeptides 5-7. The partial regions of $^1\text{PPTTTTPSPP}^9$ (tops) and $^{10}\text{PTSTTTLPPT}^{19}$ (bottoms) are shown in a heavy atom representation. The $^4\text{TTP}^6$ and $^{14}\text{TTL}^{16}$ regions of main chains, GalNAc, and Gal residues are represented in blue, red, and yellow, respectively. Disordered peptide regions are given in white.

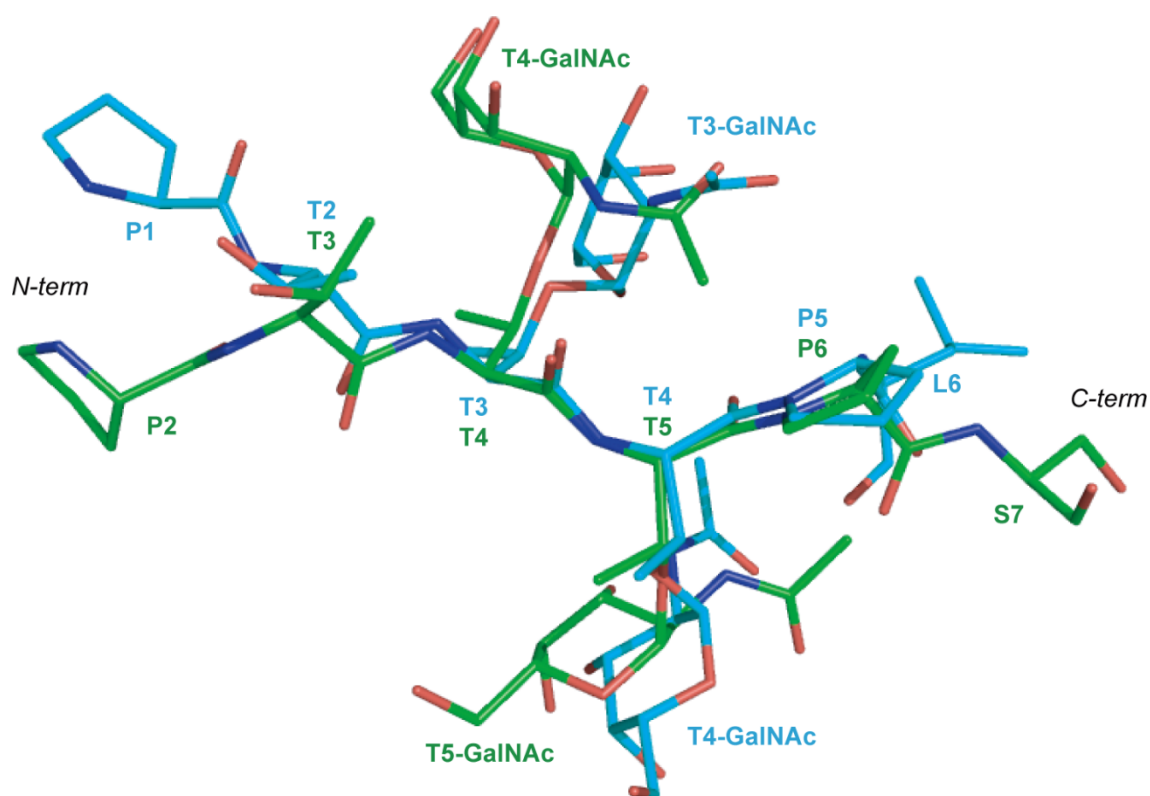


Figure 2-12. Superposition of the lowest-energy structures of Tn-MUC2 glycopeptide 5 (green) and the NMR structure of MUC2 glycopeptide Ac-PTT(Tn)T(Tn)LK-NH₂ deposited in the Protein Data Bank (PDB ID: 2LHX) (blue). The regions of ²PTT(Tn)T(Tn)PS⁷ of 5 and ¹PTT(Tn)T(Tn)PL⁶ of 2LHX are aligned with the backbone atoms of the underlined moieties with an rmsd of 0.362 Å on a PyMOL program. Both structures are represented in heavy atoms.

2-2-5-1. Peptide backbone conformations.

The backbone ϕ and ψ torsion angles and the side chain χ^1 torsion angles were obtained from the thirty lowest-energy structures of the individual MUC2 glycopeptides **5-7** (**Table 2-6**). The relationships between rigidity and extended conformation on the peptide backbones were investigated from the mean ϕ and ψ torsion angles and their standard deviations (**Table 2-7**). Increasing in the size of sugar chains results in decreasing the backbone rigidity. In particular, the conversion from **6** to **7** leads to increase in the conformational lability on the ϕ and ψ torsion angles of Pro6, which might be reflected by the lost of the NOE between T5-GalNAc and Pro6. Tn-bound MUC2 glycopeptide **5** indicates highest dihedral constrains in the regions of ${}^4\text{T}^*\text{T}^*\text{PS}^7$ and ${}^{13}\text{TT}^*\text{T}^*\text{LP}^{17}$. The peripheral β 1,3-galactose residues in **6** make $\phi, \psi(\text{Thr4})$, $\phi(\text{Thr5})$, $\phi(\text{Ser7})$, and $\phi(\text{Thr14})$ relax. The terminal α 2,3-sialic acid residues in **7** further increase the torsion angular variations on $\psi(\text{Thr4})$, $\psi(\text{Thr5})$, $\phi, \psi(\text{Pro6})$, $\phi, \psi(\text{Thr13})$, and $\psi(\text{Thr15})$. Eventually, the conserved rigidity between **5-7** were observed only on $\psi(\text{Thr14})$, $\phi(\text{Thr15})$, and $\phi, \psi(\text{Leu16})$. On the other hand, extended natures are relatively well conserved rather than rigidity. The consistent dihedral extensions were detected on the ψ torsion angles of Thr4, Thr5, Thr14, Thr15, and Leu16. β 1,3-galactosylation for **5** results in adding extended nature only on $\phi(\text{Thr5})$ and in removing it from $\phi(\text{Thr15})$, while α 2,3-sialiation for **6** enhances extended nature only on $\phi(\text{Thr13})$ and removes it from $\phi(\text{Thr5})$ and $\phi(\text{Thr14})$. Interestingly, the $\phi, \psi(\text{Thr13})$ are drastically shifted after α 2,3-sialylation. Indeed, the mean value of Thr13(ϕ, ψ) in **7** ($-93.7^\circ, 165.3^\circ$) is largely different from that in **5** ($-150.2^\circ, 51.5^\circ$) and in **6** ($-149.5^\circ, 54.9^\circ$) (**Table 2-6**). Sugar elongation is considered as a potential trigger event not only for decreasing backbone rigidity but also for shifting the backbone conformational equilibria over the region of

T*T*X and proximal residues.

residue		5 (Tn)		6 (T)		7 (sialyl T)	
Pro1	<i>f</i>	-61.7 ±	9.2	-61.9 ±	9.4	-59.7 ±	8.6
	<i>Y</i>	131.3 ±	61.0	148.0 ±	64.1	148.5 ±	78.3
Pro2	<i>f</i>	-64.2 ±	9.3	-61.6 ±	27.2	-64.7 ±	8.2
	<i>Y</i>	118.0 ±	94.6	-74.0 ±	57.0	-69.3 ±	58.5
Thr3	<i>f</i>	-95.9 ±	25.1	-74.3 ±	19.0	-69.8 ±	27.3
	<i>Y</i>	162.2 ±	15.4	134.9 ±	27.9	130.7 ±	23.2
	<i>cl</i>	8.3 ±	26.4	60.7 ±	48.4	54.9 ±	44.2
Thr4	<i>f</i>	-112.4 ±	1.6	-100.5 ±	6.4	-99.6 ±	5.7
	<i>Y</i>	162.7 ±	3.6	176.0 ±	7.1	163.8 ±	13.8
	<i>cl</i>	57.3 ±	0.4	51.4 ±	3.2	51.2 ±	3.3
Thr5	<i>f</i>	-108.3 ±	1.0	-126.3 ±	11.9	-106.6 ±	13.8
	<i>Y</i>	151.0 ±	0.5	157.9 ±	1.8	164.8 ±	8.4
	<i>cl</i>	54.8 ±	0.7	76.5 ±	4.0	49.1 ±	12.4
Pro6	<i>f</i>	-83.8 ±	1.2	-51.5 ±	2.4	-73.1 ±	25.1
	<i>Y</i>	-165.0 ±	0.4	-160.3 ±	3.6	-109.9 ±	45.8
Ser7	<i>f</i>	-79.3 ±	10.4	153.9 ±	29.5	-93.8 ±	32.8
	<i>Y</i>	130.9 ±	15.1	101.2 ±	72.1	162.6 ±	72.1
	<i>cl</i>	-172.1 ±	5.0	100.1 ±	67.2	-138.5 ±	25.7
Pro8	<i>f</i>	-62.2 ±	8.7	-61.1 ±	10.2	-67.4 ±	14.1
	<i>Y</i>	146.2 ±	77.3	161.7 ±	90.2	159.1 ±	77.9
Pro9	<i>f</i>	-77.7 ±	2.2	-77.3 ±	6.7	-78.1 ±	6.3
	<i>Y</i>	133.9 ±	20.1	138.6 ±	25.2	145.8 ±	25.6
Pro10	<i>f</i>	-65.4 ±	6.9	-74.7 ±	4.4	-74.3 ±	5.7
	<i>Y</i>	172.2 ±	9.3	175.4 ±	10.2	176.1 ±	9.8
Thr11	<i>f</i>	-98.2 ±	49.6	-106.9 ±	64.1	-95.3 ±	56.1
	<i>Y</i>	61.5 ±	79.9	133.4 ±	41.5	100.3 ±	42.1
	<i>cl</i>	-19.7 ±	56.2	111.6 ±	55.2	117.7 ±	57.4
Ser12	<i>f</i>	-114.3 ±	77.5	-51.0 ±	17.3	-63.1 ±	21.7
	<i>Y</i>	121.9 ±	57.1	125.4 ±	47.0	4.0 ±	40.2
	<i>cl</i>	-18.8 ±	98.5	178.1 ±	19.2	174.0 ±	5.8
Thr13	<i>f</i>	-150.2 ±	0.1	-149.5 ±	2.0	-93.7 ±	6.7
	<i>Y</i>	51.5 ±	1.6	54.9 ±	3.4	165.3 ±	11.7
	<i>cl</i>	1.4 ±	5.8	150.6 ±	43.3	62.6 ±	32.6
Thr14	<i>f</i>	-144.6 ±	3.2	-134.6 ±	11.5	-119.8 ±	8.1
	<i>Y</i>	161.0 ±	0.8	176.3 ±	3.7	-177.3 ±	2.8
	<i>cl</i>	83.2 ±	1.5	78.8 ±	6.4	70.3 ±	2.4
Thr15	<i>f</i>	-135.1 ±	1.1	-100.0 ±	1.3	-101.1 ±	2.6
	<i>Y</i>	156.0 ±	0.2	153.3 ±	1.7	168.2 ±	7.2
	<i>cl</i>	74.7 ±	0.8	60.8 ±	1.2	50.5 ±	1.0
Leu16	<i>f</i>	-96.2 ±	1.6	-110.3 ±	2.4	-101.4 ±	3.8
	<i>Y</i>	158.7 ±	2.2	163.0 ±	1.0	164.4 ±	2.3
	<i>cl</i>	-89.0 ±	1.1	-87.4 ±	1.3	-91.4 ±	1.7
Pro17	<i>f</i>	-74.9 ±	1.6	-86.7 ±	1.0	-75.1 ±	14.1
	<i>Y</i>	133.7 ±	18.4	128.4 ±	20.5	139.8 ±	24.9
Pro18	<i>f</i>	-65.8 ±	9.0	-65.5 ±	8.7	-65.3 ±	8.7
	<i>Y</i>	129.1 ±	29.1	-71.0 ±	80.5	162.2 ±	29.1
Thr19	<i>f</i>	-111.9 ±	74.9	-119.6 ±	62.7	-107.8 ±	87.7
	<i>cl</i>	11.3 ±	46.4	13.1 ±	55.6	-0.9 ±	44.6

Table 2-6. Average dihedral angles in the thirty lowest energy structures of MUC2 glycopeptides 5-7.

							*		*											
	P1		P2		T3		T4		T5		P6		S7		P8		P9		P10	
	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
5 (Tn)						<i>E</i>		<i>E</i>		<i>E</i>		<i>E</i>								<i>E</i>
6 (T)							<i>E</i>		<i>E</i>		<i>E</i>		<i>E</i>			<i>E</i>				<i>E</i>
7 (ST)							<i>E</i>		<i>E</i>				<i>E</i>		<i>E</i>					<i>E</i>

							*		*											
	T11		S12		T13		T14		T15		L16		P17		P18		T19			
	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ		
5 (Tn)						<i>E</i>		<i>E</i>		<i>E</i>		<i>E</i>								
6 (T)						<i>E</i>		<i>E</i>		<i>E</i>		<i>E</i>								
7 (ST)						<i>E</i>		<i>E</i>		<i>E</i>		<i>E</i>				<i>E</i>				

* O-glycosylation site

E $150^\circ < |\phi| \text{ or } |\psi| \leq +180^\circ$

$\sigma \geq 15$

$10 \leq \sigma < 15$

$5 \leq \sigma < 10$

$\sigma < 5$

Table 2-7. Relationships between rigidity and extended nature of the backbone torsion angles obtained from the thirty lowest-energy structures of MUC2 glycopeptides 5-7

2-2-5-2. Carbohydrate arrangements.

The χ^1 , ϕ_s , and ψ_s torsion angles indicative of the sugar chain arrangements and the $d1$, θ , and $d2$ parameters associated with intramolecular interaction geometries were obtained at the O-glycosylated Thr4, Thr5, Thr14, and Thr15 from the thirty lowest-energy structures of **5-7**. These allow for assessment of the effects from carbohydrate structures clustered on the T*T*X motifs.

2-2-5-2-1. χ^1 torsion angles.

The individual coordinates of $\chi^1(\text{Thr4, Thr5})$ (**Figure 2-13A**) and $\chi^1(\text{Thr14, Thr15})$ (**Figure 2-13B**) in all compounds are highly constrained between **5-7**. The mean values of $\chi^1(\text{Thr4})$ torsion angles are in a narrow range of $54.3 \pm 3.1^\circ$, while those of $\chi^1(\text{Thr5})$ torsion angles are in a wider range of $62.8 \pm 13.7^\circ$. Indeed, the fluctuation of $\chi^1(\text{Thr5})$ is suggested by larger than 20° of the increase from **5** to **6** and the decrease and from **6** to **7**. On the other hand, the mean values of $\chi^1(\text{Thr14})$ and $\chi^1(\text{Thr15})$ torsion angles similarly and progressively decrease in a range of $66.9 \pm 16.4^\circ$ with increase in the carbohydrate size. The shifts of $\chi^1(\text{Thr14, Thr15})$ along a diagonal line are different from those of $\chi^1(\text{Thr4, Thr5})$ with Thr5 preference, suggesting the influences on the χ^1 torsion angles of the T* residues from carbohydrate types are inconsistent between $^4\text{T}^*\text{T}^*\text{P}^6$ and $^{14}\text{T}^*\text{T}^*\text{L}^{16}$ motifs.

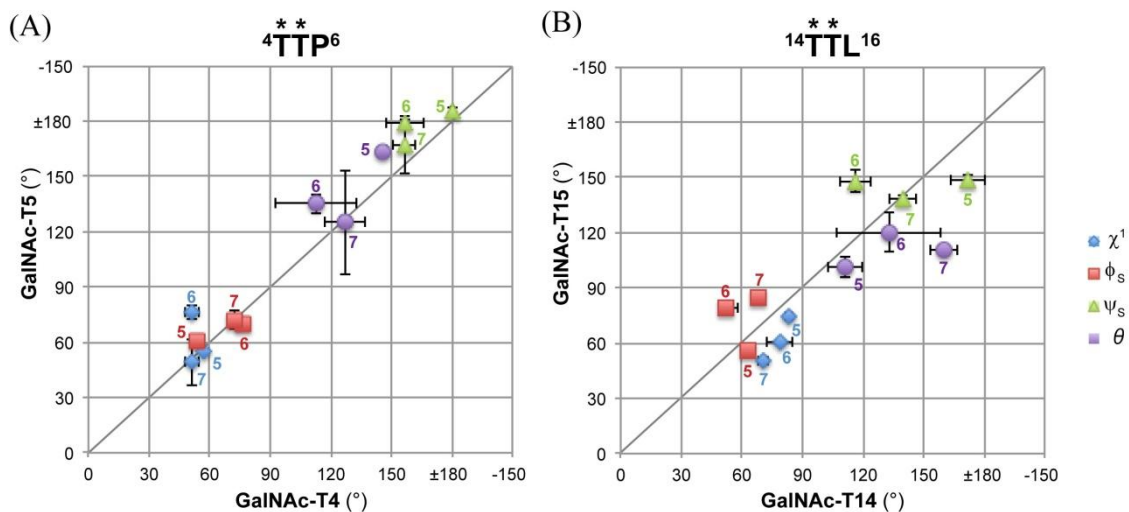


Figure 2-13. Distributions of the χ^1 , ϕ_s , and ψ_s torsion angles and the θ angle for the O-glycosylated threonines (T4, T5, T14 and T15) obtained from the thirty lowest-energy structures of mucin model glycopeptides 5-7. (A) ${}^4\text{TTP}^6$ region and (B) ${}^{14}\text{TTL}^{16}$ region.

2-2-5-2-2. ϕ_s and ψ_s torsion angles.

The individual coordinates of $\phi_s(\text{Thr4, Thr5})$ (**Figure 2-13A**) and $\phi_s(\text{Thr14, Thr15})$ (**Figure 2-13B**) are also restricted. The every $\phi_s(\text{Thr4, Thr5})$ is roughly aligned on a diagonal. The mean values of $\phi_s(\text{Thr4})$ and $\phi_s(\text{Thr5})$ in **5** are around 60° , while those in **6** and **7** are shifted to around $70\sim 75^\circ$. The mean $\phi_s(\text{Thr14, Thr15})$ in **5** is located at $(62.7^\circ, 56.0^\circ)$ with little Thr14 dominance. The $\phi_s(\text{Thr14, Thr15})$ in **6** is moved to $(52.3^\circ, 79.3^\circ)$ in the Thr15 dominant area ($\text{Thr14} < \text{Thr15}$) and that in **7** still remains in the same area with the further shift to $(68.1^\circ, 84.7^\circ)$. $\beta 1,3$ -Galactosylation causes the relatively large alterations for both $\phi_s(\text{Thr4, Thr5})$ and $\phi_s(\text{Thr14, Thr15})$ rather than $\alpha 2,3$ -sialylation.

On the other hand, the ψ_s torsion angles of O-glycosylated threonines are somewhat flexible relative to the χ^1 and the ϕ_s torsion angles of them in **5-7**. The mean

$\psi_s(\text{Thr4, Thr5})$ in **5** at around $(180^\circ, 180^\circ)$ is moved only on the $\psi_s(\text{Thr4})$ to 156.7° in **6**, while further conversion from **6** to **7** leads to decrease only on the $\psi_s(\text{Thr5})$ to 167.3° . The $\psi_s(\text{Thr14, Thr15})$ in **5-7** are shifted on the $\psi_s(\text{Thr14})$ side with remarkable preference. The mean $\psi_s(\text{Thr14, Thr15})$ in **5** at $(171.8^\circ, 148.2^\circ)$ is changed on the $\psi_s(\text{Thr14})$ to 116.4° with maintenance of the $\psi_s(\text{Thr15})$ in **6**. Conversion from **6** to **7** results in the shift of the $\psi_s(\text{Thr14, Thr15})$ to the almost even position at $(139.4^\circ, 137.9^\circ)$. Both $\beta 1,3$ -galactosylation and $\alpha 2,3$ -sialylation result in pronounced large moving on the $\psi_s(\text{Thr14})$ as compared to the ψ_s of other *O*-glycosylated threonines.

2-2-5-2-3. Intramolecular non-covalent interactions between GalNAc and peptide.

We performed quantitative analyses on structural geometries indicative of the non-covalent intramolecular interactions in the thirty lowest-energy structures in **5-7**.

2-2-5-2-4. *d1* distances and θ angles for Hydrogen bonding geometries.

As similar to **4**, the backbone ψ torsion angles of Thr5 on the ${}^4\text{T}^*\text{T}^*\text{P}^6$ motif in MUC2 glycopeptides are conserved with high mean values: 151.0° for **5**, 157.9° for **6**, and 164.8° for **7**. In addition, the following Pro6 are consistently fixed in the *trans* conformations. The arrangements of *N*-acetyl N2 and HN2 of T5-GalNAc and carbonyl O of Thr5 in **5-7** are therefore considered to adopt acceptable geometries to form hydrogen bonds.³⁵ Indeed, the θ angle ($\text{N2-HN2}\cdots\text{O}$) and *d1* distance ($\text{HN2}\cdots\text{O}$) of Thr5 are around $125\sim 163^\circ$ and $2.0\sim 3.4$ Å in average, respectively. These are also in good agreement with the reported geometries for hydrogen bond.¹⁷

The mean *d1*(Thr4, Thr5) coordinate in **5** is located at $(2.0$ Å, 2.4 Å) with high restriction (**Figure 2-14A**). The mean *d1*(Thr4, Thr5) in **6** shows the significant shift to

(3.4 Å, 3.4 Å) with high fluctuation on the Thr4 side, and that in **7** moves to (2.5 Å, 3.2 Å) with increasing in flexibility especially on the Thr5 side. The mean $\theta(\text{Thr4}, \text{Thr5})$ of **5** is located at (145.9°, 162.9°) with high convergence, while that is shifted to (112.5°, 135.1°) in **6** and to (127.1°, 125.1°) in **7** with increasing in flexibility on both sides. The increment patterns of fluctuation on the $\theta(\text{Thr4}, \text{Thr5})$ are with increasing in carbohydrate size, which seem to be correlated with those on $d1(\text{Thr4}, \text{Thr5})$. The most stabilized core structure on the ${}^4\text{T}^*\text{T}^*\text{P}^6$ motif could be formed through a probable hydrogen bond in **5** because of the combination of shortest $d1(\text{Thr4}, \text{Thr5})$ and highest $\theta(\text{Thr4}, \text{Thr5})$ coordinates with less fluctuation. They appear to reflect the highest backbone rigidity of **5** in **5-7** (**Table 2-7**).

With respect to the ${}^{14}\text{T}^*\text{T}^*\text{L}^{16}$ motif, the $d1(\text{Thr14}, \text{Thr15})$ and $\theta(\text{Thr14}, \text{Thr15})$ coordinates show T14-dominant shifts between **5-7**. The conversion from **5** to **6** results in the shift of the mean $\theta(\text{Thr14}, \text{Thr15})$ from (111.3°, 101.3°) to (132.7°, 120.2°), although the mean $d1(\text{Thr14}, \text{Thr15})$ are in maintenance of the similar positions at (3.2 Å, 2.8 Å) for **5** and (3.3 Å, 2.8 Å) for **6** (**Figure 2-14B**). These shifts are also accompanied by increases in fluctuation on the Thr14. After conversion from **6** to **7**, the mean $d1(\text{Thr14}, \text{Thr15})$ and $\theta(\text{Thr14}, \text{Thr15})$ of **7** is moved to (2.3 Å, 2.9 Å) and (159.7°, 110.3°), respectively. It suggests that the probable hydrogen bond between T14-GalNAc and Thr14 could strongly stabilize the local scaffold after $\alpha 2,3$ -sialylation. Because **7** has the combination of shortest $d1(\text{Thr14})$ and highest $\theta(\text{Thr14})$ in **5-7**. These data may reflect the maintenance of relatively high rigidity on Thr14 of the ${}^{14}\text{T}^*\text{T}^*\text{L}^{16}$ motif even after $\alpha 2,3$ -sialylation as compared to the cases of the ${}^4\text{T}^*\text{T}^*\text{P}^6$ motif (**Table 2-7**).

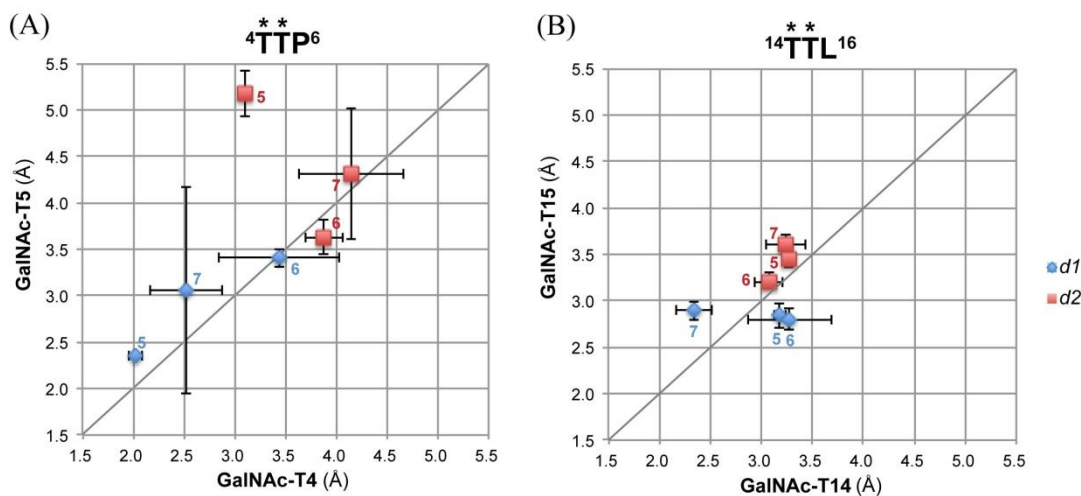


Figure 2-14. Distributions of the $d1$ and $d2$ distances for T4, T5, T14 and T15 obtained from the thirty lowest-energy structures of MUC2 glycopeptides 5-7.

2-2-5-2-5. $d2$ distances for hydrophobic interaction geometries.

The weak NOE contacts between *N*-acetyl methyl group of T(*i*)-GalNAc and the side chain of the (*i*+2) amino acid residue were detected on the both T*T*X motifs in 5-7, indicating the hydrophobic interaction. In order to define the $d2$ distance, on the basis of the NOE interactions, we selected the methyl or methylene carbon atoms of C δ (Pro6), C β (Ser7), C δ (Leu16), and C δ (Pro17) on the (*i*+2) residues as the partner atoms of *N*-acetyl methyl C_{Me} of GalNAc residues attached to Thr4, Thr5, Thr14, and Thr15, respectively. The $d2$ (Thr4,Thr5) and $d2$ (Thr14,Thr15) coordinates obtained from the thirty lowest-energy structures of 5-7 are indicated in **Figure 2-14A** and **3B**, respectively.

The mean $d2$ (Thr4,Thr5) at (3.1 Å, 5.2 Å) in **5** is moved to (3.9 Å, 3.6 Å) in **6** and to (4.1 Å, 4.3 Å) in **7**. Large fluctuation was observed only in the case of **7**. The mean $d2$ (Thr14,Thr15) at (3.3 Å, 3.4 Å) in **5** is slightly decreased to (3.1 Å, 3.2 Å) in **6**, while that is somewhat increased to (3.2 Å, 3.6 Å) in **7**. The most of them are consistent with

the pattern of medium-range interactions¹⁷, although the mean $d2(\text{Thr5})$ in **5** and in **7** seem to be little far from them. These suggest the relatively weak interactions on the ${}^4\text{T}^*\text{T}^*\text{P}^6$ motif rather than the ${}^{14}\text{TTL}^{16}$ motif. The hydrophobic interactions on the ${}^{14}\text{T}^*\text{T}^*\text{L}^{16}$ motif appear to exist nearly even between Thr14 and Thr15 with little alterations between **5-7**.

2-2-5-3. Differences in shapes on the TTX motifs between MUC2 glycopeptides.

The lowest-energy structures of **5-7** can be superimposed on the peptide backbones of ${}^4\text{T}^*\text{T}^*\text{P}^6$ and ${}^{14}\text{T}^*\text{T}^*\text{L}^{16}$ motifs with rmsd values of 0.496 Å and 0.500 Å, respectively (**Figure 2-15**). The peptide backbones in the ${}^4\text{T}^*\text{T}^*\text{PS}^7$ regions are conformationally similar to each other (**Figure 2-15A**, left), while the peptide backbones in the ${}^{13}\text{TT}^*\text{T}^*\text{LP}^{17}$ regions in **6** and **7** show definite differences over residue 13-15 in comparison to **5** (**Figure 2-15B**, left). It is mainly attributed to the shifts of the mean $\phi(\text{Thr15})$ torsion angle from -135.1° to -100.0° after $\beta 1,3$ -galactosylation (from **5** to **6**) (**Table 2-8**). Moreover, $\alpha 2,3$ -sialylation (from **6** to **7**) makes the mean Thr13(ϕ, ψ) coordinates shift from $(-149.5^\circ, 54.9^\circ)$ to $(-93.7^\circ, 165.3^\circ)$. These alterations result in distinguished movements of the upstream peptide chain (see the ${}^{10}\text{PTST}^{13}$ regions in **Figure 2-11**). In this context, sugar elongations also lead to substantial differences in the carbohydrate arrangements on Thr14 as contrasted with those on Thr15 with small deviations (**Figure 2-15B**, middle). These appear to reflect the shifts of the ψ_s torsion angles with Thr14 dominance (**Figure 2-13B**). On the other hand, the carbohydrate arrangements on Thr4 and Thr5 show only little differences between **5-7**. From these findings, the structural alterations related to Thr14 are regarded as the consequences of proper accommodation for elongated O-glycan clusters. In this context, the coupled

motions of multiple dihedral angles play an important role in rational rearrangements of peptide and sugar chains for minimizing steric crowding. Thus, carbohydrate-induced conformational changes of T*T*X motif and its surrounding would become a potential factor to influence on larger mucin structures.

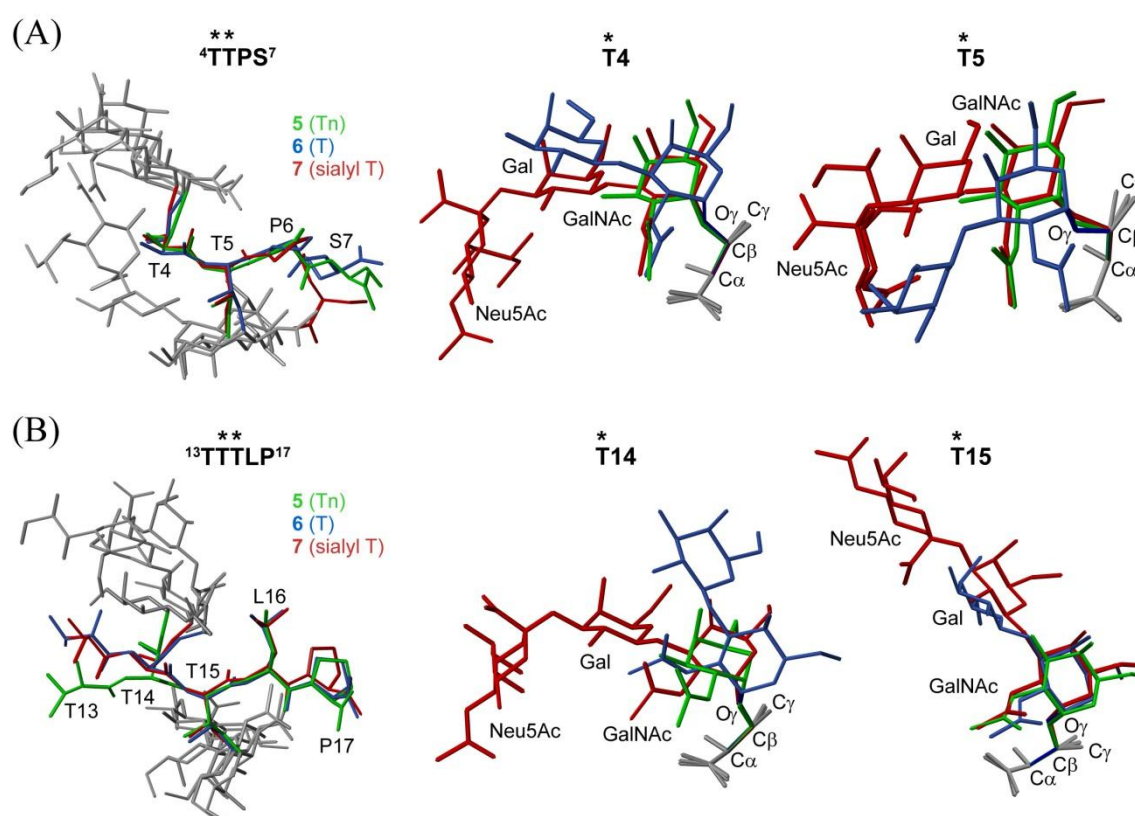


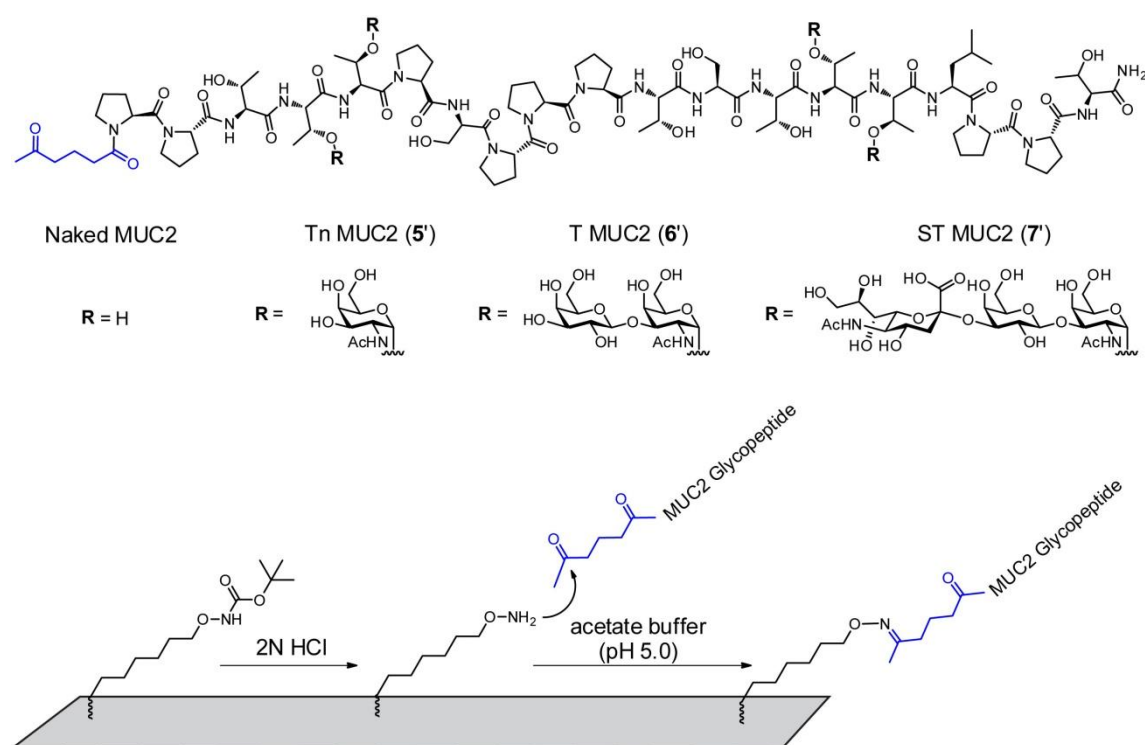
Figure 2-15. Comparison in the lowest-energy structure between MUC2 glycopeptides 5-7. (A) Overlays in the ${}^4\text{TTPS}^7$ regions. (B) Overlays in the ${}^{13}\text{TTTLP}^{17}$ regions. The glycan chains attached to T4 or T14 (middle) and T5 or T15 (right) are illustrated with the underlying threonine residues, which are superimposed with the individual threonine backbone atoms. The forced moieties in all structures are colored in green for 5, blue for 6, and red for 7. The other moieties are masked in grey.

residue		5 (Tn)		6 (T)		7 (sialyl T)	
Pro1	<i>f</i>	-61.7 ±	9.2	-61.9 ±	9.4	-59.7 ±	8.6
	<i>y</i>	131.3 ±	61.0	148.0 ±	64.1	148.5 ±	78.3
Pro2	<i>f</i>	-64.2 ±	9.3	-61.6 ±	27.2	-64.7 ±	8.2
	<i>y</i>	118.0 ±	94.6	-74.0 ±	57.0	-69.3 ±	58.5
Thr3	<i>f</i>	-95.9 ±	25.1	-74.3 ±	19.0	-69.8 ±	27.3
	<i>y</i>	162.2 ±	15.4	134.9 ±	27.9	130.7 ±	23.2
	<i>cl</i>	8.3 ±	26.4	60.7 ±	48.4	54.9 ±	44.2
Thr4	<i>f</i>	-112.4 ±	1.6	-100.5 ±	6.4	-99.6 ±	5.7
	<i>y</i>	162.7 ±	3.6	176.0 ±	7.1	163.8 ±	13.8
	<i>cl</i>	57.3 ±	0.4	51.4 ±	3.2	51.2 ±	3.3
Thr5	<i>f</i>	-108.3 ±	1.0	-126.3 ±	11.9	-106.6 ±	13.8
	<i>y</i>	151.0 ±	0.5	157.9 ±	1.8	164.8 ±	8.4
	<i>cl</i>	54.8 ±	0.7	76.5 ±	4.0	49.1 ±	12.4
Pro6	<i>f</i>	-83.8 ±	1.2	-51.5 ±	2.4	-73.1 ±	25.1
	<i>y</i>	-165.0 ±	0.4	-160.3 ±	3.6	-109.9 ±	45.8
Ser7	<i>f</i>	-79.3 ±	10.4	153.9 ±	29.5	-93.8 ±	32.8
	<i>y</i>	130.9 ±	15.1	101.2 ±	72.1	162.6 ±	72.1
	<i>cl</i>	-172.1 ±	5.0	100.1 ±	67.2	-138.5 ±	25.7
Pro8	<i>f</i>	-62.2 ±	8.7	-61.1 ±	10.2	-67.4 ±	14.1
	<i>y</i>	146.2 ±	77.3	161.7 ±	90.2	159.1 ±	77.9
Pro9	<i>f</i>	-77.7 ±	2.2	-77.3 ±	6.7	-78.1 ±	6.3
	<i>y</i>	133.9 ±	20.1	138.6 ±	25.2	145.8 ±	25.6
Pro10	<i>f</i>	-65.4 ±	6.9	-74.7 ±	4.4	-74.3 ±	5.7
	<i>y</i>	172.2 ±	9.3	175.4 ±	10.2	176.1 ±	9.8
Thr11	<i>f</i>	-98.2 ±	49.6	-106.9 ±	64.1	-95.3 ±	56.1
	<i>y</i>	61.5 ±	79.9	133.4 ±	41.5	100.3 ±	42.1
	<i>cl</i>	-19.7 ±	56.2	111.6 ±	55.2	117.7 ±	57.4
Ser12	<i>f</i>	-114.3 ±	77.5	-51.0 ±	17.3	-63.1 ±	21.7
	<i>y</i>	121.9 ±	57.1	125.4 ±	47.0	4.0 ±	40.2
	<i>cl</i>	-18.8 ±	98.5	178.1 ±	19.2	174.0 ±	5.8
Thr13	<i>f</i>	-150.2 ±	0.1	-149.5 ±	2.0	-93.7 ±	6.7
	<i>y</i>	51.5 ±	1.6	54.9 ±	3.4	165.3 ±	11.7
	<i>cl</i>	1.4 ±	5.8	150.6 ±	43.3	62.6 ±	32.6
Thr14	<i>f</i>	-144.6 ±	3.2	-134.6 ±	11.5	-119.8 ±	8.1
	<i>y</i>	161.0 ±	0.8	176.3 ±	3.7	-177.3 ±	2.8
	<i>cl</i>	83.2 ±	1.5	78.8 ±	6.4	70.3 ±	2.4
Thr15	<i>f</i>	-135.1 ±	1.1	-100.0 ±	1.3	-101.1 ±	2.6
	<i>y</i>	156.0 ±	0.2	153.3 ±	1.7	168.2 ±	7.2
	<i>cl</i>	74.7 ±	0.8	60.8 ±	1.2	50.5 ±	1.0
Leu16	<i>f</i>	-96.2 ±	1.6	-110.3 ±	2.4	-101.4 ±	3.8
	<i>y</i>	158.7 ±	2.2	163.0 ±	1.0	164.4 ±	2.3
	<i>cl</i>	-89.0 ±	1.1	-87.4 ±	1.3	-91.4 ±	1.7
Pro17	<i>f</i>	-74.9 ±	1.6	-86.7 ±	1.0	-75.1 ±	14.1
	<i>y</i>	133.7 ±	18.4	128.4 ±	20.5	139.8 ±	24.9
Pro18	<i>f</i>	-65.8 ±	9.0	-65.5 ±	8.7	-65.3 ±	8.7
	<i>y</i>	129.1 ±	29.1	-71.0 ±	80.5	162.2 ±	29.1
Thr19	<i>f</i>	-111.9 ±	74.9	-119.6 ±	62.7	-107.8 ±	87.7
	<i>cl</i>	11.3 ±	46.4	13.1 ±	55.6	-0.9 ±	44.6

Table 2-8. Average dihedral angles for peptide backbone in MUC2 glycopeptides
5-7

2-2-6. Auto-antibody analysis against MUC2 glycopeptides in cancer patients

NMR conformational analyses demonstrated that tumor associated O-GlcNAc modifications at consecutive threonine residues induced extended and rigid core TTX structure by hydrogen bond and hydrophobic interaction between carbohydrate and peptide chain. From these results, we hypothesized that this core TTX motif could become good acceptor motif against antibody, in other words, it could be epitope. To check the antigenicity of synthetic MUC2 glycopeptides (5-7), microarray plate displaying each MUC2 glycopeptide was prepared.



Scheme 2-2. Schematic protocol for chemoselective immobilization of keto-functionalized MUC2 glycopeptides and on the basis of a concept of “Glycoblotting method”

For the preparation of glycopeptide-immobilized microarray plate, keto-functionalized MUC2 glycopeptides which were modified by 5-oxohexanoic acid at *N*-terminal instead of acetyl group as **5-7** were synthesized using solid-phase synthesis (Scheme 2-2).³⁹ A microarray slide surface were coated by N-protected aminooxy (AO)/phosphorylcholine (PC)-copolymer. This N-protected AO groups were deprotected by acid treatment. The keto-functionalized MUC2 glycopeptides were covalently immobilized on the slide surface forming oxime bond with coated aminooxy group in acetate buffer (pH 5.0). Four types of MUC2 glycopeptides, Naked, Tn, T, or sialyl T modified at consecutive Thr, were designed by focusing on glycan structures, and they were immobilized on the microarray surface at the concentration of 0.05 mM, 0.5 mM, and 5 mM, respectively.

To check the antigenicity of MUC2 glycopeptides, human sera from eight colon cancer patients and eight healthy volunteers were tested with MUC2 microarray because MUC2 have been known as biomarker for colon cancer. Eight patient serums and eight healthy human serums were obtained (**Table 2-8**). As shown in **Table 2-8**, serums from stage-1 and stage-4 patients were tested in this study to compare stage-specific auto-antibody reactivity against synthetic MUC2 glycopeptides. These human serum solutions at 100-fold dilution were applied to microarray slide displaying MUC2 glycopeptides, following that the solution of CyTM3-labeled goat anti-human IgG was added onto the slide to monitor level of IgG auto-antibodies.

Entry	Name	Age	Sex	Status	Stage		Entry	Name	Age	Sex	Status
1	Stage4_M40	40s	Male	Colorectal	4		9	Healthy_M40	40s	Male	Healthy
2	Stage4_M60	60s			4		10	Healthy_M60	60s		
3	Stage4_M60	60s			4		11	Healthy_M60	60s		
4	Stage1_M60	60s			1		12	Healthy_M60	60s		
5	Stage1_F50	50s	Female	Cancer	1		13	Healthy_F50	50s	Female	
6	Stage1_F60	60s			1		14	Healthy_F60	60s		
7	Stage4_F60	60s			4		15	Healthy_F60	60s		
8	Stage4_F60	60s			4		16	Healthy_F60	60s		

Table 2-8. List of tested human serum from healthy and colorectal cancer patients

As anticipated, microarray experiments clearly indicated that MUC2 glycopeptides having rigid TTX motif showed significant level of IgG auto-antibody in both sera from healthy and patient. In **Figure 2-16**, microarray arrangement and mean value of fluorescence intensities with standard deviation (n=6) were summarized. In this results, significant level of auto-antibody against naked and Tn MUC2 were observed in all eight samples in healthy human serum, while T and sialyl T MUC2 showed less auto-antibody level. In other words, naked and Tn-modified MUC2 glycopeptide might be unnatural form and recognized as antigen in healthy human body, and MUC2 glycopeptide carrying relatively matured O-glycan, T and sialyl T antigen, might not be recognized as antigen. It suggested that O-glycan elongation might regulate peptide antigenicity by shielding peptide backbone.

Interestingly, we are able to find stage-specific auto-antibody reactivity in cancer patient test. In stage4 patients, all four types of MUC2 glycopeptide showed no or quite

less level of auto-antibody recognition except for one patient (Stage4_M40). However, all three stage1 patient serums indicated that higher level of auto-antibody recognition against naked, Tn, and T MUC2 compared with healthy or stage4 serum. In all cases, sialyl T MUC2 cannot become antigen of auto-antibody, suggesting that MUC2 glycopeptide carrying sialyl T glycans might be naturally present and have biological functions. These microarray results suggested that auto-antibody against naked, Tn, or T MUC2 glycopeptide might be important to slow the cancer progression. From this view point, these synthetic MUC2 glycopeptides can be applicable for colon cancer vaccine.

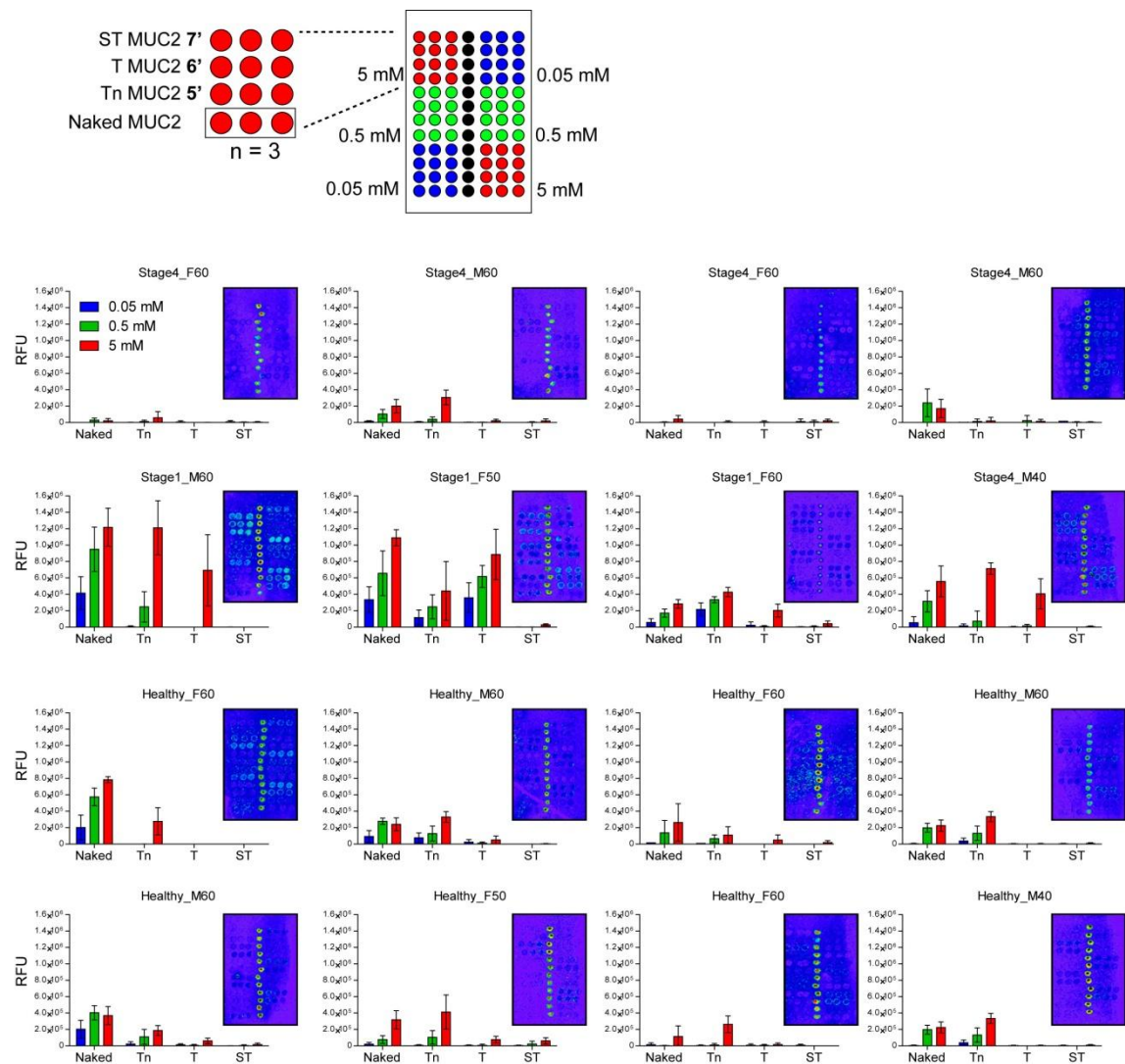


Figure 2-16. Microarray analysis of IgG auto-antibody from healthy or cancer patients against synthetic MUC2 glycopeptides

2-3. Conclusion

We carried out the NMR-based conformational study for two series of mucin-type glycopeptides focused on the T*T*X motif carrying clustered O-glycans at consecutive threonine. The consistent structures with high rigidity and extended nature have been elucidated on this segment with the additional residue at the C-terminal side, which is one of hydrophobic amino acids (X = A, V, L, or P) with high abundance in native mucin tandem repeats. The allowable geometries strongly suggest the existences of hydrogen bonds and hydrophobic interactions between the GalNAc residues and the proximal peptide, which would help for establishing the unique structural feature. In 9-mer mucin-like model glycopeptides, bulkiness of hydrophobic X affect the conformational equilibria lying on the T*T*X motif, which results in different promotion of the structural propensities on the peptide and sugar chains. In 19-mer MUC2 tandem repeat glycopeptides, sugar elongations to the clustered O-glycans increase peptide backbone lability on the T*T*X motif as well as its adjacent residues. Along with this, drastic conformational changes are made on the peptide backbone, which accompany alterations of the carbohydrate orientations. High-level coordination among multiple torsion angles related to the peptide backbone and the GalNAc-peptide linkages would aid in the proper accommodation for elongated O-glycan clusters on the mucin-type glycopeptides carrying the T*T*X motif.

Microarray analyses with human sera from healthy and colon cancer patients revealed that synthetic MUC2 glycopeptides having rigid T*T*X motif showed significant level of IgG auto-antibody. Importantly, these auto-antibody levels were decreased in stage4 patients although healthy and stage1 patients showed certain level of auto-antibody. These significant differences of auto-antibody production would help for development

of new synthetic vaccines toward improved efficiency in presentation of clustered carbohydrate epitopes to the immune system.

2-4. Experimental Section

General methods and materials

All commercially available solvents and reagents were of the highest purity available and used without further purification. TentaGel® S RAM resin was purchased from Hipec Laboratories (Kyoto, Japan) and Rink Amide-ChemMatrix® resin was purchased from Biotage Japan Ltd. (Tokyo, Japan). *N*^α-Fmoc-amino acid derivatives except for glycosylated amino acids and recombinant rat α 2,3-(O)-sialyltransferase (α 2,3-SiaT) were purchased from Merck Millipore (Darmstadt, Germany). Fmoc-Thr(Ac₃GalNAc α 1)-OH (**8**) and Fmoc-Thr(Ac₄Gal β 1-3Ac₂GalNAc α 1)-OH (**9**) were synthesized according to the method reported previously.⁴⁰ 1-[Bis(dimethylamino)methyliumyl]-1H-benzotriazole-3-oxide hexafluorophosphate (HBTU), 1-hydroxybenzotriazole monohydrate (HOBt) and *N,N*-diisopropylethylamine (DIEA) were purchased from Kokusan Chemical Co., Ltd (Tokyo, Japan). *N,N*-dimethylformamide (DMF) and 20% piperidine in DMF was purchased from Watanabe Chemical Industries, Ltd. (Hiroshima, Japan). Acetic anhydride and 2,2,2-trifluoroacetic acid (TFA) were from Wako Pure Chemical Industries, Ltd. (Osaka, Japan). Cytidine-5'-monophospho-*N*-acetylneuraminic acid, disodium salt (CMP-NANA) was obtained from Yamasa Corporation (Chiba, Japan). All solid-phase reactions for glycopeptides synthesis were performed manually in a polypropylene tube equipped with a filter (LibraTube®, Hipec Laboratories) under microwave irradiation. The microwave was irradiated during coupling reactions and Fmoc removal using temperature control at 50 °C. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was obtained by a Bruker UltraFLEX I using 2,5-dihydroxybenzoic acid as matrix. High-resolution electrospray ionization mass

spectrometry (ESI-HRMS) was measured by JEOL JMS-700TZ. Reverse-phase HPLC (RP-HPLC) separation and analysis were performed on a Hitachi system equipped with L-6250 intelligent pump and L-7400 UV detector using a reverse-phase C18 column Inertsil ODS-3 250×20 mmI.D. (for **1-6**) or 250×4.6 mmI.D. (for **7**) (GL Sciences Inc., Tokyo, Japan) or on a Hitachi system equipped with L-7100 pump and L-7405 UV detector using Resolux 120 C18 250×4.6 mmI.D. (Biotage Japan Ltd., Tokyo, Japan).

CyTM3-labeled goat anti-human IgG (H + L) was from Kirkegaard & Perry Laboratories, Inc. (Gaithersburg, MD, USA). Amersham CyTM3 mono-reactive dye pack were from GE Healthcare Bio-Sciences KK (Tokyo, Japan). Microarray slides were basically manufactured by Sumitomo Bakelite Co., Ltd. (Tokyo, Japan) as follows: Glass-like raw materials made of cyclic polyolefin were molded into slides. The surface of naked slides were coated by methacrylic polymer, (2-methacryloyloxyethylphosphorylcholinecyclohexylmethacrylate-N-[2-[2-[2-(t-butoxycarbonylaminoxyacetylaminomethoxy)ethoxy]ethyl]-methacrylamido copolymer (N-protected AO/PC-copolymer).

“Reaction buffer” that specified in the present study is a solution of 50 mM Tris–HCl, 100 mM NaCl, 1 mM CaCl₂, 1 mM MnCl₂, 1 mM MgCl₂, 0.05% Tween 20, pH 7.4. “Washing buffer” that specified in this paper is a solution of 50 mM Tris–HCl, 100 mM NaCl, 1 mM CaCl₂, 1 mM MnCl₂, 1 mM MgCl₂, 0.05% Triton X-100, pH 7.4. Fluorescence images of microarray slides after assays were measured at 10 μm resolution on a Typhoon Trio plus variable mode imager (GE Healthcare) with a green (532 nm) laser and a 580 BP 30 filter at a PMT voltage of 600 V and normal sensitivity. The digital images of fluorescence responses were analyzed using ArrayVisionTM software version 8.0 (GE Healthcare).

Solid-phase synthesis of glycopeptides (1-6).

Solid-phase syntheses of glycopeptide **1-6** were performed as described below. TentaGel S RAM resin (0.25 mmol/g, 12 μ mol, for **1-4**) or Rink amide ChemMatrix resin (0.48 mmol/g, 20 μ mol, for **5** and **6**) was swollen with dichloromethane for 1 h. Fmoc group was removed with 20% piperidine in DMF (1 mL) for 3 min under microwave irradiation. Standard N^{α} -Fmoc-amino acid (4.0 equiv.) was coupled with HBTU (4.0 equiv.), HOBt (4.0 equiv.) and DIEA (8.0 equiv.) in DMF (230 μ l for **1-4**, 300 μ l for **5** and **6**) for 9 min under microwave irradiation. In the case of glycosylated amino acid, N^{α} -Fmoc-Thr(Ac₄Gal β 1-3Ac₂GalNAc α 1)-OH (1.2 equiv. for **1-4** and **6**) or Fmoc-Thr(Ac₃GalNAc α 1)-OH (1.2 equiv. for **5**) was coupled with PyBOP (1.2 equiv.), HOAt (1.2 equiv.) and DIEA (3.0 equiv.) in DMF (70 μ l). After agitation for 9 min under microwave irradiation, the extra PyBOP and HOAt (1.2 equiv., each) were treated for another 9 min period in the same microwave-assisted fashion. At the end of solid-phase synthesis, an *N*-terminal amino group of peptidyl-resin was capped with acetic anhydride with DIEA in DMF for 30 sec. For cleavage from resin and removal of acid-labile protective groups, a cocktail of TFA/triisopropylsilane/H₂O (95:2.5:2.5, v/v/v, 1 mL) was treated to resin for 1 h at ambient temperature. The resin was filtered off and the glycopeptide was precipitated using cold *tert*-butylmethyl ether (5 ml). After centrifugation (3,000 rpm, 4 °C, 1 min), supernatant was carefully removed by decantation. The precipitate was dissolved in 50% acetonitrile aqueous (5 mL), and then lyophilized. The dried precipitate was dissolved in methanol and the solution was adjusted and kept to pH 12.5 with 1 M NaOH. After stirring for 1 h at ambient temperature, the solution was neutralized with 20% acetic acid in methanol and the solvent was removed by air blow. Preparative RP-HPLC gave purified compound **1-6** in

24% (4.8 mg, 2.9 μmol , t_R = 24.7 min; **1**), 24% (4.8 mg, 2.9 μmol , t_R = 21.4 min; **2**), 17% (3.7 mg, 2.0 μmol , t_R = 26.4 min; **3**), 10% (2 mg, 1.2 μmol , t_R = 24.8 min; **4**), 23% (12.6 mg, 4.6 μmol , t_R = 32.3 min; **5**) and 14% (15.6 mg, 2.8 μmol , t_R = 24.7 min; **6**) overall yields (calculated from the loading amount of initial resin), respectively. RP-HPLC conditions are shown as below: ambient temperature and flow rate of 5 mL/min, UV detection at 220 nm, eluent A: 0.1% TFA in water, eluent B: 0.1% TFA in acetonitrile, and the liner gradient from 3 to 40% B over 50 min (for **1-4**) or from 3 to 30% B over 50 min (for **5** and **6**). The purified **1-6** were characterized by analytical RP-HPLC (**Figure S1**, Supporting Information) and high-resolution ESI-MS. ESI-HRMS:

- (1) $\text{C}_{63}\text{H}_{108}\text{N}_{12}\text{Na}_2\text{O}_{38}$ $[\text{M}+2\text{Na}]^{2+}$ calcd (m/z) 843.3341, found (m/z) 843.3358;
- (2) $\text{C}_{65}\text{H}_{112}\text{N}_{12}\text{Na}_2\text{O}_{38}$ $[\text{M}+2\text{Na}]^{2+}$ calcd (m/z) 857.3498, found (m/z) 857.3520;
- (3) $\text{C}_{66}\text{H}_{114}\text{N}_{12}\text{Na}_2\text{O}_{38}$ $[\text{M}+2\text{Na}]^{2+}$ calcd (m/z) 864.3576, found (m/z) 864.3584;
- (4) $\text{C}_{65}\text{H}_{110}\text{N}_{12}\text{Na}_2\text{O}_{38}$ $[\text{M}+2\text{Na}]^{2+}$ calcd (m/z) 856.3420, found (m/z) 856.3382;
- (5) $\text{C}_{118}\text{H}_{190}\text{N}_{24}\text{Na}_3\text{O}_{50}$ $[\text{M}+3\text{Na}]^{3+}$; calcd (m/z) 937.4252, found (m/z) 937.42353.
- (6) $\text{C}_{142}\text{H}_{230}\text{N}_{24}\text{Na}_3\text{O}_{70}$ $[\text{M}+3\text{Na}]^{3+}$ calcd (m/z) 1153.4962, found (m/z) 1153.4946.

Enzymatic synthesis of sialyl glycopeptide (7).

Compound **6** (3.4 mg, 1 μmol) was dissolved in a total volume of 1 mL of 50 mM HEPES buffer (pH 7.0) and 10 mM MnCl_2 . To the solution were added $\alpha 2,3\text{-SiaT}$ (60 mU) and CMP-NANA (7.6 mg, 12 μmol) and the mixture was incubated for 72 h at ambient temperature. Preparative RP-HPLC gave purified compound **7** in 46% (2.1 mg, 0.46 μmol , t_R = 19.07 min) yield. Preparative RP-HPLC condition is shown as below: ambient temperature and flow rate of 1 mL/min, UV detection at 220 nm, eluent A:

0.1% 25 mM ammonium acetate buffer (pH 5.0), eluent B: 10% eluent A in acetonitrile, and the linear gradient from 3 to 12% B over 20 min and then 12% of B eluted constantly over 30 min. The purified **7** was analyzed by RP-HPLC and high-resolution ESI-HRMS. ESI-HRMS (**7**): $C_{186}H_{298}N_{28}O_{102}$ $[M-3H]^{3-}$ calcd (m/z) 1138.97480, found (m/z) 1138.9731.

Nuclear magnetic resonance (NMR) spectroscopic experiments.

NMR experiments of **1-7** were performed on a Bruker Avance 600 spectrometer at 600.03 MHz for proton frequency equipped with a cryoprobe. Glycopeptide **1-7** were dissolved in 99.9% D_2O (300 μ l) or in a mixed solution of 90% H_2O and 10% D_2O (300 μ l) at the concentration of 1.5~2 mM respectively. The solution was adjusted to pH 5.8 with 0.1 M NaOH. *Shigemi* thin-walled micro NMR tubes were used to contain samples in the NMR spectroscopy. All NMR spectra were collected at 300 K without spinning. Data acquisition was performed with the Bruker TopSpin 2.1 software package. Two-dimensional⁴¹ homonuclear DQF-COSY,^{42,43} TOCSY⁴⁴ with MLEV-17 sequence⁴⁵ and NOESY⁴⁶ spectra were recorded in the indirect dimension using States-TPPI phase cycling. Two-dimensional heteronuclear ^{13}C -edited HSQC and HSQC-TOCSY measurements were recorded with echo-antiecho mode. TOCSY experiments were applied for a spin-locking time of 60 ms, and NOESY experiments were carried out with mixing time of 100, 200 and 400 ms. The suppression of water signal was performed by presaturation during 2 s relaxation delay and by a 3-9-19 WATERGATE pulse sequence with a field gradient.⁴⁷ TOCSY and NOESY spectra were acquired with 2048 by 512 frequency data point and were zero-filled to yield 2048 by 2048 data matrices. DQF-COSY with 16384 by 512 frequency data points was also recorded and

zero-filled to yield 16384 by 16384 matrix in order to measure the coupling constant. The sweep widths of 7183.91 Hz were applied. Time domain data in both dimensions were multiplied by a sine bell window function with a 90° phase shift prior to Fourier transformation. All NMR data were processed by and NMRPipe⁴⁸ software and analyzed using a Sparky program.⁴⁹

Structure Calculation

Three-dimensional structures of synthesized glycopeptide **1-7** were calculated using the CNS 1.1 program⁵⁰ with standard protocols for distance geometry-simulated annealing and refinement. Distance restraints for calculations were estimated from the cross-peak intensities in NOESY spectra with a mixing time of 400 ms. The estimated restraints were classified into four categories: strong (1.6-2.6 Å), medium (1.6-3.5 Å), weak (1.6-5.0 Å) and very weak (1.6-6.0 Å). In the first stage of structure determination, the structures of glycopeptides were calculated using only interproton distance information. After the validation of fulfilling distance restraints for the obtained structures, the restraints of the dihedral angle ϕ and χ^1 were adopted for further structural calculation. When the coupling constant $^3J_{HN\alpha}$ was more than 8.0 Hz and less than 6.0 Hz, the dihedral angle ϕ was constricted to $-120 \pm 30^\circ$ and $-60 \pm 30^\circ$, respectively. The conformation of sugar ring was fixed to the chair conformation. All analyses of rmsd values and the solution structures of glycopeptides were performed with PROCHECK⁵¹ and MOLMOL⁵² program.

Microarray analyses of synthetic MUC2 glycopeptides

Human serum (3 μ l) were diluted with reaction buffer (297 μ l) and the solutions were

subjected to centrifugal filtration (12,000 $\times$ g, 3 min, 4 °C) with Amicon Ultrafree®-MC centrifugal filter units (pore size 0.45 μ m, Durapore PVDF membrane) that purchased from Merck Millipore (Billerica, MA, USA). Triplicates of 0.05 mM, 0.5 mM and 5 mM of keto-functionalized MUC2 in 25 mM AcOH/Pyridine (pH 5.4) with 0.005% (w/v) Triton X-100 were spotted on BioChip Arrayer (Filgen Inc., Aichi, Japan) with a 0.8-mm pitch using a Filgen solid pin (400 μ m pin diameter). Amersham Cy3mono-reactive dye (1 μ g/ml) that dissolved in Milli-Q water was also spotted on the same slides as grid markers. The slides after spotting were placed in a dry oven for 1 h at 80 °C to complete oxime bond formation according to the general condition used in glycoblotting of various hemiacetals. After washing once with Milli-Q water, the slides were immersed in an aqueous solution of 10 mg ml⁻¹ succinic anhydride for 1 h at ambient temperature for capping free aminooxy groups remained on the slide surface. The slides after washing twice with Milli-Q water were dried by centrifugation (2000 rpm, 2 min, at ambient temperature). After the slides were outlined with Super PAP Pen to create a hydrophobic barrier for processing slides individually, 70 μ l of human serum solution was applied to each well and the slides were allowed to remain for 2 h at ambient temperature. After removal of the solutions by aspiration, each well was washed with washing buffer (50 μ l) for 2 min three times and 1 μ g/ml CyTM3-labeled goat anti-human IgG conjugate in reaction buffer (70 μ l) was added. After standing for 1 h at ambient temperature, each well was washed as the same manner, and the slides were rinsed with Milli-Q water for 1 $\times$ 1 min. The slides dried by centrifugation (2000 rpm, 2 min, ambient temperature) were subjected to scan of fluorescent images on a Typhoon Trio plus and the fluorescent responses were analyzed using an ArrayVision software.

2-5. Reference

- (1) Rose, M. C.; Voynow, J. a. Respiratory Tract Mucin Genes and Mucin Glycoproteins in Health and Disease. *Physiol. Rev.* **2006**, *86* (1), 245–278.
- (2) Tarp, M. A.; Clausen, H. Mucin-Type O-Glycosylation and Its Potential Use in Drug and Vaccine Development. *Biochim. Biophys. Acta - Gen. Subj.* **2008**, *1780*, 546–563.
- (3) Hatstrup, C. L.; Gendler, S. J. Structure and Function of the Cell Surface (tethered) Mucins. *Annu. Rev. Physiol.* **2008**, *70*, 431–457.
- (4) Apostolopoulos, V.; McKenzie, I. F. Cellular Mucins: Targets for Immunotherapy. *Crit Rev Immunol.* **1994**, *14* (3-4), 293–309.
- (5) Ogata, S.; Koganty, R.; Reddish, M.; Longenecker, B. M.; Chen, A.; Perez, C.; Itzkowitz, S. H. Different Modes of Sialyl-Tn Expression during Malignant Transformation of Human Colonic Mucosa. *Glycoconj. J.* **1998**, *15* (1), 29–35.
- (6) Reddish, M. A.; Jackson, L.; Koganty, R. R.; Qiu, D.; Hong, W.; Longenecker, B. M. Specificities of Anti-Sialyl-Tn and Anti-Tn Monoclonal Antibodies Generated Using Novel Clustered Synthetic Glycopeptide Epitopes. *Glycoconj J* **1997**, *14* (5), 549–560.
- (7) Zhang, S.; Walberg, L. a.; Ogata, S.; Itzkowitz, S. H.; Koganty, R. R.; Reddish, M.; Gandhi, S. S.; Longenecker, B. M.; Lloyd, K. O.; Livingston, P. O. Immune Sera and Monoclonal Antibodies Define Two Configurations for the Sialyl Tn Tumor Antigen. *Cancer Res.* **1995**, *55* (C), 3364–3368.
- (8) Tanaka, N.; Nakada, H.; Inoue, M.; Yamashina, I. Binding Characteristics of an Anti-Siaalpha2-6GalNAcalpha-Ser/Thr (sialyl Tn) Monoclonal Antibody (MLS 132). *Eur. J. Biochem.* **1999**, *263* (1), 27–32.
- (9) Inoue, M.; Nakada, H.; Tanaka, N.; Yamashina, I. Tn Antigen Is Expressed on Leukosialin from T-Lymphoid Cells. *Cancer Res* **1994**, *54* (1), 85–88.
- (10) Nakada, H.; Inoue, M.; Numata, Y.; Tanaka, N.; Funakoshi, I.; Fukui, S.; Mellors, A.; Yamashina, I. Epitopic Structure of Tn Glycophorin A for an Anti-Tn Antibody (MLS 128). *Proc Natl Acad Sci U S A* **1993**, *90* (6), 2495–2499.

- (11) Nakada, H.; Numata, Y.; Inoue, M.; Tanaka, N.; Kitagawa, H.; Funakoshi, I.; Fukui, S.; Yamashina, I. Elucidation of an Essential Structure Recognized by an Anti-GalNAc alpha-Ser(Thr) Monoclonal Antibody (MLS 128). *J. Biol. Chem.* **1991**, 266 (19), 12402–12405.
- (12) MacLean, G. D.; Reddish, M.; Koganty, R. R.; Wong, T.; Gandhi, S.; Smolenski, M.; Samuel, J.; Nabholz, J. M.; Longenecker, B. M. Immunization of Breast Cancer Patients Using a Synthetic Sialyl-Tn Glycoconjugate plus Detox Adjuvant. *Cancer Immunol. Immunother.* **1993**, 36 (4), 215–222.
- (13) Reddish, M. A.; Maclean, G. D.; Poppema, S.; Berg, A.; Longenecker, B. M. Pre-Immunotherapy Serum CA27.29 (MUC-1) Mucin Level and CD69+ Lymphocytes Correlate with Effects of Theratope Sialyl-Tn-KLH Cancer Vaccine in Active Specific Immunotherapy. *Cancer Immunol Immunother* **1996**, 42, 303–309.
- (14) Lo-man, R.; Bay, S.; Vichier-Guerre, S.; Deriaud, E.; Canctacuzene, D.; Leclerc, C. A Fully Synthetic Immunogen Carrying a Carcinoma-Associated Carbohydrate for Active Specific Immunotherapy. *Cancer Res.* **1999**, 59, 1520–1524.
- (15) Ragupathi, G.; Howard, L.; Cappello, S.; Koganty, R. R.; Qiu, D.; Longenecker, B. M.; Reddish, M. a; Lloyd, K. O.; Livingston, P. O. Vaccines Prepared with Sialyl-Tn and Sialyl-Tn Trimers Using the 4-(4-Maleimidomethyl)cyclohexane-1-Carboxyl Hydrazide Linker Group Result in Optimal Antibody Titers against Ovine Submaxillary Mucin and Sialyl-Tn-Positive Tumor Cells. *Cancer Immunol. Immunother.* **1999**, 48 (1), 1–8.
- (16) Gilewski, T. A; Ragupathi, G.; Dickler, M.; Powell, S.; Bhuta, S.; Panageas, K.; Koganty, R. R.; Chin-Eng, J.; Hudis, C.; Norton, L.; Houghton, A.N.; Livingston, P. O. Immunization of High-Risk Breast Cancer Patients with Clustered sTn-KLH Conjugate plus the Immunologic Adjuvant QS-21. *Clin. Cancer Res.* **2007**, 13 (10), 2977–2985.
- (17) Coltart, D. M.; Royyuru, A. K.; Williams, L. J.; Glunz, P. W.; Sames, D.;

- Kuduk, S. D.; Schwarz, J. B.; Chen, X.-T.; Danishefsky, S. J.; Live, D. H. Principles of Mucin Architecture: Structural Studies on Synthetic Glycopeptides Bearing Clustered Mono-, Di-, Tri-, and Hexasaccharide Glycodomains. *J. Am. Chem. Soc.* **2002**, *124* (33), 9833–9844.
- (18) Borgert, A.; Heimbürg-Molinaro, J.; Song, X.; Lasanajak, Y.; Ju, T.; Liu, M.; Thompson, P.; Ragupathi, G.; Barany, G.; Smith, D. F.; Cummings, R. D.; Live, D. Deciphering Structural Elements of Mucin Glycoprotein Recognition. *ACS Chem. Biol.* **2012**, *7*, 1031–1039.
- (19) Hashimoto, R.; Fujitani, N.; Takegawa, Y.; Kurogochi, M.; Matsushita, T.; Naruchi, K.; Ohyabu, N.; Hinou, H.; Gao, X. D.; Manri, N.; Satake, H.; Kaneko, A.; Sakamoto, T.; Nishimura, S.-I. An Efficient Approach for the Characterization of Mucin-Type Glycopeptides: The Effect of O-Glycosylation on the Conformation of Synthetic Mucin Peptides. *Chem. - A Eur. J.* **2011**, *17* (8), 2393–2404.
- (20) Itoh, Y.; Kamata-Sakurai, M.; Denda-Nagai, K.; Nagai, S.; Tsuiji, M.; Ishii-Schrade, K.; Okada, K.; Goto, A.; Fukayama, M.; Irimura, T. Identification and Expression of Human epiglycanin/MUC21: A Novel Transmembrane Mucin. *Glycobiology* **2008**, *18* (1), 74–83.
- (21) Matsushita, T.; Hinou, H.; Kurogochi, M.; Shimizu, H.; Nishimura, S.-I. Rapid Microwave-Assisted Solid-Phase Glycopeptide Synthesis. *Org. Lett.* **2005**, *7* (5), 877–880.
- (22) Garcia-Martin, F.; Hinou, H.; Matsushita, T.; Hayakawa, S.; Nishimura, S.-I. An Efficient Protocol for the Solid-Phase Synthesis of Glycopeptides under Microwave Irradiation. *Org. Biomol. Chem.* **2012**, *10* (8), 1612.
- (23) Pardi, A.; Billeter, M.; Wüthrich, K. Calibration of the angular dependence of the amide proton-C α proton coupling constants, $^3J_{\text{HN}\alpha}$, in a globular protein. Use of $^3J_{\text{HN}\alpha}$ for identification of helical secondary structure. *J Mol Biol.* **1984**, *180*, 741–751.
- (24) Wang, A. C.; Bax, A. Determination of the backbone dihedral angles ϕ in human ubiquitin from reparametrized empirical Karplus equations. *J. Am. Chem. Soc.* **1996**, *118*, 2483–2494.

- (25) Wüthrich, K. NMR of proteins and nucleic acids; Wiley-Interscience, **1986**.
- (26) Dyson, H. J.; Wright, P. E. Defining solution conformations of small linear peptides. *Annu Rev Biophys Biophys Chem* **1991**, 20, 519–538.
- (27) Liang, R.; Andreotti, A. H.; Kahne, D. Sensitivity of glycopeptide conformation to carbohydrate chain length. *J. Am. Chem. Soc.* **1995**, 117, 10395–10396.
- (28) Live, D. H.; Williams, L. J.; Kuduk, S. D.; Schwarz, J. B.; Glunz, P. W.; Chen, X. T.; Sames, D.; Kumar, R. A.; Danishefsky, S. J. Probing cell-surface architecture through synthesis: an NMR-determined structural motif for tumor-associated mucins. *Proc Natl Acad Sci USA* **1999**, 96, 3489–3493.
- (29) Schuman, J.; Qiu, D.; Koganty, R. R.; Longenecker, B. M.; Campbell, A. P. Glycosylations versus conformational preferences of cancer associated mucin core. *Glycoconj. J.* **2000**, 17, 835–848.
- (30) Kirnarsky, L.; Prakash, O.; Vogen, S. M.; Nomoto, M.; Hollingsworth, M. A.; Sherman, S. Structural effects of O-glycosylation on a 15-residue peptide from the mucin (MUC1) core protein. *Biochemistry* **2000**, 39, 12076–12082.
- (31) Dziadek, S.; Griesinger, C.; Kunz, H.; Reinscheid, U. M. Synthesis and structural model of an alpha(2,6)-sialyl-t glycosylated MUC1 eicosapeptide under physiological conditions. *Chemistry* **2006**, 12, 4981–4993.
- (32) Corzana, F.; Busto, J. H.; García de Luis, M.; Jiménez-Barbero, J.; Avenoza, A.; Peregrina, J. M. The Nature and Sequence of the Amino Acid Aglycone Strongly Modulates the Conformation and Dynamics Effects of Tn Antigen's Clusters. *Chem. - A Eur. J.* **2009**, 15 (15), 3863–3874.
- (33) Mimura, Y.; Yamamoto, Y.; Inoue, Y.; Chûjô, R. N.m.r. study of interaction between sugar and peptide moieties in mucin-type model glycopeptides. *Int. J. Biol. Macromol.* **1992**, 14, 242–248.
- (34) Tachibana, Y.; Monde, K.; Nishimura, S.-I. Sequential glycoproteins: Practical method for the synthesis of antifreeze glycoprotein models containing base labile groups. *Macromolecules* **2004**, 37, 6771–6779.
- (35) Narimatsu, Y.; Kubota, T.; Furukawa, S.; Morii, H.; Narimatsu, H.; Yamasaki, K. Effect of glycosylation on cis/trans isomerization of prolines in IgA1-hinge

- peptide. *J. Am. Chem. Soc.* **2010**, 132, 5548–5549.
- (36) Corzana, F.; Busto, J. H.; Jimenez-Oses, G.; Asensio, J. L.; Jimenez-Barbero, J.; Peregrina, J. M.; Avenoza, A. New insights into alpha-GalNAc-Ser motif: influence of hydrogen bonding versus solvent interactions on the preferred conformation. *J. Am. Chem. Soc.* **2006**, 128, 14640–14648.
- (37) Corzana, F.; Busto, J. H.; Jimenez-Oses, G.; García de Luis, M.; Asensio, J. L.; Jimenez-Barbero, J.; Peregrina, J. M.; Avenoza, A. Serine versus threonine glycosylation: the methyl group causes a drastic alteration on the carbohydrate orientation and on the surrounding water shell. *J. Am. Chem. Soc.* **2007**, 129, 9458–9467.
- (38) Grinstead, J. S.; Koganty, R. R.; Krantz, M. J.; Longenecker, B. M.; Campbell, A. P. Effect of Glycosylation on MUC1 Humoral Immune Recognition: NMR Studies of MUC1 Glycopeptide–Antibody Interactions. *Biochemistry* **2002**, 41, 9946–9961.
- (39) Matsushita, T.; Takada, W.; Igarashi, K.; Naruchi, K.; Miyoshi, R.; Garcia-Martin, F.; Amano, M.; Hinou, H.; Nishimura, S.-I. A Straightforward Protocol for the Preparation of High Performance Microarray Displaying Synthetic MUC1 Glycopeptides. *Biochim. Biophys. Acta - Gen. Subj.* **2014**, 1840 (3), 1105–1116.
- (40) Cai, H.; Huang, Z. H.; Shi, L.; Zou, P.; Zhao, Y. F.; Kunz, H.; Li, Y. M. Synthesis of Tn/T Antigen MUC1 Glycopeptide BSA Conjugates and Their Evaluation as Vaccines. *Eur. J. Org. Chem.* **2011**, 2011, 3685–3689.
- (41) States, D. J.; Haberkorn, R. A.; Ruben, D. J. A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. *Journal of Magnetic Resonance* **1982**, 48, 286–292.
- (42) Piantini, U.; Sorensen, O. W.; Ernst, R. R. Multiple quantum filters for elucidating NMR coupling networks. *J. Am. Chem. Soc.* **1982**, 104, 6800–6801.
- (43) Rance, M.; Sorensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wüthrich, K. Improved spectral resolution in cosy ¹H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Co.* **1983**, 117,

479–485.

- (44) Braunschweiler, L.; Ernst, R. R. Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *Journal of Magnetic Resonance* **1983**, 53, 521–528.
- (45) Bax, A. D.; Davis, D. G. MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *Journal of Magnetic Resonance* **1985**, 65, 355–360.
- (46) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. Investigation of exchange processes by two - dimensional NMR spectroscopy. *J. Chem. Phys.* **1979**, 71, 4546.
- (47) Piotto, M.; Saudek, V.; Sklenář, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of Biomolecular NMR* **1992**, 2, 661–665.
- (48) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **1995**, 6, 277–293.
- (49) Godderd, T. D.; Kneller, D. G. SPARKY 3; University of California, San Francisco, **1999**.
- (50) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D: Biological Crystallography* **1998**, 54, 905–921.
- (51) Laskowski, R. A.; Rullmannn, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of biomolecular NMR* **1996**, 8, 477–486.
- (52) Koradi, R.; Billeter, M.; Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **1996**, 14, 51–55.

Chapter 3
O-Fucosylation and O-Glucosylation
on Human Noctch1 EGF12

3-1. Introduction

The Notch signaling pathway is highly conserved cell-cell communication system, which involves essential roles in cellular development processes. In particular, Notch signaling strongly influences cancer biology including cancer stem cells, angiogenesis, cancer immunity, and embryonic differentiations.^{1,2} Notch signaling pathway is mediated by five-transmembrane ligand [Jagged1 and Jagged2, and Delta-like ligand 1(DLL-1), DLL-3 and DLL-4] on signal-sending cells and Notch (NOTCH1~4) extracellular domain (NECD) on signal-receiving cells. The Notch signaling is triggered by ligand-receptor binding which induces proteolysis by ADAM10/17 metalloproteases and γ -secretase to cleave Notch intercellular domain (NICD). Cleaved NICD moves into nucleus and regulates the expression of Notch target gene.^{3,4}

It has been reported that NECDs are modified by *O*-glycans at threonine and serine residue, which contributes to regulate the Notch signaling activity.⁵ Human Notch1 extracellular domain is composed of 36 epidermal growth factor (EGF) like repeats. Each EGF domain has six cysteine residues forming three disulfide bonds (C1-C3, C2-C4, and C5-C6). Unusual *O*-glycan modifications such as *O*-glucosylation⁶, *O*-fucosylation⁷, and *O*-*N*-acetylglucosaminylation⁸ have been reported in consensus sequence between two cysteine residues, C1-X-S-X-(P/A)-C2⁹, C2-X-X-X-X-(T/S)-C3¹⁰, and C5-X-X-G-X-(T/S)-G-X-X-C6¹¹, respectively (where X is any amino acid). In mammalian Notch, these α -*O*-fucose and β -*O*-glucose residues are further modified by specific carbohydrate residues, forming unusual glycan structure, Neu5Ac α 2 $\rightarrow$ 3/6Gal β 1 $\rightarrow$ 4GlcNAc β 1 $\rightarrow$ 3Fuc α 1 $\rightarrow$ ¹² and Xyl α 1 $\rightarrow$ 3Xyl α 1 $\rightarrow$ 3Glc β 1 $\rightarrow$ ¹³, respectively. *O*-Fucosylation of human Notch1 EGF12 is one of the most important modification to Notch signaling activity because EGF11-12 are known as ligand binding

region of Notch extracellular domain^{14,15}, and only EGF12 is modified by O-fucosylation in this binding region.¹⁶ In addition, it has been described that glycan structure of O-fucose glycan on EGF12 regulates specificity and affinity of ligand binding.^{12,17,18} Especially, *Fringe*-mediated O-GlcNAc modification at fucose residue has critical roles for regulation of these Notch ligand binding.¹² A part of O-fucosylation effects on backbone structure of Notch EGF12 domain are getting clear by solution NMR study of synthetic mouse EGF12 glycopeptide¹⁹ and crystallography of recombinant human¹⁸ and mouse EGF11-13 complex.²⁰ These conformational studies have revealed that O-fucose locating on anti-parallel β -sheet region show hydrophobic interaction with opposite side of anti-parallel β -sheet to form “sugar bridge” structure.

O-glucosylation has been found in several EGF domains including EGF12 in ligand binding domain. According to previous mass spectral-based glycoproteomics, O-glucose modification is found at high stoichiometry, and trisaccharide is found as the major form in mouse Notch1.⁹ *Drosophila* Notch showed that xylosyl-glucose disaccharide is the major form, and trisaccharide was not found in mass analysis.²¹ The details of the O-glucosylation function are still unclear. Some experiments have revealed that the lack of O-glucosylation induced loss of Notch activation while Notch proteins are expressed at normal level. In addition, O-glucose modification is independent of ligand binding affinity even though O-glucosylation is found in EGF12 which is known as binding region.²²⁻²⁴ In recent study, over expression of human xylosyltransferase (GXYL1), which is responsible for addition of first xylose residue to O-glucose, cause negative regulation of Notch signaling in *Drosophila*, proposing that elongation of O-glucosylation with xylose residue could inhibit proper trafficking of Notch protein.²⁵ These results suggest that O-glucosylation affect backbone

conformation of Notch extracellular domain. However, conformational impacts of *O*-glucosylation on EGF domains are still unclear because only a few conformational studies of *O*-glucosylation have been reported.²⁰ Indeed, conformational impacts of Xyl-Xyl-Glc trisaccharide have not been reported while this form is found majorly in mammals Notch. Structural characterization of EGF domains carrying this trisaccharide would be required to reveal *O*-glucose effects on Notch signaling pathway.

Despite of crucial roles of *O*-glycosylation in Notch signaling, molecular mechanism how *O*-glycans contribute to this signaling is still unclear. In this chapter, chemical synthesis, NMR conformational analysis, and cellular assay of human Notch1 EGF12 domain were carried out to reveal the detailed *O*-glycosylation effects on Notch peptide conformation as well as on Notch function.

3-2. Result and Discussion

3-2-1. Design of human Notch1 EGF12 glycopeptide analogues

38-mer human Notch1 EGF12, 452-489 residues (swissprot **P46531**), were employed as target sequence. This EGF12 peptide contains six cysteine residues forming three disulfide bonds, Cys456-Cys467, Cys461-Cys476, and Cys478-Cys487. In addition, Notch1 EGF12 has two glycosylation sites, O-glucosylation at Ser458 and O-fucosylation at Thr466. According to that, 38-mer synthetic EGF12 glycopeptide analogues carrying O-fucose- or O-glucose-type glycans were designed (**Figure 3-1**). To elucidate the glycosylation impacts on peptide conformation, naked EGF12 (**10**), mono-fucosylated EGF12 (Fuc EGF12, **11**), GlcNAc-fucosylated EGF12 (GnF EGF12, **12**), mono-glucosylated EGF12 (Glc EGF12, **13**), Xyl-Xyl-Glc EGF12 (XXG EGF12, **14**), and disubstituted EGF12 (XXG GnF EGF12, **15**) were designed for NMR study. Furthermore, enzymatic sugar elongations were carried out for total synthesis of human Notch1 EGF12 with matured glycosylation for XXG GalGnF EGF12 (**16**) and XXG SiaGalGnF EGF12 (**17**).

The synthetic strategy was illustrated in **Figure 3-1**. The most challenging steps of synthetic strategy of O-glycosylated EGF12 analogs will be preparation of new Fmoc-protected glycoamino acids carrying unusual $\text{Xyl}\alpha 1 \rightarrow 3\text{Xyl}\alpha 1 \rightarrow 3\text{Glc}\beta 1 \rightarrow$ structure (**21**) and folding reaction to generate accurate three intramolecular disulfide bonds.

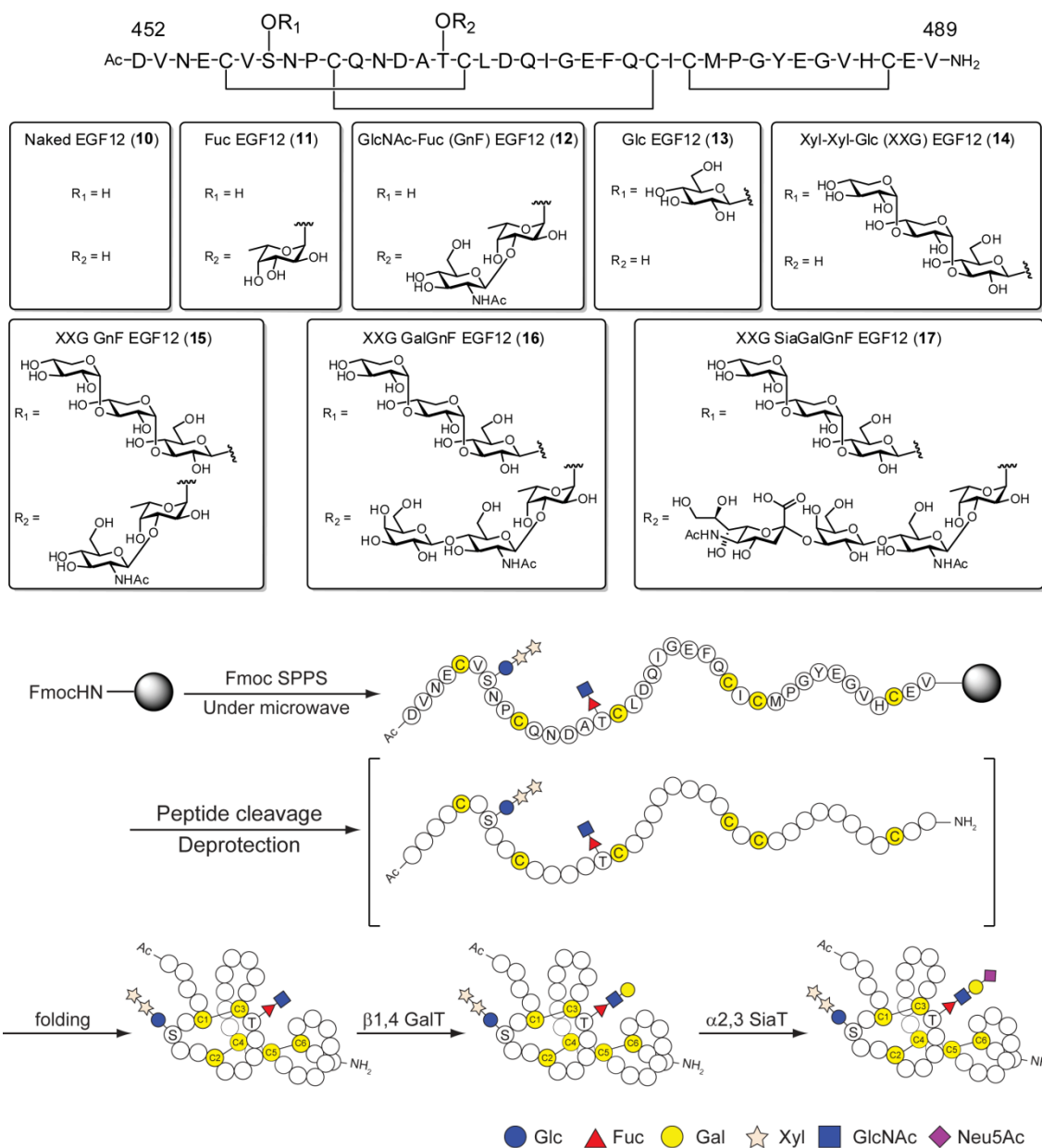


Figure 3-1. Synthetic human Notch1 EGF12 glycopeptide analogues

3-2-2. Synthesis of Fmoc-Ser[Ac₃Xyl(α1→3)Ac₂Xyl(α1→3)Ac₃Glcβ1→]-OH

Prior to preparation of the glycosylated EGF12 analogues with Fmoc SPPS strategy, Fmoc-protected glycosylated amino acids are needed to synthesize as building block. For the synthesis of designated glycosylated human Notch1 EGF12 analogues, Fmoc-Thr(Ac₃Fucα1→)-OH (**18**), Fmoc-Thr[Ac₃GlcNAc(β1→3)Fucα1→]-OH (**19**), Fmoc-Ser(Ac₄Glcβ1→)-OH (**20**) and Fmoc-Ser[Ac₃Xyl(α1→3)Ac₂Xyl(α1→3)Ac₃Glcβ1→]-OH (**21**) can be employed as glycoaminoacid building blocks. It was demonstrated that GlcNAcβ1→3Fucα1→ glycan was elongated to Siaα2→3Galβ1→4GlcNAcβ1→3Fucα1→ with corresponding glycosyltransferases.¹⁹ The compounds **18**, **19**,¹⁹ and **20**²⁶ were already synthesized in previous reports. The compound **21** carrying matured O-glucosyl trisaccharide is needed to be synthesized.

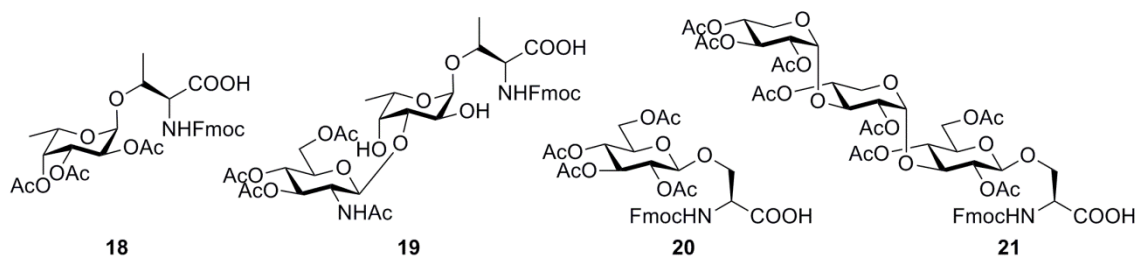
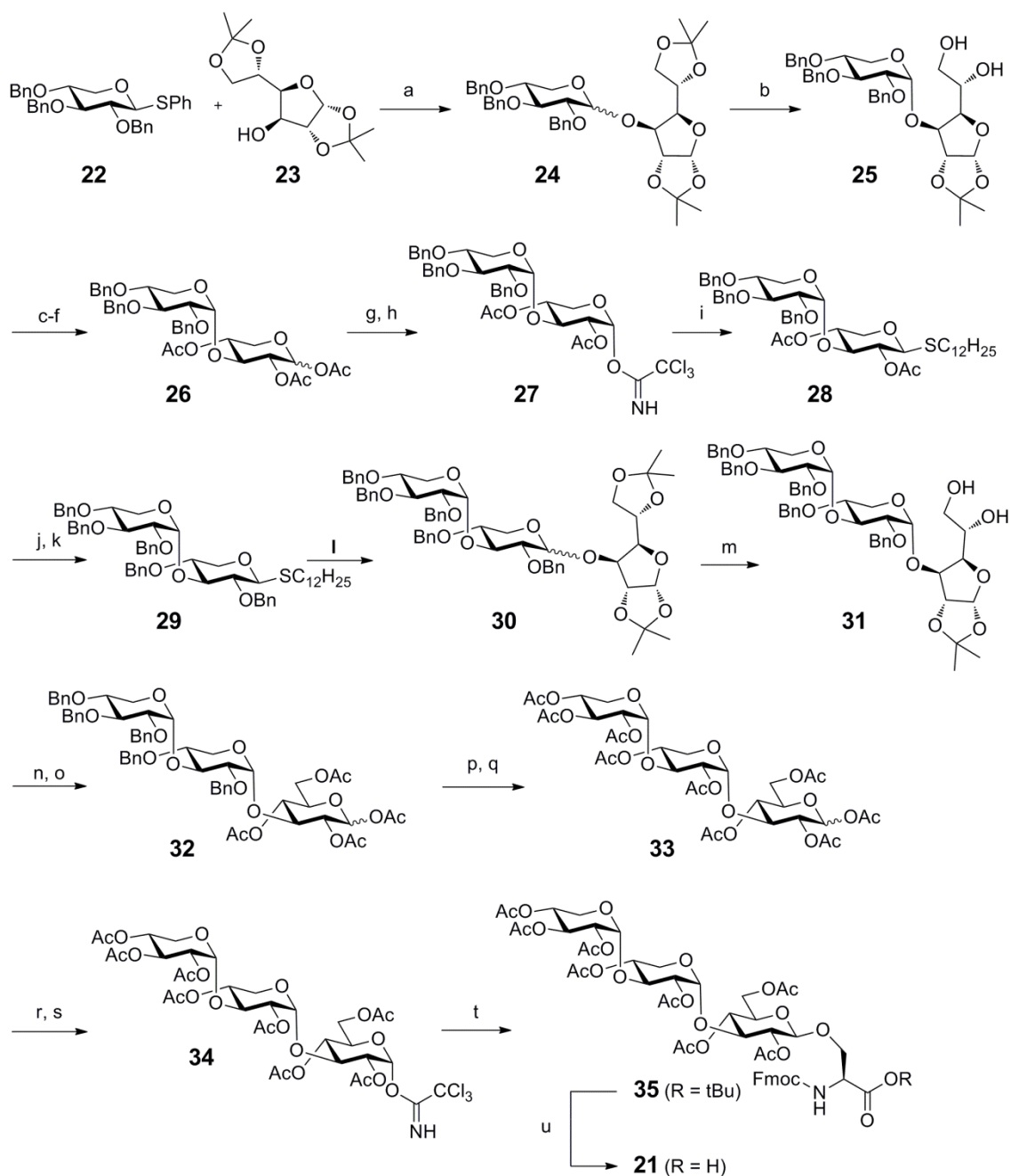


Figure 3-2. Fmoc-protected glycoamino acid derivatives

For the synthesis of compound **21**, we employed commercially available D-xylose and 1,2;5,6-O-diisopropylidene-α-D-glucofuranose **23** as starting materials.²⁷ This compound **23** can be a good acceptor substrate because OH groups except for 3rd-position where glycosylated position of trisaccharide are protected by isopropylidene group. Xylα1→3Glcβ1→-fructose **25** was given by selective deprotection of 5,6-O-isopropylidene of anomeric mixture **24** by mild-acidic condition, which allow

α/β separation on the silica gel column. After selective deprotection of 5,6-*O*-isopropylidene group in **24**, glucofranoside **25** can be easily converted to xylofranoside by periodate oxidation of 5,6-diol and reduction to afford disaccharide **26** containing Xyl α 1 $\rightarrow$ 3Xyl (94 % yield, 4 steps overall yield). Due to acid-labile property of α -glycoside linkage between two xylose residues, conversion of compound **10** into penta-*O*-benzyl derivative **29** as a glycosyldonor was performed *via* the stable thioalkyl glycoside **28**.²⁸ The deprotection of 5,6-*O*-isopropylidene group of anomeric mixture of trisaccharide **30** afford clear separation of α - and β -isomers on the silica gel column to give **31**, and converted into the per-*O*-acetate **33**. Finally, glycosylation of trisaccharide imidate **34** with Fmoc-Ser-OtBu proceeded stereoselectively to afford β -glycoside **35** and the removal of *t*-butyl protection gave new building block **21** in total 21 steps and 2.1 % total yield (**Scheme 3-1**).



Reagents and conditions: a) NIS (1.3 eq.), TfOH (0.1 eq.), THF/DCM (2:1), -20 °C, 92%; b) 80% AcOH_{aq}, 40 °C, 41%; c) NaIO₄ (2.0 eq.), MeOH, 0 °C; d) NaBH₄ (4.0 eq.), MeOH, 0 °C; e) 60% TFA_{aq}, r.t.; f) Ac₂O, pyridine, r.t., 94%; g) H₂NNH₂·H₂O (1.3 eq.), AcOH (1.3 eq.), 0 °C; h) CCl₃CN (5.0 eq.), DBU (0.1 eq.), r.t., 74%; i) HSC₁₂H₂₅ (3.0 eq.), TMSOTf (0.1 eq.), DCM, -20 °C, 64%; j) NaOMe (0.1 eq.), MeOH, r.t.; k) BnBr (5.0 eq.), NaH (5.0 eq.), r.t., 96%; l) **23** (1.2 eq.), NIS (1.5 eq.), TfOH (0.1 eq.), THF/DCM (2:1), -20 °C, 81%; m) 80% AcOH_{aq}, 40 °C, 40%; n) 60% TFA_{aq}, r.t.; o) Ac₂O, pyridine, r.t., 63%; p) 20% Pd(OH)₂/C (50w%), H₂, THF/MeOH (1:3), r.t.; q) Ac₂O, pyridine, DMAP (0.1 eq.), 98%; r) H₂NNH₂·H₂O (1.3 eq.), AcOH (1.3 eq.), 0 °C; s) CCl₃CN (5.0 eq.), DBU (0.1 eq.), r.t., 84%; t) Fmoc-Ser-OtBu (1.2 eq.), TMSOTf (0.1 eq.), DCM, -20 °C, 80%; u) 90% TFA_{aq}, r.t., 97%

Scheme 3-1. Synthetic scheme of Fmoc-(XXG)Ser-OH

3-2-3. Synthesis of human Notch1 EGF12 glycopeptide analogues

38-mer linear human Notch1 EGF12 glycopeptide analogues were prepared by microwave-assisted SPPS protocol.²⁹ Rink Amide-ChemMatrix resin which produce amide structure at C-terminus (-CONH₂) was employed for solid support. Fmoc amino acid derivatives were coupled by HBTU/HOBt activation system under microwave irradiation. For the glycopeptides synthesis, synthetic Fmoc-Thr(Ac₃Fucα1→)-OH (**18**), Fmoc-Thr[Ac₃GlcNAc(β1→3)Fucα1→]-OH (**19**), Fmoc-Ser(Ac₄Glcβ1→)-OH (**20**) or Fmoc-Ser[Ac₃Xyl(α1→3)Ac₂Xyl(α1→3)Ac₃Glcβ1→]-OH (**21**) were employed as glycoamino acid building blocks. These glycoamino acid residues were coupled by double activation method using PyBOP/HOAt coupling cocktail, which gave efficient yield of each glycosylated EGF12 analogue with fewer equivalents of glycoamino acids (i.e. 1.2 eq.).³⁰ After completion of peptide elongation and acetylation of *N*-terminal amino group, resulted peptides were cleaved from resin using cleavage cocktail containing TFA:EDT:H₂O:TIS = 94:2.5:2.5:1 for 2 h at room temperature. For cleavage of **15**, reagent H (TFA:phenol:thioanisole:EDT:H₂O:DMSO:NH₄I = 81:5:5:2.5:3:2:1.5)³¹ which reduce methionine oxidation was used because oxidation was observed with TFA:EDT:H₂O:TIS cleavage cocktail. TFA solution containing cleaved peptide were directly precipitated by mixture solution of hexane and *t*-butyl methyl ether in ice-bath to avoid formation of undesired oxidative disulfide bond formation and methionine oxidation during evaporation of TFA solution by air flowing. After lyophilization of centrifuged material, removal of acetyl group on carbohydrate moiety were conducted under basic methanol in the presence of 6.5 mM DTT (pH 12.5 adjusted by 1 N NaOH) at ambient temperature for 0.5-1 h. Reaction proceeding were monitored by MALDI-TOF MS. With respect to EGF12 analogues carrying

Xyl-Xyl-Glu trisaccharide (**14**, **15**), de-acetylation reaction was performed for 24 h at ambient temperature to remove a sterically hindered *O*-acetyl group. The crude materials were injected to RP-HPLC to remove excess of DTT after deprotection.

3-2-4. Glutathione-mediated folding reaction

The synthetic linear glycopeptide analogues were subjected to glutathione-mediated folding reaction. 1.0 mg EGF12 analogues were dissolved in 10 mL of 50 mM Tris-HCl buffer (pH 8.0) containing 1.0 mM of reduced glutathione and 0.2 mM of oxidized glutathione, mixed for 24 h at room temperature.³² The reaction proceeding were monitored by analytical HPLC after reaction mixture were acidified by TFA (2% of total volume) and lyophilized. Folded peptides with native disulfide bond would be obtained with these glutathione-mediated strategy based on thermodynamic stability of native peptide conformation.

In HPLC analysis of folding reaction of naked EGF12, three major peaks were observed at 20, 22 and 23 min during linear gradient of acetonitrile containing 0.1% TFA, 25 to 35% for 30 min. These three HPLC peaks at 20, 22 and 23 min were named as #1, #2 and #3, respectively (**Figure 3-3**).

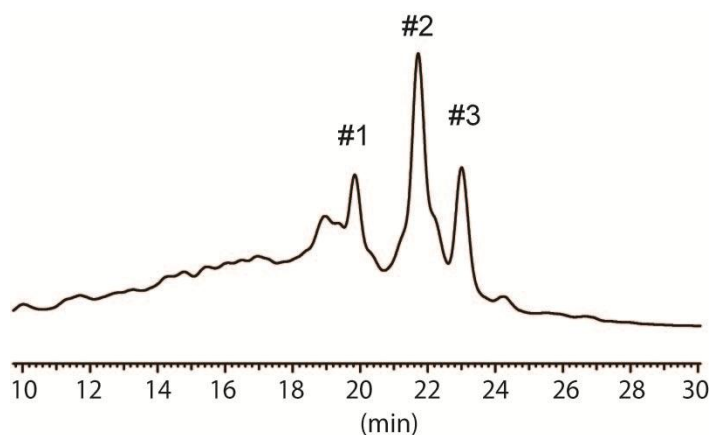


Figure 3-3. HPLC result of folding reaction of naked EGF12

According to MALDI-TOF MS analyses of eluted substrate, these three eluted compounds (#1~3) have the same molecular weight corresponding to naked EGF12 analogue with three disulfide bond. After isolation of each eluted peptide, the disulfide bond positions were determined by enzymatic digestion using Thermolysin (from *Bacillus thermoproteolyticus eokko*) as previous mouse EGF12.¹⁹ Digested peptide fragments were detected by MALDI-TOF mass spectrometry after lyophilization. MALDI-TOF MS analyses revealed that fragment C2-C3 (m/z $[M+Na]^+$: 1171.673) and C4-C6 (m/z $[M+Na]^+$: 903.526) were found from #1, fragment C2-C3 (m/z $[M+Na]^+$: 1171.750) and C5-C6 (m/z $[M+Na]^+$: 1375.863) were found from #2, and C2-C4 (m/z $[M+H]^+$: 1270.768) and C5-C6 (m/z $[M+H]^+$: 1353.846) were found from #3, respectively (**Table 3-1**).

Peak No. (Eluted time in HPLC)			
	#1 (20 min)	#2 (22 min)	#3 (23 min)
Fragment 1	$\begin{array}{c} \text{H-V-S-N-P-C-Q-N-D-A-T-C-OH} \\ \text{C2} \quad \quad \quad \uparrow \quad \quad \quad \text{C3} \end{array}$	$\begin{array}{c} \text{H-V-S-N-P-C-Q-N-D-A-T-C-OH} \\ \text{C2} \quad \quad \quad \uparrow \quad \quad \quad \text{C3} \end{array}$	$\begin{array}{c} \text{C2} \\ \text{H-V-S-N-P-C-Q-N-D-OH} \\ \\ \text{H-F-Q-C-OH} \\ \text{C4} \end{array}$
	calcd: 1148.4227	calcd: 1148.4227	calcd: 1269.4754
	found: 1171.672 $[M+Na]^+$	found: 1171.672 $[M+Na]^+$	found: 1270.768 $[M+H]^+$
Fragment 2	$\begin{array}{c} \text{C4} \\ \text{H-F-Q-C-OH} \\ \\ \text{H-V-H-C-E-OH} \\ \text{C6} \end{array}$	$\begin{array}{c} \text{C5} \\ \text{H-I-C-M-P-G-Y-E-G-OH} \\ \\ \text{H-V-H-C-E-OH} \\ \text{C6} \end{array}$	$\begin{array}{c} \text{C5} \\ \text{H-I-C-M-P-G-Y-E-G-OH} \\ \\ \text{H-V-H-C-E-OH} \\ \text{C6} \end{array}$
	calcd: 880.3208	calcd: 1352.5199	calcd: 1352.5199
	found: 903.526 $[M+Na]^+$	found: 1375.863 $[M+Na]^+$	found: 1353.896 $[M+H]^+$

Table 3-1. MALDI-TOF MS analysis of enzymatic digestion

Although only two fragments were found in each effluent, it would be sufficiently enough for determining total disulfide pairs. Since native human Notch1 EGF12 have C1-C3, C2-C4, and C5-C6 cross-linking, the peptide eluted at 23 min (#3) which has the second highest intensity in the HPLC analysis was desired product having proper disulfide bond position, while misfolded peptide eluted at 22 min (#2) was generated as the major product of this glutathione-mediated folding reaction. This result can be attributed to unnatural condition, such as the lack of covalent attachment of neighboring EGF 11 and 13 which can form hydrophobic packing between Y411 in EGF11 and I471 in EGF12, as well as Y482 in EGF12 and I509 in EGF13, respectively.³³

To obtain the desired naked EGF12 having accurate cross link with sufficient yield, the folding reactions were conducted at 40 °C. However, the peak intensity of desired peptide seemed decreased to third highest of three major peaks in HPLC analysis (**Figure 3-4**). This result suggested that native secondary structure, anti-parallel β -sheet structure, was not able to be formed in higher temperature due to unnatural condition mentioned above. The importance of stability of secondary structure during disulfide bond formation can also be explained by another folding reaction in the presence of additive 1.0 M guanidine-HCl salt. This condition gave similar results with 40 °C condition which decrease desired product because guanidine can act as hydrogen bond donor and prevent the formation of intramolecular hydrogen bond (**Figure 3-4**) which is needed to form β -sheet structure.

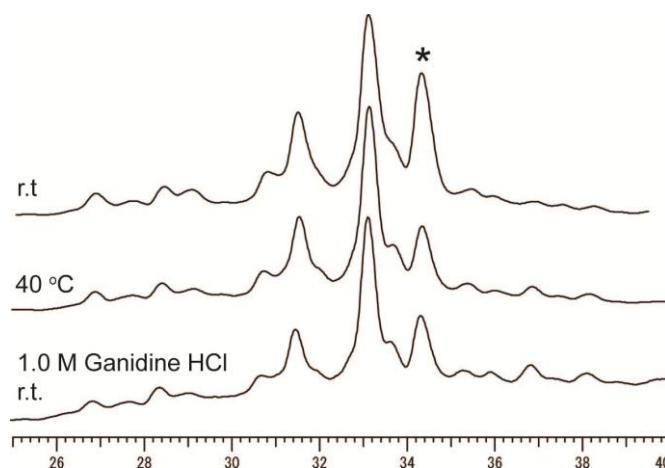


Figure 3-4. Folding of naked EGF12 at 40 °C or in the presence of guanidine

To enhance the stability of β -sheet structure, folding reaction was conducted under 4 °C for 24 h. HPLC analysis of 4 °C condition demonstrated that desired product (#3) was generated as major product but misfolded peptide (#2) was still generated as not negligible amount (**Figure 3-6**, Blue). It is suggested that lower temperature could improve reaction efficiency due to stabilization of secondary structure.

For further improvement of folding reaction, other factors to stabilize peptide conformation are required. Focusing on the properties of EGF domains, some Notch1 EGF domains are known as calcium binding protein (EGF 6, 7, 8, 11-21, 23, 25, 27, and 30-32). Calcium ion can be coordinated at junction of EGF domains (**Figure 3-5**).³³⁻³⁵

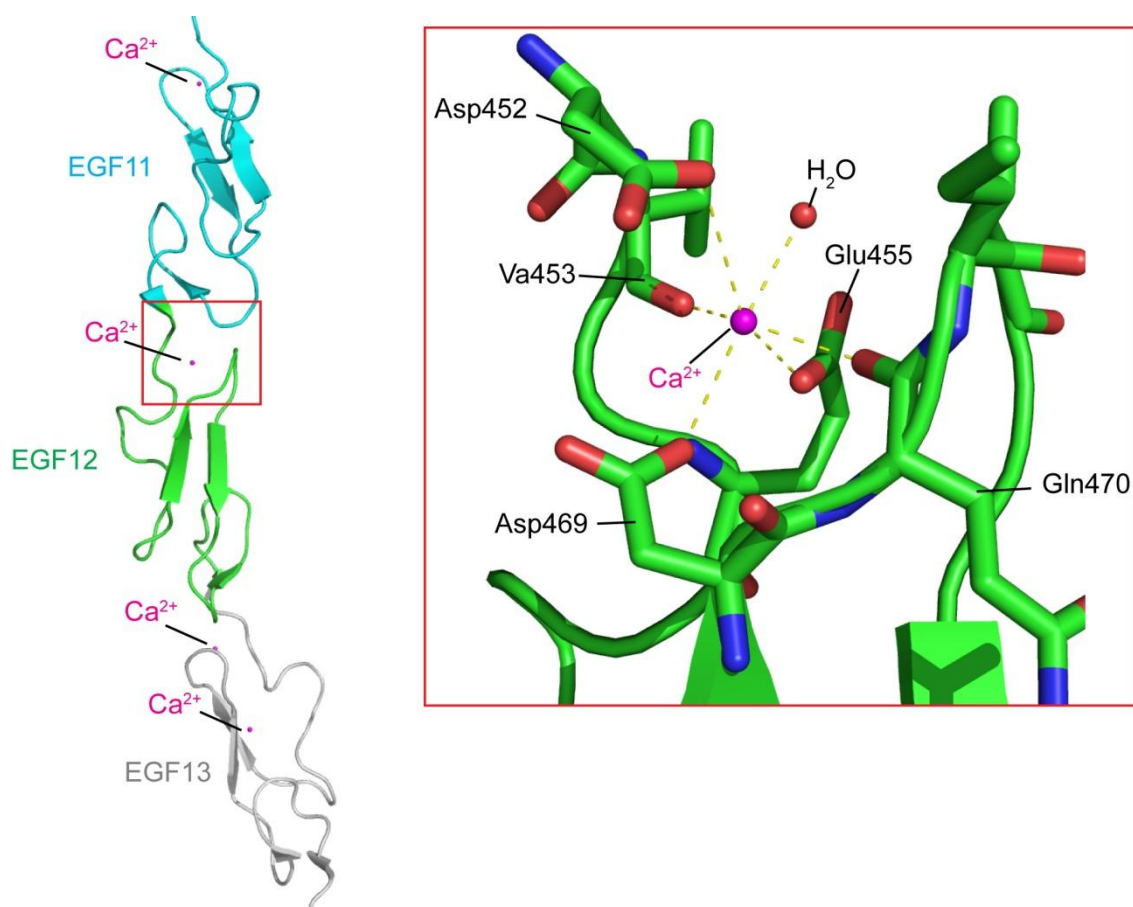


Figure 3-5. Crystal structure of human Notch1 EGF11-13 (PDB ID: 2VJ3)

Crystal structure of human Notch1 EGF12 (2VJ3) suggest that Asp452, Val453, Glu455, Asp469, and Gln470 are calcium binding residues. Asp452, Val453, and Glu455 are located at *N*-terminal region of EGF12, and Asp469 and Gln470 contributes to formation of anti-parallel β -sheet. This calcium coordination between *N*-terminus and β -hairpin structure might stabilize the native secondary structure of human Notch1 EGF12. Based on this calcium binding property of EGF12, we expected that calcium ion may stabilize the peptide conformation during disulfide bond formation reaction.

Folding reaction of naked EGF12 in the presence of 5 mM CaCl₂ in redox buffer was conducted at room temperature. The reaction proceeding was monitored by analytical HPLC. Surprisingly, reaction yield of desired product was improved similar to 4 °C

conditions, suggesting that Ca^{2+} ion can bind with EGF12 analogue alone without covalent attachment of EGF11 and 13, and it is able to stabilize their native conformation during disulfide bond formation (**Figure 3-6**, green). However, similar to results of 4 °C, misfolded peptides were still generated as not negligible amount.

Next, folding reaction using combined condition, under 4 °C and in the presence of 5 mM CaCl_2 , was demonstrated to evaluate double stabilization effects. Surprisingly, this condition produced desired naked EGF12 as almost single product in HPLC analysis. The formation of misfolded peptides at 20 and 22 min were clearly suppressed. It is likely that lower temperature and Ca^{2+} ion would stabilize native EGF12 conformation with different manner so that these two factors would gave more efficient yield than single factor.

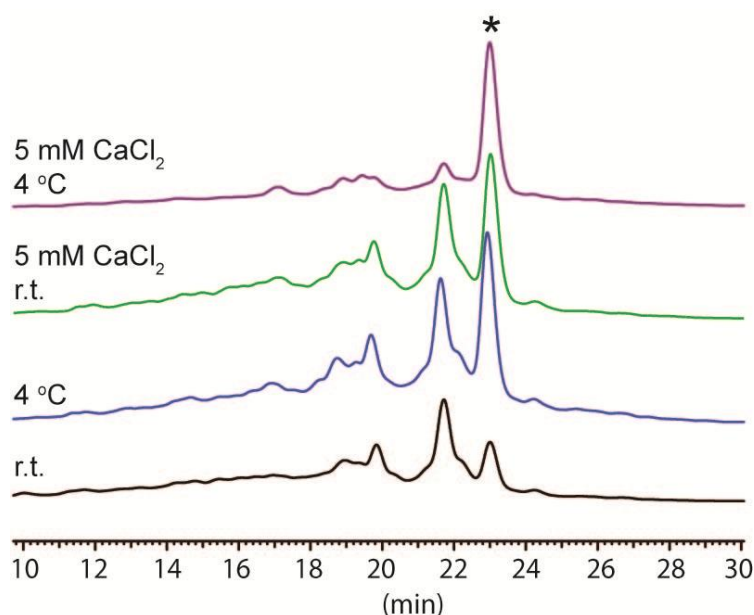


Figure 3-6. Folding reaction of naked EGF12 (**10**) at 4 °C with Ca^{2+}

Linear glycopeptides analogues were subjected to folding reaction using the glutathione-containing redox buffer. Previous NMR and crystal structure demonstrated that *O*-fucose residue on Thr466 would act as “bridge” between anti-parallel β -sheet, and support to form secondary structure after formation of three disulfide cross links. However, folding reaction of Fuc EGF12 analogue (**11**) at r.t. condition gave less desired product similar to naked case, suggesting that hydrophobic interaction between fucose residue and amino acid might be too weak to assist the formation of β -sheet conformation during glutathione-mediated folding reaction. The folding results of Fuc EGF12 (**11**) analogue with 4 °C and 5 mM CaCl_2 condition showed that the linear Fuc EGF12 (**11**) analogues were properly folded like naked one (**Figure 3-7**). These lower temperature and calcium ion can stabilize even fucosylated EGF12 during folding reaction and afford accurate folding.

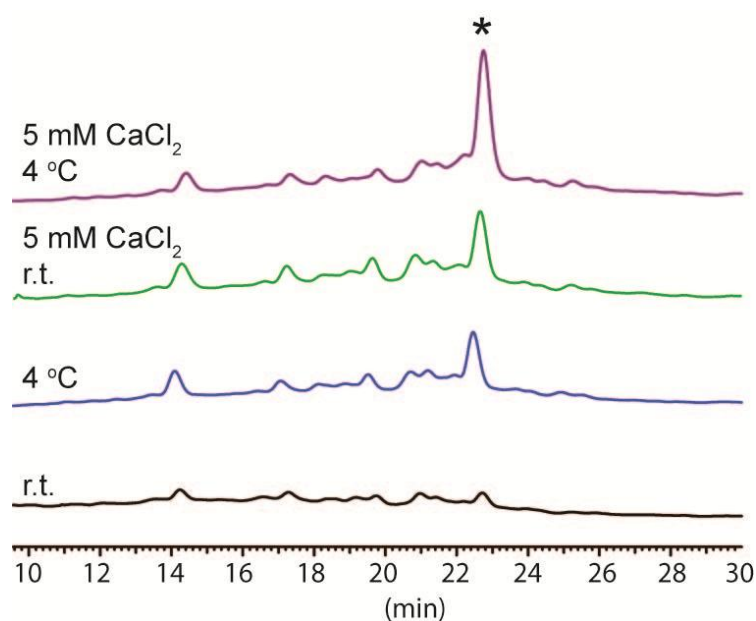


Figure 3-7. Folding reaction of Fuc EGF12 (**11**) at 4 °C with Ca^{2+}

With respect to GnF EGF12 (**12**) analogue, it showed that desired product was obtained as major product at room temperature without CaCl_2 . It is suggested that steric effects of GlcNAc elongation on fucose residue might prevent formation of misfolded peptide. As similar to naked (**10**) and Fuc EGF12 (**11**), 4°C and addition of CaCl_2 can reduce misfolded peptide (see 18~19 min in **Figure 3-8**).

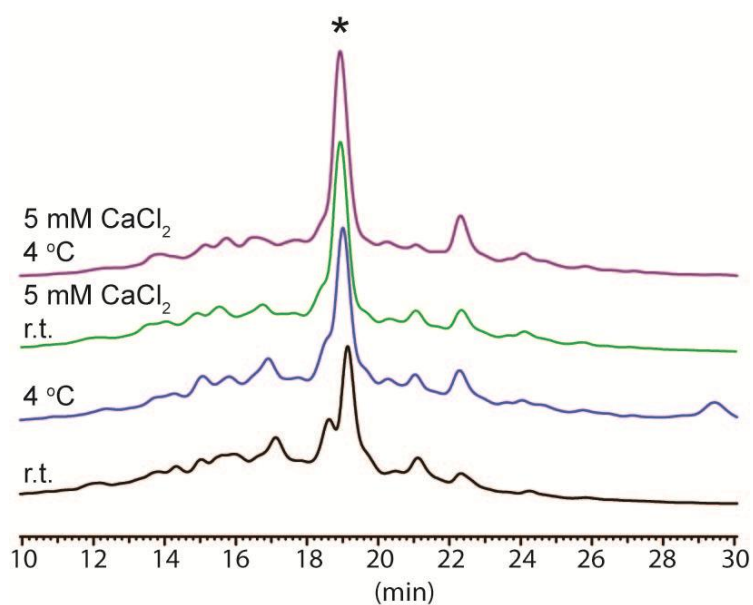


Figure 3-8. Folding reaction of GnF EGF12 (**12**) at 4°C with Ca^{2+}

In contrast to *O*-fucosylation, *O*-glucosylation effects at Ser458 on peptide conformation have not been reported. Glc EGF12 analogue (**13**) formed native disulfide bond as single product even in the absence of Ca^{2+} ion or room temperature. It suggested that *O*-glucosylation might have some stabilization ability (**Figure 3-9**). In both lower temperature and the presence of CaCl_2 , accurate folded EGF12 seemed obtained as major product.

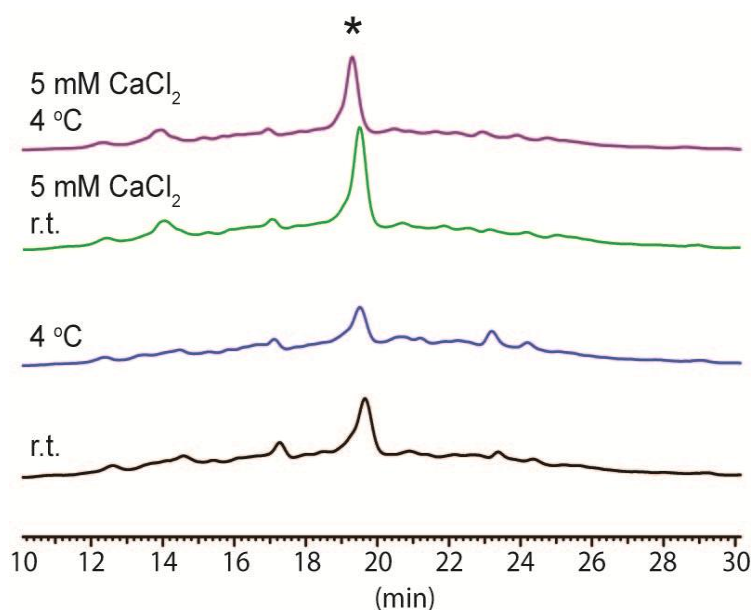


Figure 3-9. Folding reaction of Glc EGF12 (**13**) at 4 °C with Ca^{2+}

With respect to XXG EGF12 (**14**), folding reaction generated complex mixture at room temperature, although Glc EGF12 was efficiently folded in the same condition. The steric hindrance due to elongated trisaccharide might prevent proper folding. However, desired product was observed in the presence of calcium ion. Furthermore, folding reaction was properly proceeded in 4 °C and 5 mM CaCl₂ condition, suggesting that these lower temperature and calcium ion can overcome the barrier of carbohydrate steric hindrance. (O) indicates a oxidized methionine residue. (**Figure 3-10**).

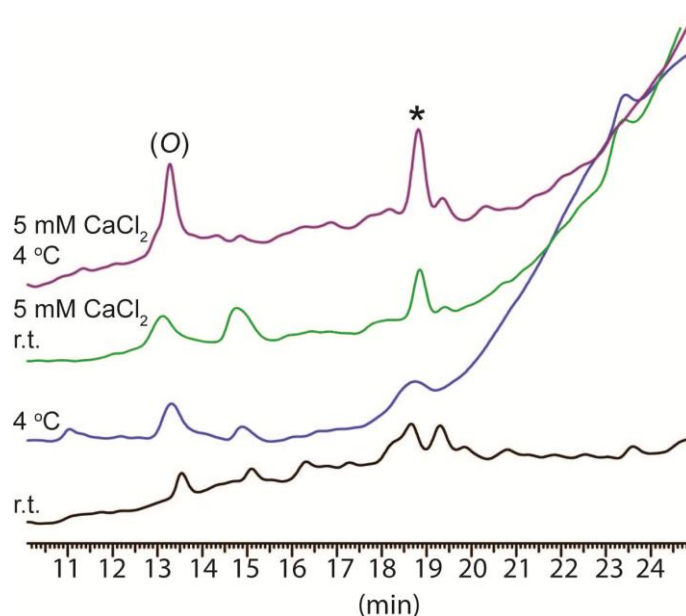


Figure 3-10. Folding reaction of XXG EGF12 (**14**) at 4 °C with Ca²⁺

Finally, synthetic XXG GnF EGF12 analogue (**15**) which has two glycan modifications was subjected to folding reaction. The effects of calcium ion and reaction temperature on folding efficiency were evaluated. In contrast with XXG EGF12, proper folding was proceeded at room temperature. The presence of GlcNAc-Fuc glycan might enhance the folding efficiency, while sterically hindered XXG trisaccharide was attached. Unlike the naked or mono-glycosylated EGF12 analogues (**10-14**), folding efficiency of di-glycosylated EGF12 (**15**) at 4 °C gave worse results compared with the result of room temperature. This result indicates that two glycan modifications might decrease folding reaction rate due to steric hindrance at low temperature. The presence of calcium ion improved folding reaction at room temperature and 4 °C. In naked and mono-substituted EGF12 (**10~14**), 5 mM CaCl₂ and 4 °C condition was the most efficient condition to produce folded peptide. HPLC result of XXG GnF EGF12 showed the most efficient condition was 5 mM CaCl₂ and room temperature judging from HPLC intensity.

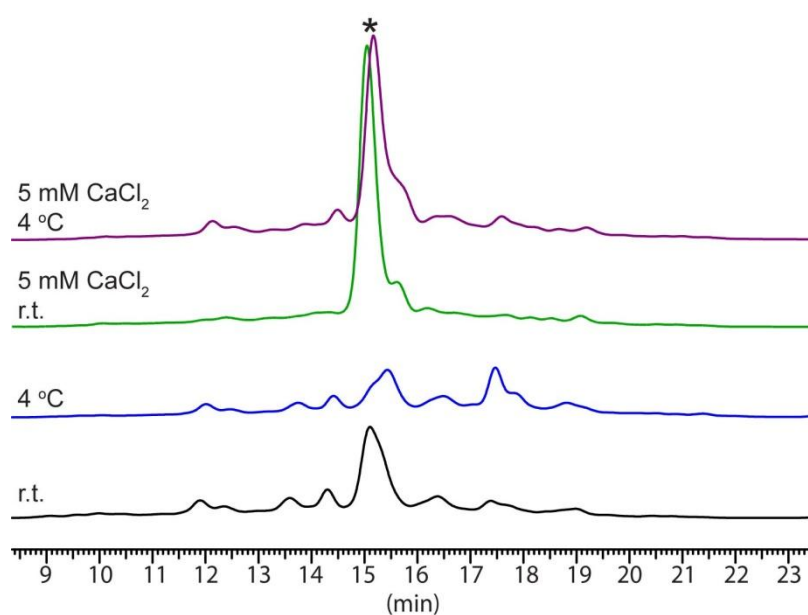


Figure 3-11. Folding reaction of XXG GnF EGF12 (**15**) at 4 °C with Ca²⁺

In summary, glutathione-mediated folding reaction were performed to obtain folded synthetic Notch1 EGF12 glycopeptide analogues. All disulfide cross linked pairs were confirmed by NOESY experiments (see 3-2-6). Lack of covalent attachment with EGF11 and 13 might prevent formation of desired products. However, novel condition such as lower temperature and in the presence of Ca^{2+} in reaction mixture was able to dramatically enhance the yield of desired product owing to conformational stabilization during folding reaction. It is surprised that Ca^{2+} which is known to be required for Notch function can contribute to folding reaction. This condition can be applied to naked and mono-substituted EGF12 analogues (**10-14**). But, folded di-substituted EGF12 (**15**) was obtained in the presence of CaCl_2 at room temperature condition because two glycans might cause steric hindrance, and it makes difficult to form proper six disulfide bonds at 4 °C situation. Hence, addition of calcium ion and temperature controlling is quite important for accurate folding reaction of calcium binding EGF domains.

3-2-5. Enzymatic sugar elongation toward total synthesis of fully modified EGF12

Enzymatic sugar elongations were performed to di-substituted XXG GnF EGF12 analogue **15** for total synthesis of human Notch1 EGF12 carrying fully matured O-glycans.^{19,36} Precursor **15** carrying GlcNAc β 1 $\rightarrow$ 3Fuc α 1 $\rightarrow$ glycan at Thr466 can become a substrate of β 1,4-galactosyltransferase. A human recombinant β 1,4-galactosyltransferase and UDP-Gal were applied to XXG GnF EGF12 (**15**). The reaction mixture was subjected to HPLC analysis and it showed that precursor **15** was consumed and completely converted to XXG GalGnF EGF12 (**16**). The formation of desired product was confirmed by MALDI-TOF MS analysis (**Figure 3-12**).

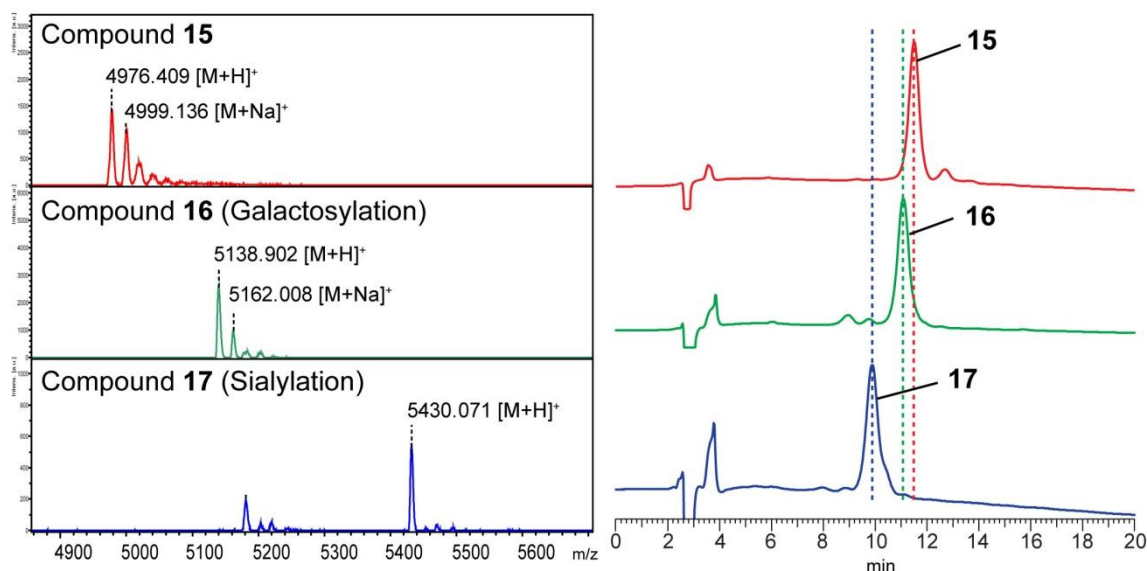


Figure 3-12. Enzymatic glycosylations monitored by HPLC and MALDI-TOF MS

Calcd M.W. 4975.3172 for **15**, 5137.4578 for **16**, 5428.7124 for **17**.

Subsequently, sialylation of fucose glycan was performed. α 2,3-sialyltransferase from *photobacterium phosphoreum* was used for sialylation of EGF12 analogue **16**. This enzyme transfer Neu5Ac from CMP-NANA to LacNAc (Gal β 1 $\rightarrow$ 4GlcNAc) motif via α 2,3 linkage. To tris buffer (pH 6.5) containing NaCl and Triton-X 100 were added

galactosylated EGF12 analogue **16**, α 2,3 sialyltransferase, and CMP-NANA, and reaction mixture was incubated at ambient temperature for 24 h. HPLC analysis indicated that sialylated EGF12 **17** was generated, but starting material **16** was also found. This result might be due to sialidase activity of this α 2,3 sialyltransferase enzyme. For that reason, additional CMP-NANA was mixed in reaction mixture after 24 h incubation, and extra 5 h incubation produced clear HPLC profile (**Figure 3-13**).

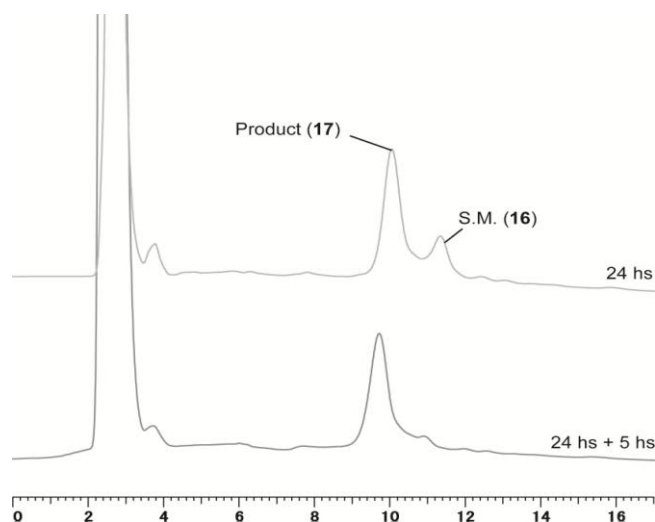


Figure 3-13. Enzymatic sialylation for XXG Sia-Gal-GnF EGF12 **17**

The alteration of HPLC profile and MALDI-TOF MS during enzymatic sugar elongations were summarized in **Figure 3-12**. These results indicated that even bacterial enzyme can catalyze the sugar elongation of human Notch1 EGF12 and Xyl-Xyl-Glc trisaccharide did not disturb these enzymatic sugar elongation reactions.

Here, standard protocols for total synthesis of fully glycosylated human Notch1 EGF12 was established with microwave-assisted Fmoc-SPPS, Ca^{2+} -assisted folding reaction, and enzymatic sugar elongation. This strategy will allow rapid and highly efficient synthesis of desired EGF domains carrying structure-defined glycans.

3-2-6. NMR-based conformational analysis of EGF12 analogues

Toward conformational characterization of synthetic human Notch1 EGF12 analogues, naked, Fuc, GnF, Glc, XXG, and XXG GnF EGF12 analogues (**10-15**) were subjected to NMR analyses. First, ^1H and TOCSY experiments were performed to identify NMR signals. Second, NOESY experiments to reveal disulfide bond pairs and specific interaction between *O*-glycans and peptide chain. Chemical shift differences were compared between naked and *O*-glycosylated EGF12 analogues to estimate *O*-glycosylation impacts on peptide conformation. Finally, the conformational impacts of *O*-glycosylations on Notch protein were discussed using three-dimensional NMR structure of synthetic EGF12 analogues.

^1H and TOCSY experiments

^1H spectra of synthetic glycopeptides analogues were measured in 90% $\text{H}_2\text{O}/10\% \text{D}_2\text{O}$ (pH 6.0) to observe amide protons of peptide backbone using 3-9-19 WATERGATE pulse sequence.^{37,38} In Fuc EGF12 (**11**), $\text{H}\alpha$ and HN protons of β -sheet contributing residues (Cys16, Leu17, Gln24, Phe23, Cys25 and Ile26) were found at 5.0 to 5.5 ppm and at 8.8 to 9.5 ppm, respectively. These lower field shifts were due to electron-withdrawing effect by hydrogen bond to form β -sheet, however, these peaks were found only in Fuc- (**11**) and GnF EGF12 (**12**) analogues but did not find in naked- (**10**), Glc EGF12 (**13**) and XXG EGF12 (**14**) (**Figure 3-14 A**). Furthermore, TOCSY spectrum of naked- (**10**) and Glc EGF12 (**13**) gave broad signals although Fuc- (**11**) and GnF EGF12 (**12**) gave sharp signals (**Figure 3-14 B**). These results suggested that conformational equilibrium might exist between proper β -sheet conformation and unnatural conformation in naked- (**10**) and Glc EGF12 (**12**) because of no covalent

attachment of neighboring EGF11 and 13, which form hydrophobic packing with EGF12.³³ In addition, the lack of *O*-fucosylation bridge between anti-parallel β -sheet might disturb conformational stability. In this case, addition of 25 mM CaCl_2 (5.0 equivalents of 5.0 mM Glc EGF12 peptide) to NMR solution improved NMR signal resolution (**Figure 3-14 C**). These facts indicated that Ca^{2+} ion as well as *O*-fucosylation might be able to stabilize the β -sheet conformation after proper disulfide bond formation, and *O*-glucose modification might have no effect on β -sheet stability.

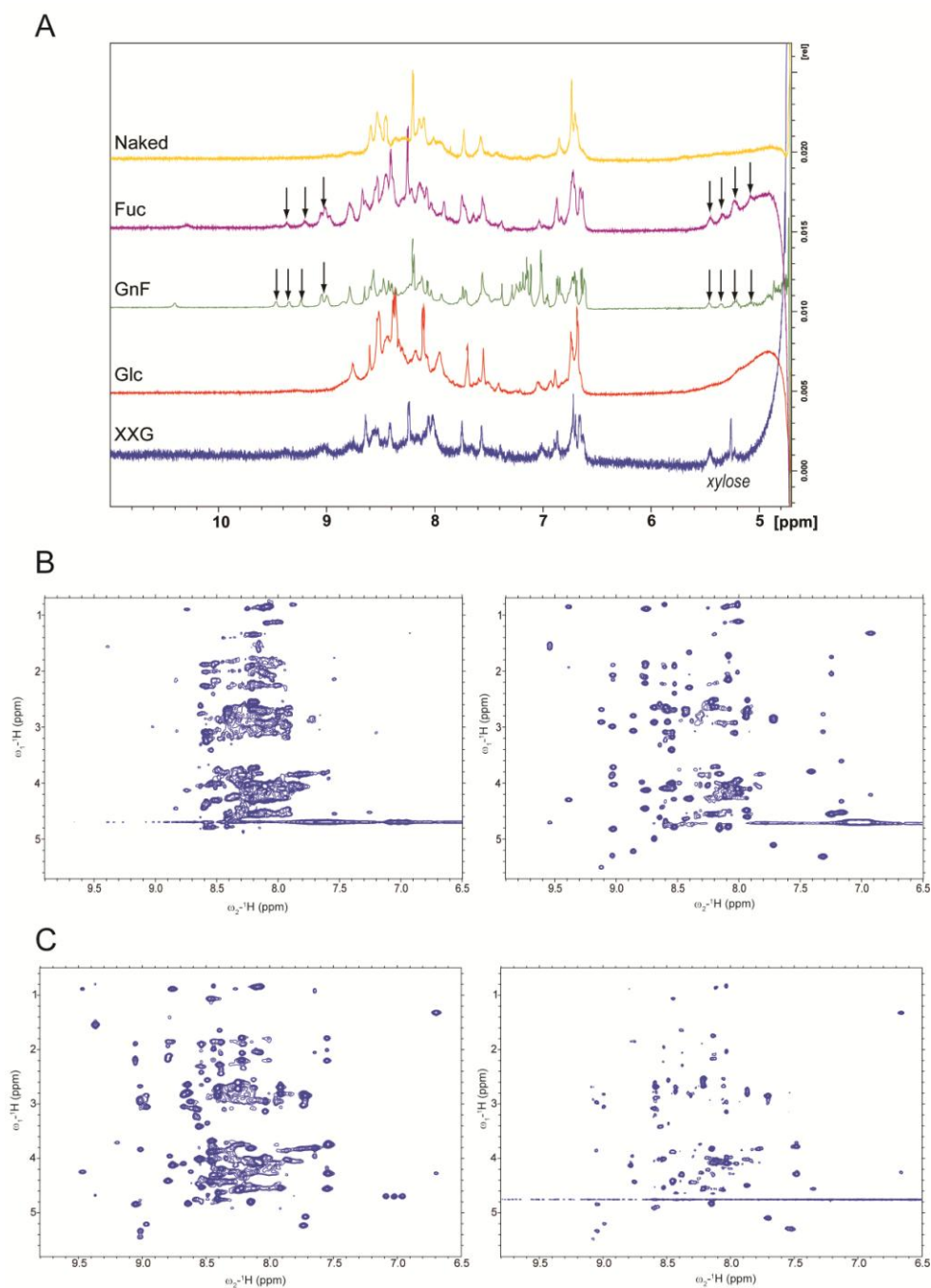


Figure 3-14. ^1H and 2D NMR spectra of synthetic EGF12 analogues
 (A) ^1H NMR of naked (**10**), Fuc (**11**), GnF (**12**) Glc EGF12 (**13**) and XXG (**14**). Bold arrow indicates HN or H α corresponding to β -sheet structure. (B) TOCSY spectrum of Glc EGF12 (**13**) showing HN region. left: without addition of CaCl_2 . right: with 25 mM of CaCl_2 . (C) TOCSY spectrum of Fuc EGF12 (**11**) showing HN region. left: without addition of CaCl_2 . right: with 25 mM of CaCl_2

NOESY experiments

In order to obtain further conformational characters of *O*-glycosylated EGF12 analogues, two dimensional NMR spectra, TOCSY, NOESY, DQF-COSY, and ¹³C-edited HSQC were collected in 90% H₂O/10% D₂O or in 99.9% D₂O using Bruker AVANCE 800 MHz. The 25 mM CaCl₂ was added to each NMR solution to obtain stable EGF12 conformation. In NMR assignment, the sequential connectivity of peptide backbone between HN (*i*) and H α (*i*+1) were detected with strong or medium intensities for all amino acid residues except for proline in NOESY spectra. All side chain protons were assigned using TOCSY and NOESY connection.

NOESY spectra revealed the disulfide bond pairs through NOE correlation between β -protons in cross-linked cystein residues. NOE correlation of C1-C3, C2-C4 and C5-C6 were found in naked EGF12 (**10**) so that the formation of proper disulfide pairs were confirmed. In *O*-glycosylated EGF12 (**11-15**) analogues, NOE signals related to C2-C4 and C5-C6 were found. Although NOE interaction of C1-C3 was notable to be found due to peak overlap, these two NOE correlations are sufficient to determine the correct disulfide bond pairs (**Figure 3-16**). The proper disulfide bond pairs were detected in all EGF12 analogues using NOESY spectra.

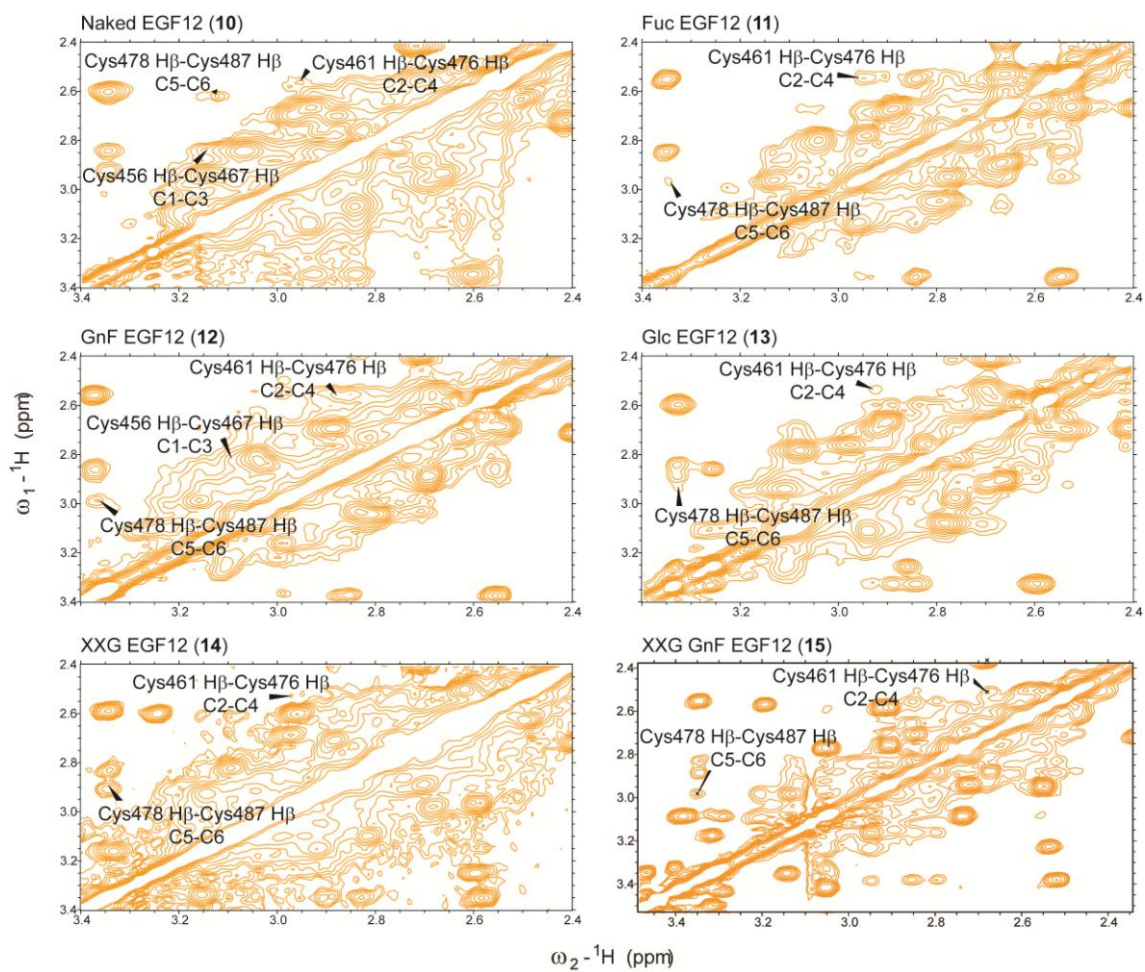


Figure 3-15. NOESY spectra indicating disulfide bond pairs

Further NOESY experiments revealed that *O*-glucose as well as *O*-fucose glycans have specific NOE interactions with peptide chain. With respect to Fuc EGF12 (**11**), *O*-fucose residue have NOE interactions with Ala465, Thr466, Ile477, and Met479. Thr466 residue showed NOE connection with anomeric proton of *O*-fucose residue because *O*-fucose residue attached to Thr466. Ala465, Ile477 and Met479 were found to be interacted with methyl protons of *O*-fucose residue, suggesting that these residues interacted with *O*-fucose residue through hydrophobic interaction. These hydrophobic interactions were also found in NMR study of mouse EGF12, and it suggested that *O*-fucose acted as bridge structure between anti-parallel β -sheet also in human Notch1 EGF12. From these NOE connections, it is expected that *O*-fucosylation of human Notch1 EGF12 have similar functions with mouse case in terms of stabilization of β -sheet moiety.¹⁹ In GnF EGF12 (**12**), *O*-fucose have the same NOE connection as Fuc EGF12. However, GlcNAc residue showed NOE interaction only with proximal fucose residue but not with peptide moiety, suggesting that elongated GlcNAc residue might have little effects on peptide conformation (**Figure 3-16**, blue).

With respect to Glc EGF12 (**13**), NOE interactions of *O*-glucose residue were only found with underlying Ser7 residue. In detail, NOE signals of anomeric proton of *O*-glucose residue and β -protons of Ser458 were observed due to covalent attachment of *O*-glucose and serine. No NOE signals with other amino acids were observed. On the other hand, specific interaction between Xyl-Xyl-Glc trisaccharide and peptide chain was observed in XXG EGF12 (**14**). The NOE connection between anomeric proton of xylose at nonreducing end and H α of Glu473 on β -sheet region was found although these residues located far away in primary sequence (**Figure 3-16**, red). This specific NOE contacts between carbohydrate and peptide chain should be important clue to infer

the *O*-glucose trisaccharide roles in Notch biological functions. In the case of di-substituted XXG/GnF EGF12 (**15**), NOE correlation patterns between *O*-glycans and peptide chain showed no difference with mono-substituted EGF12 analogues (**11-14**). It means that conformational impacts of *O*-fucose and *O*-glucose glycans on human Notch1 EGF12 would be independent of each other.

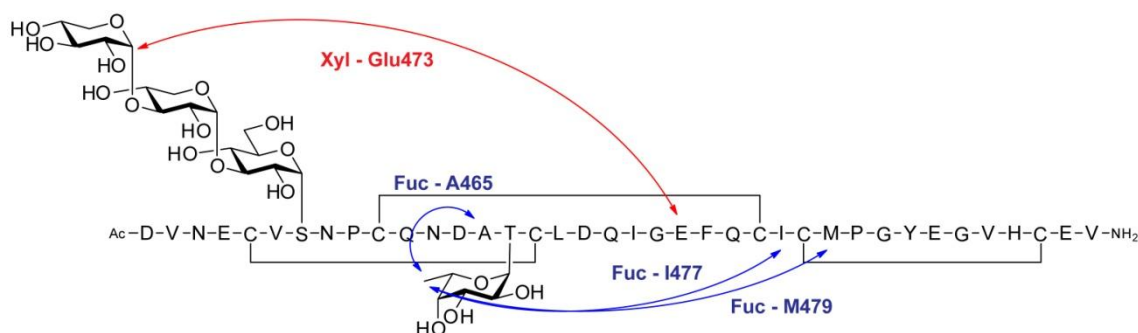


Figure 3-16. NOE interaction between carbohydrate and peptide found in NOESY experiment

¹H chemical shift difference on H α

Next, comparative analyses of chemical shift alteration in H α of peptide back bone were demonstrated between naked EGF12 (**10**) and each glycosylated EGF12 analogues (**11-15**) to predict *O*-glycosylation impacts on peptide backbone. With respect to Fuc (**11**) and GnF EGF12 (**12**), similar chemical shift changing were observed, suggesting that *O*-fucosylation affected peptide conformation but further GlcNAc elongation at fucose residue gave less conformational impacts on peptide backbone (**Figure 3-17**). Compared with naked EGF12 (**10**), *O*-fucosylation gave lower-field shift at *O*-fucosylated Thr466, and higher-field shift at Met479. This chemical shift difference of Met479 would be due to carbohydrate-peptide interaction found in NOESY experiments. These features were also found in mouse EGF12 although some amino

acid residues were changed. In other words, *O*-fucosylation effects might be conserved in human and mouse EGF12.

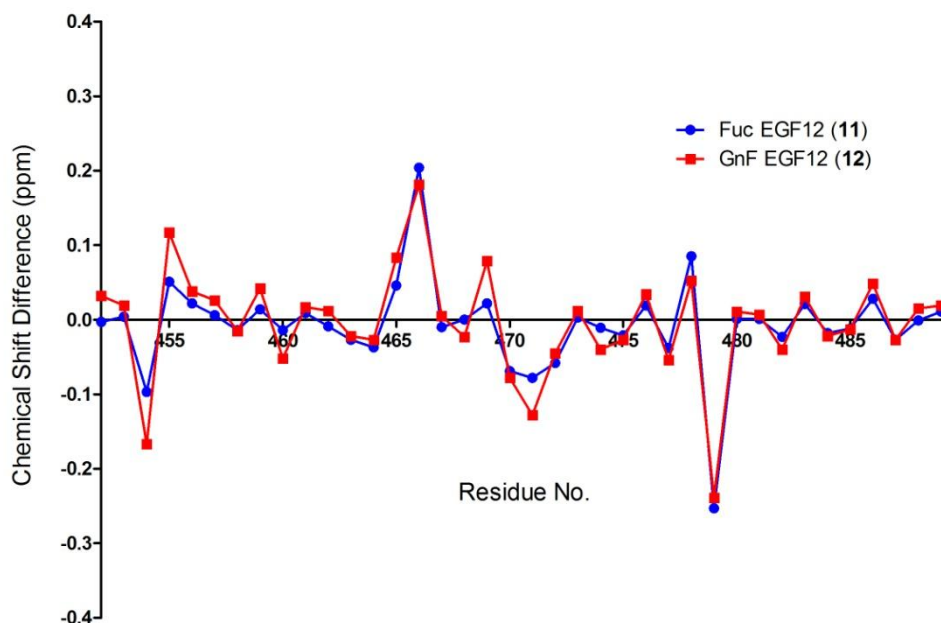


Figure 3-17. H α chemical shift difference between naked and fucosylated EGF12

On the other hand, Glc EGF12 analogue (**13**) showed lower-field shift at glycosylated Ser458 residue as found in *O*-fucosylation and similar chemical shifts of H α proton with naked EGF12 (**10**) other than Ser458 (**Figure 3-18**). This chemical shift difference indicated that *O*-glucose modification have little effect on peptide conformation. However, some residues in XXG EGF12 (**14**) showed large chemical shift difference compared to Glc EGF12 (**13**). Especially, H α of Phe474 showed lower-field shift over 0.2 ppm in XXG EGF12. In NOESY experiments, no specific glycan-peptide correlation was observed in Glc EGF12 (**13**) except for *O*-glucose with underlying serine residue, however, XXG EGF12 showed that specific NOE connection between xylose residue at nonreducing end and Glu22 residue on β -sheet moiety. The chemical shift difference as well as NOE contacts of XXG trisaccharide suggested that

Xyl-Xyl-Glc modification might produce some conformational impacts on EGF12 peptide backbone although glucose monosaccharide modification did not show any effects on backbone chemical shift. In the case of di-substituted EGF12 (**15**), chemical shift differences were observed at many amino acid residues (**Figure 3-19**). However, it has similar chemical shift changing at Ser458, Thr466, Phe474, and Met479 with mono-substituted EGF12 analogues (**11-14**), suggesting that similar conformational features might be induced by glycosylation between mono-substituted and di-substituted EGF analogues.

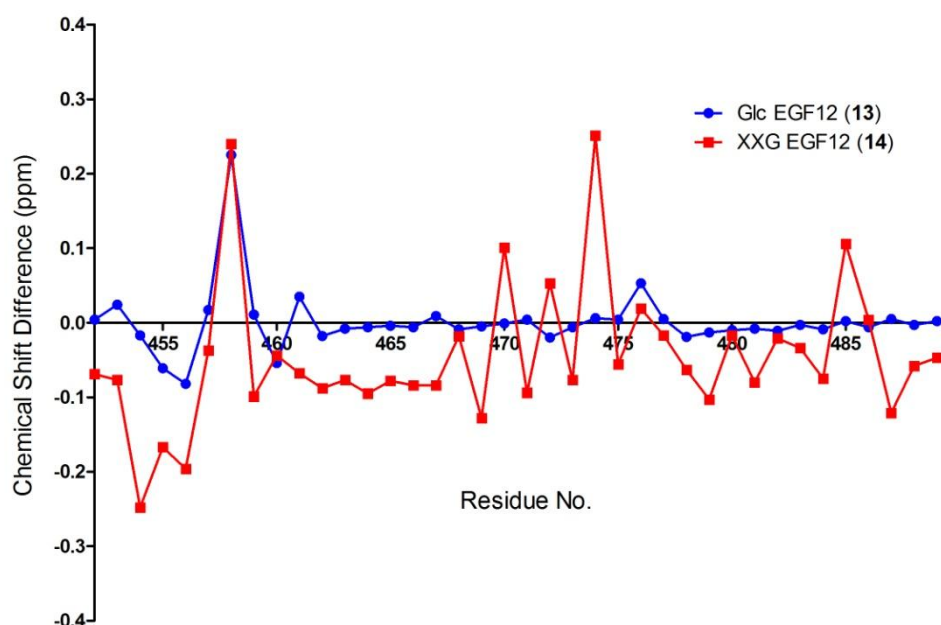


Figure 3-18. H α chemical shift difference between naked and glucosylated EGF12

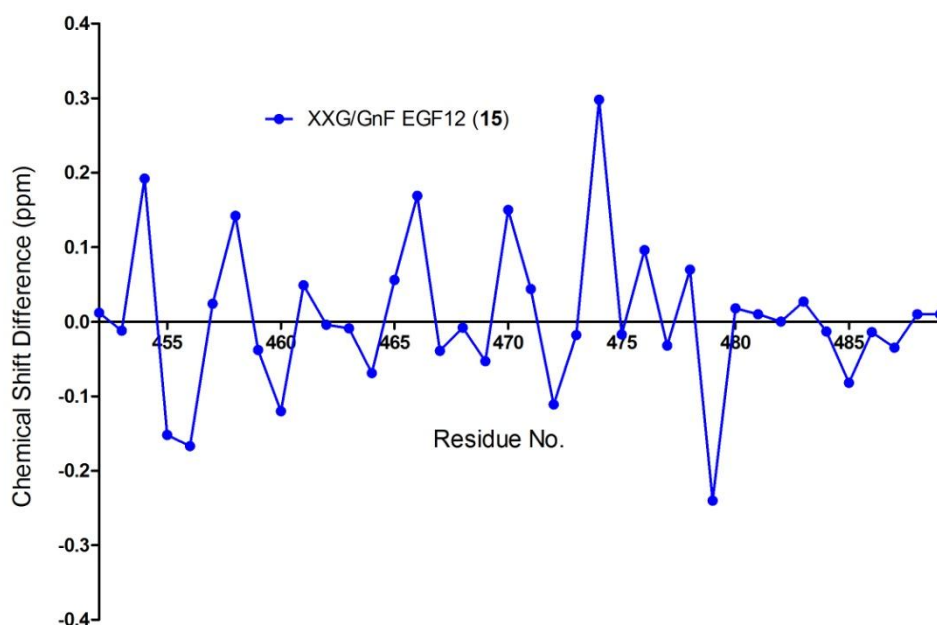


Figure 3-19. H α chemical shift difference between naked and XXG/GnF EGF12

NMR Structural analysis

To reveal *O*-glycosylation impacts on human Notch1 EGF12 peptide backbone, three-dimensional NMR structure of synthetic EGF12 analogues (**10-15**) were calculated. NMR structure will provide us clear conformational difference induced by *O*-fucosylation and *O*-glycosylation at EGF12 peptide. For NMR structural calculation, distance and geometry information were obtained from two-dimensional NMR measurement. The intensities of all NOESY signals were clarified as strong, medium, weak, and very weak, and proton-proton distance were set as 2.4, 3.5, 4.0, and 6.0 Å, respectively. When coupling constant between HN and H α determined by DQF-COSY become more than 8.0 Hz, the dihedral angle HN-N-C α -H α (ϕ) was set at $-120 \pm 30^\circ$. NMR structures of synthetic EGF12 analogues were generated with CNS 1.1 program software using these distance and angle information. The lowest 20 energy structures of synthetic EGF12 analogues (**10-15**) were superimposed by Cys456-Asp467 and

Glu473-Pro480 residues with mean structure as bold line (**Figure 3-20**, left). The most favored structure of each EGF12 analogue (**10-15**) are presented in ribbon structure, showing that each EGF12 analogue have β -sheet structure in the center region of peptide backbone (**Figure 3-20**, right).

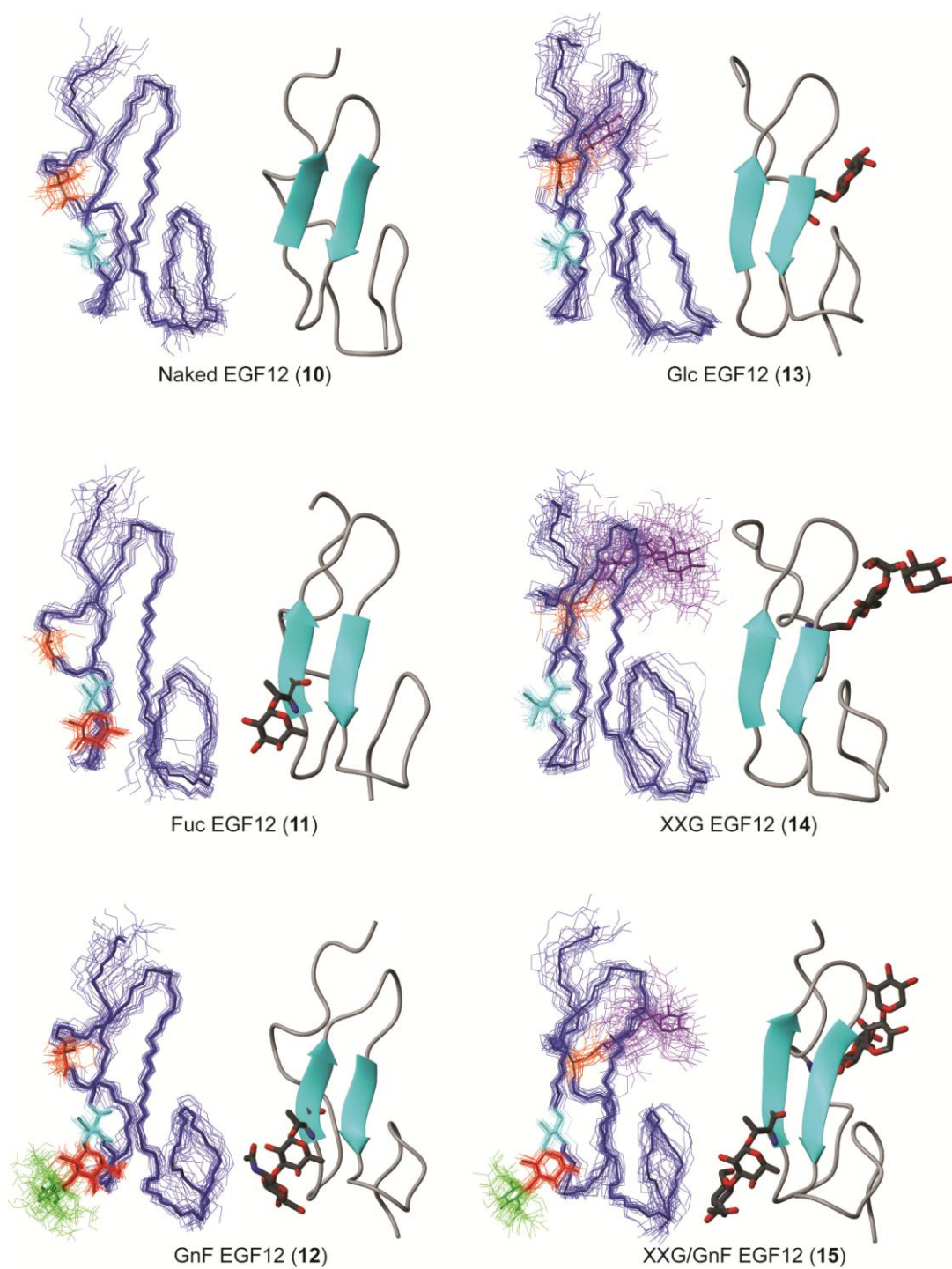


Figure 3-20. NMR structure of synthetic human Notch1 EGF12

The 20 lowest energy structure (left) and ribbon structure (right), Orange: Ser458 (O-glucosylation site), Cyan: Thr466 (O-fucosylation site), Red: Fuc, Green: GlcNAc, Purple: Glc or XXG, Blue: peptide backbone (bold: mean structure)

Each EGF12 analogue showed similar entire peptide conformation adopting anti-parallel β -sheet structure, suggesting that anti-parallel β -sheet structure on EGF12 can be formed in the presence of Ca^{2+} and it might be independent of *O*-glycosylation. According to the lowest 20 energy structures, EGF domain adopted well converged β -sheet moiety which is reported as ligand binding region, but other region, including β -hairpin and terminus region, showed relatively disordered structure. The lack of covalent attachment of EGF11 and 13 might induce these conformational disorders because of no hydrophobic packing with neighboring EGF domains.³³ The RMSD values of superimposed region, Cys456-Asp467 and Glu473-Pro480, including the anti-parallel β -sheet, were determined 1.03 Å for naked naked EGF12 (**10**), 0.98 Å for Fuc EGF12 (**11**), 1.10 Å for GnF EGF 12 (**12**), 0.91 Å for Glc EGF12 (**13**), 0.94 Å for XXG EGF12 (**14**), and 0.89 Å for XXG/GnF EGF12 (**15**), respectively, which indicated that each EGF12 analogue had good convergence in β -sheet region.

NMR Structural analysis of O-fucosylated EGF domain

Next, we focused on *O*-fucosylated and *O*-glucosylated region of NMR structures to explore *O*-glycosylation effects on the peptide structure of synthetic EGF12 analogues. In NOESY experiments of fucosylated EGF12 **11** and **12**, methyl group of fucose residue showed correlation with Ile477 and Met479. NMR structure of Fuc EGF12 as well as GnF EGF12 showed that the methyl group of fucose faced to Ile477 and Met479 on the opposite side of anti-parallel β -sheet (**Figure 3-21**). Atomic distances of methyl carbon of fucose with C β of Ile477, and with C α of Met479 in mean structure of fucosylated EGF12 (**11**, **12**) were calculated as 3.51 Å and 4.17 Å (**11**), or 3.52 Å and 3.55 Å (**12**), respectively. These close distances indicate the presence of hydrophobic

interaction between *O*-fucose residue and Ile477 as well as Met479, which fix the spatial arrangement of *O*-fucose residue (**Figure 3-20**). This *O*-fucose bridge between anti-parallel β -sheet provide conformational stability in Fuc and GnF EGF12 analogues found in TOCSY experiments (**Figure 3-14**). In GnF EGF12 (**12**), any interactions between GlcNAc residue and peptide chain was not observed in neither NOESY experiment nor NMR structure, which induced disordered GlcNAc structure in **Figure 3-20** (green).

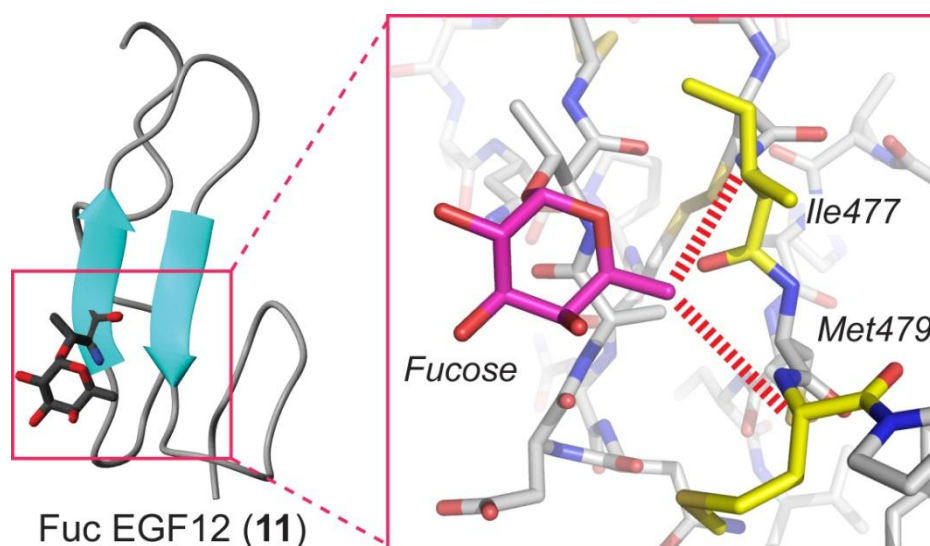


Figure 3-21. Fucose residue interacted with peptide chain in Fuc EGF12

To understand the conformational impacts on peptide backbone induced by *O*-fucosylation, backbone mean structure of naked, Fuc, and GnF EGF12 (**10-12**) were superimposed (**Figure 3-22**). As seen in **Figure 3-22**, there are no significant difference between naked and fucosylated EGF12. *O*-fucosylation as well as GlcNAc elongation did not affect the conformation of Thr466, *O*-fucosylation site, and I477 and M479 which is interacted with *O*-fucose residue, suggesting that fucosylation does not induce significant conformational alteration but it can stabilize anti-parallel β -sheet structure to

form “fucose bridge”. Furthermore GlcNAc elongation provides no conformational impacts on peptide backbone. From these conformational comparisons, it is suggested that alteration of binding affinity with Notch ligands induced by GlcNAc elongation might be due to additional sugar-ligand interaction but not due to GlcNAc-induced conformational alteration. Modeling analysis with co-crystal structure of mouse Notch1 EGF11-13 and Delta-like 4 suggested that GlcNAc residue is likely to interact with the amino acid residue on Delta-like 4.²⁰ Therefore, the major O-fucoylation impacts revealed in this study is stabilization of anti-parallel β -sheet by forming bridge structure, but not alteration of peptide conformation from naked peptide. Although carbohydrate-induced conformational changes were not observed, sugar modification on Notch1 EGF12 will increase Notch-ligand interactions to regulate binding affinity with ligand.

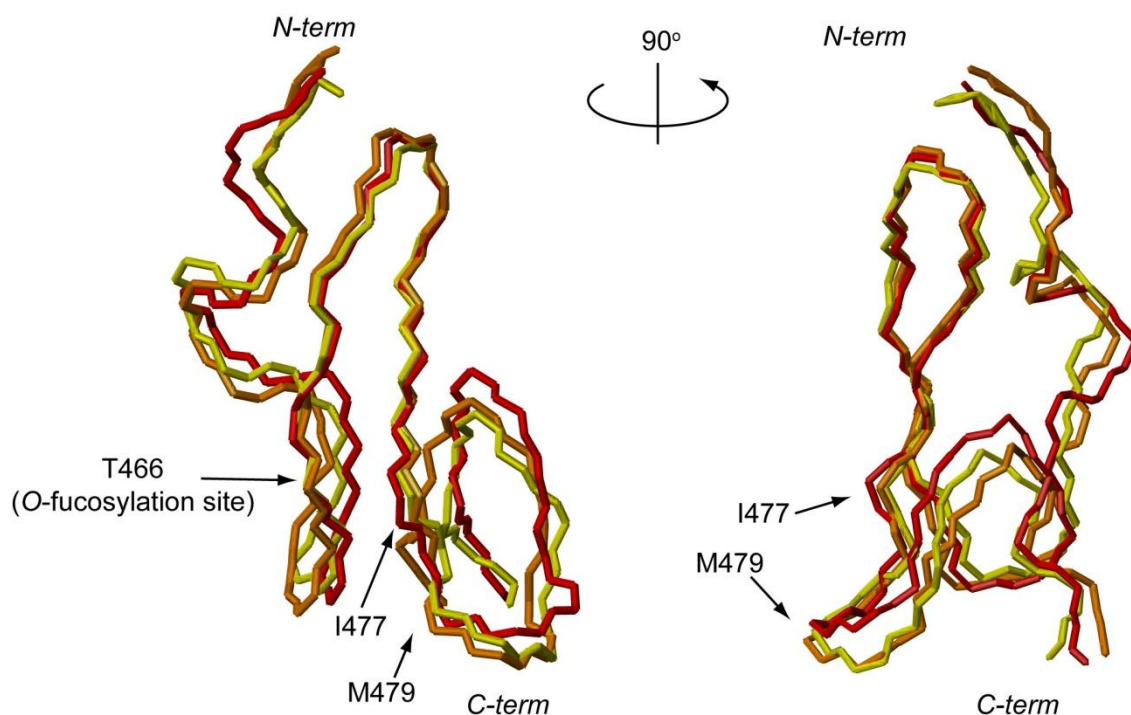


Figure 3-22. Comparisons of mean structure of naked and fucosylated EGF12. Backbone mean structures were represented. T466 is O-fucosylated residue. I477 and M479 interact with O-fucose. Red: naked (**10**), Orange: Fuc (**11**), Yellow: GnF (**12**).

NMR Structural analysis of O-glucosylated EGF domain

O-glucosylation impacts on peptidic backbone were also characterized by NMR structural analysis. It is important to note that, although O-fucosylation has crucial effect on Notch ligand binding, O-glucose does not affect ligand binding but regulate Notch signaling activation. Judging from three-dimensional NMR structure of O-glucosylation site, these O-glucosylations seemed to be independent of Notch-ligand binding because the binding region of Notch1 EGF12 was reported at O-fucosylation face of anti-parallel β -sheet (**Figure 3-20**).³⁹ With respect to NMR analyses of O-glycosylated EGF domain, in contrast to O-fucosylation, sugar-peptide correlation was not observed in Glc EGF12 (**14**) in NOESY experiments, however, double xyloes

elongations afford the NOE connection between xylose residue at nonreducing end and Glu473 residue on β -strand. Three-dimensional NMR structure demonstrated that *O*-glucose trisaccharide on Ser458 reached to β -sheet region, while glucose monosaccharide showed no interaction because of remoteness (**Figure 3-23A**). The distance between Ser458 and Glu473 seems far in primary sequence, but these residues locate moderate distance for carbohydrate bridging after adopting tertiary structure. This trisaccharide bridge might affect the conformational stability of anti-parallel β -strands. Compared to fixed conformation of *O*-fucose glycan, *O*-glucose glycan showed highly dispersed structure in both monosaccharide and trisaccharide. No specific NOE interactions between glucose residue and peptide chain as mentioned in 2D NMR experiments provide flexibility to this *O*-glucose moiety. Importantly, NMR structure of *O*-glucosylated EGF analogues demonstrated that the *O*-glucose and XXG trisaccharide cover the hydrophobic pocket formed by Pro460, Phe474 and Val485 residues although NOE signals between carbohydrate residue and these hydrophobic residues were not found. It is likely that XXG glycan in compound **15** will be able to cover larger area of this hydrophobic surface than *O*-glucose monosaccharide in **13** because of carbohydrate bulkiness and the spacial disorder. In addition, this carbohydrate shield might relate to chemical shift difference at Phe474 and Val485 in XXG EGF12 (**14**) shown in **Figure 3-18**. This *O*-glucosylation effect would enhance the peptide solubility and conformational stability by shielding this hydrophobic region from solvent exposure. These conformational stabilization effects given from *O*-glucosylation would be important for Notch signaling activation although detailed molecular mechanism is still unclear.

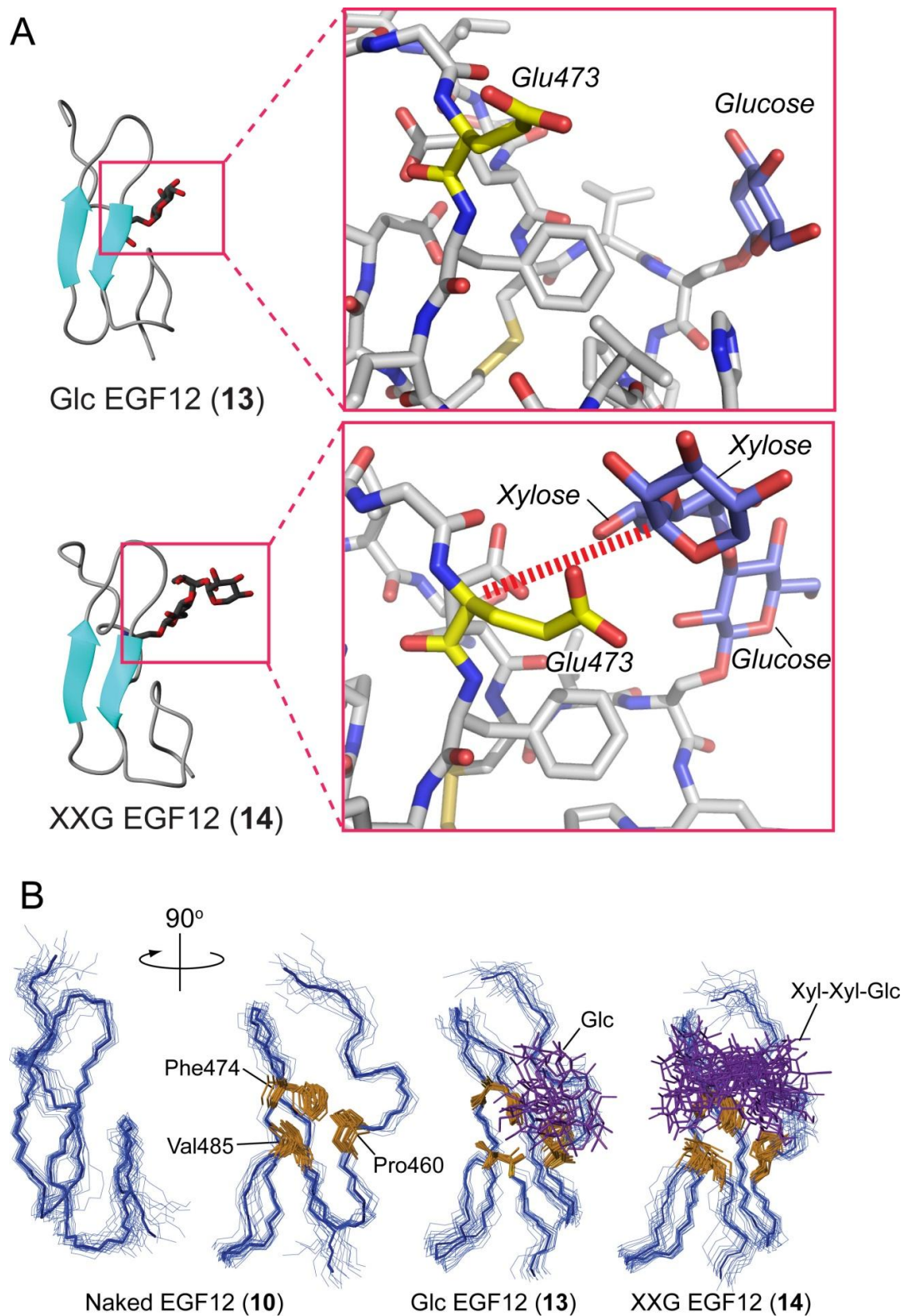


Figure 3-23. O-glucos glycan interact with peptide moiety

To reveal *O*-glucosylation impacts of peptidic backbone, mean structures were superimposed (**Figure 3-24**). It was revealed that *O*-glucosylation including xlyose elongations have no influence on conformation of anti-parallel β -sheet region, meaning that *O*-glucosylation is independent of ligand binding process. This fact is in agreement with previous result determining ligand binding affinity of *O*-glycosylated EGF domains. In addition, it was noteworthy that local conformational changes at Ser458, which is *O*-glucosylated residue, and Pro460 as well as Val485 were observed. These alterations were induced by *O*-glucosylation bridge and hydrophobic shielding. Because *O*-glucosylation did not affect ligand binding, these findings would be related to conformational stabilization induced by *O*-glucosylation.

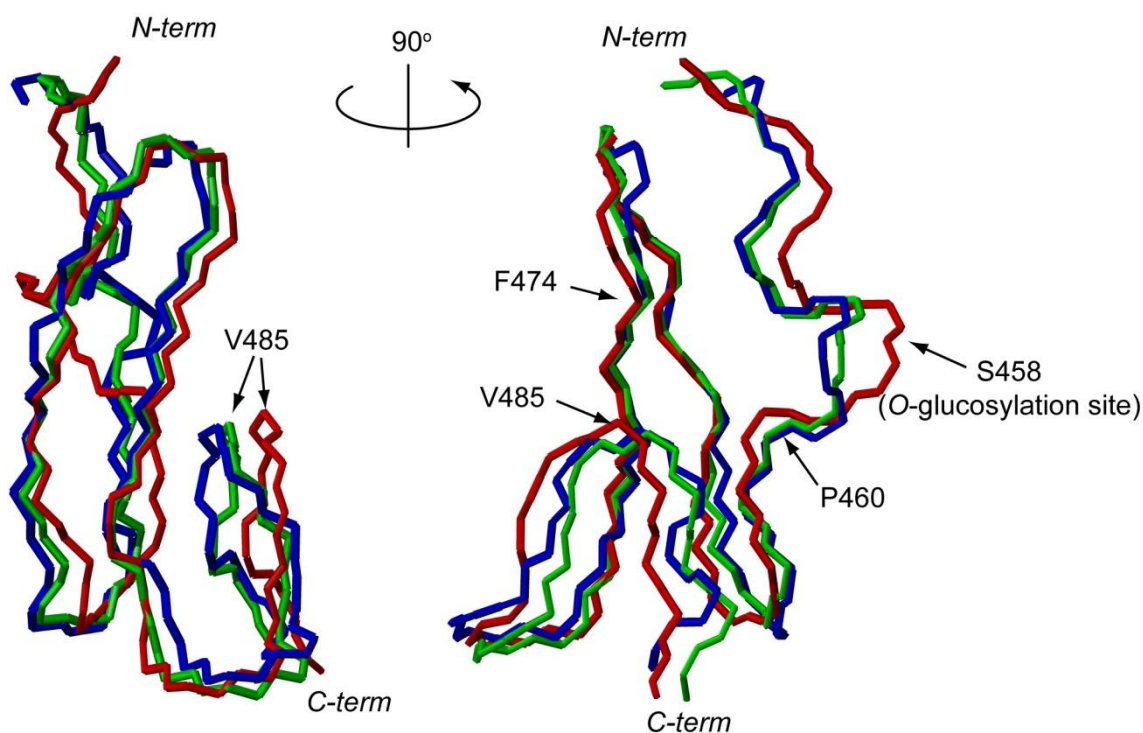


Figure 3-24. Comparison of mean structure of naked and glucosylated EGF12. Backbone mean structures were represented. Red: naked (**10**), Green: Glc EGF12 (**13**), Blue: XXG EGF12 (**14**).

NMR Structural analysis of di-substituted EGF domain

NMR structure of XXG/GnF EGF12 (**15**) carrying Xyl α 1 $\rightarrow$ 3Xyl α 1 $\rightarrow$ 3Glc β 1 $\rightarrow$ and GlcNAc β 1 $\rightarrow$ 3Fuc α 1 $\rightarrow$ was generated to assess conformational impacts of di-modification on peptide backbone. From NOE results showing no significant difference between mono-substituted and di-substituted EGF domain, we expected that O-glucosylation as well as O-fucosylation contribute to peptide conformation independently. As expected, XXG/GnF EGF12 adopted anti-parallel β -sheet structure similar to naked and mono-substituted EGF domains. In addition, O-fucose and O-glucose type glycans demonstrated similar behavior to conformational stabilization (**Figure 3-25**).

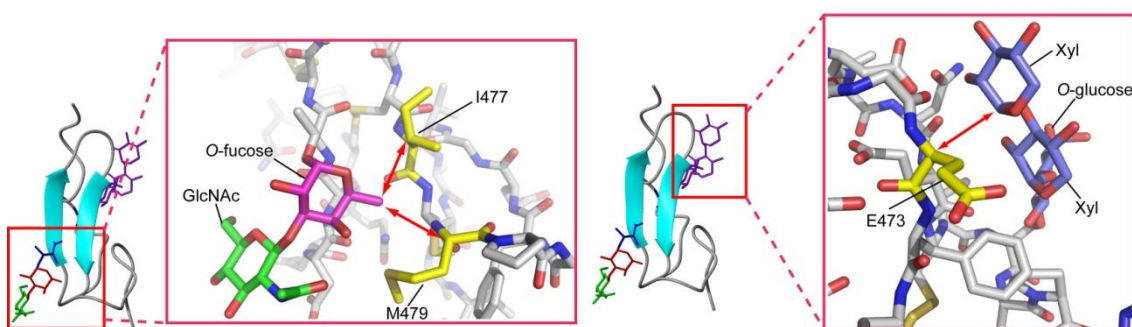


Figure 3-25. O-glycosylation interact with peptide moiety.

Furthermore, conformational comparison with naked EGF domain was performed. **Figure 3-26** indicates that O-fucosylated Thr466 as well as fucose-interacted residues (Ile477 and Met479) of XXG/GnF EGF12 adopts similar conformation to naked EGF, however O-glucosylated Ser458 as well as glucose-interacted residues (Pro460 and V485) showed significant difference during O-glucosylation. These observations were in agreement with mono-glycosylated EGF domains. Thus, we can conclude that

O-fucosylation and O-glucosylation affect independently on conformational stabilization of Notch1 EGF12 analogues.

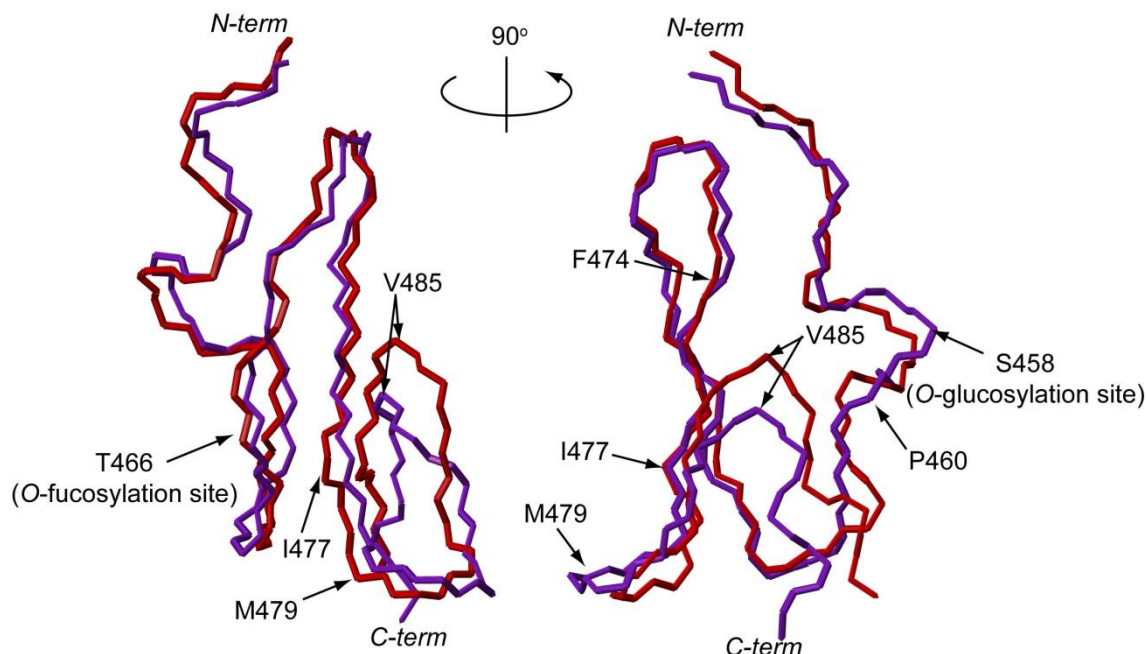


Figure 3-26. Comparison of mean structure of naked and di-substituted EGF12. Backbone mean structures were represented. Red: naked (**10**), Purple: XXG/GnF EGF12 (**15**).

Conformational comparison with reported 3D structure of Notch1 EGF12

Comparison of NMR structure of synthetic human Notch1 EGF12 with reported human and mouse Notch1 EGF12 revealed the importance of Ca^{2+} coordination as well as O-glycans for conformational stabilization. NMR structure of O-glycosylated human Notch1 EGF12 carrying $\text{Xyl}\alpha 1 \rightarrow 3\text{Xly}\alpha 1 \rightarrow 3\text{Glc}\beta 1 \rightarrow \text{Ser458}$ and $\text{GlcNAc}\beta 1 \rightarrow 3\text{Fuc}\alpha 1 \rightarrow \text{Thr466}$ (**15**) is superimposed with synthetic mouse Notch1 EGF12 carrying $\text{GlcNAc}\beta 1 \rightarrow 3\text{Fuc}\alpha 1 \rightarrow \text{Thr466}$ resolved by solution NMR (**Figure 3-27A**).¹⁹ It is noteworthy that solution NMR structure of mouse EGF12 was obtained in the absence of Ca^{2+} . Interestingly, conformational alterations were found not only in N-terminus and β -hairpin region that is responsible for Ca^{2+} binding, but also in the

adjacent *O*-glycosylation site at Thr466, demonstrating that the coordination of Ca^{2+} has crucial roles in the construction of total EGF12 structure. In addition, comparison of NMR structure of human Notch1 EGF12 (**15**) and the corresponding domain of EGF11-13 obtained by X-ray crystallography²⁰ indicated well-converged β -hairpin as well as β -sheet region while *C*-terminus showed large difference due to the lack of covalent attachment of EGF13 in NMR structure (**Figure 3-27B**). Based on accumulating evidence that human Notch1 EGF11-13 have the ability to interact with ligand in a Ca^{2+} -dependent manner, these conformational stabilization factors of Ca^{2+} as well as glycosylation contribute to the Notch signaling in synergistic manner.

To further understand *O*-glycosylation as well as calcium effects on Notch signaling, conformational study for other EGF domains having *O*-fucose, *O*-glucose as well as *O*-GlcNAc type glycans will be necessary because these *O*-glycosylations are widely found in other EGF domains which might participate in Notch signaling pathway. Present strategy for synthesis and NMR experiments of *O*-glycosylated EGF domain are first attempts to elucidate the functions of *O*-glycosylation on Notch EGF domain. *O*-Glucosylation, *O*-fucosylation and calcium effects discussed here should be important information to understand the molecular mechanism of Notch signaling pathway.

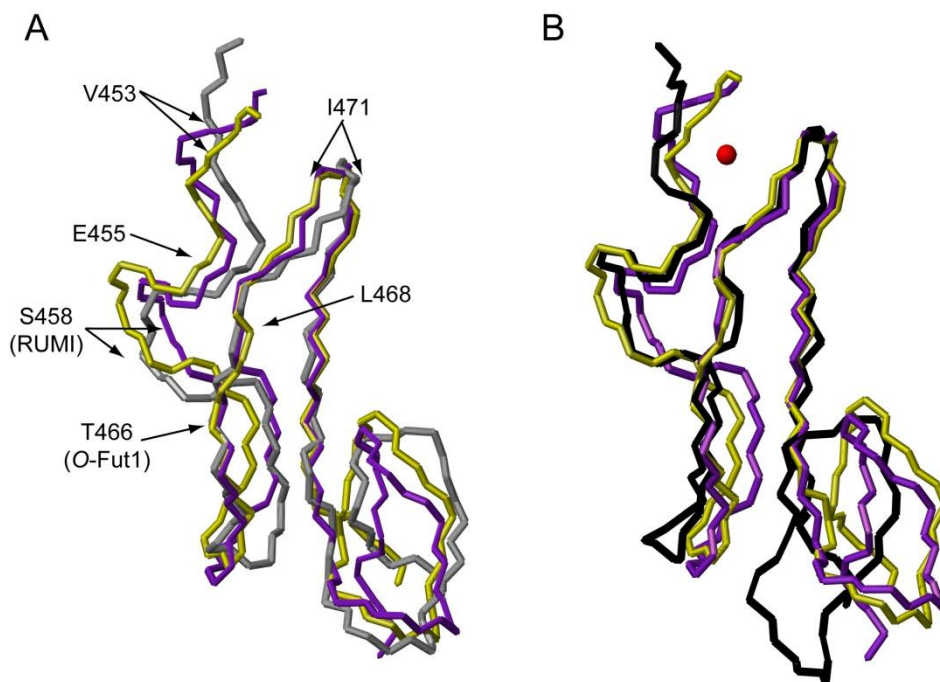


Figure 3-27. Conformational comparison of human Notch1 EGF12 (**15**) and reported Nothc1 EGF12. (A) Comparison of human Notch1 EGF12 carrying GlcNAc β 1 $\rightarrow$ 3Fuc α 1 $\rightarrow$ Thr466 (**15**) and human Notch1 EGF12 carrying Xyl α 1 $\rightarrow$ 3Xyl α 1 $\rightarrow$ 3Glc β 1 $\rightarrow$ Ser456 and GlcNAc β 1 $\rightarrow$ 3Fuc α 1 $\rightarrow$ Thr466 (**16**) represented in yellow and purple, respectively, and NMR structure of mouse EGF12 having GlcNAc β 1 $\rightarrow$ 3Fuc α 1 $\rightarrow$ Thr466 folded in the absence of Ca²⁺ (Gray) (PDB ID: 2RQZ). (B) Comparison of human Notch1 EGF12 (**15**) and (**16**) represented in yellow and purple, respectively, and EGF12 region of human Notch1 EGF11-13 complex carrying GlcNAc β 1 $\rightarrow$ 3Fuc α 1 $\rightarrow$ Thr466 in the presence of Ca²⁺ resolved by X-ray crystallography (black) (PDB ID: 4D0E).

3-2-7. Cellular assay of Synthetic Notch EGF domain

Finally, our interest was focused on feasibility of synthetic EGF domains as new tools to investigate functional roles of posttranslational O-glycosylations in the Notch signaling pathway. Our synthetic EGF domains adopting anti-parallel β -sheet conformation are expected to be able to act as soluble antagonist against Notch ligands expressed on cell surface. This antagonist molecules might module Notch mechanism both *trans*-activation and *cis*-inhibition.^{40,41} To test the ability of our synthetic EGF glycopeptide towards cell proliferation, the naked EGF12 (**10**) and XXG/GnF EGF12 (**15**) were added to cultured breast cancer cell MCF7 and lung adenocarcinoma epithelial cells A549 because Notch signaling can initiate and activate these cancer and is negatively correlated to clinical outcome.⁴²⁻⁴⁵ Preliminary experiments demonstrated that synthetic EGF domain inhibit the formation of large aggregation of cancer cells compared with control cells which is cultured without these molecules (**Figure** 3-28 and 3-29). These results suggest that synthetic EGF12 module influence cellular adhesiveness of MCF7 and A549 while they did not cause serious changes in the cellular morphology when compared with those of the dead cells treated by cisplatin.

MCF-7

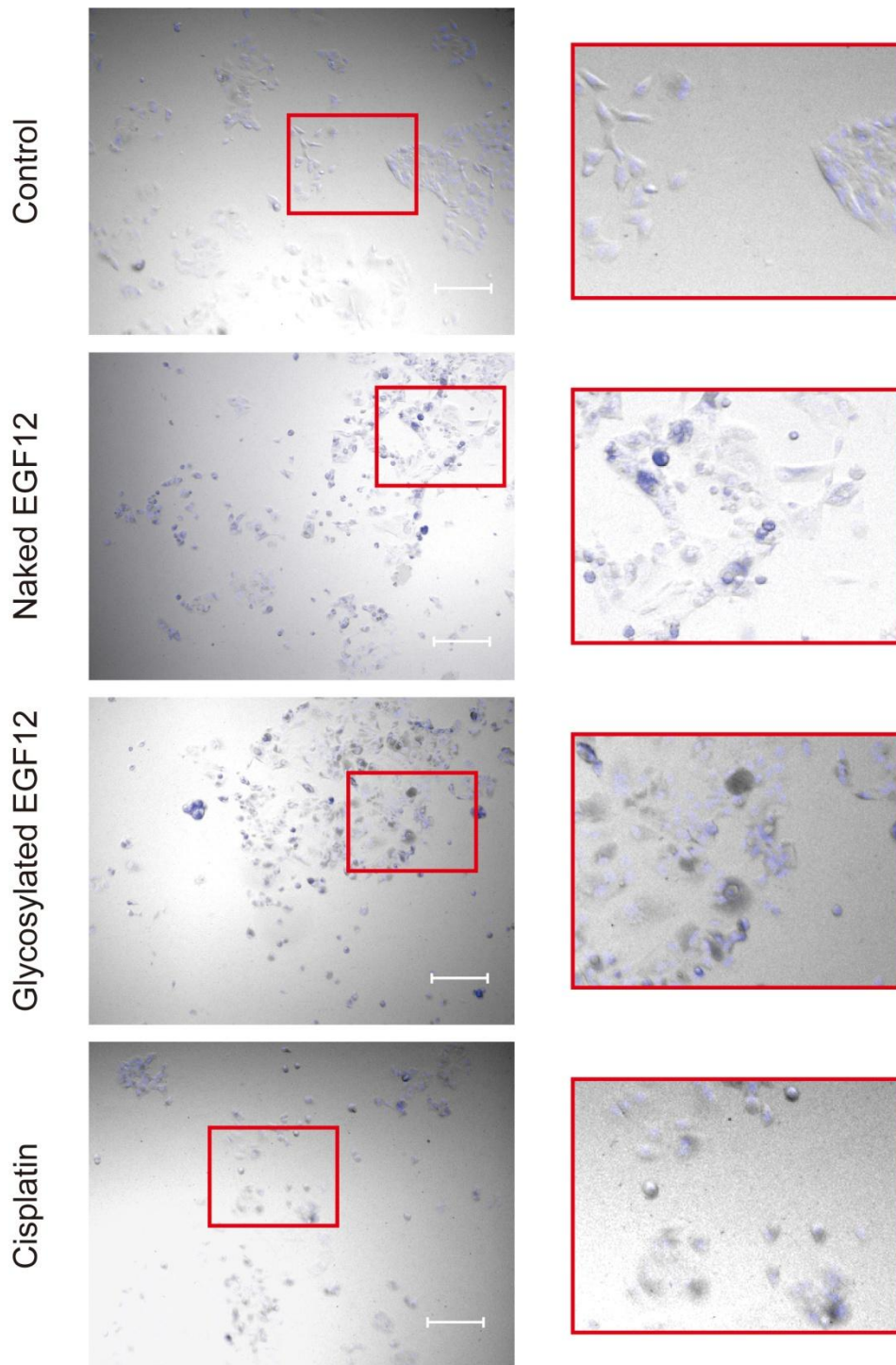


Figure 3-28. Effects of exogenously applied synthetic NOTCH1 EGF12 modules (100 μ M, 96 h) on the proliferation of human breast cancer cell lines (MCF-7). White bar: 200 μ m

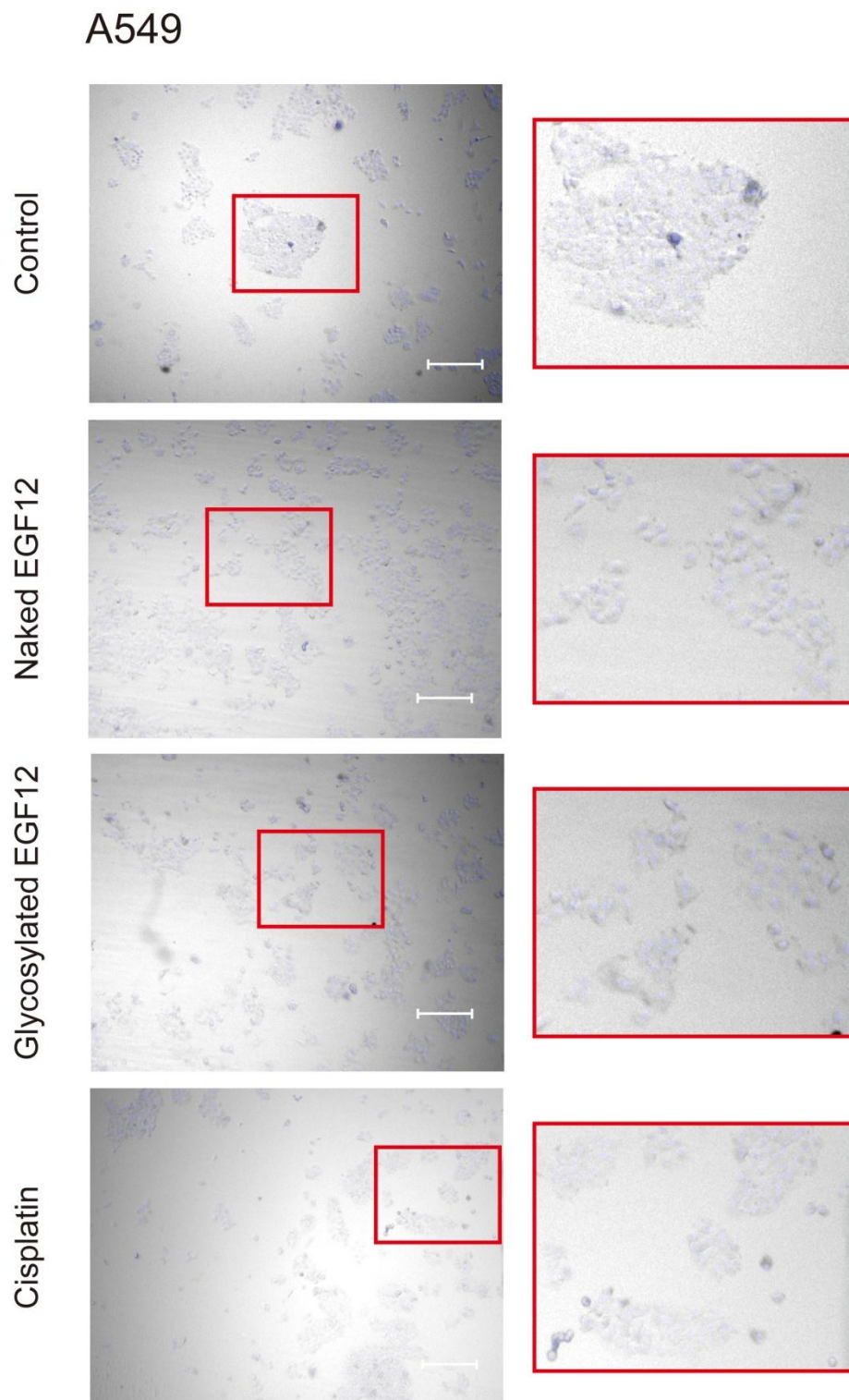


Figure 3-29. Effects of exogenously applied synthetic NOTCH1 EGF12 modules (100 μ M, 48 h) on the proliferation of human lung cancer cell lines (A549). White bar: 200 μ m

In addition, proliferation of these cancer cells was decreased significantly in the presence of 100 μ M of synthetic EGF12 analogues (**10, 15**) at 96 h for MCF7 and 48 h for A549 compared to controls (**Figure 3-30A and B**). Interestingly, the inhibitory effects of naked and XXG GnF EGF12 on these cancer cells were almost the same and cell proliferation after seeding the cells appeared to approximately 80% when compared with the results in the control experiments. As indicated in **Figure 3-30C**, synthetic EGF12 modules inhibited cell proliferation in a dose-dependent manner. These results suggested that the saturation of this inhibitory effect may represent Notch signaling states in the single tumor cell type generated by the integration of the expression levels between Notch receptors and ligands. However, the present preliminary experiments using single tumor type did not show specific functional roles of *O*-glycans in EGF12 domain by comparing with naked EGF12. The molecular mechanisms for the effects of the glycosylations in each EGF domain on Notch receptor signaling remain unclear.

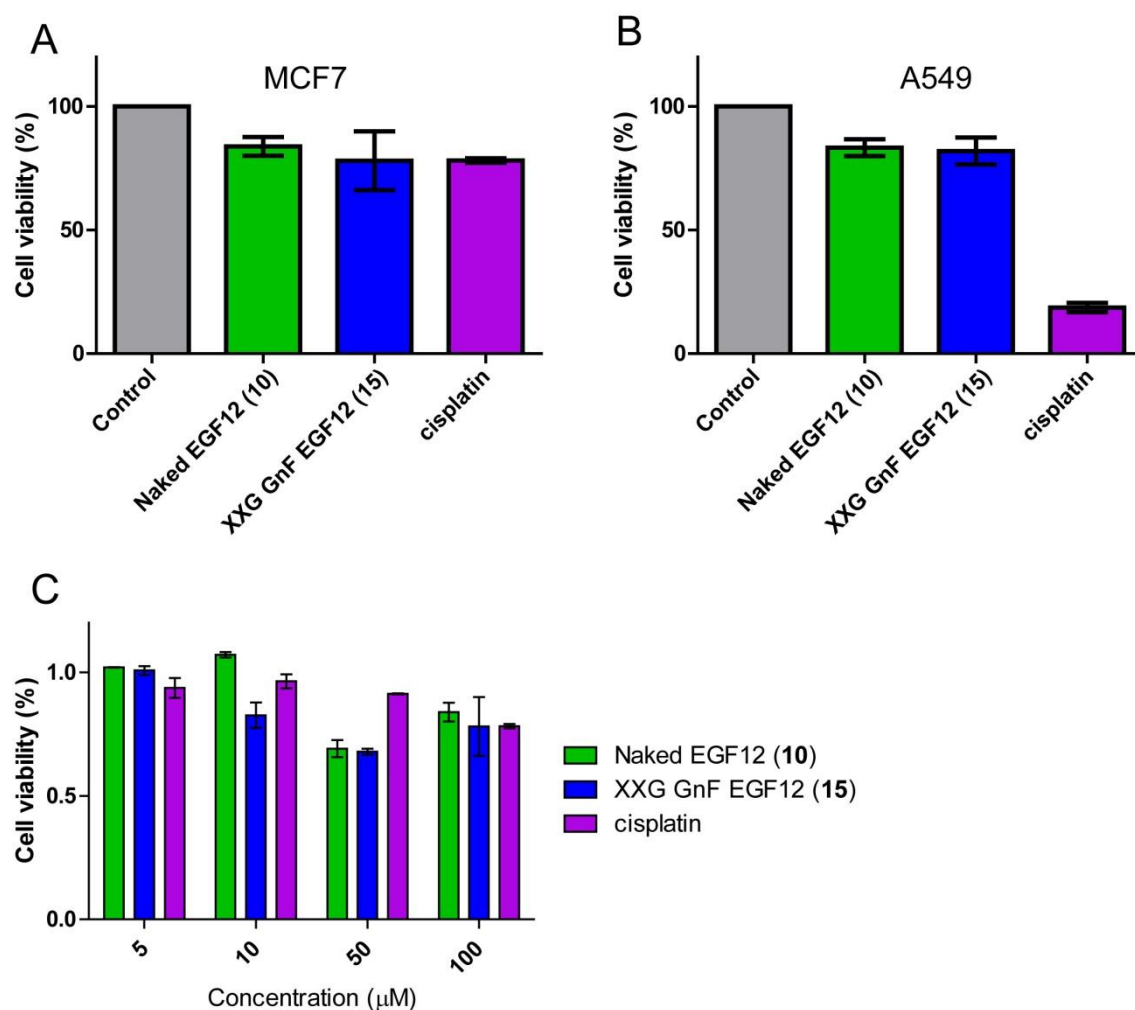


Figure 3-30. Cell viability of MCF7 and A549 with synthetic EGF12

(A) Cell viability of MCF-7 incubated with 100 μ M of synthetic naked EGF12 (**10**), glycosylated EGF12 (**15**), or cisplatin after 96 h. (B) Cell viability of A549 incubated with 100 μ M of synthetic naked EGF12 (**10**), glycosylated EGF12 (**15**), or cisplatin after 48 h. (C) The experiment described in (A) was repeated with various concentrations of reagents, showing the dose-dependency on Notch1 signaling. Proliferation was measured using the MTT assay and the graph shows the mean values of triplicate experiments with standard error of mean (S.E.M.) in (A)~(C). The control shows cell viability without any specific treatment.

3-3. Conclusion

A standard method for rapid and highly efficient synthesis of multiple *O*-glycosylated EGF domains were established in this thesis. A new synthetic Fmoc-glycoamino acid building block, Ca^{2+} -induced disulfide bond formation, and enzymatic sugar elongations afford various kind of human Notch1 EGF 12 analogues carrying unusual multiple *O*-glycans, $\text{Xyl}\alpha 1 \rightarrow 3\text{Xyl}\alpha 1 \rightarrow 3\text{Glc}\beta 1 \rightarrow$ at Ser458 and $\text{GlcNAc}\beta 1 \rightarrow 3\text{Fuc}\alpha 1 \rightarrow$ at Thr466. Surprisingly, it is found that Ca^{2+} contribute not only Notch function but also protein folding. Galactosylation and sialylation of fucose-initiated glycan were carried out with general enzymatic glycosylation methods, which afford tetrasaccharide ($\text{Neu5Ac}\alpha 2 \rightarrow 3\text{Gal}\beta 1 \rightarrow \text{GlcNAc}\beta 1 \rightarrow 3\text{Fuc}\alpha 1 \rightarrow$) at Thr466. NMR analyses of six kinds of synthetic EGF12 modules demonstrated that all peptide backbone adopted anti-parallel β -sheet structure in Notch ligand binding region in the presence of Ca^{2+} ion. In addition, it is revealed that *O*-glucose trisaccharide as well as *O*-fucose interacted with peptide chain while *O*-glucose monosaccharide did not show specific sugar-peptide interactions. These carbohydrate-peptide interactions are important factors to stabilize anti-parallel β -sheet in ligand binding region. Furthermore, Ca^{2+} -coordination is also found to stabilize peptide conformation. The inhibitory effects of soluble synthetic EGF modules on cancer cell lines are the first steps to decipher structural and functional significance of Notch individual domains and posttranslational modification in Notch receptor. Versatility of present approach will provide nice tools for the discovery of novel approach to target the Notch pathway in cancer.

3-4. Experimental Section

Reagents and General Methods

All commercially available solvents and reagents were used without further purification. *N*^α-Fmoc-amino acid derivatives except for glycosylated amino acids were purchased from Merck Millipore (Darmstadt, Germany). 1-[Bis(dimethylamino)methyl]-1H-benzotriazole-3-oxide hexafluorophosphate (HBTU), 1-hydroxybenzotriazole monohydrate (HOBt), *N,N*-diisopropylethylamine (DIEA), *N,N*-dimethylformamide (DMF) and 2,2,2-trifluoroacetic acid (TFA) were purchased from Watanabe Chemical Industries, Ltd. (Hiroshima, Japan). 1H-Benzotriazol-1-yloxy-tri(pyrrolidino)phosphonium hexafluorophosphate (PyBOP) was purchased from Merck Millipore (Darmstadt, Germany). 1-hydroxybenzotriazole (HOAt) was purchased from GenScript. (Piscataway, USA). Thermolysin was purchased from Sigma-Aldrich Japan (Tokyo, Japan). All mixing operation in peptide synthesis was performed by a vortex mixer. All solid-phase reactions for glycopeptides synthesis were performed manually in a polypropylene tube equipped with a filter (LibraTube®, Hipep Laboratories) under microwave irradiation. The microwave was irradiated during coupling reactions and Fmoc removal using temperature control at 50 °C.

A recombinant human β 1,4-Galactosyltransferase (β 1,4-GalT) was purchased from TOYOBO (Osaka, Japan), a recombinant α 2,3-sialyltransferase from *Photobacterium phosphoreum* (α 2,3-SiaT) was purchased from JT (Shizuoka, Japan), and Uridine diphosphate galactose, disodium salt (UDP-Gal) and Cytidine-5'-monophospho-*N*-acetylneuraminic acid, disodium salt (CMP-NANA) were obtained from Yamasa Corporation (Chiba, Japan)

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) were recorded with a Bruker UltraFlex I mass spectrometer in reflector positive or linear positive mode using matrix as 2,5-dihydroxybenzoic acid (DHB) or 2,4,6-trihydroxyacetophenone (2,4,6-THAP). Typically, the sample were dissolved in 1 μ l of 50% (v/v) aqueous acetonitrile and mixed with the same volume of 10 mg/ml DHB or 2,4,6-THP in 50% (v/v) aqueous acetonitrile containing 0.1% TFA. All ^1H - and ^{13}C -NMR spectra for identification of synthetic peptides were collected with 500 MHz Varian UnityInova (Agilent Inc, USA) and 800 MHz Bruker AVANCE (Bruker Biospin Co., Germany). Two-dimensional homonuclear double-quantum-filtered scalar-correlated spectroscopy (DQF-COSY), TOCSY with MLEV-17 sequence, and nuclear Overhauser enhancement spectroscopy (NOESY) spectra were recorded in the indirect dimension using State-TPPI phases cycling.

Reverse-phase HPLC (RP-HPLC) separation and analysis were performed on a Hitachi system equipped with L-6250 intelligent pump and L-7400 UV detector using a reverse-phase C18 column Inertsil ODS-3 250 $\times$ 20 mmI.D. (GL Sciences Inc., Tokyo, Japan) or on a Hitachi system equipped with L-7100 pump and L-7405 UV detector using 250 $\times$ 4.6 mmI.D. (GL Sciences Inc., Tokyo, Japan)

Solid Phase Peptide Synthesis of EGF12 analogues^{29,30}

Solid-phase syntheses of EGF12 analogs (**10-15**) were performed as described below. Rink amide ChemMatrix resin (0.48 mmol/g, 24 μ mol) was swollen with dichloromethane (DCM) for 1 h at ambient temperature. Fmoc group was removed with 20% piperidine in DMF (1 mL) for 3 min under microwave irradiation. *N* $^\alpha$ -Fmoc-amino acid (4.0 equiv.) was coupled with HBTU (4.0 equiv.), HOBt (4.0 equiv.) and DIEA (8.0

equiv.) in DMF (240 μ l) for 10 min under microwave irradiation. In the case of glycoamino acids, Fmoc-Ser(Ac₄Glc β 1 $\rightarrow$)-OH or Fmoc-Ser[Ac₃Xyl(α 1 $\rightarrow$ 3)Ac₂Xyl(α 1 $\rightarrow$ 3)Ac₃Glc β 1 $\rightarrow$]-OH at Ser7 (1.2 equiv., each), and Fmoc-Thr(Ac₃Fuc α 1 $\rightarrow$)-OH or Fmoc-Thr[Ac₃GlcNAc(β 1 $\rightarrow$ 3)Fuc α 1 $\rightarrow$]-OH at Thr15 (1.2 equiv., each) was coupled with PyBOP (1.2 equiv.), HOAt (1.2 equiv.) and DIEA (3.0 equiv.) in DMF/DCM (70/70 μ l) for 10 min under microwave irradiation. After agitation for 10 min under microwave irradiation, the extra PyBOP and HOAt (1.2 equiv., each) were treated for another 10 min period in the same microwave-assisted fashion. At the end of solid-phase synthesis, an *N*-terminal amino group of peptidyl-resin was capped with acetic anhydride with DIEA in DMF for 1 min at ambient temperature. For peptide cleavage from resin and removal of acid-labile protective groups, 1.5 mL of TFA/ethanedithiol/H₂O/triisopropylsilane (94:2.5:2.5:1, v/v/v/v, 1.5 mL) was treated to resin for 2 h at ambient temperature for compound **10-14**. For cleavage of di-substituted EGF analogue (**15**), 1.5 mL of Reagent H (TFA:phenol:thioanisole:1,2-ethanedithiol:H₂O: dimethylsulphide:ammonium iodide = 81:5:5:2.5:3:2:1.5)³¹ was treated to resin for 2 h at ambient temperature. The resin was filtered off and the peptide was precipitated using mixture solution of *tert*-butylmethyl ether (20 ml) and hexane (20 ml) in ice bath without removal of TFA solution. After centrifugation (3,000 rpm, 4 °C, 10 min), supernatant was carefully removed by decantation and repeated three times. The precipitate was dissolved in 50% acetonitrile aqueous (5 mL), and then lyophilized.

With regard to glycosylated EGF12 analogs (**11-15**), the dried precipitate was dissolved in methanol containing 6.5 mM of dithiothreitol (DTT) and the solution was adjusted and kept to pH 12.5 with 1 N NaOH aqueous solution. The reaction mixture

was stirred for 30 min. For deprotection of XXG trisaccharide (**14**, **15**), reaction mixture was stirred for 24 h in the same fashion. The solution was neutralized with 20% acetic acid in methanol, and solvent were removed by evaporator. The excess of DTT was removed by RP-HPLC. When the peptide was precipitated during neutralization after acetyl deprotection, these precipitates were directly subjected to oxidative folding reaction.

Disulfide bond formation of EGF analog

The linear peptide was dissolved in redox buffer (0.1 mg/ml) containing 50 mM Tris-HCl (pH 8.0), 5 mM CaCl₂, 1 mM reduced glutathione and 0.2 mM oxidized glutathione. The reaction mixtures were routinely stirred for 24 h at 4 °C. For di-substituted **15**, the reaction mixture was routinely stirred for 24 h at ambient temperature. Reaction progression was monitored by analytical RP-HPLC. After the reaction proceeding, reaction mixture was acidified by addition of TFA (1% volume of reaction mixture), then lyophilized and subjected to RP-HPLC for purification.

RP-HPLC conditions are shown as below: ambient temperature and flow rate of 5 ml/min, UV detection at 220 nm, eluent A: 0.1% TFA in water, eluent B: 0.1% TFA in acetonitrile, and the liner gradient from 25 to 35% B over 50 min for all EGF12 analogues.

The analytical RP-HPCL gave purified compound naked EGF (**10**) in 10% (10 mg, 2.4 µmol, t_R = 21.08 min), Fuc EGF12 (**11**) in 7.4% (7.8 mg, 1.8 µmol, t_R = 21.37 min), GnF EGF12 (**12**) in 8.5% (9.3 mg, 2.04 µmol, t_R = 18.17 min), Glc EGF12 (**13**) in 6.4% (6.5 mg, 1.5 µmol, t_R = 19.36 min), XXG EGF12 (**14**) in 2.4% (2.8 mg, 0.6 µmol, t_R = 19.10 min), XXG/GnF EGF12 (**15**) in 4.1% (4.9 mg, 0.98 µmol, t_R = 14.88 min).

ESI-HRMS for

10: C₁₇₂H₂₅₇N₄₇Na₃O₆₂S₇ [M+3Na]³⁺ calcd (*m/z*) 1421.8731, found (*m/z*) 1421.8721

11: C₁₇₈H₂₆₇N₄₇Na₃O₆₆S₇ [M+3Na]³⁺ calcd (*m/z*) 1470.5573, found (*m/z*) 1470.5579

12: C₁₈₆H₂₈₀N₄₈Na₃O₇₁S₇ [M+3Na]³⁺ calcd (*m/z*) 1538.2504, found (*m/z*) 1538.2511

13: C₁₇₈H₂₆₇N₄₇Na₃O₆₇S₇ [M+3Na]³⁺ calcd (*m/z*) 1474.2222, found (*m/z*) 1474.2214

14: C₁₈₈H₂₈₃N₄₇Na₃O₇₅S₇ [M+3Na]³⁺ calcd (*m/z*) 1563.9171, found (*m/z*) 1563.9168

15: C₂₀₂H₃₀₆N₄₈Na₃O₈₄S₇ [M+3Na]³⁺ calcd (*m/z*) 1680.2962, found (*m/z*) 1680.2916

*Thermolysin digestion*¹⁹

Thermolysin was dissolved in 50 mM Tris-HCl buffer (pH 8.0), and concentration was adjusted to 0.03 mM. Each resulted peptide (100 µg) was dissolved in 100 µl of 50 mM Tris-HCl buffer (pH 8.0), then diluted enzyme (10 µl) was added to peptide solution. After 24 h incubation at 35 °C, reaction mixture was lyophilized. The dried material was dissolved in 10 µl of water, and then subjected to MALDI-TOF MS analysis.

Enzymatic carbohydrate elongation

Compound **15** (1.0 mg, 0.2 µmol) was dissolved in a total volume of 100 µL of 25 mM HEPES buffer (pH 7.0), 10 mM MnCl₂ and 0.1% Triton-X 100. To the solution were added β1,4-GalT (4 µU) and UDP-Gal (283 µg, 0.5 µmol) and the mixture was incubated for 24 h at ambient temperature. Analytical RP-HPLC gave purified compound **16** in quantitative yield (*t_R* = 11.10 min).

Galactosylated compound **16** (0.1 mg, 0.02 µmol) was dissolved in total 20 µl of 50 mM Tris-HCl (pH 6.5), 500 mM NaCl and 0.1% Triton-X 100. To the solution were

added α 2,3-SiaT (0.2 μ U) and CMP-NANA (127 μ g, 0.2 μ mol) and the mixture was incubated for 24 h at ambient temperature. After 24 h incubation, to a reaction mixture was added extra CMP-NANA (127 μ g, 0.2 μ mol) and reaction mixture was incubated for another 5 h. Analytical RP-HPLC gave purified compound **17** in quantitative yield (t_R = 9.90 min).

RP-HPLC condition is shown as below: ambient temperature and flow rate of 1 mL/min, UV detection at 220 nm, eluent A: 25 mM ammonium acetate buffer (pH 5.0), eluent B: 10% eluent A in acetonitrile, and the linear gradient from 20 to 35% B over 25 min and. ESI-HRMS:

(**16**) $C_{208}H_{316}N_{48}Na_3O_{89}S_7$ $[M+3Na]^{3+}$ calcd (m/z) 1734.3183, found (m/z) 1734.3169.

(**17**) $C_{219}H_{330}N_{49}O_{97}S_7$ $[M-3H]^{3-}$ calcd (m/z) 1807.3481, found (m/z) 1807.3491.

Nuclear magnetic resonance (NMR) spectroscopic experiment

NMR experiments of EGF12 analogues (**10-15**) were performed on a Bruker AVANCE 800 MHz spectrometer for proton frequency. EGF12 analogues were dissolved in 99.9% D_2O (300 μ l) or in a mixture solution of 90% H_2O and 10% D_2O (300 μ l) containing 25 mM $CaCl_2$ at the concentration of 1.5~3.0 mM respectively. These solutions were adjusted at pH 6.0 by 0.1 N NaOH_{aq}. These peptide solutions were packed in *Shigemi* thin-walled micro NMR tubes in NMR spectroscopy. All spectra were measured at 298 K except for temperature-dependent NMR experiments. Data acquisition was performed with the Bruker TopSpin 3.1 software package, Two-dimensional⁴⁶ homonuclear DQF-COSY^{47,48}, TOCSY⁴⁹ with MLEV-17 sequence⁵⁰ and NOESY spectra⁵¹ were

recorded in indirect dimension using States-TPPI phase cycling. Two-dimensional heteronuclear ^{13}C -edited HSQC and HSQC-TOCSY measurements were recorded with echo-antiecho mode. TOCSY experiments were applied for a spin-locking time of 80 msec, and NOESY experiments were carried out with mixing time of 100, 200 and 400 msec. The suppression of water signal was performed by presaturation during 2 sec relaxation delay and by a 3-9-19 WATERGATE pulse sequence with a field gradient.³⁸ TOCSY and NOESY spectra were acquired with 2048 by 512 frequency data point and were zero-filled to yield 2048 by 2048 data matrices. DQF-COSY with 16384 by 512 frequency data points was also recorded and zero-filled to yield 16384 by 16384 matrix in order to measure the coupling constant. The sweep widths of 799.714 Hz were applied. Time domain data in both dimensions were multiplied by a sine bell window function with a 90° phase shift prior to Fourier transformation. All NMR data were analyzed using a Sparky program.⁵²

Structure Calculation

Three-dimensional structures of synthetic EGF12 analogues (**10-15**) were calculated using the CNS 1.1 program⁵² with standard protocols for distance geometry-simulated annealing and refinement. Distance restraints for calculations were estimated from the cross-peak intensities in NOESY spectra with a mixing time of 400 msec. The estimated restraints were classified into four categories: strong (1.6-2.6 Å), medium (1.6-3.5 Å), weak (1.6-5.0 Å) and very weak (1.6-6.0 Å). In the first stage of structure determination, the structures of glycopeptides were calculated using only interproton distance information. After the validation of fulfilling distance restraints for the obtained structures, the restraints of the dihedral angle ϕ were adopted for further structural

calculation. When the coupling constant $^3J_{HN\alpha}$ was more than 8.0 Hz, the dihedral angle ϕ was constricted to $-120 \pm 30^\circ$. The conformation of sugar ring was fixed to the chair conformation. All analyses of rmsd values and the solution structures of glycopeptides were performed with PROCHECK⁵³ and MOLMOL⁵⁴ program. The NMR structures of synthetic EGF12 analogues (**10-15**) were analyzed by MOLMOL and PyMOL program.

*Synthesis of Fmoc-Ser[Ac₃Xyl α (1 $\rightarrow$ 3)Ac₂Xyl α (1 $\rightarrow$ 3)Ac₃Glc β 1 $\rightarrow$]-OH (**21**)*

*1,2;5,6-O-diisopropyliden-3-O-(2,3,4-tri-O-benzyl-D-xylopyranosyl)- α -D-glucopyranoside (**24**)*

Thiophenyl xyloside derivative **22** (15.8 g, 30.9 mmol) and di-acetone D-glucose **23** (6.69 g, 25.7 mmol) were co-evaporated with toluene, and residues were dried-up in vacuum with NIS (7.52 g, 33.4 mmol) and MS4A (5.0 g) for overnight. To the dried residue was added dried THF (200 ml) and dried DCM (100 ml) under Ar gas atmosphere, and stirred for 15 min. TfOH (227 μ l, 2.57 mmol) was added to reaction mixture at -20°C , and mixture was stirred for 2 h. After consuming the donor substrate checked by TLC, TEA was added for neutralization, and then filtered by Celite. Reaction mixture was washed with H₂O, sat aqueous Na₂S₂O₃, aqueous NaHCO₃ and brine. The separated organic layer was dried over MgSO₄, and then the solvent was evaporated. The residue was subjected to column chromatography on silica gel to give **24** as anomeric mixture of xyloside residue (15.6 g, 92%; $\alpha:\beta = 1:1$). α -isomer; ^1H NMR (500 MHz, CHLOROFORM-*d*, δ): 7.30 (m, 15 H, Bn Ar), 5.91 (d, $J=3.7$ Hz, 1 H, H-1), 5.14 (d, $J=3.7$ Hz, 1 H, H-1'), 4.87 (m, 2 H, Bn CH₂), 4.72 (m, 4 H, Bn CH₂), 4.61 (d, $J=3.1$ Hz, 1 H, H-2), 4.50 (m, 1 H, H-5), 4.11 (dd, $J=8.1, 2.7$ Hz, 1 H, H-4), 4.04 (m, 2 H, H-6), 3.87 (t, $J=9.0$ Hz, 1 H, H-3'), 3.59 (m, 1 H, H-4'), 3.46 (dd, $J=9.7,$

3.4 Hz, 1 H, H-2'), 3.19 (dd, $J=11.7, 9.7$ Hz, 1 H, H-5'a), 1.47 (s, 3 H, CH₃), 1.41 (s, 3 H, CH₃), 1.35 (s, 3 H, CH₃), 1.24 (s, 3 H, CH₃); ¹³C NMR (125 MHz (HSQC), CHLOROFORM-*d*, δ): 127.95 (Ar), 105.22 (C-1), 93.38 (C-1'), 84.12 (C-2), 81.33 (C-4), 80.81 (C-3'), 80.44 (C-3), 79.98 (C-2'), 77.94 (C-4'), 75.53-75.69 (Bn CH₂), 72.29 (C-5), 66.98 (C-6), 64.18 (C-5'), 25.46-26.91 (CH₃); β -isomer; ¹H NMR (500 MHz, CHLOROFORM-*d*, δ): 7.30 (m, 15 H, Ar H), 5.72 (d, $J=3.7$ Hz, 1 H, H-1), 4.87 (m, 4 H, Bn CH₂), 4.72 (m, 6 H, Bn CH₂), 4.64 (m, $J=3.1$ Hz, 2 H, Bn CH₂), 4.48 (d, $J=3.7$ Hz, 1 H, H-2), 4.41 (d, $J=7.4$ Hz, 1 H, H-1'), 4.36 (m, 1 H, H-5), 4.32 (m, 1 H, H-4), 4.27 (d, $J=2.9$ Hz, 1 H, H-3), 4.04 (m, 4 H, H-6), 3.72 (dd, $J=9.3, 4.1$ Hz, 1 H, H-5'b), 3.59 (m, 2 H, H-4', H-3'), 3.53 (m, 1 H, H-5'a), 3.30 (t, $J=8.1$ Hz, 1 H, H-2'), 1.50 (s, 3 H, CH₃), 1.41 (s, 3 H, CH₃), 1.31 (s, 3 H, CH₃), 1.23 (s, 3 H, CH₃); ¹³C NMR (125 MHz (HSQC), CHLOROFORM-*d*, δ): 127.95 (Ar), 104.88 (C-1), 101.88 (C-1'), 83.70 (C-3'), 82.56 (C-2), 81.62 (C-2'), 80.15 (C-3), 80.04 (C-4), 77.94 (C-4'), 75.53-75.69 (Bn CH₂), 73.45 (C-5), 66.98 (C-6), 61.01 (C-5'), 25.46-26.91 (CH₃); ESI-TOF MS (m/z) [$M+Na$]⁺ calcd for C₃₈H₄₆NaO₁₀, 685.2989; found, 685.2978

1,2-O-isopropyliden-3-O-(2,3,4-tri-O-benzyl- α -D-xylopyranosyl)- α -D-glucopyranoside (25)

The compound **24** (28.8 g, 43.5 mmol) was dissolved in 80% AcOH aqueous solution (200 ml), and stirred at 40 °C for 3 hs. α - and β -anomers were separated in TLC. The product was extracted by EtOAc, and washed by H₂O, sat NaHCO_{3aq} and brine. The separated organic layer was dried over MgSO₄, and then the solvent was evaporated. The residue was subjected to column chromatography on silica gel to give **25** as α -isomer (12.0 g, 41 %); ¹H NMR (500 MHz, CHLOROFORM-*d*, δ): 7.27-7.32 (m, 15

H, Ar H), 5.92 (d, $J=3.7$ Hz, 1 H, H-1), 4.95 (m, 1 H, Bn CH₂), 4.86 (d, $J=6.9$ Hz, 1 H, Bn CH₂), 4.86 (d, $J=15.8$ Hz, 1 H, Bn CH₂), 4.73 (d, $J=11.5$ Hz, 1 H, Bn CH₂), 4.64 (d, $J=3.7$ Hz, 1 H, H-1'), 4.62 (d, $J=11.2$ Hz, 2 H, Bn CH₂), 4.59 (d, $J=3.7$ Hz, 1 H, H-2), 4.18 (m, 1 H, H-4), 4.10 (d, $J=3.7$ Hz, 1 H, H-3), 3.90 (m, 1 H, H-5), 3.86 (m, 1 H, H-3'), 3.81 (br, 1 H, H-6b), 3.67 (m, 1 H, H-5'b), 3.62 (m, 1 H, H-6a), 3.58 (m, 1 H, H-4'), 3.54 (m, 1 H, H-5'a), 3.43 (dd, $J=9.6, 3.6$ Hz, 1 H, H-2'), 1.47 (s, 3 H, CH₃), 1.29 (s, 3 H, CH₃); ¹³C NMR (125 MHz (HSQC), CHLOROFORM-*d*, δ): 128.40 (Bn CH₂), 105.81 (C-1), 101.12 (C-1'), 86.05 (C-3), 83.37 (C-2), 81.18 (C-3'), 80.47 (C-2), 78.73 (C-2'), 78.23 (C-4'), 73.73, 73.79, 74.72, 74.65, 74.95, 75.71, 75.73 (Bn CH₂), 69.54 (C-5), 64.68 (C-6), 61.17 (C-5'), 26.53, 26.9 (CH₃); ESI-TOF MS (m/z) [M+Na]⁺ calcd for C₃₅H₄₂NaO₁₀, 645.2676; found, 645.26910

1,2,4-tri-O-acetyl-3-O-(2,3,4-tri-O-benzyl- α -D-xylopyranosyl)-D-xylopyranoside
(26)

To a solution of **25** (4.57 g; 7.34 mmol) in MeOH (80 ml) was added sodium periodate (3.14 g, 14.7 mmol) dissolved in H₂O at 0 °C, and then stirred for 1 h at ambient temperature. Ethylene glycol (613 μ l, 11.0 mmol) was added to reaction mixture to consume excess of sodium periodate. The reaction mixture was diluted by EtOAc, and washed with H₂O and Brine. The organic layer was dried over MgSO₄, and then solvent was evaporated. The dried residue was dissolved in MeOH (80 ml). NaBH₄ (1.11 g, 29.4 mmol) was added to the solution at 0 °C, and stirred for 1 h at ambient temperature. Acetone (8 ml) was added to the solution to consume excess of NaBH₄. The reaction mixture was diluted by EtOAc, and washed with H₂O and brine. The organic layer was dried over MgSO₄, and then solvent was evaporated. The residue was dissolved in 60%

TFA aqueous solution (100 ml), and stirred for 3 hs at ambient temperature. The solvent was evaporated with toluene, and dried residue was dissolved in pyridine (10 ml). To the solution was added Ac₂O (7ml, 73.4 mmol) and stirred for 2 hs. The reaction mixture was evaporated, diluted by EtOAc, and washed with H₂O, 1 N HCl solution, and sat NaHCO_{3aq} and brine. The separated organic layer was dried over MgSO₄, and then the solvent was evaporated. The residue was subjected to column chromatography on silica gel to give **26** (4 step) as anomeric mixture (4.98 g, 94%; α : β = 3:10); α -isomer; ¹H NMR (500 MHz, CHLOROFORM-*d* δ): 7.28–7.35 (m, 15H, Ar H), 6.21 (d, *J*=3.4 Hz, 1 H, H-1), 5.09 (m, 1 H, H-4), 5.07 (m, 1 H, H-2'), 4.99 (m, 1 H, H-1'), 4.84 (m, 2 H, Bn CH₂), 4.72 (m, 2 H, Bn CH₂), 4.62 (m, 2 H, Bn CH₂), 4.10 (m, 1 H, H-3'), 3.96 (m, 1 H, H-5b), 3.82 (t, *J*=6.2 Hz, 1 H, H-3), 3.62 (m, 2 H, H-5'), 3.56 (m, 3 H, H-5a, H-4'), 3.42 (1 H, H-2), 2.08 (s, 3 H, CH₃), 2.06 (s, 3 H, CH₃), 1.98 (s, 3 H, CH₃); ¹³C NMR (125 MHz (HSQC), CHLOROFORM-*d*, δ): 128.00 (Bn Ar), 97.93 (C-1'), 89.71 (C-1), 80.05 (C-3'), 79.90 (C-2'), 77.90 (C-4'), 75.70 (Bn CH₂), 74.19 (C-3), 73.60, 73.51 (Bn CH₂), 70.35 (C-4), 70.33 (C-2), 61.06 (C-5), 60.81 (C-5'), 20.92 (Ac CH₃); β -isomer; ¹H NMR (500 MHz, CHLOROFORM-*d* δ): 7.28–7.35 (m, 15H, Ar H), 5.80 (d, *J*=4.6 Hz, 1 H, H-1), 4.98 (m, 1 H, H-4), 4.94 (d, *J*=4.0 Hz, 1 H, H-2), 4.90 (m, 1 H, H-1'), 4.84 (m, 2 H, Bn CH₂), 4.67 (m, 4 H, Bn CH₂), 4.21 (dd, *J*=12.4, 3.6 Hz, 1 H, H-5b), 3.87 (t, *J*=6.0 Hz, 1 H, H-3), 3.81 (m, 1 H, H-3'), 3.63 (m, 2 H, H-5'), 3.56 (m, 3 H, H-5a, H-4'), 3.42 (dd, *J*=9.7, 3.4 Hz, 1 H, H-2'), 2.08 (s, 3 H, CH₃), 2.06 (s, 3 H, CH₃), 1.98 (s, 3 H, CH₃); ¹³C NMR (125 MHz (HSQC), CHLOROFORM-*d*, δ): 127.80-128.05 (Bn Ar), 98.24 (C-1'), 91.40 (C-1), 81.00 (C-3'), 79.73 (C-2'), 77.90 (C-4'), 75.77 (Bn CH₂), 74.36 (C-3), 73.59, 73.40 (Bn CH₂), 69.32 (C-4), 68.63 (C-2), 61.60 (C-5), 60.69 (C-5'), 20.92 (Ac CH₃); ESI-TOF MS (*m/z*)

$[M+Na]^+$ calcd for $C_{35}H_{42}NaO_{12}$, 701.2574; found, 701.2572

1-(2,2,2-trichloroacetimido)-2,4-di-O-acetyl-3-O-(2,3,4-tri-O-benzyl- α -D-xylopyranosyl)-D-xylopyranoside (27)

To a solution of **26** (4.69 g, 6.90 mmol) in DMF (70 ml) was added AcOH (513 μ l, 8.97 mmol) and $H_2NNH_2 \cdot H_2O$ (436 μ l, 8.97 mmol) at 0 °C, and stirred for 1 h. The reaction mixture was diluted by EtOAc, and washed with H_2O , sat $NaHCO_{3aq}$ and Brine. The separated organic layer was dried over $MgSO_4$, and then the solvent was evaporated. To the solution of dried material in DCM (70 ml) was added CCl_3CN (3.46 ml, 34.5 mmol) and DBU (103 μ l, 0.69 mmol) at ambient temperature, and stirred for 1 h. The reaction mixture was evaporated, and dried residue was subjected to column chromatography on silica gel to give **27** (2 step) as α -isomer (4.0 g, 74%); ^1H NMR (500 MHz, $CHLOROFORM-d$ δ): 8.63 (s, 1 H, $-O(C=NH)Cl_3$), 7.31 (m, 15 H, Ar H), 6.42 (d, $J=3.7$ Hz, 1 H, H-1), 5.18 (ddd, $J=10.6, 9.2, 5.8$ Hz, 1 H, H-4), 5.09 (dd, $J=9.3, 3.7$ Hz, 1 H, H-2), 4.95 (d, $J=3.3$ Hz, 1 H, H-1'), 4.84 (d, $J=2.0$ Hz, 2 H, Bn CH_2), 4.73 (m, 2 H, Bn CH_2), 4.60 (m, 2 H, Bn CH_2), 4.21 (t, $J=9.3$ Hz, 1 H, H-3), 3.98 (dd, $J=11.2, 5.8$ Hz, 1 H, H-5b), 3.82 (m, 1 H, H-3'), 3.73 (t, $J=11.0$ Hz, 1 H, H-5a), 3.67 (m, 1 H, H-5'b), 3.54 (m, 2 H, H-4', H-5'a), 3.40 (dd, $J=9.8, 3.5$ Hz, 1 H, H-2'), 2.00 (s, 3 H, CH_3), 1.89 (s, 3 H, CH_3); ^{13}C NMR (125 MHz (HSQC), $CHLOROFORM-d$ δ): 127.87 (Bn Ar), 98.08 (C-1'), 93.75 (C-1), 80.97 (C-3'), 80.13 (C-2'), 78.02 (C-4'), 75.87 (Bn CH_2), 74.64 (C-3), 73.69, 73.59 (Bn CH_2), 71.04 (C-2), 70.20 (C-4), 60.92 (C-5'), 60.82 (C-5), 20.20, 20.77 (Ac CH_3); ESI-TOF MS (m/z) $[M+Na]^+$ calcd for $C_{37}H_{40}Cl_3NaO_{11}$, 802.1565; found, 802.1584

Dodecyl-1-thio-2,4-di-O-acetyl-3-O-(2,3,4-tri-O-benzyl- α -D-xylopyranosyl)- β -D-xylopyranoside (28)

The glycosylimidate **27** (4.0 g, 5.12 mmol) was co-evaporated with toluene, and dried in vacuum with MS4A (3.0 g) for overnight. To the solution of dried material in dry DCM (60 ml) was added 1-dodecanethiol (3.68 ml, 15.36 mmol) under Ar gas atmosphere, and stirred for 15 min. TMSOTf (94 μ l, 0.51 mmol) was added to reaction mixture at -20 °C, and stirred for 1 h. TEA was added for neutralization of reaction mixture. The reaction mixture was washed with H₂O, sat NaHCO_{3aq} and brine. The separated organic layer was dried over MgSO₄, and then the solvent was evaporated. The residue was subjected to column chromatography on silica gel to give **28** as β -isomer (3.0 g, 64%).

¹H NMR (500 MHz, CHLOROFORM-*d* δ): 7.31 (m, 15 H, Ar H), 4.96 (m, 2 H, H-2, H-4), 4.88 (d, *J*=3.4 Hz, 1 H, H-1'), 4.85 (s, 2 H, Bn CH₂), 4.66 (m, 4 H, Bn CH₂), 4.56 (d, *J*=7.2 Hz, 1 H, H-1), 4.29 (dd, *J*=11.9, 4.4 Hz, 1 H, H-5b), 3.87 (t, *J*=9.3 Hz, 1 H, H-3'), 3.80 (t, *J*=7.2 Hz, 1 H, H-3), 3.60 (m, 2 H, H-5'), 3.51 (m, 1 H, H-4'), 3.37 (m, 3 H, H-2' and H-5), 2.64 (m, 2 H, -SCH₂-), 2.07 (s, 3 H, Ac CH₃), 1.92 (s, 3 H, Ac CH₃), 1.58 (m, 2 H, -SCH₂CH₂-), 1.36 (m, 3 H, CH₂CH₃), 1.28 (m, 16 H, CH₂), 0.88 (m, 3 H, CH₃) ; ¹³C NMR (125 MHz (HSQC), CHLOROFORM-*d*, δ): 127.88 (Bn Ar), 98.50 (C-1'), 83.59 (C-1), 80.92 (C-3'), 79.92 (C-2'), 77.97 (C-4'), 77.13 (C-3), 75.76, 73.47 (Bn CH₂), 70.95 (C-2), 69.50 (C-4), 64.21 (C-5), 61.17 (C-5'), 30.16 (-SCH₂-), 29.88 (-SCH₂CH₂-), 29.67 (CH₂), 20.88, 21.15 (Ac CH₃), 22.98 (CH₂CH₃), 14.23 (CH₃); ESI-TOF MS(*m/z*) [M+Na]⁺ calcd for C₄₇H₆₄NaO₁₀S, 843.4118 ; found, 843.4132

Dodecyl-1-thio-2,4-di-O-benzyl-3-O-(2,3,4-tri-O-benzyl- α -D-xylopyranosyl)- β -D-xylopyranoside (29)

To a solution of **28** (3.0 g, 3.66 mmol) in MeOH (40 ml) was added sodium methoxide (20 mg, 0.37 mmol), and stirred for 30 min. Dowex 50W was added to reaction mixture for neutralization of reaction mixture. After filtration of resin, the reaction mixture was evaporated. To a solution of dried material in DMF (40 ml) was added NaH (440 mg, 18.3 mmol) at 0 °C, and stirred for 10 min. BnBr (2.17 ml, 18.3 mmol) was added to reaction mixture at 0 °C, and stirred for overnight at ambient temperature. The reaction mixture was diluted by EtOAc and ice-water was added for quench. The organic layer was washed with H₂O and brine, and dried over MgSO₄. The solvent was evaporated, and the residue was subjected to column chromatography on silica gel to give **29** as β -isomer (3.24 g, 96%).

¹H NMR (500 MHz, CHLOROFORM-*d* δ): 7.28 (m, 25 H, Ar), 5.65 (d, *J*=3.5 Hz, 1 H, H'-1), 4.70 (m, 15 H, Bn CH₂), 4.43 (d, *J*=9.3 Hz, 1 H, H-1), 3.96 (m, 4 H, H-3, H-4, H-5b, H'-5b), 3.77 (td, *J*=9.3, 5.1 Hz, 1 H, H'-4), 3.50 (m, 4 H, H-2, H'-2, H'-3, H'-5a), 3.18 (t, *J*=10.7 Hz, 1 H, H-5a), 2.71 (m, 2 H, -SCH₂-), 1.64 (m, 2 H, -SCH₂CH₂-), 1.40 (m, 3 H, CH₂CH₃), 1.29 (m, 16 H, CH₂), 0.88 (t, *J*=6.9 Hz, 3 H, CH₃); ¹³C NMR (125 MHz (HSQC), CHLOROFORM-*d* δ): 127.83-129.04 (Bn Ar), 96.41 (C-1'), 86.19 (C-1), 81.14 (C-4), 79.55 (C-4'), 79.45 (C-2, C-2'), 78.60 (C-3'), 78.07 (C-3), 75.77, 75.74, 73.09, 73.04, 72.51 (Bn CH₂), 66.89 (C-5), 60.31 (C-5'), 30.89 (-SCH₂-), 29.96 (-SCH₂CH₂-), 29.11 (CH₂), 22.79 (CH₂CH₃), 14.23 (CH₃); ESI-TOF MS (*m/z*) [M+Na]⁺ calcd for C₅₇H₇₂NaO₈S, 939.4846; found, 939.4893

*1,2;5,6-O-diisopropylidene-3-O-[2,4-di-O-benzyl-3-O-(2,3,4-tri-O-benzyl- α -D-xyl
opryanosyl)-D-xylopyranosyl]- α -D-glucofranoside (30)*

The disaccharide donor **29** (3.24 g, 3.53 mmol) and di-acetone glucose **23** (1.1 g, 4.2 mmol) was co-evaporated with toluene. The residue was dried by vacuum with NIS (1.2 g, 5.3 mmol) and MS4A (2.0 g) for overnight. To the dried residue was added dried THF (30 ml) and dried DCM (15 ml) under Ar gas atmosphere, and stirred for 15 min. TfOH (31.2 μ l, 0.35 mmol) was added to reaction mixture at -20 °C, and mixture was stirred for 2 hs. After consuming the donor substrate in TLC, TEA was added for neutrization, and then filtered by Celite. Reaction mixture was washed with H₂O, NaS₂O_{3aq}, NaHCO_{3aq} and brine. The separated organic layer was dried over MgSO₄, and then the solvent was evaporated. The residue was subjected to column chromatography on silica gel to give **30** as α/β mixture of xyloside residue (2.8 g, 81%; $\alpha:\beta$ = 1:1); α -isomer, ¹H NMR (500 MHz, CHLOROFORM-*d* δ): 7.36 (m, 25 H, Ar H), 5.98 (d, *J*=3.8 Hz, 1 H, H-1), 5.63 (t, *J*=3.2 Hz, 1 H, H-1''), 5.20 (d, *J*=3.5 Hz, 1 H, H-1'), 4.96 (d, *J*=7.5 Hz, 2 H, Bn CH₂), 4.89 (d, *J*=11.0 Hz, 1 H, Bn CH₂), 4.66 (m, 1 H, H-2), 4.63 (m, 7 H, Bn CH₂), 4.38 (m, 1 H, H-4), 4.22 (m, 2 H, H-3, H-3'), 4.12 (dd, *J*=6.4, 2.0 Hz, 2 H, H-6), 4.08 (m, 1 H, H-5'b), 3.99 (m, 3 H, H-5, H-5''b, H-3''), 3.81 (m, 1 H, H-4'), 3.58 (m, 2 H, H-5', H-4''), 3.51 (m, 1 H, H-2'), 3.44 (s, 1 H, H-2''), 3.23 (dd, *J*=11.9, 9.6 Hz, 1 H, H-5''a), 1.55 (s, 3 H, CH₃), 1.49 (s, 3 H, CH₃), 1.42 (s, 3 H, CH₃), 1.25 (s, 3 H, CH₃); ¹³C NMR (125 MHz (HSQC), CHLOROFORM-*d*, δ): 127.78-128.41 (Bn Ar), 105.64 (C-1), 97.91 (C-1'), 96.81 (C-1''), 84.15 (C-2), 81.12 (C-3''), 80.75 (C-3), 80.34 (C-4), 79.87 (C-2''), 79.62 (C-4'), 79.34 (C-2'), 78.53 (C-4''), 76.83 (C-5), 74.10 (C-3'), 72.71-75.77 (Bn CH₂), 66.15 (C-6), 63.51 (C-5''), 60.60 (C-5'), 26.79, 26.50, 25.62, 25.31 (Ac CH₃); β -isomer, ¹H NMR (500 MHz, CHLOROFORM-*d* δ): 7.36 (m, 25 H,

Ar H), 5.71 (d, $J=3.8$ Hz, 1 H, H-1), 5.63 (t, $J=3.2$ Hz, 1 H, H-1''), 4.96 (d, $J=7.5$ Hz, 2 H, Bn CH₂), 4.89 (d, $J=11.0$ Hz, 1 H, Bn CH₂), , 4.52 (m, 1 H, H-1'), 4.49 (m, 1 H, H-2), 4.63 (m, 7 H, Bn CH₂), 4.44 (m, 1 H, H-3'), 4.34 (d, $J=3.2$ Hz, 1 H, H-4), 4.18 (m, 1 H, H-3), 4.08 (m, 2 H, H-6), 3.99 (m, 5 H, H-5, H-5'b, H-5''b, H-3''), 3.74 (m, 1 H-4'), 3.58 (m, 3 H, H-5', H-4''), 3.51 (dd, $J=9.9, 3.5$ Hz, 1 H, H-2'), 3.44 (s, 2 H, H-2''), 3.23 (dd, $J=11.9, 9.6$ Hz, 1 H, H-5''a), 1.51 (s, 3 H, CH₃), 1.46 (s, 3 H, CH₃), 1.36 (s, 3 H, CH₃), 1.22 (s, 3 H, CH₃); ¹³C NMR (125 MHz (HSQC), CHLOROFORM-*d*, δ): 127.78-128.41 (Bn Ar), 105.30 (C-1), 101.91 (C-1'), 96.81 (C-1''), 82.39 (C-2), 81.12 (C-3''), 80.75 (C-3), 80.34 (C-4), 79.87 (C-2''), 79.62 (C-4'), 79.34 (C-2'), 78.53 (C-4''), 76.83 (C-5), 73.30 (C-3'), 72.71-75.77 (Bn CH₂), 66.15 (C-6), 63.51 (C-5''), 60.60 (C-5'), 26.79, 26.50, 25.62, 25.31 (Ac CH₃); ESI-TOF MS (m/z) [M+Na]⁺ calcd for C₅₇H₆₆NaO₁₄, 997.4350; found, 997.4361

1,2-O-isopropylidene-3-O-[2,4-di-O-benzyl-3-O-(2,3,4-tri-O-benzyl- α -D-xylopyranosyl)- α -D-xylopyranosyl]- α -D-glucofranoside (31)

The compound **30** (2.8 g, 2.86 mmol) was dissolved in 80% AcOH aqueous solution (50 ml) and stirred for 1 h at 40 °C. After removal of 5,6-isopropylidene group, α/β mixture were separated by TLC. The reaction mixture was diluted by EtOAc and washed with H₂O, sat NaHCO_{3aq} and brine. The separated organic layer was dried over MgSO₄, and then the solvent was evaporated. The residue was subjected to column chromatography on silica gel to give **31** as α -isomer (1.1 g, 40%); ¹H NMR (500 MHz, CHLOROFORM-*d* δ): 7.29 (m, 25 H, Ar), 5.90 (d, $J=3.4$ Hz, 1 H, H-1), 5.61 (d, $J=3.4$ Hz, 1 H, H-1'), 4.92 (m, 4 H, Bn CH₂), 4.57 (m, 2 H, H-1', H-2), 4.60 (m, 6 H, Bn CH₂), 4.19 (m, 2 H, H-3', H-4), 4.07 (m, 1 H, H-3), 3.94 (m, 4 H, H-5'b, H-5, H-3''),

3.80 (d, $J=9.2$ Hz, 1 H, H-6b), 3.70 (m, 1 H, H-4'), 3.66 (m, 1 H, H-5'a), 3.60 (m, 4 H, H-5''b, H-6a, H-4''), 3.49 (m, 3 H, H-5''a, H-2', H-2''), 1.46 (m, 3 H, CH₃), 1.28 (m, 3 H, CH₃); ¹³C NMR (125 MHz (HSQC), CHLOROFORM-*d*, δ): 127.69-128.97 (Bn Ar), 105.76 (C-1), 100.70 (C-1'), 96.58 (C-1''), 85.61 (C-3), 82.93 (C-2), 80.98 (C-3''), 80.32 (C-4), 79.38 (C-4'), 79.10 (C-2''), 78.54 (C-2'), 78.40 (C-4''), 75.54, 75.46 (Bn CH₂), 73.69 (C-3'), 73.09 (Bn CH₂), 69.20 (C-5), 64.65 (C-6), 60.77 (C-5''), 60.23 (C-5'), 26.84, 26.49 (CH₃); ESI-TOF MS (m/z) [M+Na]⁺ calcd for C₅₄H₆₂NaO₁₄, 957.4037; found, 957.4017

1,2,4,6-tetra-O-acetyl-3-O-[2,4-di-O-benzyl-3-O-(2,3,4-tri-O-benzyl- α -D-xylopyranosyl)- α -D-xylopyranosyl]-D-glucopyranoside (32)

The trisaccharide **31** (2.89 g, 3.1 mmol) was dissolved in 60% TFA aqueous solution (30 ml) and stirred for 2 hs at ambient temperature. The solvent was evaporated with toluene. To a solution of dried residue in pyridine (30 ml) was added Ac₂O (30 ml), and stirred for overnight at ambient temperature. The solution was diluted by EtOAc, and washed with H₂O, 1 N HCl_{aq}, sat NaHCO_{3aq} and brine. The separated organic layer was dried over MgSO₄, and then the solvent was evaporated. The residue was subjected to column chromatography on silica gel to give **32** as anomeric mixture (2.1g, 63%; α : β = 3:1); α -isomer, ¹H NMR (500 MHz, CHLOROFORM-*d* δ): 7.29 (m, 25 H, Ar), 6.23 (d, $J=3.8$ Hz, 1 H, H-1), 5.58 (m, 1 H, H''-1), 5.25 (m, 1 H, H-4), 5.09 (m, 1 H, H-2), 4.92 (m, 2 H, Bn CH₂), 4.74 (m, 1 H, H'-1), 4.70 (m, 2 H, Bn CH₂), 4.67 (m, 2 H, Bn CH₂-), 4.56 (m, 4 H, Bn CH₂), 4.24 (m, 2 H, H-6b), 4.12 (m, 2 H, H-6a), 4.06 (m, 2 H, H-5, H'-3), 3.94 (m, 3 H, H'-5b, H-3, H''-3), 3.66 (m, 1 H, H'-4), 3.55 (m, 4 H, H'-5a, H''-5, H''-2), 3.41 (m, 1 H, H'-2), 2.19 (m, 3 H, Ac CH₃), 2.10 (s, 3 H, Ac CH₃), 1.97 (s,

3 H, Ac CH₃), 1.75 (s, 3 H, Ac CH₃); ¹³C NMR (125 MHz (HSQC), CHLOROFORM-*d*, δ): 127.53-128.53 (Bn Ar), 99.07 (C-1'), 96.50 (C-1''), 89.75 (C-1), 81.78 (C-4'), 81.14 (C-3''), 79.78 (C-4), 79.24 (C-2'), 78.72 (C-2''), 78.69 (C-3), 74.02, 74.13, 75.57 (Bn CH₂), 73.31 (C-3'), 72.49, 72.72 (Bn CH₂), 70.74 (C-2), 69.99 (C-5), 67.88 (C-4), 61.85 (C-6), 60.39 (C-5', C-5''), 21.00, 20.88, 20.76, 20.49 (Ac CH₃); β-isomer, ¹H NMR (500 MHz, CHLOROFORM-*d* δ): 7.29 (m, 25 H, Ar), 5.62 (*J*=8.4 Hz, 1 H, H-1), 5.58 (m, 1 H, H'-1), 5.25 (m, 1 H, H-4), 5.09 (m, 1 H, H-2), 4.92 (m, 2 H, Bn CH₂), 4.63 (m, 1 H, H'-1), 4.70 (m, 2 H, Bn CH₂), 4.67 (m, 2 H, Bn CH₂-), 4.56 (m, 4 H, Bn CH₂), 4.24 (m, 2 H, H-6b), 4.12 (m, 2 H, H-6a), 4.06 (m, 2 H, H-5, H'-3), 3.94 (m, 3 H, H'-5b, H-3, H'-3), 3.66 (m, 1 H, H'-4), 3.55 (m, 4 H, H'-5a, H'-5, H'-2), 3.41 (m, 1 H, H'-2), 2.19 (m, 3 H, Ac CH₃), 2.10 (s, 3 H, Ac CH₃), 1.97 (s, 3 H, Ac CH₃), 1.75 (s, 3 H, Ac CH₃); ¹³C NMR (125 MHz (HSQC), CHLOROFORM-*d*, δ): 127.53-128.53 (Bn Ar), 99.81 (C-1'), 96.50 (C-1''), 92.17 (C-1), 81.78 (C-4'), 81.14 (C-3''), 79.78 (C-4), 79.24 (C-2'), 78.72 (C-2''), 78.69 (C-3), 74.02, 74.13, 75.57 (Bn CH₂), 73.31 (C-3'), 72.49, 72.72 (Bn CH₂), 70.74 (C-2), 69.99 (C-5), 67.88 (C-4), 61.85 (C-6), 60.39 (C-5', C-5''), 21.00, 20.88, 20.76, 20.49 (Ac CH₃); ESI-TOF MS (*m/z*) [M+Na]⁺ calcd for C₅₉H₆₆NaO₁₈, 1085.4147; found, 1085.4191

1,2,4,6-tetra-O-acetyl-3-O-[2,4-di-O-acetyl-3-O-(2,3,4-tri-O-acetyl-α-D-xylopyranosyl)-α-D-xylopyranosyl]-D-glucopyranoside (33)

To a solution of **32** (270 mg, 0.25 mmol) in THF (1 ml) and MeOH (3 ml) mixture was added 20% Pd(OH)₂C (135 mg), and stirred under 0.1 MPa hydrogen gas for 2 hs at ambient temperature. The reaction mixture was filtered by Celite, and evaporated. To a solution of residue in pyridine (3 ml) was added Ac₂O (3 ml) and DMAP (3.6 mg, 30

μmol), and stirred for overnight at ambient temperature. The solution was diluted by EtOAc, and washed with H_2O , 1 N HCl_{aq} , sat $\text{NaHCO}_{3\text{aq}}$ and brine. The separated organic layer was dried over MgSO_4 , and then the solvent was evaporated. The residue was subjected to column chromatography on silica gel to give **33** as anomeric mixture (206 mg, 98% $\alpha:\beta = 3:1$); α -isomer, ^1H NMR (500 MHz, $\text{CHLOROFORM-}d$ δ): 6.29 (d, $J=3.7$ Hz, 1 H, H-1), 5.34 (m, 1 H, H''-3), 5.30 (d, $J=3.7$ Hz, 1 H, H''-1), 5.23 (m, 2 H, H-4, H'-1), 5.05 (dd, $J=10.1, 3.8$ Hz, 1 H, H-2), 4.94 (m, 2 H, H'-4, H''-4), 4.72 (dd, $J=10.3, 3.8$ Hz, 1 H, H'-2), 4.68 (dd, $J=10.1, 3.5$ Hz, 1 H, H''-2), 4.17 (m, 3 H, H-6b, H-3), 4.05 (m, 3 H, H-5, H-6a, H'-3), 3.98 (m, 1 H, H''-3), 3.73 (m, 3 H, H'-5, H''-5b), 3.61 (m, 1 H, H''-5a), 2.16 (s, 3 H, CH_3), 2.11 (s, 3 H, CH_3), 2.09 (m, 9 H, CH_3), 2.05 (s, 3 H, CH_3), 2.01 (m, 9 H, CH_3); ^{13}C NMR (125 MHz (HSQC), $\text{CHLOROFORM-}d$ δ): 95.73 (C-1'), 95.56 (C-1''), 89.16 (C-1), 73.47 (C-3), 72.46 (C-3'), 72.05 (C-2''), 71.47 (C-4''), 71.24 (C-2'), 70.42 (C-2), 70.13 (C-5), 69.38 (C-4), 69.28 (C-4'), 69.12 (C-3''), 61.49 (C-6), 58.61 (C-5, C-5''), 21.45, 20.88, 20.74, 20.70 (Ac CH_3); β -isomer, ^1H NMR (500 MHz, $\text{CHLOROFORM-}d$ δ): 5.59 (d, $J=8.2$ Hz, 1 H, H-1), 5.34 (m, 1 H, H''-3), 5.30 (d, $J=3.7$ Hz, 1 H, H''-1), 5.23 (m, 2 H, H-4, H'-1), 5.13 (m, 1 H, H-2), 4.94 (m, 2 H, H'-4, H''-4), 4.72 (dd, $J=10.3, 3.8$ Hz, 1 H, H'-2), 4.68 (dd, $J=10.1, 3.5$ Hz, 1 H, H''-2), 4.17 (m, 1 H, H-6b), 4.05 (m, 2 H, H-6a, H-3'), 3.98 (m, 1 H, H''-3), 3.90 (m, 1 H, H-3), 3.73 (m, 4 H, H-5, H'-5, H''-5b), 3.61 (m, 1 H, H''-5a), 2.16 (s, 3 H, CH_3), 2.11 (s, 3 H, CH_3), 2.09 (m, 9 H, CH_3), 2.05 (s, 3 H, CH_3), 2.01 (m, 9 H, CH_3); ^{13}C NMR (125 MHz (HSQC), $\text{CHLOROFORM-}d$ δ): 95.73 (C-1'), 95.56 (C-1''), 91.92 (C-1), 77.22 (C-3), 72.80 (C-5), 72.46 (C-3'), 72.05 (C-2''), 71.47 (C-4''), 71.24 (C-2'), 70.99 (C-2), 69.38 (C-4), 69.28 (C-4'), 69.12 (C-3''), 61.49 (C-6), 58.61 (C-5, C-5''), 21.45, 20.88, 20.74, 20.70 (Ac CH_3); ESI-TOF MS (m/z) $[\text{M}+\text{Na}]^+$ calcd for

C₃₄H₄₆NaO₂₃, 845.2328; found, 845.2368

1-(2,2,2-trichloroacetimido)-2,4,6-tetra-O-acetyl-3-O-[2,4-di-O-acetyl-3-O-(2,3,4-tri-O-acetyl- α -D-xylopyranosyl)- α -D-xylopyranosyl]- α -D-glucopyranoside (34)

34 was prepared with the same manner in **27. 33** (760 mg, 0.92 mmol) gave **34** as α -isomer (715 mg, 84%); ¹H NMR (500 MHz, CHLOROFORM-*d* δ): 8.67 (s, 1 H, -O(C=NH)Cl₃), 6.53 (d, *J*=3.7 Hz, 1 H, H-1), 5.34 (t, *J*=9.9 Hz, 1 H, H''-3), 5.31 (d, *J*=3.7 Hz, 1 H, H''-1), 5.26 (m, 1 H, H-4), 5.24 (s, 1 H, H'-1), 5.09 (dd, *J*=9.9, 3.5 Hz, 1 H, H-2), 4.95 (m, 2 H, H'-4, H''-4), 4.70 (ddd, *J*=13.3, 10.1, 3.7 Hz, 2 H, H'-2 and H''-2), 4.27 (t, *J*=9.7 Hz, 1 H, H-3), 4.18 (m, 1 H, H-6b), 4.08 (m, 4 H, H-5, H-6a, H'-3, H''-3), 3.71 (m, 4 H, H'-5, H''-5), 2.11 (s, 3 H, CH₃), 2.10 (s, 3 H, CH₃), 2.07 (s, 3 H, CH₃), 2.04 (s, 3 H, CH₃), 2.02 (m, 12 H, CH₃) ; ¹³C NMR (125 MHz (HSQC), CHLOROFORM-*d* δ): 95.56 (C-1'), 95.58 (C-1''), 93.14 (C-1), 73.42 (C-3), 72.41 (C-3'), 72.11 (C-2''), 71.57 (C-4), 71.09 (C-2'), 71.01 (C-2), 70.28 (C-5, C-3''), 69.27 (C-4'), 69.18 (C-3''), 69.17 (C-4), 61.50 (C-6), 58.59 (C-5', C-5''), 20.75 (AcCH₃); ESI-TOF MS (*m/z*) [M+Na]⁺ calcd for C₃₄H₄₄Cl₃NNaO₂₂, 946.1421; found, 946.1436

N-(9-Fluorenylmethoxycarbonyl)-O-{2,4,6-tetra-O-acetyl-3-O-[2,4-di-O-acetyl-3-O-(2,3,4-tri-O-acetyl- α -D-xylopyranosyl)- α -D-xylopyranosyl]- β -D-glucopyranosyl }-L-serine tert-butyl ester(35)

The trisaccharide donor **34** (565 mg, 0.61 mmol) and Fmoc-Ser-OtBu (280 mg, 0.73 mmol) was co-evaporated with toluene, and dried in vacuum for overnight. The mixture was dissolved in DCM (10 ml), and to this solution was added TMSOTf (11 μ l, 61

μmol) at $-20\text{ }^{\circ}\text{C}$ for 1 h. To the reaction mixture was added TEA, and reaction mixture was filtered by Celite. The reaction mixture was washed H_2O , sat $\text{NaHCO}_{3\text{aq}}$ and brine. The separated organic layer was dried over MgSO_4 , and then the solvent was evaporated. The residue was subjected to column chromatography on silica gel to give **35** as β -isomer (560 mg, 80%); ^1H NMR (500 MHz, $\text{CHLOROFORM-}d$ δ): 7.77 (m, 2 H, Fmoc Ar), 7.61 (m, 2 H, Fmoc Ar), 7.41 (m, 2 H, Fmoc Ar), 7.32 (m, 2 H, Fmoc Ar), 5.55 (d, $J=7.5$ Hz, 1 H, HN), 5.33 (m, 1 H, $\text{H}''\text{-3}$), 5.28 (d, $J=3.6$ Hz, 1 H, $\text{H}''\text{-1}$), 5.17 (d, $J=3.2$ Hz, 1 H, $\text{H}'\text{-1}$), 5.12 (t, $J=9.7$ Hz, 1 H, H-4), 4.94 (m, 3 H, H-2, $\text{H}'\text{-4}$, $\text{H}''\text{-4}$), 4.72 (m, 1 H, $\text{H}''\text{-2}$), 4.67 (dd, $J=10.1, 3.5$ Hz, 1 H, $\text{H}'\text{-2}$), 4.43 (m, 2 H, $\text{H}\beta$), 4.33 (d, $J=7.5$ Hz, 2 H, $\text{H}\alpha$, H-1), 4.22 (m, 2 H, Fmoc CH), 4.13 (m, 3 H, H-6, Fmoc CH_2), 4.05 (m, 1 H, $\text{H}'\text{-3}$), 3.84 (m, 1 H, H-3), 3.79 (m, 1 H, Fmoc CH_2), 3.75 (m, 2 H, $\text{H}''\text{-5}$), 3.67 (m, 1 H, $\text{H}'\text{-5a}$), 3.52 (m, 1 H, H-5), 3.47 (m, 1 H, $\text{H}'\text{-5b}$), 2.09 (s, 3 H, Ac CH_3), 2.08 (s, 3 H, Ac CH_3), 2.07 (s, 3 H, Ac CH_3), 2.05 (s, 3 H, Ac CH_3), 2.04 (s, 3 H, Ac CH_3), 2.02 (s, 6 H, Ac CH_3), 2.01 (s, 3 H, Ac CH_3), 1.46 (m, 9 H, *t*-Bu); ^{13}C NMR (125 MHz (HSQC), $\text{CHLOROFORM-}d$ δ): 120.13, 125.12, 127.31, 128.04 (Fmoc Ar), 101.26 (C-1), 96.07 (C-1''), 95.55 (C-1'), 76.87 (C-3), 72.62 (C-3'), 72.05 (C-2'), 71.94 (C-5), 71.67 (C-4''), 71.23 (C-2''), 70.07 (C-4), 69.71 (Fmoc CH_2), 69.46 (C-2, C-4'), 69.38 (C-3''), 67.06 (C β), 62.06 (C-6), 58.68 (C-5', C-5''), 54.81 (C α), 47.10 (Fmoc CH), 27.95 (*t*Bu), 20.92 (Ac CH_3); ESI-TOF MS (m/z) $[\text{M}+\text{Na}]^+$ calcd for $\text{C}_{54}\text{H}_{67}\text{NNaO}_{26}$, 1168.3849; found, 1168.3844

N-(9-Fluorenylmethoxycarbonyl)-*O*-{2,4,6-tetra-*O*-acetyl-3-*O*-[2,4-di-*O*-acetyl-3-*O*-(2,3,4-tri-*O*-acetyl- α -D-xylopyranosyl)- α -D-xylopyranosyl]- β -D-glucopyranosyl}-*L*-serine (**21**)

The glycoamino derivative **35** (560 mg, 0.49 mmol) was treated with 90% TFA aqueous solution (10 ml) at ambient temperature for 1 h. The solvent was evaporated, and the residue was subjected to column chromatography on silica gel to give **21** (520 mg, 97%).

^1H NMR (500 MHz, DMSO- d_6 , δ): 7.89 (d, $J=7.7$ Hz, 2 H, Ar Fmoc), 7.72 (dd, $J=6.9$, 5.1 Hz, 2 H, Fmoc Ar), 7.42 (m, 2 H, Fmoc Ar), 7.38 (br. s., 1 H, HN), 7.33 (m, 2 H, Fmoc Ar), 5.17 (d, $J=3.7$ Hz, 1 H, H''-1), 5.14 (m, 1 H, H''-3), 5.06 (d, $J=3.4$ Hz, 1 H, H'-1), 4.94 (m, 1 H, H-4), 4.89 (m, 3 H, H-2, H'-4, H''-4), 4.75 (m, 2 H, H'-2, H''-2), 4.71 (m, 1 H, H-1), 4.32 (m, 2 H, H β), 4.23 (m, 2 H, Fmoc CH), 4.14 (m, 3 H, H α , H-3, H-6b), 3.96 (m, 1 H, H-6a), 3.85 (m, 3 H, Fmoc CH₂, H-5, H'-3), 3.78 (m, 3 H, Fmoc CH₂, H'-5b), 3.60 (m, 2 H, H'-5a, H''-5b), 3.36 (d, $J=11.3$ Hz, 1 H, H''-5a), 2.06 (s, 3 H, Ac CH₃), 2.03 (s, 3 H, Ac CH₃), 2.00 (s, 3 H, Ac CH₃), 1.99 (s, 3 H, Ac CH₃), 1.98 (s, 3 H, Ac CH₃), 1.97 (s, 3 H, Ac CH₃), 1.95 (m, 3 H, Ac CH₃), 1.93 (s, 3 H, Ac CH₃); ^{13}C NMR (125 MHz (HSQC), DMSO- d_6 , δ): 120.70, 127.65, 125.89, 128.34 (Fmoc Ar), 99.75 (C-1), 95.98 (C-1'), 95.94 (C-1''), 75.80 (C-3), 72.64 (C'-3), 71.40 (C'-2), 71.15 (C''-4), 71.05 (C-5), 70.96 (C''-2), 70.54 (C-4), 69.03 (C''-3), 68.92 (C-2, C'-4), 68.69 (Fmoc CH₂), 66.30 (C β), 62.12 (C-6), 58.54 (C'-5, C''-5), 54.34 (C α), 47.24 (Fmoc CH), 20.88, 20.95, 20.99, 21.19, 21.30 (Ac CH₃); ESI-TOF MS (m/z) [$\text{M}+\text{Na}$] $^+$ calcd for C₅₀H₅₉NNaO₂₆, 1123.3223; found, 1123.3235

3-5. Reference

- (1) Artavanis-Tsakonas, S.; Rand, M. D.; Lake, R. J. Notch Signaling: Cell Fate Control and Signal Integration in Development. *Science*. **1999**, *284* (5415), 770–776.
- (2) Bray, S. J. Notch Signalling: A Simple Pathway Becomes Complex. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (9), 678–689.
- (3) Kopan, R.; Ilagan, M. X. G. The Canonical Notch Signaling Pathway: Unfolding the Activation Mechanism. *Cell* **2009**, *137* (2), 216–233.
- (4) Andersson, E. R.; Sandberg, R.; Lendahl, U. Notch Signaling: Simplicity in Design, Versatility in Function. *Development* **2011**, *138* (17), 3593–3612.
- (5) Haines, N.; Irvine, K. D. Glycosylation Regulates Notch Signalling. *Nat. Rev. Mol. Cell Biol.* **2003**, *4* (10), 786–797.
- (6) Moloney, D. J.; Shair, L. H.; Lu, F. M.; Xia, J.; Locke, R.; Matta, K. L.; Haltiwanger, R. S. Mammalian Notch1 Is Modified with Two Unusual Forms of O-Linked Glycosylation Found on Epidermal Growth Factor-like Modules. *J. Biol. Chem.* **2000**, *275* (13), 9604–9611.
- (7) Harris, R. J.; Spellman, M. W. O-Linked Fucose and Other Post-Translational Modifications Unique to EGF Modules. *Glycobiology* **1993**, *3* (3), 219–224.
- (8) Matsuura, A.; Ito, M.; Sakaidani, Y.; Kondo, T.; Murakami, K.; Furukawa, K.; Nadano, D.; Matsuda, T.; Okajima, T. O-Linked N-Acetylglucosamine Is Present on the Extracellular Domain of Notch Receptors. *J. Biol. Chem.* **2008**, *283* (51), 35486–35495.
- (9) Rana, N. a.; Nita-Lazar, A.; Takeuchi, H.; Kakuda, S.; Luther, K. B.; Haltiwanger, R. S. O-Glucose Trisaccharide Is Present at High but Variable Stoichiometry at Multiple Sites on Mouse Notch1. *J. Biol. Chem.* **2011**, *286* (36), 31623–31637.
- (10) Shao, L.; Moloney, D. J.; Haltiwanger, R. Fringe Modifies O-Fucose on Mouse Notch1 at Epidermal Growth Factor-like Repeats within the Ligand-Binding Site and the Abruption Region. *J. Biol. Chem.* **2003**, *278* (10), 7775–7782.
- (11) Alfaro, J. F.; Gong, C.; Monroe, M. E.; Aldrich, J. T.; Clauss, T. R. W.; Purvine, S. O.; Wang, Z.; Camp, D. G.; Shabanowitz, J.; Stanley, P.; Hart, G.

- W.; Hunt, D. F.; Yang, F.; Smith, R. D. Tandem Mass Spectrometry Identifies Many Mouse Brain O -GlcNAcylated Proteins Including EGF Domain-Specific O -GlcNAc Transferase Targets. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (19), 7280–7285.
- (12) Brückner, K.; Perez, L.; Clausen, H.; Cohen, S. Glycosyltransferase Activity of Fringe Modulates Notch-Delta Interactions. *Nature* **2000**, *406* (6794), 411–415.
- (13) Whitworth, G. E.; Zandberg, W. F.; Clark, T.; Vocadlo, D. J. Mammalian Notch Is Modified by D-Xyl-alpha1-3-D-Xyl-alpha1-3-D-Glc-beta1-O-Ser: Implementation of a Method to Study O-Glucosylation. *Glycobiology* **2010**, *20* (3), 287–299.
- (14) Rebay, I.; Fleming, R. J.; Fehon, R. G.; Cherbas, L.; Cherbas, P.; Artavanis-Tsakonas, S. Specific EGF Repeats of Notch Mediate Interactions with Delta and Serrate: Implications for Notch as a Multifunctional Receptor. *Cell* **1991**, *67* (4), 687–699.
- (15) Xu, A.; Lei, L.; Irvine, K. D. Regions of Drosophila Notch That Contribute to Ligand Binding and the Modulatory Influence of Fringe. *J. Biol. Chem.* **2005**, *280* (34), 30158–30165.
- (16) Rana, N. a; Haltiwanger, R. S. Fringe Benefits: Functional and Structural Impacts of O-Glycosylation on the Extracellular Domain of Notch Receptors. *Current Opinion in Structural Biology*, 2011, *21*, 583–589.
- (17) Moloney, D. J.; Panin, V. M.; Johnston, S. H.; Chen, J.; Shao, L.; Wilson, R.; Wang, Y.; Stanley, P.; Irvine, K. D.; Haltiwanger, R. S.; Vogt, T. F. Fringe Is a Glycosyltransferase That Modifies Notch. *Nature* **2000**, *406* (6794), 369–375.
- (18) Taylor, P.; Takeuchi, H.; Sheppard, D.; Chillakuri, C.; Lea, S. M.; Haltiwanger, R. S.; Handford, P. A. Fringe-Mediated Extension of O-Linked Fucose in the Ligand-Binding Region of Notch1 Increases Binding to Mammalian Notch Ligands. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (20), 7290–7295.
- (19) Hiruma-Shimizu, K.; Hosoguchi, K.; Liu, Y.; Fujitani, N.; Ohta, T.; Hinou, H.; Matsushita, T.; Shimizu, H.; Feizi, T.; Nishimura, S. I. Chemical Synthesis, Folding, and Structural Insights into O-Fucosylated Epidermal Growth Factor-like Repeat 12 of Mouse Notch-1 Receptor. *J. Am. Chem. Soc.* **2010**,

- 132 (42), 14857–14865.
- (20) Luca, V. C.; Jude, K. M.; Pierce, N. W.; Nachury, M. V.; Fischer, S.; Garcia, K. C. Structural Basis for Notch1 Engagement of Delta-like 4. *Science* (80-.). **2015**, 347 (6224), 847–853.
 - (21) Aoki, K.; Porterfield, M.; Lee, S. S.; Dong, B.; Nguyen, K.; McGlamry, K. H.; Tiemeyer, M. The Diversity of O-Linked Glycans Expressed during *Drosophila Melanogaster* Development Reflects Stage- and Tissue-Specific Requirements for Cell Signaling. *J. Biol. Chem.* **2008**, 283 (44), 30385–30400.
 - (22) Lee, T. V.; Takeuchi, H.; Jafar-Nejad, H. Regulation of Notch Signaling via O-Glcucosylation. Insights from *Drosophila* Studies. *Methods Enzymol.* **2010**, 480 (C), 375–398.
 - (23) Acar, M.; Jafar-Nejad, H.; Takeuchi, H.; Rajan, A.; Ibrani, D.; Rana, N. a; Pan, H.; Haltiwanger, R. S.; Bellen, H. J. Rumi Is a CAP10 Domain Glycosyltransferase That Modifies Notch and Is Required for Notch Signaling. *Cell* **2008**, 132 (2), 247–258.
 - (24) Fernandez-Valdivia, R.; Takeuchi, H.; Samarghandi, A.; Lopez, M.; Leonardi, J.; Haltiwanger, R. S.; Jafar-Nejad, H. Regulation of Mammalian Notch Signaling and Embryonic Development by the Protein O-Glcucosyltransferase Rumi. *Development* **2011**, 138 (10), 1925–1934.
 - (25) Lee, T. V.; Sethi, M. K.; Leonardi, J.; Rana, N. a.; Buettner, F. F. R.; Haltiwanger, R. S.; Bakker, H.; Jafar-Nejad, H. Negative Regulation of Notch Signaling by Xylose. *PLoS Genet.* **2013**, 9 (6).
 - (26) Lin, H.; Thayer, D. a; Wong, C.-H.; Walsh, C. T. Macrolactamization of Glycosylated Peptide Thioesters by the Thioesterase Domain of Tyrocidine Synthetase. *Chem. Biol.* **2004**, 11 (12), 1635–1642.
 - (27) Faure, R.; Saura-Valls, M.; Brumer 3rd, H.; Planas, a; Cottaz, S.; Driguez, H. Synthesis of a Library of Xylogluco-Oligosaccharides for Active-Site Mapping of Xyloglucan Endo-Transglycosylase. *J Org Chem* **2006**, 71 (14), 5151–5161.
 - (28) Matsui, H.; Furukawa, J.; Awano, T.; Nishi, N.; Sakairi, N. Lauryl and Stearyl Thioglycosides : Preparation and Reactivity of the Glycosyl Donor. *Chem. Lett.* **2000**, 12, 326–327.

- (29) Matsushita, T.; Hinou, H.; Kurogochi, M.; Shimizu, H.; Nishimura, S. I. Rapid Microwave-Assisted Solid-Phase Glycopeptide Synthesis. *Org. Lett.* **2005**, 7 (5), 877–880.
- (30) Garcia-Martin, F.; Hinou, H.; Matsushita, T.; Hayakawa, S.; Nishimura, S.-I. An Efficient Protocol for the Solid-Phase Synthesis of Glycopeptides under Microwave Irradiation. *Org. Biomol. Chem.* **2012**, 10 (8), 1612.
- (31) Huang, H.; Rabenstein, D. L. A Cleavage Cocktail for Methionine-Containing Peptides. *J. Pept. Res.* **1999**, 53 (5), 548–553.
- (32) Zamborelli, T. J.; Dodson, W. S.; Harding, B. J.; Zhang, J.; Bennett, B. D.; Lenz, D. M.; Young, Y.; Haniu, M.; Liu, C. F.; Jones, T.; Jarosinski, M. A. A Comparison of Folding Techniques in the Chemical Synthesis of the Epidermal Growth Factor-like Domain in Neu Differentiation Factor Alpha/beta. *J. Pept. Res.* **2000**, 55 (5), 359–371.
- (33) Hambleton, S.; Valeyev, N. V.; Muranyi, A.; Knott, V.; Werner, J. M.; McMichael, A. J.; Handford, P. a.; Downing, a. K. Structural and Functional Properties of the Human Notch-1 Ligand Binding Region. *Structure* **2004**, 12 (12), 2173–2183.
- (34) Selander-sunnerhagensq, M.; Ullners, M.; Perssonll, E.; Telemanii, O. How an Epidermal Growth Factor (EGF) -like Domain Binds Calcium. *Biochemistry* **1992**, 267 (27), 19642–19649.
- (35) Cordle, J.; Redfield, C.; Stacey, M.; Van Der Merwe, P. A.; Willis, A. C.; Champion, B. R.; Hambleton, S.; Handford, P. a. Localization of the Delta-like-1-Binding Site in Human Notch-1 and Its Modulation by Calcium Affinity. *J. Biol. Chem.* **2008**, 283 (17), 11785–11793.
- (36) Mine, T.; Miyazaki, T.; Kajiwar, H.; Tateda, N.; Ajisaka, K.; Yamamoto, T. A Recombinant α -(2→3)-Sialyltransferase with an Extremely Broad Acceptor Substrate Specificity from Photobacterium Sp. JT-ISH-224 Can Transfer N-Acetylneuraminic Acid to Inositols. *Carbohydr. Res.* **2010**, 345 (17), 2485–2490.
- (37) Piotta, M.; Saudek, V.; Sklenář, V. Gradient-Tailored Excitation for Single-Quantum NMR Spectroscopy of Aqueous Solutions. *J. Biomol. NMR*

- 1992**, 2 (6), 661–665.
- (38) Sklenar, V.; Piotto, M.; Leppik, R.; Saudek, V. Gradient-Tailored Water Suppression for ¹H-¹⁵N HSQC Experiments Optimized to Retain Full Sensitivity. *Journal of Magnetic Resonance, Series A*, 1993, 102, 241–245.
 - (39) Whiteman, P.; De Madrid, B. H.; Taylor, P.; Li, D.; Heslop, R.; Viticheep, N.; Tan, J. Z.; Shimizu, H.; Callaghan, J.; Masiero, M.; Li, J. L.; Banham, A. H.; Harris, A. L.; Lea, S. M.; Redfield, C.; Baron, M.; Handford, P. A. Molecular Basis for Jagged-1/serrate Ligand Recognition by the Notch Receptor. *J. Biol. Chem.* **2013**, 288 (10), 7305–7312.
 - (40) Cordle, J.; Johnson, S.; Zi Yan Tay, J.; Roversi, P.; Wilkin, M. B.; de Madrid, B. H.; Shimizu, H.; Jensen, S.; Whiteman, P.; Jin, B.; Redfield, C.; Baron, M.; Lea, S. M.; Handford, P. A. A Conserved Face of the Jagged/Serrate DSL Domain Is Involved in Notch Trans-Activation and Cis-Inhibition. *Nat. Struct. & Mol. Biol.* **2008**, 15 (8), 849–857.
 - (41) Sprinzak, D.; Lakhanpal, A.; Lebon, L.; Santat, L. a; Fontes, M. E.; Anderson, G. a; Garcia-Ojalvo, J.; Elowitz, M. B. Cis-Interactions between Notch and Delta Generate Mutually Exclusive Signalling States. *Nature* **2010**, 465 (7294), 86–90.
 - (42) Ranganathan, P.; Weaver, K. L.; Capobianco, A. J. Notch Signalling in Solid Tumours: A Little Bit of Everything but Not All the Time. *Nat. Rev. Cancer* **2011**, 11 (5), 338–351.
 - (43) Koch, U.; Radtke, F. Notch and Cancer: A Double-Edged Sword. *Cell. Mol. Life Sci.* **2007**, 64 (21), 2746–2762.
 - (44) Westhoff, B.; Colaluca, I. N.; D’Ario, G.; Donzelli, M.; Tosoni, D.; Volorio, S.; Pelosi, G.; Spaggiari, L.; Mazzarol, G.; Viale, G.; Pece, S.; Fiore, P. P. D. Alterations of the Notch Pathway in Lung Cancer. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 106 (52), 22293–22298.
 - (45) Zhang, X.; Zhao, X.; Shao, S.; Zuo, X.; Ning, Q.; Luo, M.; Gu, S.; Zhao, X. Notch1 Induces Epithelial-Mesenchymal Transition and the Cancer Stem Cell Phenotype in Breast Cancer Cells and STAT3 Plays a Key Role. *Int. J. Oncol.* **2014**.

- (46) States, D. .; Haberkorn, R. .; Ruben, D. . A Two-Dimensional Nuclear Overhauser Experiment with Pure Absorption Phase in Four Quadrants. *J. Magn. Reson.* **1982**, 48 (2), 286–292.
- (47) Piantini, U.; Sorensen, O. W.; Ernst, R. R. Multiple Quantum Filters for Elucidating NMR Coupling Networks. *J. Am. Chem. Soc.* **1982**, 104 (24), 6800–6801.
- (48) Rance, M.; Sørensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wüthrich, K. Improved Spectral Resolution in Cosy ¹H NMR Spectra of Proteins via Double Quantum Filtering. *Biochem. Biophys. Res. Commun.* **1983**, 117 (2), 479–485.
- (49) Braunschweiler, L.; Ernst, R. . Coherence Transfer by Isotropic Mixing: Application to Proton Correlation Spectroscopy. *Journal of Magnetic Resonance*, 1983, 53, 521–528.
- (50) Bax, A.; Davis, D. G. MLEV-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy. *Journal of Magnetic Resonance* (1969), 1985, 65, 355–360.
- (51) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. Investigation of Exchange Processes by Two-Dimensional NMR Spectroscopy. *J. Chem. Phys.* **1979**, 71
- (52) Lee, W.; Westler, W. M.; Bahrami, A.; Eghbalnia, H. R.; Markley, J. I. PINE-SPARKY: Graphical Interface for Evaluating Automated Probabilistic Peak Assignments in Protein NMR Spectroscopy. *Bioinformatics* **2009**, 25 (16), 2085–2087.

Chapter 4
Concluding Remarks

Glycosylation is one of the most remarkable posttranslational modifications because of their functional and structural diversity and close relationship with several diseases. Protein glycosylation, which is given in template-independent manner, can regulate several protein functions. Aberrant glycosylation as well as overexpression of some glycoproteins have been found in tumor cells. Therefore, glycoproteins as well as glycopeptides have been applied as cancer therapeutic and diagnostic targets. Particularly, glycopeptide carrying *O*-glycans have been developed as antigen determinant region, called as epitope.

Understanding of glycosylation impacts on protein function and application of glycopeptide to therapeutic and diagnostic targets, organic chemical approach is necessary for preparation of structure-defined glycopeptide as well as chemical modification to functional study. In this thesis, chemical synthesis, NMR conformational analysis, and functional analysis with microarray or cellular assay were applied for general *O*-GalNAc-modified and unusual *O*-Fuc- or *O*-Glc-modified glycopeptides to elucidate *O*-glycosylation impacts on protein structures as well as functions.

In chapter 2, mucin-like model glycopeptides and MUC2 glycopeptides containing consecutive *O*-GalNAcylated threonine residues were efficiently synthesized with double-activation method. NMR analysis demonstrated that *O*-glycosylated TTX sequence can be key conformational motif which adopts extend and rigid structure. Auto-antibody analyses against synthetic MUC2 glycopeptides with microarray represented that stage-specific antibody recognition. These NMR studies and microarray analyses suggested that these conformational motifs can be applied to cancer vaccine design.

In chapter 3, conformational and functional study of unusual *O*-glycans was demonstrated. 38-mer human Notch1 EGF12 domain carrying *O*-fucose- and *O*-glucose-type glycans were synthesized and successfully folded into native structure having three disulfide bonds in the presence of Ca^{2+} . NMR analyses indicated that *O*-glycosylations as well as Ca^{2+} coordination have crucial roles to stabilize backbone conformation including β -hairpin and anti-parallel β -sheet structure. These synthetic EGF12 modules were treated to cancer cells, which suggested that these conformational modules can inhibit cancer cell proliferation in dose-dependent manner and cell aggregation which is seen in control cancer cells.

These chemical approaches of *O*-glycosylated peptides will facilitate further investigations of molecular mechanism of glycobiology and cancer biology. Furthermore, these approaches will be applied for cancer therapy and diagnosis.

Acknowledgement

I would like to express my sincere gratitude to Professor Shin-Ichiro Nishimura for his helpful suggestions, discussions, and encouragements. This work described herein would not have been successful without him.

I would like to express my special thanks to Professor Kenji Monde, Professor Makoto Demura and associate professor Hiroshi Hinou in Hokkaido University for reading this thesis, their critical suggestions, and tenderhearted encouragements.

I am deeply indebted to Assistant Professor Takahiko Matsushita for his helpful support, kind assistance and affectionate encouragements.

I would like to express grateful thanks to Assistant Professor Fayna Garcia-Martin and Dr. Kentaro Naruchi for their kind assistance and supports. And I want to express grateful thanks to Mr. Risho Miyoshi for microarray experiments in chapter 2, Mr. Ryosuke Koide for cancer cell assay in chapter 3, and my colleagues in Nishimura lab for daily supports, discussions, and encouragement.

Finally, I wish to express my sincere gratitude to my parents for their invaluable mental supports and affectionate encouragement.

2016 Sapporo

Shun HAYAKAWA