Glycoside cluster effects on antibody recognition of MUC1 glycopeptides

(抗体によるMUC1糖ペプチド認識機構における糖鎖クラスター効果)

Doctoral Thesis

2016

Shobith Rangappa

Graduate School of Life Science, Hokkaido University

CONTENTS

Abbreviation

Ac	Acetyl
Ala/A	Alanine
Arg/R	Arginine
Asp/D	Aspartic acid
C terminal	Carbonyl terminus
DCM	Dichloromethane
DHB	Dihydroxybenzoic acid
DIEA	N,N-diisopropylethylamine
DMF	N,N-dimethyformamide
DMSO	Dimethysulfoxide
DNA	Deoxyribonucleic acid
ESI	Electro spray ionization
Fmoc	9-flourenylmethylcarbonyl
Gal	Galactose
GalNAc	N-acetyl-D-galactosamine
Gly	Glycine (G)
HBTu	2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HOBt	1-Hydroxybenzotriazole
His	Histidine
Hrs	Hours
HPLC	High Performance Liquid Chromatography
HSQC	^{1}H-detected Single Quantum Coherence Spectrum
Hz	Hertz

mAb	Monoclonal Antibody
MALDI	Matrix Assisted Laser Desorption/Ionization
mM	Mili Molar
ml	MIli Litre
MUC1	Mucin 1
N terminus	Amino terminus
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
Pro/P	Proline
rpm	Revolution per minute
RT	Room Temperature
Ser/S	Serine
SPPS	Solid Phase Peptide Synthesis
T antigen	Thomsen-Friedenreich Antigen
TFA	Triflouroacetic acid
Thr	Threonine
TOF	Time Of Flight
Val	Valine (V)
α	Alpha
β	Beta
°C	Degree Centigrade
%	Percentage
µM	Micro Molar

CHAPTER 1

General Introduction

1-1Mucins

Mucins are the series of glycopeptides of higher molecular weight and are widely distributed on the epithelial layer[1] . Mucins are commonly glycosylated and known to form gel like structures which are seen as a major component of the mucosa. It plays a major role in lubrication and are also involved in cancer metastasis, aging, carcinogenesis, cell differentiation [2–5]. Various types of mucins are known to be present in salivary glands, oesophagus, stomach, duodenum, small intestine, colorectum, pancreas, male and female reproductive tract[6]. Whereas, only few types of mucins are known to be present in many parts such as mucin1 (MUC1) (Fig 1-1)[6]. Hence, MUC1 is a significant type of mucin which plays a vital role in many types of cancers.

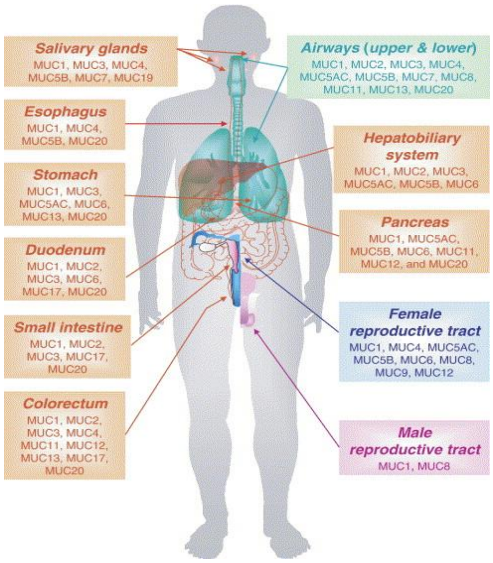


Fig 1-1. Distribution of Types of Mucins in the Human Body.

MUC1 is a highly *O*-glycosylated glycoprotein comprising of three major structural domains. Membrane spanning region of 31 amino acids residues as C terminal and 60 amino acids residues as cytoplasmic tail, the N terminus known as the signal sequence comprises of a 20 amino acid sequence which is hydrophobic in nature[7–10]. During carcinogenesis remarkable

changes in the glycans associated with the MUC1 peptide are altered (Fig1-2), hence shedding of cell surface into the blood stream are observed [11].

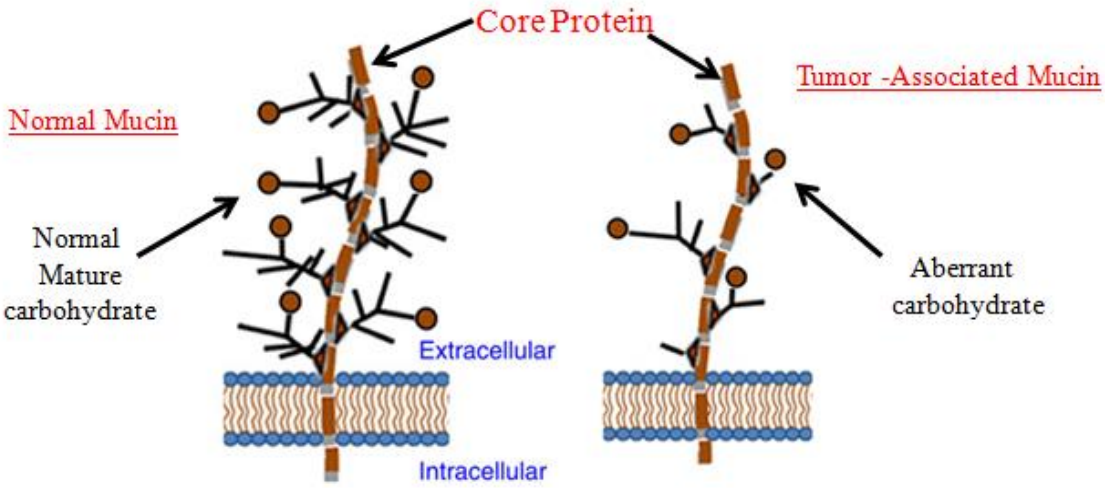


Fig 1-2. Normal and tumor associated Mucin structures.

Various studies are seen relating to the effect of *O*-glycosylation changes. It is shown that that impact of *O*-glycosylation with GalNAc residue at the Thr/Ser amino acid is different when compared with *O*-glycosylation with a disaccharide at the Thr/Ser residue. Such changes relate to various cancers and immune disorders. Certain alterations in glycosylation bring about conformational changes during antigen binding, which are considered to be significant biomarkers and are also treated to be a target to attain specific and effective mAb's.

Comparing other mucins, MUC1 is targeted due to its significant alterations in tumor cells. The sequence studied comprises of the repeated 20 amino acid sequence, i.e. His-Gly-Val-Thr-Ser-Ala-Pro-Asp-Thr-Arg-Pro-Ala-Pro-Gly-Ser-Thr-Ala-Pro-Pro-Ala[12,13]. The above mentioned amino acid sequence comprises of five *O*-glycosylation sites which are altered during various progression, differentiation and metastasis.

MUC1 could be identified by mAb derived from various cancers such as prostate cancer, hepatocellular carcinoma, pancreatic cancer, colon cancer and ovarian cancer comprising of the same tandem repeating sequence. Insight into the tandem repeating sequence revealed that Pro-Asp-Thr-Arg is the most dominant epitope characterised[14,15]. This epitope has more affinity after GalNAc residue being introduced at the Thr residue which can be observed in various cancers. *O*-glycosylation at the Thr amino acid by minimum GalNAc residue at the PDTR improves binding with various mAbs[16–21]. In certain cancer cells aberrant *O*-glycosylation is observed. Abnormal human mucin configuration was determined to be highly sialylated *O*-glycans like core 1 Sialyl T, or core 2 disialyl T based structures[22–24].

Commonly known antibodies is the DF3 mAb which was raised against metastatic breast carcinoma, SM3 mAb raised against deglycosylated purified milk mucin and KL6 mAb was derived from the human lung adenocarcinoma derived cell line, showed various interactions with different MUC1 associated glycans. Recently researchers have shown keen interest about KL6 antibodies due to its important role as a diagnostic marker for interstitial pneumonia, lung adenocarcinoma, breast carcinoma, colorectal adenocarcinoma and hepatocellular carcinoma[25–30].It has been revealed in previous reports that the heptapeptide Pro-Asp-Thr-Arg-Pro-Ala-Pro bearing a Sialyl T antigenic trisaccharide (Neu5Acα2, 3Galβ1, 3GalNAcα1→) is the minimum epitope detected by KL6 antibody. Whereas, the same KL6 antibody recognises the core1 disialylated hepta peptide similar as the core 1 sialylated heptapeptide with no discrimination[7] . Many researchers have shown that anti MUC1 antibodies requires a minimal GalNAc residue at the Thr residue at the heptapeptide where as certain other antibodies require a minimum of core 1 Sialyl T residue at the Thr residue. This conceives an important target to be observed and to understand the conformational changes that exist in the glycans present of MUC1 peptide.

Recent structural studies regarding the core 2 MUC1 glycopeptide showed remarkable changes in structural conformations between naked, asialylated and completely sialylated core 2 glycopeptide. Core 2 sialylated MUC1 glycopeptide have shown conserved structure at the PDTR sequence due to which KL6 interaction was possible, whereas the naked and asialylated core 2 glycopeptide have shown extended or complete changes at the PDTR sequence and also proved that the attachment of Sialyl group played a very important role for the KL6 mAb interaction. Taking this into consideration, the structural conformation during MUC1 peptide with completely glycosylated by core 1 T antigen, and MUC1 peptide completely glycosylated by core 1 sialylated and a slight change in MUC1 peptide comprising of four *O*-glycosylation sites with core 1 sialylated T glycans except the Thr at the PDTR motif comprising of Tn antigen, MUC1 comprising of T antigen at the PDTR Thr amino acid surrounded by Tn antigen at neighbouring 4 glycosylation sites and MUC1 completely glycosylated by Tn antigen would be an required assessment to be a good target for future MUC1 studies and as an immunotherapeutic agent.

1-2. Objective of Thesis

To elucidate the structural correction and alteration between binding of MUC1 glycopeptide and certain known antibody this demanded for a long time. However the understanding of in-depth details about the changes seen in the structure during certain *O* glycosylation has not been studied.

In this thesis, the author has challenged to elucidate the importance of certain glycosylation patterns which involves changing the structural orientation due to which certain compounds bind with antibodies and why certain glycopeptides do not bind with same antibodies. In chapter 2, a library of MUC1 glycopeptides bearing different glycan moieties at *O* glycosylation sites were synthesised. In chapter 3, these synthesised compounds were used

for micro array studies against 3 major mAb's. In chapter 4, elucidation of the 3D structure of 5 crucial MUC1 glycopeptides by NMR was done. Chapter 5, the structural data and performed docking studies against known crystallographic structure of SM3 antibody were performed.

1-3. References

(1) Gargano, A. F. G., Lämmerhofer, M., Lönn, H., Schoenmakers, P. J., and Leek, T. (2014) Mucin-based stationary phases as tool for the characterization of drug-mucus interaction. *J. Chromatogr. A 1351*, 70–81.

(2) Matsushita, T., Ohyabu, N., Fujitani, N., Naruchi, K., Shimizu, H., Hinou, H., and Nishimura, S. I. (2013) Site-specific conformational alteration induced by sialylation of MUC1 tandem repeating glycopeptides at an epitope region for the anti-KL-6 monoclonal antibody. *Biochemistry 52*, 402–414.

(3) Brockhausen, I. (1999) Pathways of O-glycan biosynthesis in cancer cells. *Biochim. Biophys. Acta - Gen. Subj. 1473*, 67–95.

(4) Taylor-Papadimitriou, J., Burchell, J., Miles, D. W., and Dalziel, M. (1999) MUC1 and cancer. *Biochim. Biophys. Acta - Mol. Basis Dis. 1455*, 301–313.

(5) Rosen, S. D. (2004) Ligands for L-selectin: homing, inflammation, and beyond. *Annu. Rev. Immunol. 22*, 129–156.

(6) Andrianifahanana, M., Moniaux, N., and Batra, S. K. (2006) Regulation of mucin expression: Mechanistic aspects and implications for cancer and inflammatory diseases. *Biochim. Biophys. Acta - Rev. Cancer 1765*, 189–222.

(7) Ohyabu, N., Hinou, H., Matsushita, T., Izumi, R., Shimizu, H., Kawamoto, K., Numata, Y., Togame, H., Takemoto, H., Kondo, H., and Nishimura, S. I. (2009) An essential epitope of anti-MUC1 monoclonal antibody KL-6 revealed by focused glycopeptide library. *J. Am. Chem. Soc. 131*, 17102–17109.

(8) Hang, H. C., and Bertozzi, C. R. (2005) The chemistry and biology of mucin-type O-linked glycosylation. *Bioorganic Med. Chem. 13*, 5021–5034.

(9) Tarp, M. A., and Clausen, H. (2008) Mucin-type O-glycosylation and its potential use in drug and vaccine development. *Biochim. Biophys. Acta - Gen. Subj. 1780*, 546–563.

(10) Hattrup, C. L., and Gendler, S. J. (2008) Structure and function of the cell surface (tethered) mucins. *Annu. Rev. Physiol. 70*, 431–457.

(11) Song, X., Airan, R. D., Arifin, D. R., Bar-Shir, A., Kadayakkara, D. K., Liu, G., Gilad, A. a., van Zijl, P. C. M., McMahon, M. T., and Bulte, J. W. M. (2015) Label-free in vivo molecular imaging of underglycosylated mucin-1 expression in tumour cells. *Nat. Commun. 6*, 6719.

(12) Matsushita, T., Takada, W., Igarashi, K., Naruchi, K., Miyoshi, R., Garcia-Martin, F., Amano, M., Hinou, H., and Nishimura, S. I. (2014) A straightforward protocol for the preparation of high performance microarray displaying synthetic MUC1 glycopeptides. *Biochim. Biophys. Acta - Gen. Subj. 1840*, 1105–1116.

(13) Garcia-Martin, F., Matsushita, T., Hinou, H., and Nishimura, S.-I. (2014) Fast Epitope Mapping for the Anti-MUC1 Monoclonal Antibody by Combining a One-Bead-One-Glycopeptide Library and a Microarray Platform. *Chem. - A Eur. J. 20*, 15891–15902.

(14) Price, M. R., Rye, P. D., Petrakou, E., Murray, a, Brady, K., Imai, S., Haga, S., Kiyozuka, Y., Schol, D., Meulenbroek, M. F., Snijdewint, F. G., von Mensdorff-Pouilly, S., Verstraeten, R. a, Kenemans, P., Blockzjil, a, Nilsson, K., Nilsson, O., Reddish, M., Suresh, M. R., Koganty, R. R., Fortier, S., Baronic, L., Berg, a, Longenecker, M. B., and Hilgers, J. (1998) Summary report on the ISOBM TD-4 Workshop: analysis of 56 monoclonal antibodies against the MUC1 mucin. San Diego, Calif., November 17-23, 1996. *Tumour Biol. 19 Suppl 1*, 1–20.

(15) Schol, D. J., Meulenbroek, M. F. a, Snijdewint, F. G. M., Von Mensdorff-Pouilly, S., Verstraeten, R. a., Murakami, F., Kenemans, P., and Hilgers, J. (1997) "Epitope fingerprinting" using overlapping 20-mer peptides of the MUC1 tandem repeat sequence. *Tumor Biol. 19*, 35–45.

(16) Karsten, U., Diotel, C., Klich, G., Paulsen, H., Goletz, S., Müller, S., and Hanisch, F. G. (1998) Enhanced binding of antibodies to the DTR motif of MUC1 tandem repeat peptide is mediated by site-specific glycosylation. *Cancer Res. 58*, 2541–2549.

(17) Von Mensdorff-Pouilly, S., Petrakou, E., Kenemans, P., van Uffelen, K., Verstraeten, a a, Snijdewint, F. G., van Kamp, G. J., Schol, D. J., Reis, C. a, Price, M. R., Livingston, P. O., and Hilgers, J. (2000) Reactivity of natural and induced human antibodies to MUC1 mucin with MUC1 peptides and n-acetylgalactosamine (GalNAc) peptides. *Int. J. Cancer 86*, 702–712.

(18) Hanisch, F. G., Stadie, T., and Bosslet, K. (1995) Monoclonal antibody BW835 defines a site-specific Thomsen-Friedenreich disaccharide linked to threonine within the VTSA motif of MUC1 tandem repeats. *Cancer Res. 55*, 4036–4040.

(19) Lloyd, K. O., Burchell, J., Kudryashov, V., Yin, B. W. T., and Taylor-Papadimitriou, J. (1996) Comparison of O-linked carbohydrate chains in MUC-1 mucin from normal breast epithelial cell lines and breast carcinoma cell lines: Demonstration of simpler and fewer glycan chains in tumor cells. *J. Biol. Chem. 271*, 33325–33334.

(20) Grinstead, J. S., Koganty, R. R., Krantz, M. J., Longenecker, B. M., and Campbell, a. P. (2002) Effect of glycosylation on MUC1 humoral immune recognition: NMR studies of MUC1 glycopeptide-antibody interactions. *Biochemistry 41*, 9946–9961.

(21) Takeuchi, H., Kato, K., Denda-Nagai, K., Hanisch, F. G., Clausen, H., and Irimura, T. (2002) The epitope recognized by the unique anti-MUC1 monoclonal antibody MY.1E12 involves sialyl??2-3galactosyl??1-3N-acetylgalactosaminide linked to a distinct threonine residue in the MUC1 tandem repeat. *J. Immunol. Methods 270*, 199–209.

(22) Grinstead, J. S., Schuman, J. T., and Campbell, a. P. (2003) Epitope Mapping of Antigenic MUC1 Peptides to Breast Cancer Antibody Fragment B27.29: A Heteronuclear NMR Study. *Biochemistry 42*, 14293–14305.

(23) Danielczyk, A., Stahn, R., Faulstich, D., Löffler, A., Märten, A., Karsten, U., and Goletz, S. (2006) PankoMab: A potent new generation anti-tumour MUC1 antibody. *Cancer Immunol. Immunother. 55*, 1337–1347.

(24) Tarp, M. a., Sørensen, A. L., Mandel, U., Paulsen, H., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (2007) Identification of a novel cancer-specific immunodominant glycopeptide epitope in the MUC1 tandem repeat. *Glycobiology 17*, 197–209.

(25) Ogawa, Y., Ishikawa, T., and Ikeda, K. (2000) Evaluation of Serum KL-6 , a Mucin-like Glycoprotein , as a Tumor Marker for Breast Cancer Evaluation of Serum KL-6 , a Mucin-like Glycoprotein , as a Tumor Marker for Breast Cancer. *Clin. Cancer Res. 6*, 4069–4072.

(26) Hirasawa, Y., Kohno, N., Yokoyama, A., Kondo, K., Hiwada, K., and Miyake, M. (2000) Natural autoantibody to MUC1 is a prognostic indicator for non-small cell lung cancer. *Am. J. Respir. Crit. Care Med. 161*, 589–594.

(27) Gad, A., Tanaka, E., Matsumoto, A., Wahab, M. A., Serwah, A. el-H., Attia, F., Ali, K., Hassouba, H., el-Deeb, A. el-R., Ichijyo, T., Umemura, T., Muto, H., Yoshizawa, K., and Kiyosawa, K. (2005) Assessment of KL-6 as a tumor marker in patients with hepatocellular carcinoma. *World J. Gastroenterol. 11*, 6607–6612.

(28) Guo, Q., Tang, W., Inagaki, Y., Midorikawa, Y., Kokudo, N., Sugawara, Y., Nakata, M., Konishi, T., Nagawa, H., and Makuuchi, M. (2006) Clinical significance of subcellular localization of KL-6 mucin in primary colorectal adenocarcinoma and metastatic tissues. *World J. Gastroenterol. 12*, 54–59.

(29) Inata, J., Hattori, N., Yokoyama, A., Ohshimo, S., Doi, M., Ishikawa, N., Hamada, H., and Kohno, N. (2007) Circulating KL-6/MUC1 mucin carrying sialyl Lewisa oligosaccharide is an independent prognostic factor in patients with lung adenocarcinoma. *Int. J. Cancer 120*, 2643–2649.

(30) Ishikawa, N., Hattori, N., Yokoyama, A., Tanaka, S., Nishino, R., Yoshioka, K., Ohshimo, S., Fujitaka, K., Ohnishi, H., Hamada, H., Arihiro, K., and Kohno, N. (2008) Usefulness of monitoring the circulating Krebs von den Lungen-6 levels to predict the clinical outcome of patients with advanced nonsmall cell lung cancer treated with epidermal growth factor receptor tyrosine kinase inhibitors. *Int. J. Cancer 122*, 2612–2620.

CHAPTER 2

Solid Phase Peptide Synthesis of MUC1 Glycopeptides

2.1. Introduction

Solid phase peptide synthesis or SPPS[1-3] is a method designed to synthesise peptides in which multiple amino acids are linked by amide bonds or also known as peptide bonds. SPPS allows for synthesis of natural peptides which can be either be synthesised at room temperature or by microwave irradiation for faster synthesis.

In nature, peptide synthesis commences from N terminal and terminates at the C terminal which primarily takes part in ribosome with the help of other factors, whereas in case of chemical synthesis the process is mostly initiated from the C terminal and terminates in N terminal. In case of SPPS the functionality of the resin is responsible for in the initial formation of peptide bond following with the elongation of the respective amino acid sequence formation as shown in figure below. Commercially different types of resin are available according to the functionality required. As an energy source Microwave irradiation is used to fasten the peptide bond formation along with the coupling reagents at with constant stirring.

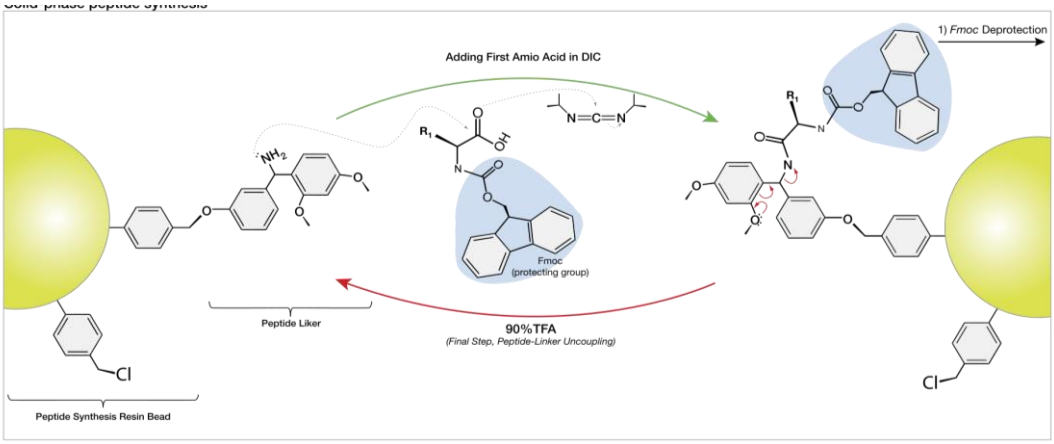


Figure 1: Solid Phase Peptide Synthesis

2-2. Results and Discussion

2-2.1. Synthesis of MUC1 Glycopeptides for Micro Array Studies:

To synthesise a library of MUC1 glycopeptides for micro array studies Microwave assisted solid phase synthesis was performed. Totally 23 MUC1 glycopeptides were synthesised from Carboxyl terminus to amine terminus of the peptide sequence. The C terminus comprises the $CONH_2$ derived from the rink amide resin. The library of glycopeptide comprises of glycosylated peptides at different *O* glycosylation sites serine/ threonine by Tn antigen (GalNAcα), core 1 (Gaβ1-3GalNAcα), sialylated core 1 (Neu5Acα2, 3 Gaβ1-3GalNAcα) and sialylated core 2 (Neu5Ac⟨2-3Galβ1−3[Galβ1-4GlcAcβ1-6]GalNAcα). Synthesis was completed by coupling 5 oxo hexanoic acid (ketone capping) as capping at the N terminal which was helpful for further micro array studies.

Glycopeptide Library Constructed on Resin: For Micro Array						
Seq: ***Linker*- H-G-V-T(R1)-S(R2)-A-P-D-T(R3)-R-P-A-P-G-S(R4)-T(R5)-A-P-P-A-*NH_2***						
No.	**Names**	**R1**	**R2**	**R3**	**R4**	**R5**
1	**Naked1**					
2	**Tn9**			**Tn**		
3	**T9**			**T**		
4	**T4**	**T**				
5	**T16**					**T**
6	**T5**		**T**			
7	**T15**				**T**	
8	**TTTTT**	**T**	**T**	**T**	**T**	**T**
9	**ST9**			**Sialyl T**		
10	**ST4**	**Sialyl T**				
11	**ST16**					**Sialyl T**
12	**ST5**		**Sialyl T**			
13	**ST15**				**Sialyl T**	
14	**STSTSTSTST**	**Sialyl T**	**Sialyl T**	**Sialyl T**	**Sialyl T**	**Sialyl T**
15	**TTTnTT**	**T**	**T**	**Tn**	**T**	**T**
16	**TnTnSTTnTn**	**Tn**	**Tn**	**Sialyl T**	**Tn**	**Tn**
17	**TnTnTTnTn**	**Tn**	**Tn**	**T**	**Tn**	**Tn**
18	**STSTTnSTST**	**Sialyl T**	**Sialyl T**	**Tn**	**Sialyl T**	**Sialyl T**
19	**TnTnTnTnTn**	**Tn**	**Tn**	**Tn**	**Tn**	**Tn**

TABLE 1:

A: Blank refers to a non-glycosylated position. Tn (Tn antigen), GalNAcα; T (T antigen), Gaβ1-3GalNAcα. Sialyl T Neu5Acα2,3 Gaβ1-3GalNAcα.

B: Numbers represent the position of amino acid starting from His residue.

Linker- 5 OXO HEXANOIC ACID

NH_2- CONH2

The following Table1 and Table 2 shows library of MUC1 glycopeptide synthesized for Micro array studies:

Glycopeptide library Constructed on Resin: For Micro Array						
Seq: ***N*-P-P-A- H-G-V-T(R1)-S(R2)-A-P-D-T(R3)-R-P-A-P-G-S(R4)-T(R5)-A-*C***						
No.	**Names**	**R1**	**R2**	**R3**	**R4**	**R5**
20	**Naked1**					
21	**ST12**			**Sialyl T**		
22	**SC2STSC2**		**SC2**	**Sialyl T**	**SC2**	
23	**TnSC2STSC2Tn**	**Tn**	**SC2**	**Sialyl T**	**SC2**	**Tn**

Table 2:

A: Blank refers to a non-glycosylated position. Tn (Tn antigen), GalNAcα; T (T antigen), Gaβ1-3GalNAcα , Sialyl T(ST) Neu5Acα2,3 Gaβ1-3GalNAcα,

sialylated core 2/SC2(Neu5Ac⟨2-3Galβ1–3[Galβ1-4GlcAcβ1-6]GalNAcα)

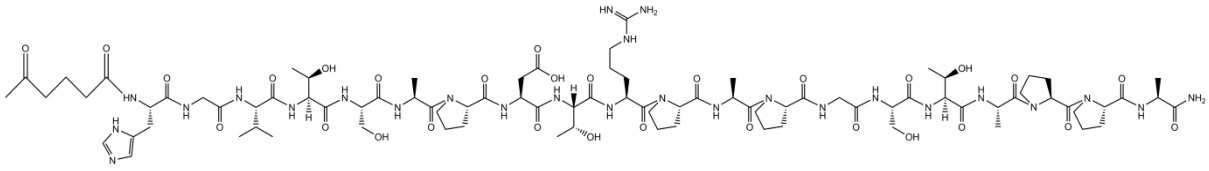
B: Numbers represent the position of amino acid starting from His residue.

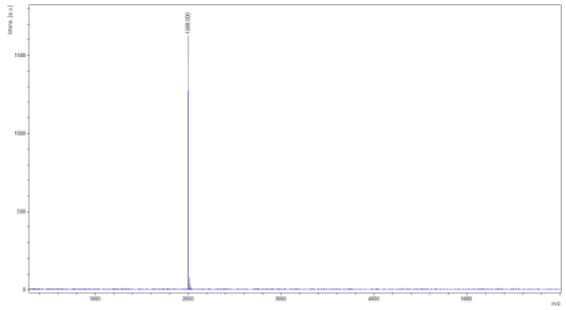
N- 5 OXO HEXONOICACID

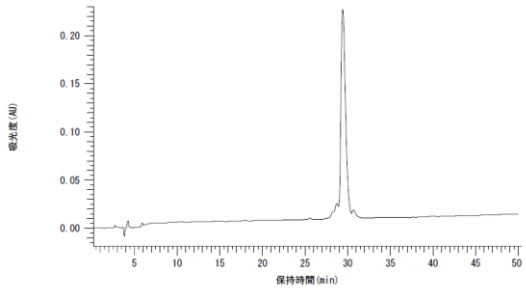
C- CONH2

The Synthesised compounds library were confirmed by MALDI TOF analysis and purified by reverse phase analytical HPLC as shown below:

1: Naked Peptide

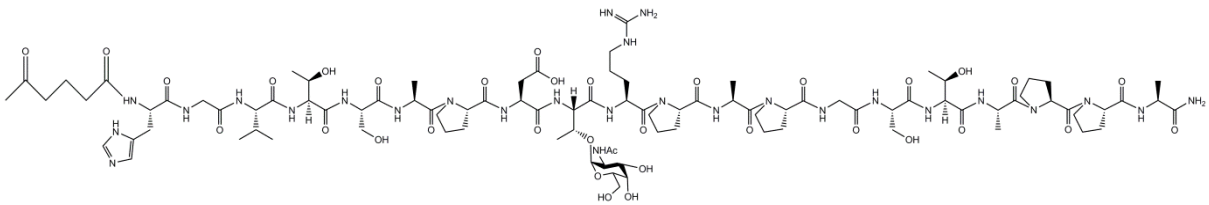


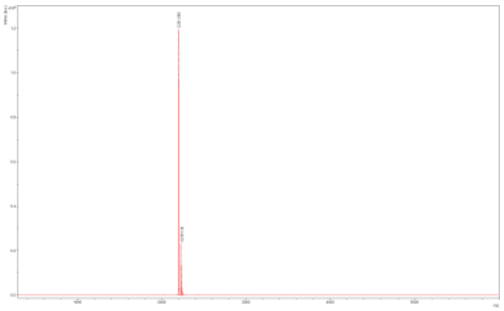


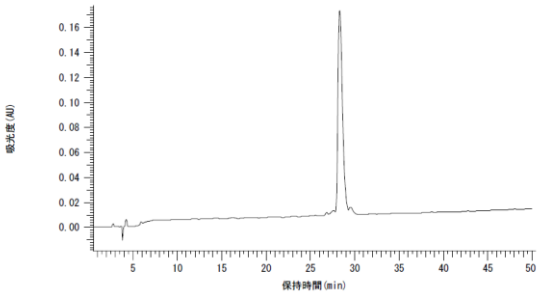


Analytical HPLC A:B(H_2O in 0.1%TFA: CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-30% (B). MALDI TOF chemical formula: $C_{86}H_{136}N_{26}O_{29}$ Exact Mass 1996.997, Observed mass 1998.000 $[M+H]^+$

2.Tn

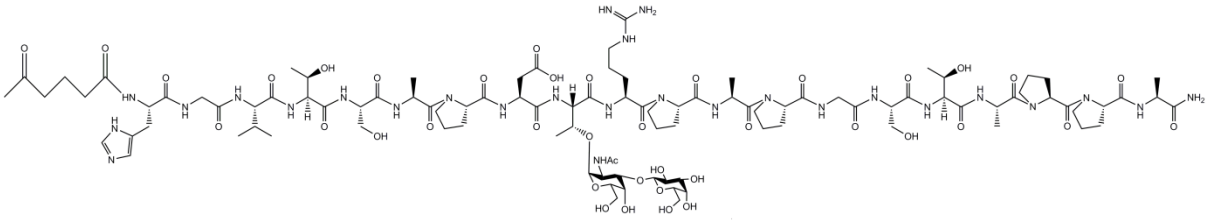


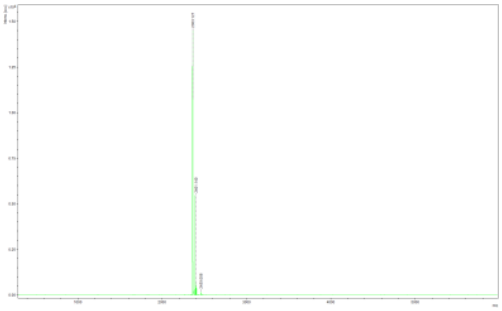


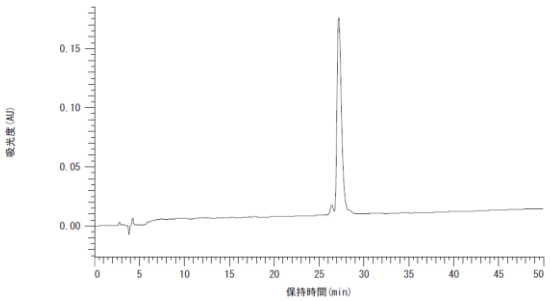


Analytical HPLC A:B(H_2O in 0.1%TFA: CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-30% (B). MALDI TOF chemical formula $C_{94}H_{149}N_{27}O_{34}$ Exact Mass 2200.076 , Observed mass 2201.088 $[M+H]^+$

3:T9

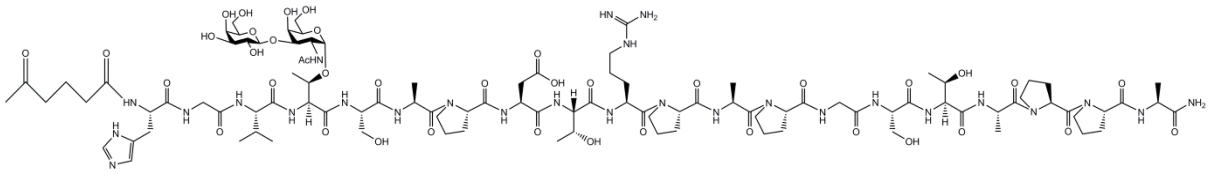


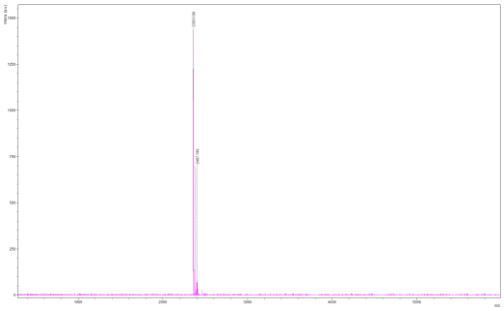


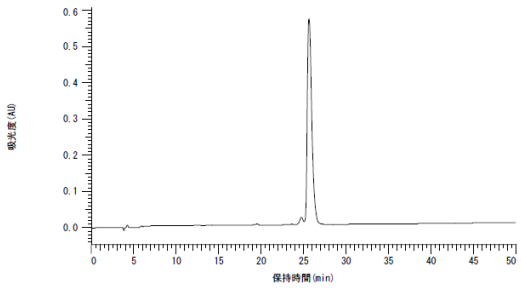


Analytical HPLC A:B(H_2O in 0.1%TFA: CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-30% (B). MALDI TOF chemical formula $C_{100}H_{159}N_{27}O_{39}$ Exact Mass 2362.129 , Observed mass 2363.127 $[M+H]^+$

4:T4

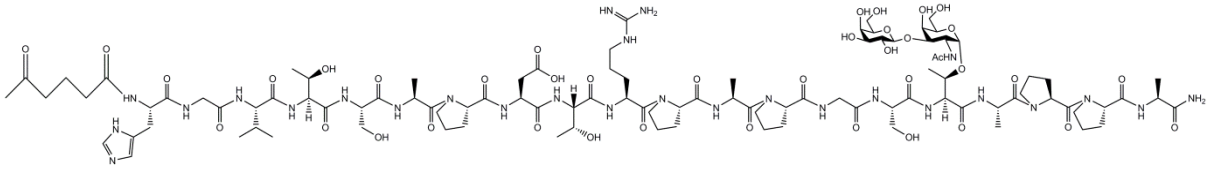


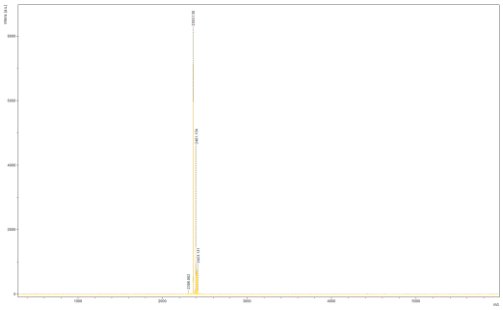


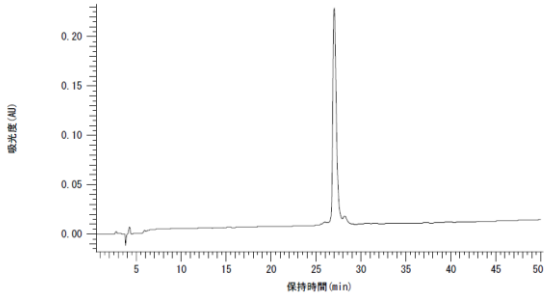


Analytical HPLC A:B(H_2O in 0.1%TFA: CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-30% (B). MALDI TOF chemical formula $C_{100}H_{159}N_{27}O_{39}$ Exact Mass 2362.129 , Observed mass 2363.109 $[M+H]^+$

5:T16

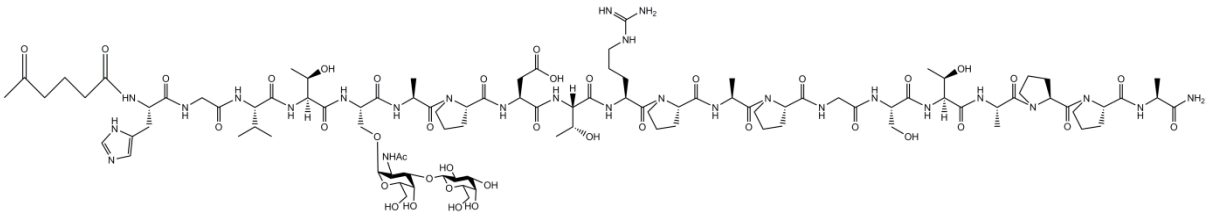


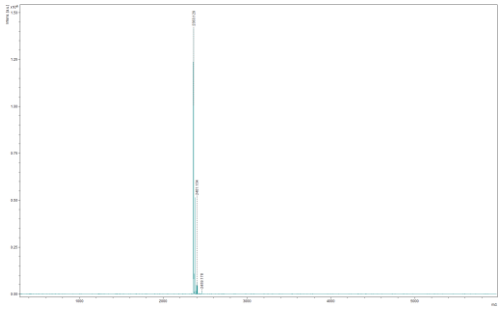


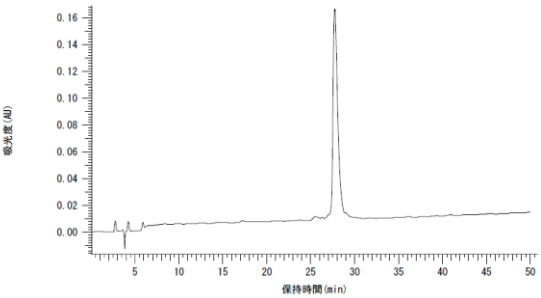


Analytical HPLC A**:**B(H_2O in 0.1%TFA**:** CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-30% (B). MALDI TOF chemical formula $C_{100}H_{159}N_{27}O_{39}$ Exact Mass 2362.129 , Observed mass 2363.130 $[M+H]^+$

6:T5

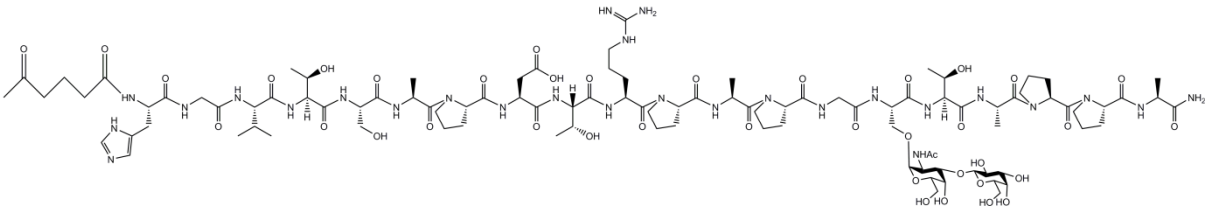


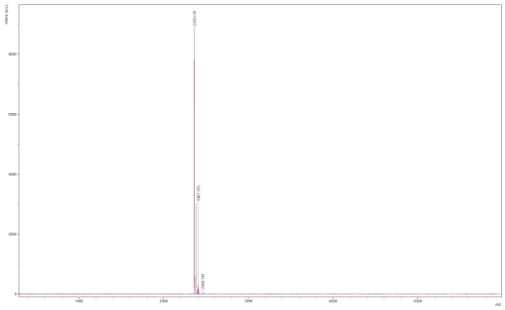


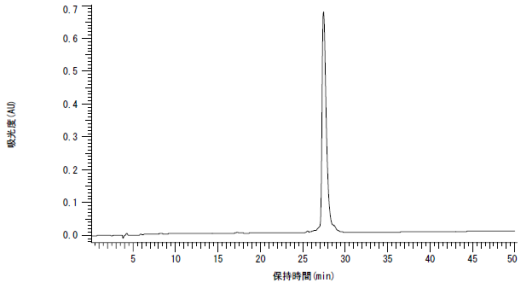


Analytical HPLC A**:**B(H_2O in 0.1%TFA**:** CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-30% (B). MALDI TOF chemical formula $C_{100}H_{159}N_{27}O_{39}$ Exact Mass 2362.129, Observed mass 2363.129 $[M+H]^+$

7:T15

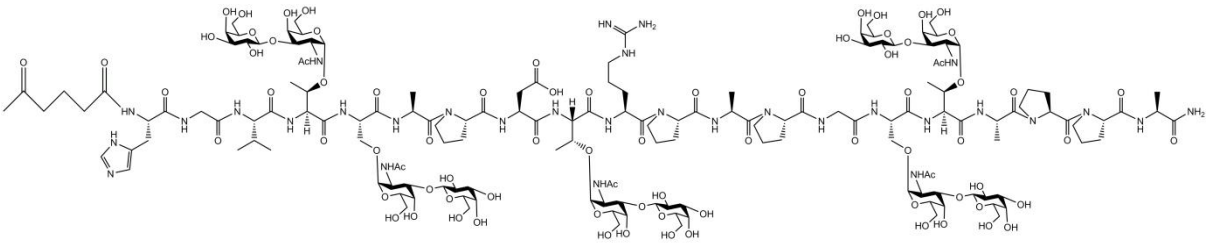


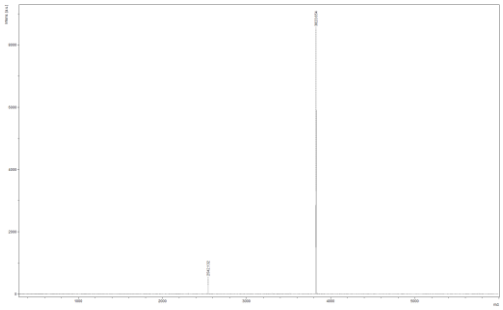


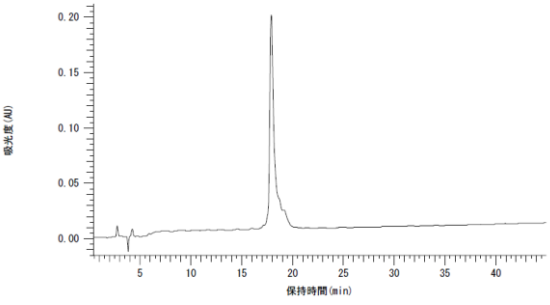


Analytical HPLC A:B(H_2O in 0.1%TFA: CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-30% (B). MALDI TOF chemical formula $C_{100}H_{159}N_{27}O_{39}$ Exact Mass 2362.129 , Observed mass 2363.135 $[M+H]^+$

8:TTTTT

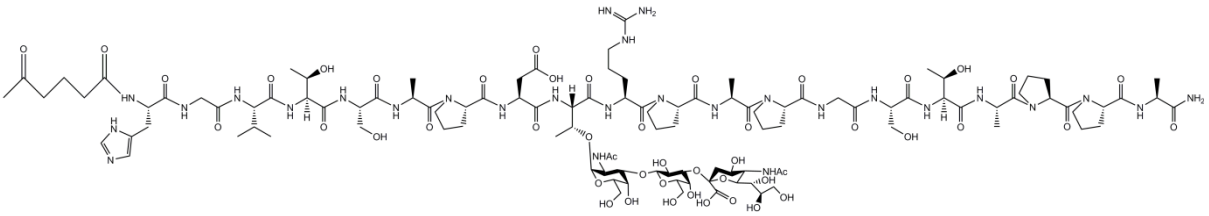


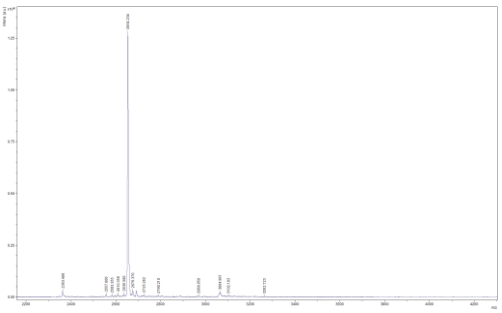


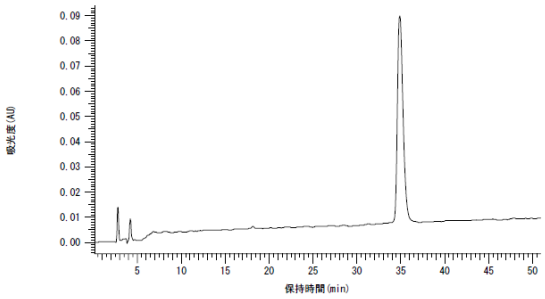


Analytical HPLC A:B(H_2O in 0.1%TFA: CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-30% (B). MALDI TOF chemical formula $C_{156}H_{251}N_{31}O_{79}$ Exact Mass 3822.658 , Observed mass 3823.654 $[M+H]^+$

9: ST9

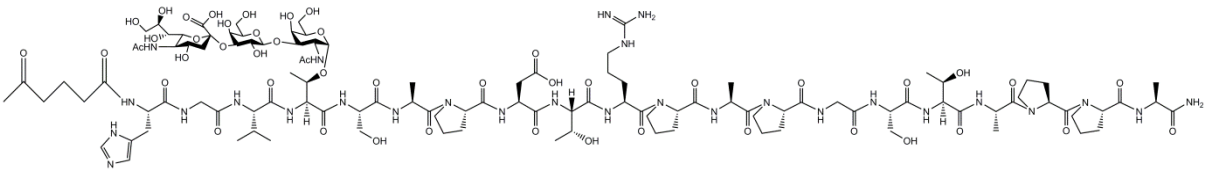


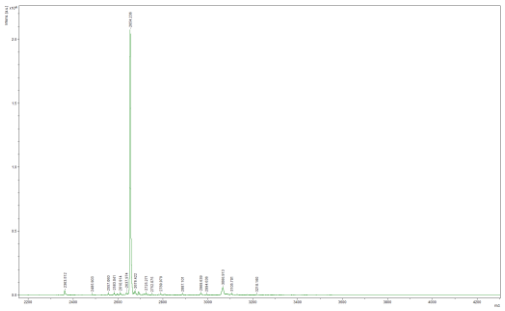


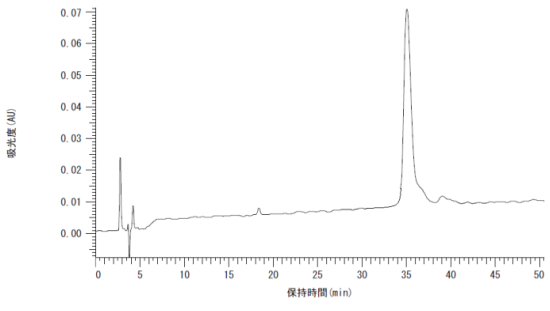


Analytical HPLC A:B(H_2O in 0.1%TFA: CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-25% (B). MALDI TOF chemical formula: $C_{111}H_{176}N_{28}O_{47}$ Exact Mass 2653.224 , Observed mass 2654.234 $[M+H]^+$

10:ST4

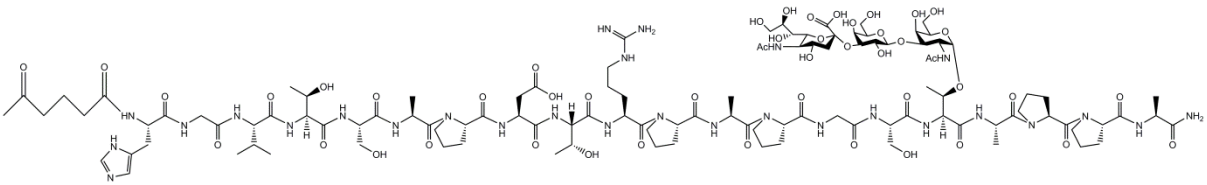


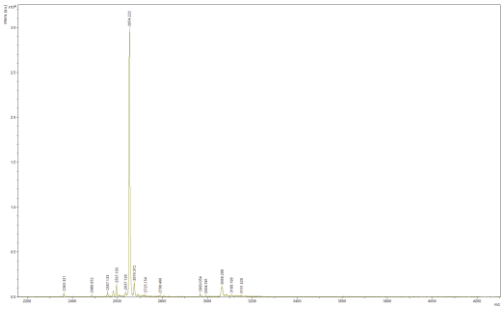


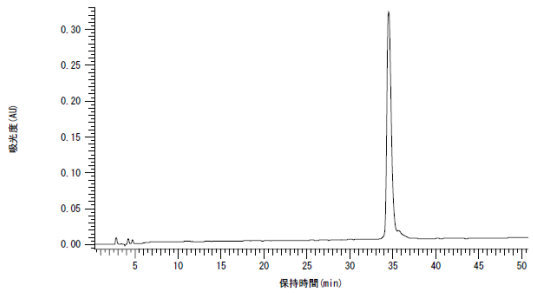


Analytical HPLC A:B(H_2O in 0.1%TFA: CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-25% (B). MALDI TOF chemical formula: $C_{111}H_{176}N_{28}O_{47}$ Exact Mass 2653.224 , Observed mass 2654.236 $[M+H]^+$

11:ST16

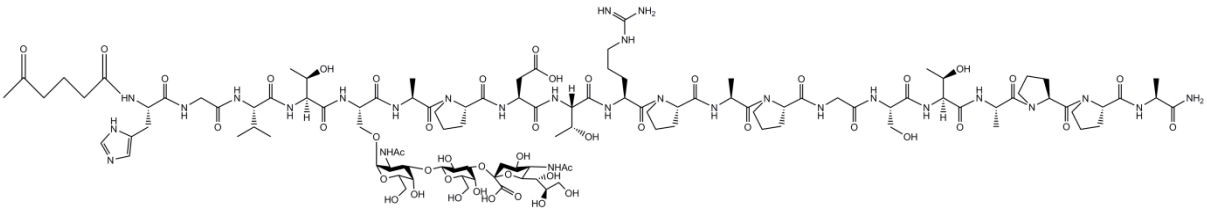


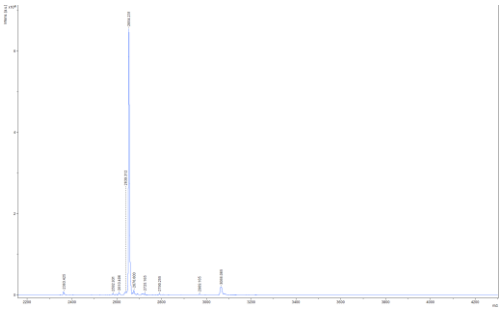


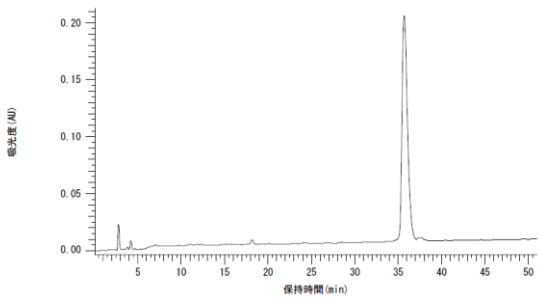


Analytical HPLC A:B(H_2O in 0.1%TFA: CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-25% (B). MALDI TOF chemical formula: $C_{111}H_{176}N_{28}O_{47}$ Exact Mass 2653.224 , Observed mass 2654.222 $[M+H]^+$

12:ST5

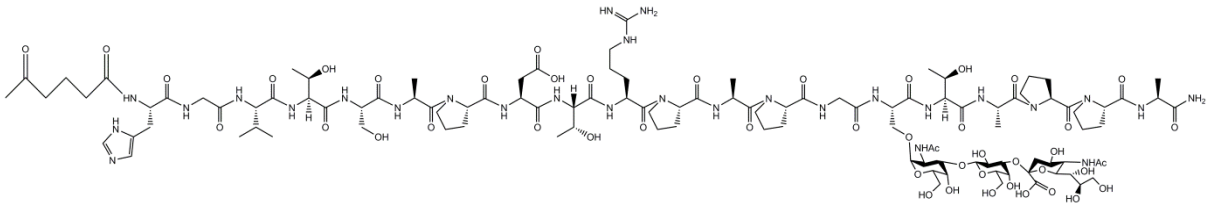


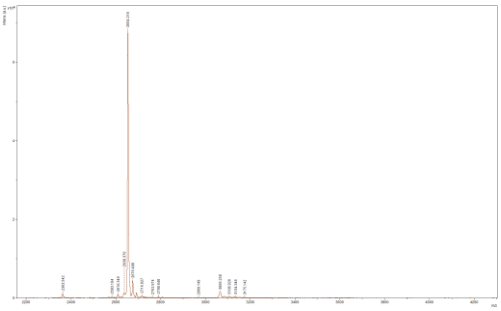


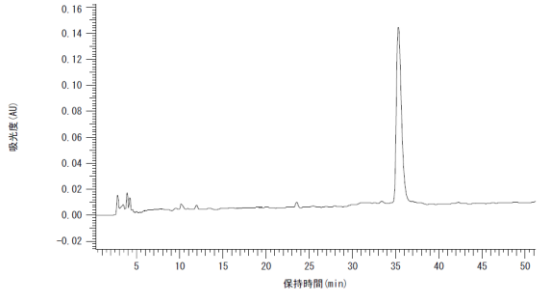


Analytical HPLC A:B(H_2O in 0.1%TFA: CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-25% (B). MALDI TOF chemical formula: $C_{111}H_{176}N_{28}O_{47}$ Exact Mass 2653.224 , Observed mass 2654.231 $[M+H]^+$

13:ST15

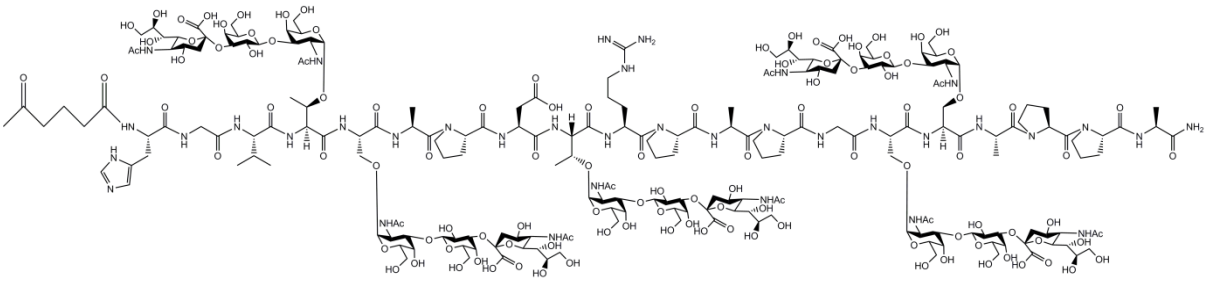


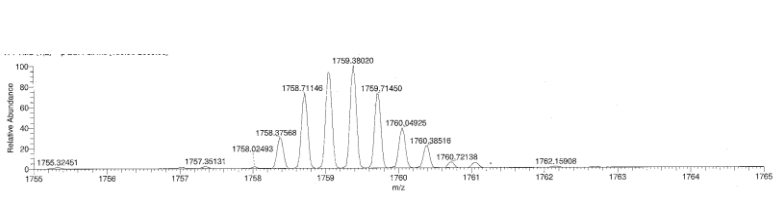


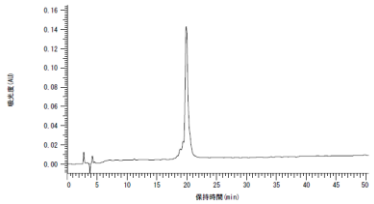


Analytical HPLC A**:**B(H_2O in 0.1%TFA**:** CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-25% (B). MALDI TOF chemical formula: $C_{111}H_{176}N_{28}O_{47}$ Exact Mass 2653.224, Observed mass 2654.218 $[M+H]^+$

14:STSTSTSTST

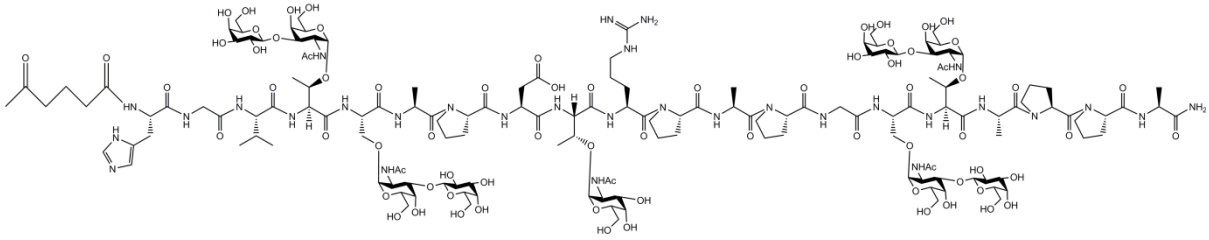


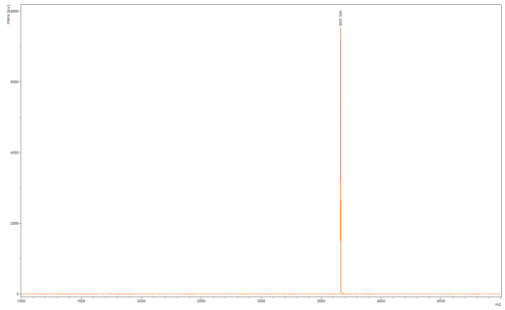


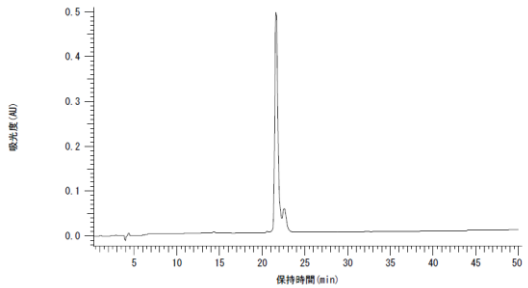


Analytical HPLC A**:**B(H_2O in 0.1%TFA**:** CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-25% (B). MALDI TOF chemical formula: $C_{211}H_{336}N_{36}O_{119}$ Exact Mass 5278.1347 ,m/z 1759.3782 Observed mass 1758.3756 $[M-3H]^{3-}$

15:TTTnTT

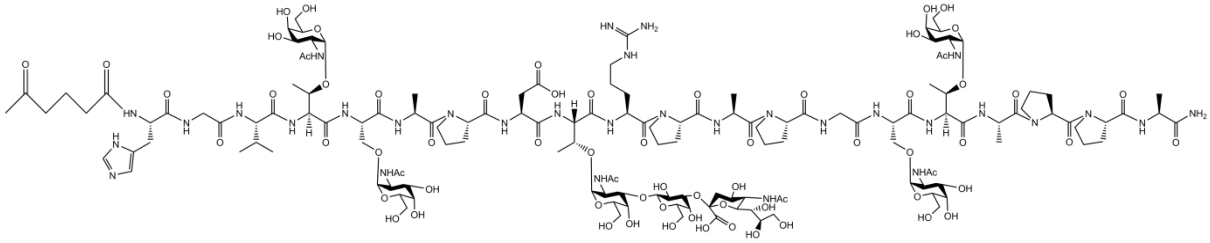


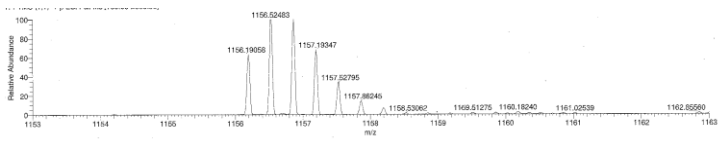


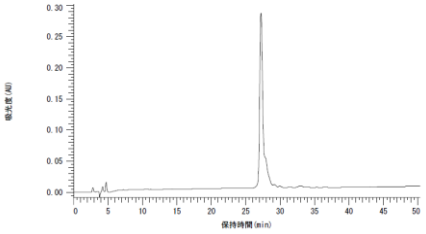


Analytical HPLC A**:**B(H_2O in 0.1%TFA**:** CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-30% (B). MALDI TOF chemical formula: $C_{150}H_{241}N_{31}O_{74}$ Exact Mass 3660.605 , Observed mass 3660.594 $[M+H]^+$

16:TnTnSTTnTn

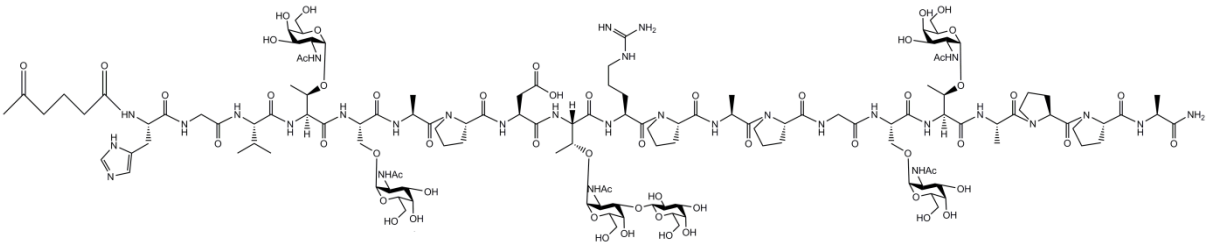


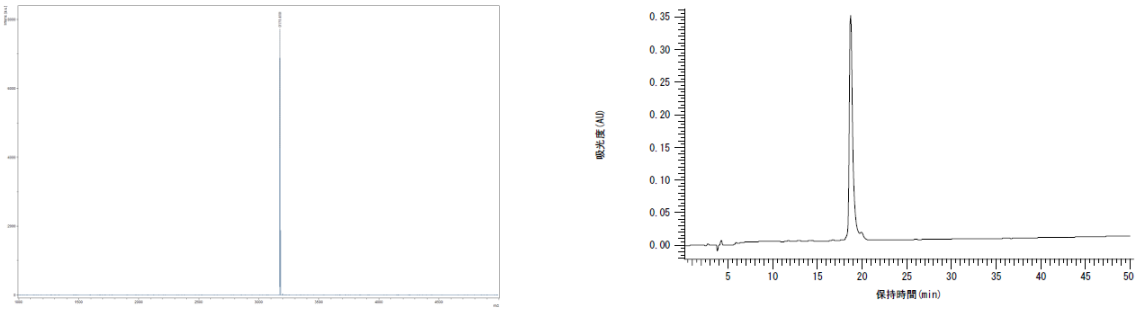




Analytical HPLC A**:**B(H_2O in 0.1%TFA**:** CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-25% (B). MALDI TOF chemical formula: $C_{143}H_{228}N_{32}O_{67}$ Exact Mass 3465.5417,m/z 1155.1805 Observed mass 1156.1905 $[M+3H]^{3+}$

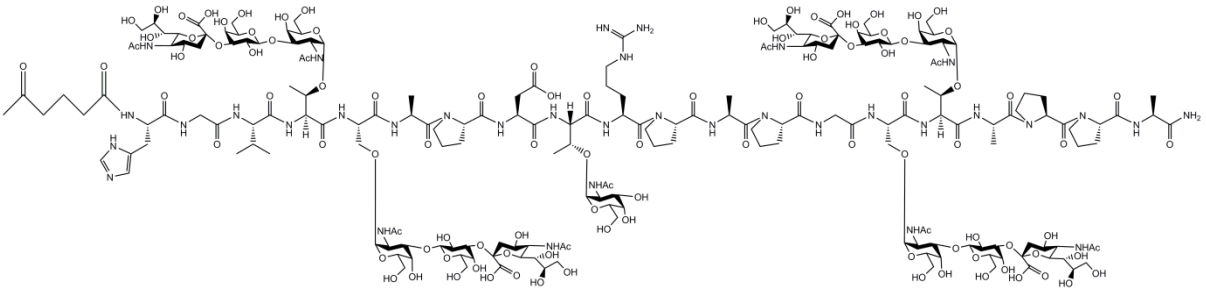
17:TnTnTTnTn

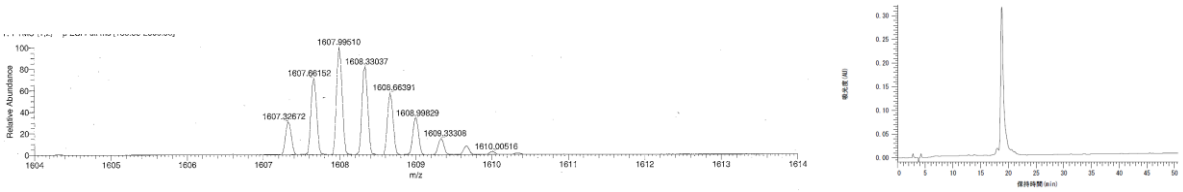




Analytical HPLC A:B(H_2O in 0.1%TFA: CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-30% (B). MALDI TOF chemical formula: $C_{132}H_{211}N_{31}O_{59}$ Exact Mass3174.446, Observed mass 3175.459 $[M+H]^+$

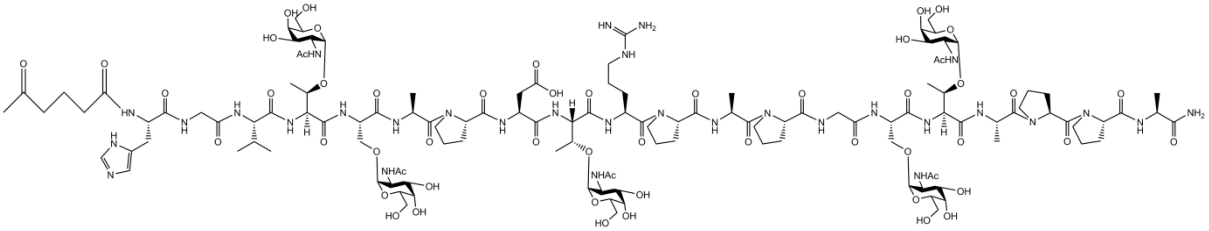
18:STSTTnSTST

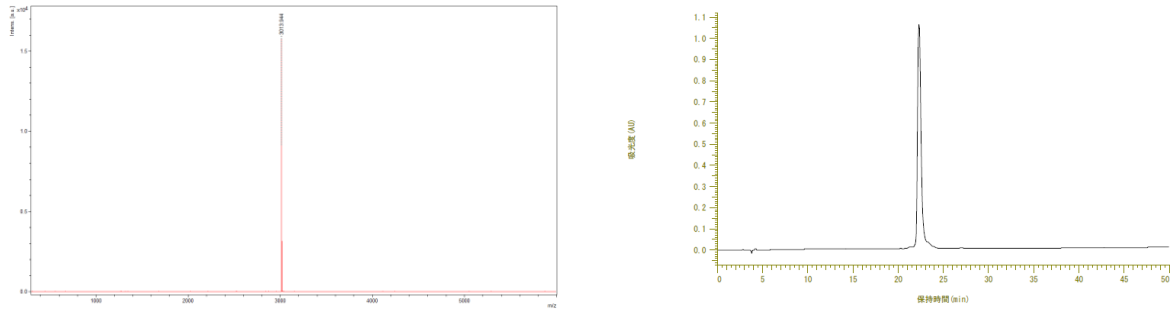




Analytical HPLC A:B(H_2O in 0.1%TFA: CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-25% (B). MALDI TOF chemical formula: $C_{194}H_{309}N_{35}O_{106}$ Exact Mass 4824.9864 , m/z 1608.3288 Observed mass 1607.3267 $[M-H]^{3-}$

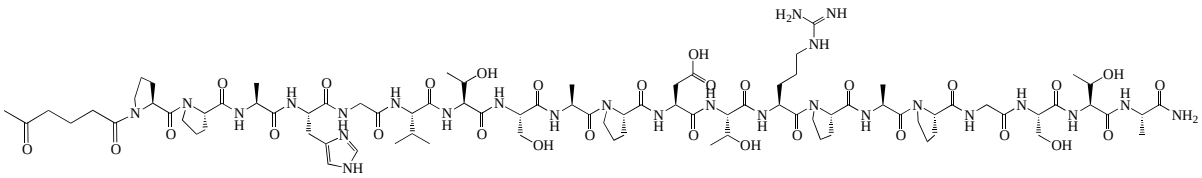
19:TnTnTnTnTn

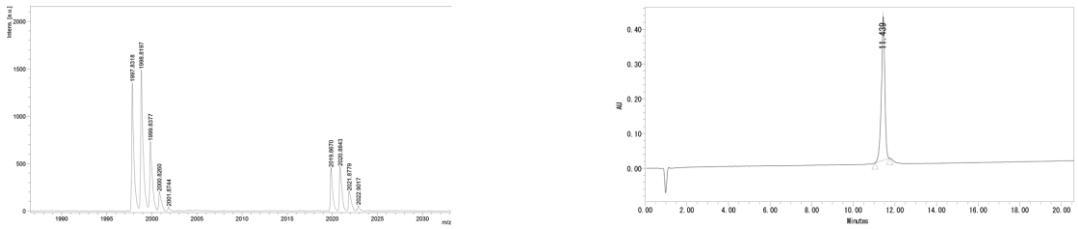




Analytical HPLC A**:**B(H_2O in 0.1%TFA**:** CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-30% (B). MALDI TOF chemical formula: $C_{131}H_{211}N_{31}O_{49}$ Exact Mass 3012.3935 Observed mass 3013.944 $[M+H]^+$

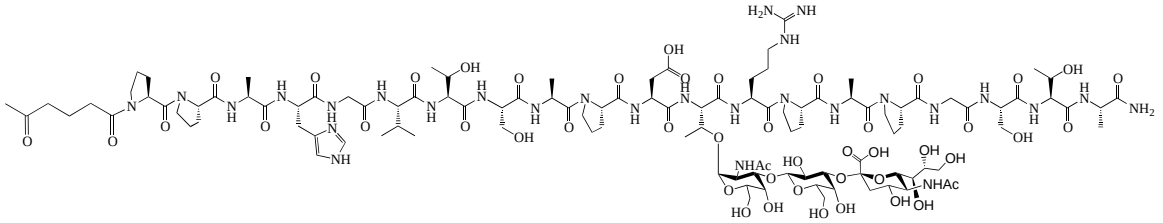
20: Naked2

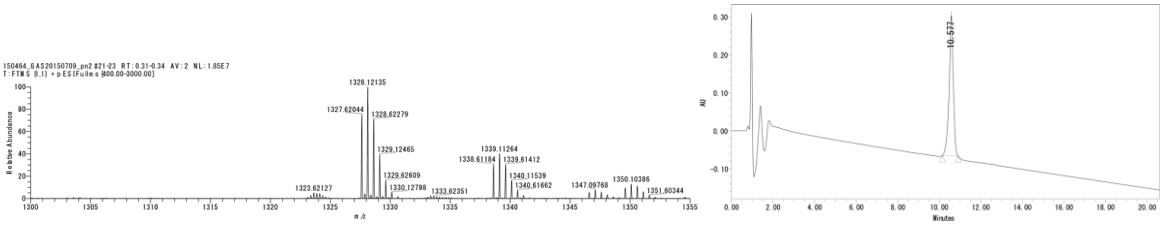




Analytical UPLC (gradient condition A): t_R = 11.439 min, peak area ratio 100%. MALDI-TOF-HRMS: $C_{86}H_{136}N_{26}O_{29}$ $[M+H]^+$ calcd (m/z) 1996,9967, found (m/z) 1997.8318.

21:ST12

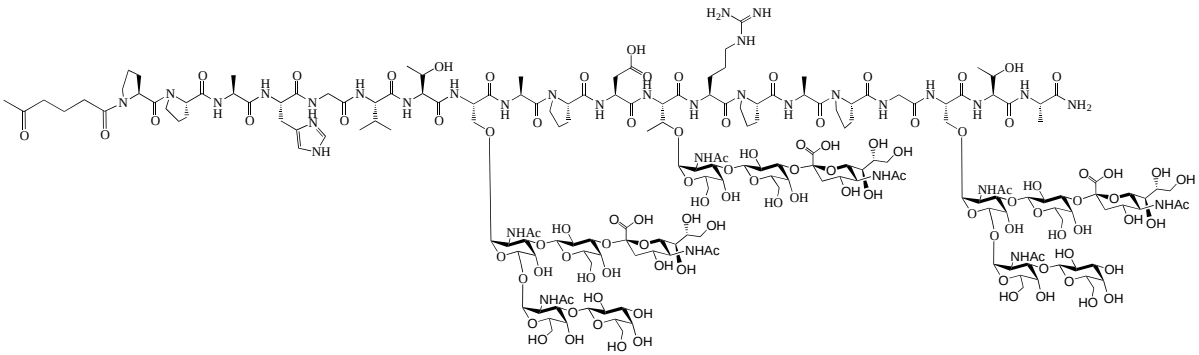


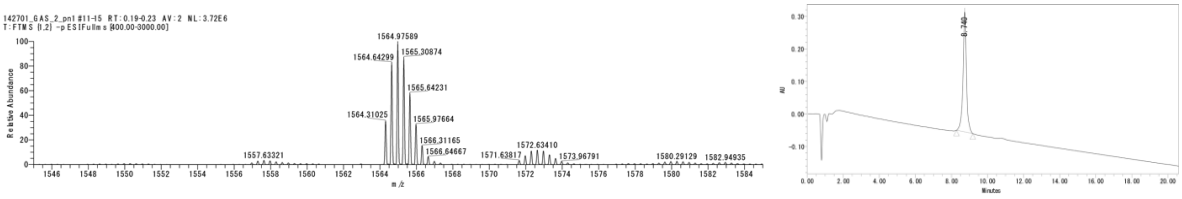


Analytical UPLC (gradient condition B): t_R = 10.577 min, peak area ratio 100%. ESI-HRMS: $C_{111}H_{176}N_{28}O_{47}$ $[M-2H]^-$ calcd (m/z) 1327.6194, found (m/z) 1327.6204

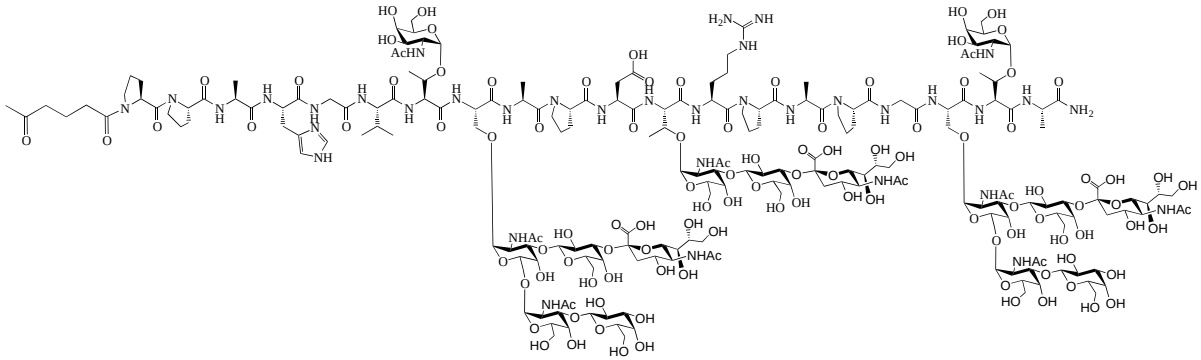
22:SC2STSC2

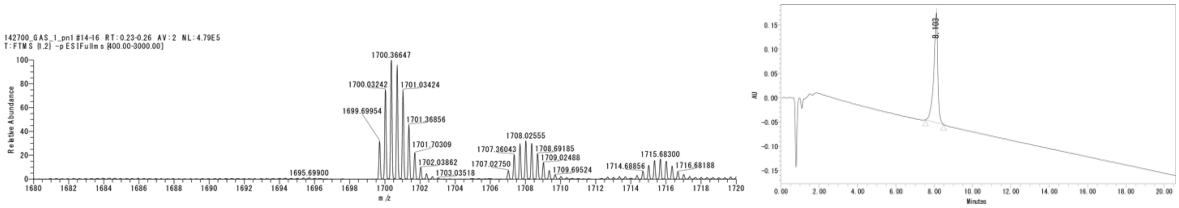
Analytical UPLC (gradient condition B): t_R = 8.740 min, peak area ratio 100%. ESI-HRMS: $C_{189}H_{302}N_{34}O_{103}$ $[M-3H]^-$ calcd (m/z) 1564.3074, found (m/z) 1564.3103





23: TnSC2STSC2Tn





Analytical UPLC (gradient condition B): t_R = 8.103 min, peak area ratio 100%. ESI-HRMS: $C_{205}H_{328}N_{36}O_{113}$ $[M-3H]^-$ calcd (m/z) 1699.6936, found (m/z) 1699.6995.

2-2.2. Synthesis of MUC2 Glycopeptides for NMR studies.

After synthesising the glycopeptides analogues for micro array studies, we synthesised 6 glycopeptides to study the structure by NMR. These glycopeptide were synthesised similarly from C terminus to N terminus. The synthesis was completed by coupling the N terminus by acetyl capping which was a preferable method for NMR data analysis.

The following table shows the compounds synthesised for NMR studies:

The Synthesised compounds library were confirmed by MALDI TOF analysis and purified by reverse phase analytical HPLC as shown below:

Glycopeptide Library Constructed on Resin: For NMR						
Seq: *N*- H-G-V-T(R1)-S(R2)-A-P-D-T(R3)-R-P-A-P-G-S(R4)-T(R5)-A-P-P-A-*C*						
No.	**Names**	**R1**	**R2**	**R3**	**R4**	**R5**
8	**TTTTT**	**T**	**T**	**T**	**T**	**T**
14	**STSTSTSTST**	**Sialyl T**	**Sialyl T**	**Sialyl T**	**Sialyl T**	**Sialyl T**
17	**TnTnTTnTn**	**Tn**	**Tn**	**T**	**Tn**	**Tn**
18	**STSTTnSTST**	**Sialyl T**	**Sialyl T**	**Tn**	**Sialyl T**	**Sialyl T**
19	**TnTnTnTnTn**	**Tn**	**Tn**	**Tn**	**Tn**	**Tn**
Seq: *N*-G-V-T(R1)-S(R2)-A-P-D-T(R3)-R-P-A-P-G-S(R4)-T(R5)-A-P-P-A-H-G-V-T -*C*						
No.	**Names**	**R1**	**R2**	**R3**	**R4**	**R5**
24	**Tn8**			**Tn**		

Table3:

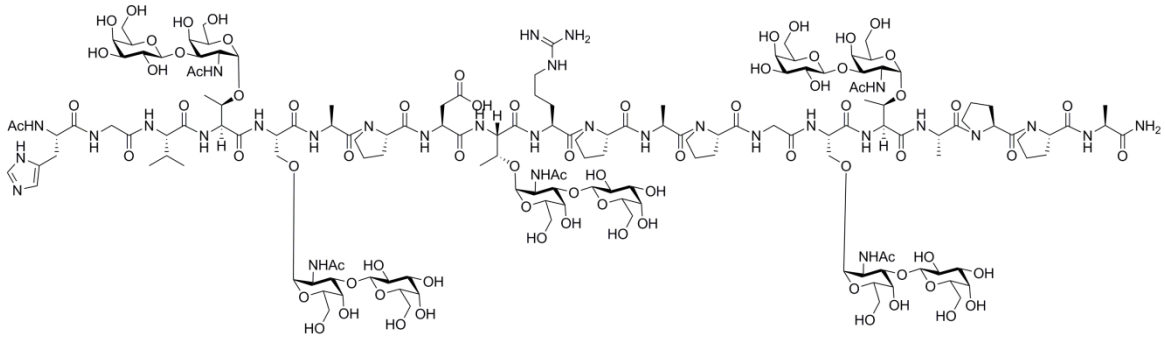
A: Blank refers to a non-glycosylated position. Tn (Tn antigen), GalNAcα; T (T antigen), Gaβ1-3GalNAcα. Sialyl T Neu5Acα2, 3 Gaβ1-3GalNAcα.

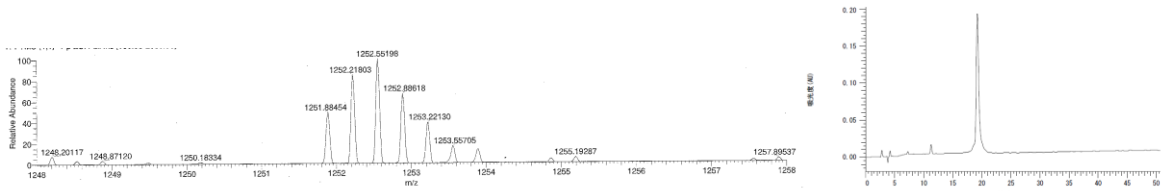
B: Numbers represent the position of amino acid starting from His residue.

N- NHAc (N terminal)

C- CONH2 (C terminal)

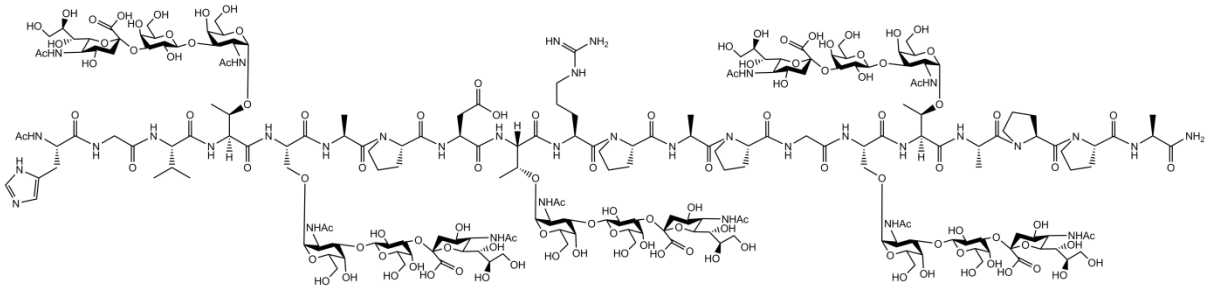
8:TTTTT

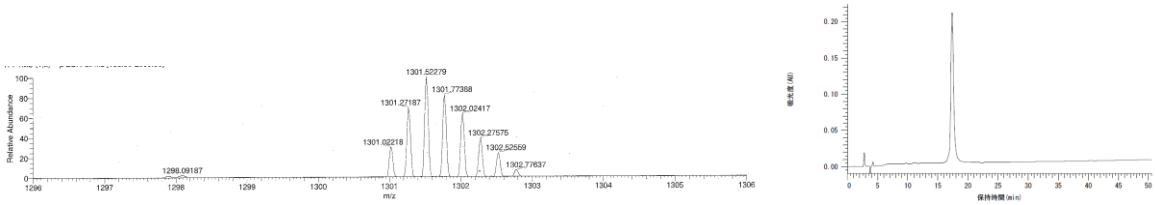




Analytical HPLC A:B(H_2O in 0.1%TFA**:** CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-25% (B). MALDI TOF chemical formula: $C_{152}H_{245}N_{31}O_{78}$ Exact Mass 3752.6157 ,m/z 1250.8719 Observed mass 1251.8845 $[M+3H]^{3+}$

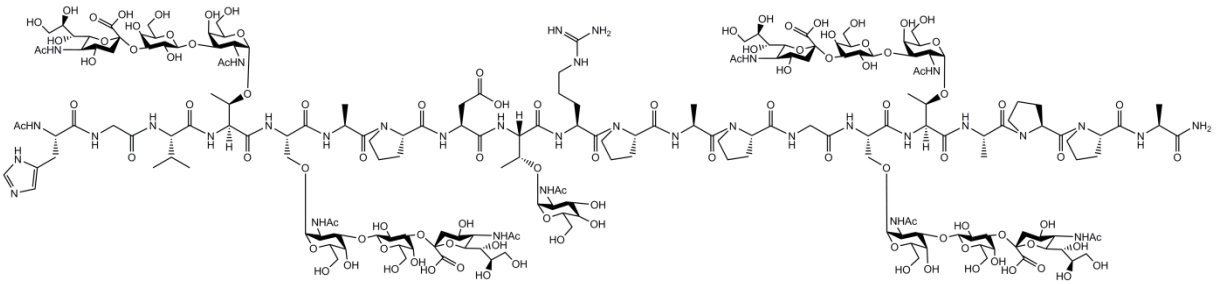
14:STSTSTSTST

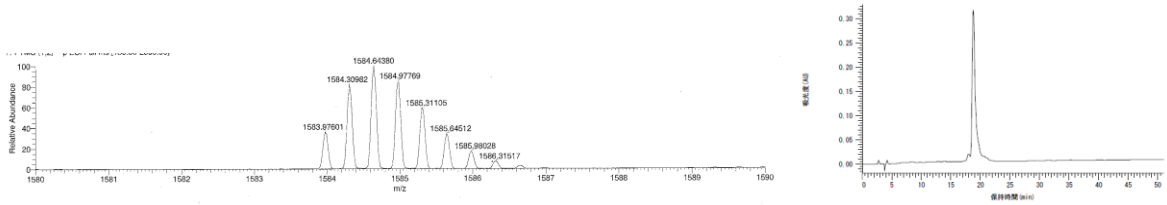




Analytical HPLC A:B(H_2O in 0.1%TFA**:** CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-25% (B). MALDI TOF chemical formula: $C_{207}H_{330}N_{36}O_{118}$ Exact Mass 5208.0928 ,m/z 1302.232 Observed mass 1301.0221 $[M-4H]^{4-}$

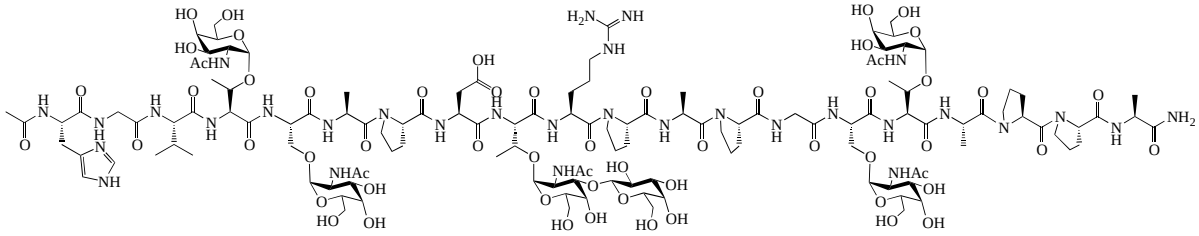
18: STSTTnSTST

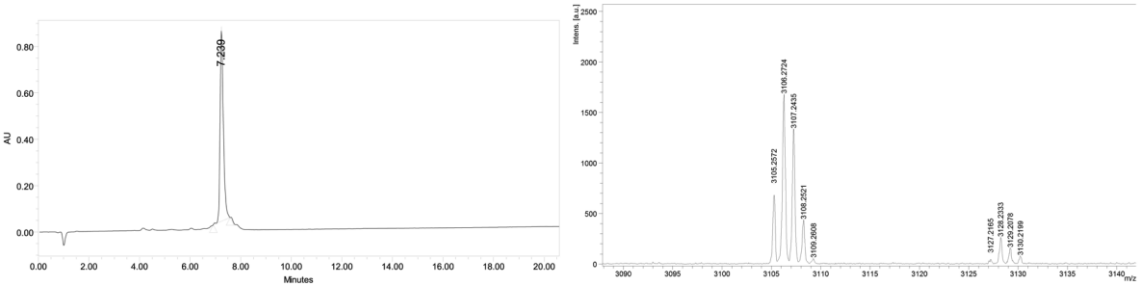




Analytical HPLC A**:**B(H_2O in 0.1%TFA**:** CH_3CN in 0.1%TFA) A/B gradient 0-50min, 5-25% (B). MALDI TOF chemical formula: $C_{190}H_{303}N_{35}O_{105}$ calculated Exact Mass 4754.9446, m/z 1584.9815 , Observed mass 1583.9760 $[M-3H]^{3-}$

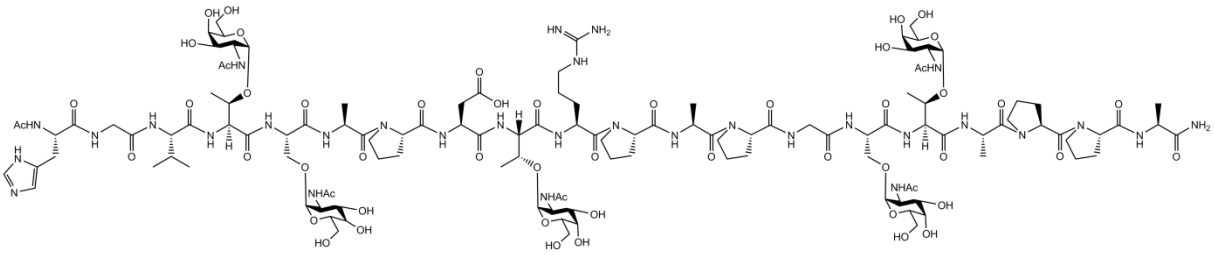
17: TnTnTTnTn

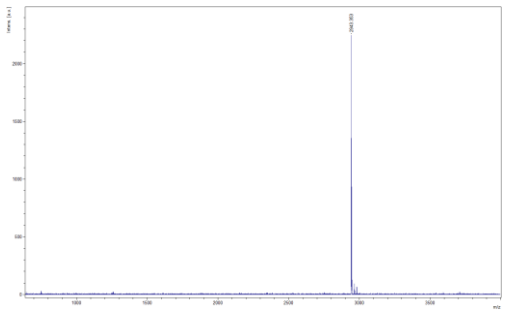


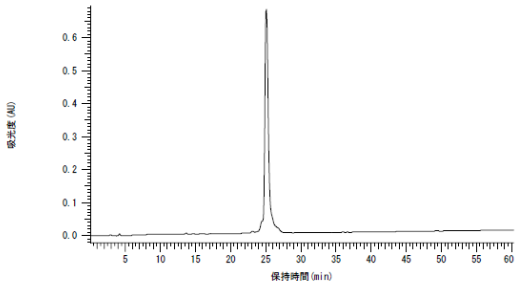


Analytical UPLC (gradient condition B): t_R = 7.239 min, peak area ratio 100%. MALDI-TOF-HRMS: $C_{128}H_{242}N_{31}O_{76}$ $[M+H]^+$ 3105.2572 calcd (m/z), found (m/z) 1540.9641.

19: TnTnTnTnTn







Analytical UPLC (gradient condition B): t_R = 7.239 min, peak area ratio 100%. MALDI-TOF-HRMS: $C_{127}H_{205}N_{31}O_{48}$ $[M+H]^+$ 2943.358 calcd (m/z), found (m/z) 2943.353

2-3. Conclusion

In summary, we successfully synthesised MUC1 peptides single glycosylated, homogeneously glycosylated and heterogeneously glycosylated bearing Tn antigen structure, core 1 structure, sialylated core 1 structure , sialylated core 2 structures. Totally 21 MUC1 glycopeptides which were used for micro array studies and 6 MUC1 glycopeptides for NMR studies. During synthesis of heavily glycosylated glycopeptides the yield was reduced when compared with naked peptide or singly glycosylated glycopeptides. All glycopeptides were purified by reverse phase analytical HPLC with >95% purity. Detection of all glycopeptides were carried out by MALDI TOF analysis, and confirmed by accepting within 5ppm difference of exact mass calculated by chemdraw software.

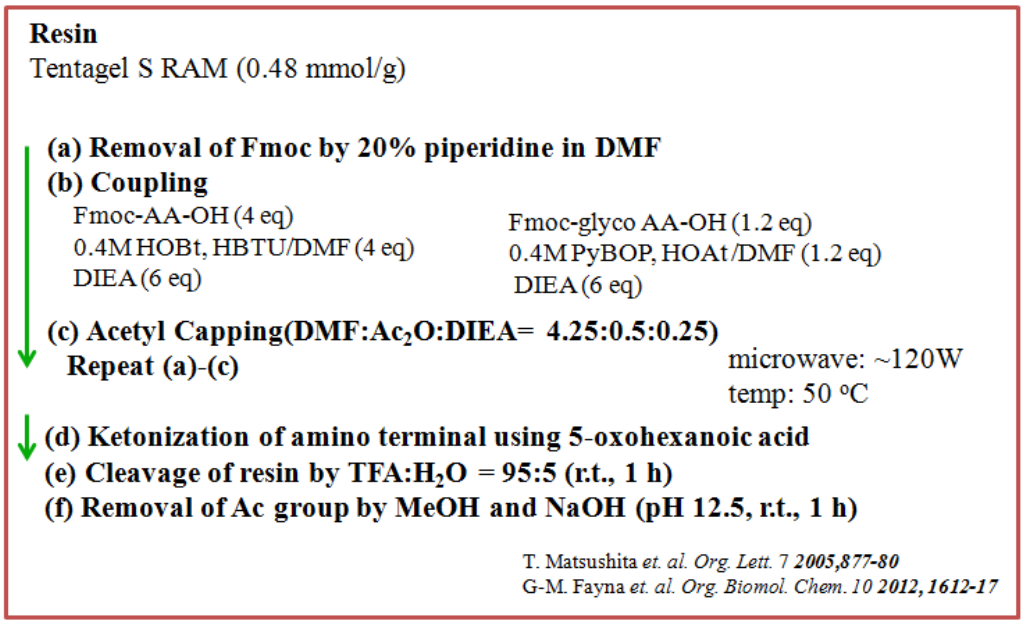
2-4. Experimental Section

Peptide Synthesis:

MUC1 glycopeptides were synthesised efficiently by chemical and enzymatic procedure[4-12]. Initially solid phase synthesis was done using a polypropylene tube (libra tube, Hipep laboratories Kyoto, Japan) which was equipped with filters. Few procedures such as washing, acetyl capping and final cleavage were done at room temperature (RT). Reaction vials were placed inside the cavity of microwave instrument comprising of microwave synthesis reactor (IDX Corp., Tochigi, Japan) (EYELA microwave synthesizer Wave Magic MWS-1000A, Tokyo Rikakikai Co., LTD., Tokyo, Japan), attached to this instrument was kept stirrer which was useful for mixing the contents. Irradiation at 2450Hz at temperatures at 50°C was maintained in the microwave[13, 14]. The solvents were purchased from Watanabe Chemical Industries, Ltd. (Hiroshima, Japan) and reagents were purchased from Kokusan Chemical Co., Ltd Tokyo, Japan and Pure Chemical Industries, Ltd. (Osaka, Japan) which were used during the experiments, no further changes were done with the solvents and reagents purchased. Reagents such as recombinant rat α2, 3-(o)-SiaT was purchased from Calbiochem (Merk Millipore, Darmstadt, Germany). Bovine D-glucose 4- β -D-galactosyltransferase (β 1,4-GalT) and rat liver α 2,3-(*O*)-sialyltransferase (α1,3-(*O*)-sialylT) were purchased from Sigma Aldrich and Calbiochem Co. Ltd (Merck Millipore). Cytidine-5'-monophospho-β-D-*N*-acetylneuraminic acid, disodium salt (CMP NANA) , Uridine-5'-diphosphogalactose, disodium salt (UDP-galactose), were purchased from Yamasa Corporation (Chiba, Japan).

All N^{α}-[9-(Fluorenylmethoxy) carbonyl (Fmoc)]-L-amino acids, (Nova biochem, merck biosciences, Darmstadt,Germany) were used. NovaPEGRink amide chem resin was used for Matrix (Biotage GB Ltd, Dyffryn, and Hengoed, UK). Initial swelling of resin was done using Dichloromethane (DCM) solution for 2-3 hrs and rinsed with N,N-dimethylformamide

(DMF) for 3-4 times. Once the swelling was complete, amino acid which was pre-activated using the coupling reagent (1-[Bis(dimethylamino)methyliumyl]-1H-benzotriazole-3-oxide hexafluorophosphate (HBTU) 4equiv, 1-hydroxybenzotriazole monohydrate (HOBt) 4 equiv. and N,N-diisopropylethylamine (DIEA) 6 equiv in DMF solution; 4 equiv of respective amino acid taken was introduced into the resin tube and placed inside the microwave for 9 min after initial coupling it was washed with DMF x3, DCM x3 and again DMF x3 at RT. The Fmoc removal was carried out using 20% piperdine solution for 3min by microwave. After Fmoc de-protection, washing was performed as earlier and continued with coupling of next amino acid followed after pre-activation step. During glycosylated amino acids double activation protocol was followed to enhance the coupling percentage. Glycosylated amino acid employed was 1.2 equiv. with the glycol amino acid coupling reagent ((benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyBOP) 1.2 equiv.,). 1-Hydroxy-7-azabenzotriazole (HOAt) 1.2 equiv., DIEA 1.8 equiv. in DMF solution). After initial coupling using microwave for 9min was completed once again the coupling reagent was added without filtering the initial mixture solution and again subjected for 9min microwave coupling. After double coupling same washing procedure was followed mentioned before. Series of peptides were synthesised for micro array studies, therefore, 5 oxohexanoic acid (Tokyo chemical industry Co.Ltd,(Tokyo Japan) was incorporated after final amino acid Fmoc de-protection as N-terminus linker (Table 1). In case of glycan synthesis for peptide 22 and 23 comprising of core 2 structure were glycosylated by using CMP-NANA (9 mM), UDP-galactose (9 mM), α1,3-(*O*)-sialylT (52.5 mU/ml), β1,4-GalT (200 mU/ml) in a total volume of 770 µl of 50 mM HEPES-NaOH buffer, pH 7.0, 10 mM $MnCl_2$, 0.1% Triton X-100 and 0.1% NaN_3. After incubation for 18h, the reaction mixture was purified it was subjected for purification. Glycopeptides synthesised for 2D NMR studies was acetyl capped at N terminus.



After complete solid phase synthesis the glycopeptide was removed by treating with a cleavage cocktail {trifluroacetic acid (TFA)/miliQ H20/tri-isopropylsilane (95/2.5, 2.5 v/v)} for 2hrs at RT. The solution was filtered from the resin and air dried followed by precipitating by adding cold tert-butyl-methyl ether in ice bath and subjected to centrifugation at 3500rpm for 10min at 0℃, the solution was discarded without disturbing the pellet and again tert-butyl-methyl ether was added and centrifuged similarly. After the second centrifugation the solution was air dried and dissolved in miliQ water (H_20)-Acetonitrile (CH_3CN) in the ratio 1:1 and lyophilized. Once the compound was lyophilised, the synthesised peptide was subjected for the removal of acetyl protecting groups using 5-10ml methanol and the pH was maintained at 12.4-12.5 using 1N NaOH solution. The de-acetylation reaction was confirmed by Mass spectrometry (BRUKER, Maldi TOF). Later the reaction was neutralised by AcOH addition. The solvent was evaporated by vacuum and lyophilised in H_20:CH_3CN (1:1) solution.

Purification process was carried out by Analytical RP-HPLC using Waters Acquity Ultra Performance LC system equipped with binary solvent delivery pump, an auto sampler and a

UV detector and an Acquity UPLC BEN®C18 column (1.7 μm, 2.1 × 50 mm, Waters) and Hitachi high performance liquid chromatography system L-6250 intelligent pump and L-7400 UV detector, using a reversed phase (RP) C18 column [Intersil ODS-3 5um, ϕ20x 250mm (GL sciences Inc.)], Preparative HPLC purifications were performed on a Prominence Shimadzu HPLC system (Shimadzu Corporation, Kyoto, Japan) equipped with two LC-6AD pumps, an SPD-20A UV/VIS detector at 220 nm for monitor, and Inertsil ODS-3 reversed-phase C-18 column (250×4.6 mm I.D., GL Sciences Inc., Tokyo, Japan). For enzymatic glycosylation products same HPLC instruments were used as used for the non glycosylated products. Purity of complete series of peptide library synthesised were checked by analytical reverse phase HPLC: elution Buffer A, H_2O containing 0.1% TFA and elution buffer B CH_3CN containing 0.1% TFA; linear A/B gradient used for the products were from 0min gradient B is 5%, and at 50min 25% or 30% as flow rate 1ml/min; uv monitoring at 220 nm having column temperature at 40°C. Purity was accomplished more than 90% for all peptides for further studies. All the eluted product peak was reconfirmed by MALDI TOF/MS (ultra flex II, bruker Daltonics,Germany). Di-hidroxy benzoic acid in TA solution was used as the matrix for MALDI. Glycopeptides with higher molecular weight glycoproteins were subjected to High-resolution electrospray ionization mass spectra (ESI-HRMS) with JEOL JMS-700TZ, and Amino acid analysis with a JEOL JLC-500/V equipped with ninhydrin detection system, performed at the Center of Instrumental Analysis at Hokkaido University. High-resolution MALDI-TOF mass spectra was performed in a Bruker Daltonics Ultraflex MALDI-TOF/TOF mass spectrometer using DHB as a matrix

2-5. References

(1) Fosgerau, K., and Hoffmann, T. (2014) Peptide therapeutics: Current status and future directions. *Drug Discov. Today 20*, 122–128.

(2) Merrifield, R. B. (1963) Solid Phase Peptide Synthesis. I. The Synthesis of. *J. Am. Chem. Soc. 85*, 2149.

(3) Walker, J. M. (2009) IN M OLECULAR B IOLOGY ™ Series Editor. *Life Sci. 531*, 588.

(4) Fumoto, M., Hinou, H., Matsushita, T., Kurogochi, M., Ohta, T., Ito, T., Yamada, K., Takimoto, A., Kondo, H., Inazu, T., and Nishimura, S. I. (2005) Molecular transporter between polymer platforms: Highly efficient chemoenzymatic glycopeptide synthesis by the combined use of solid-phase and water-soluble polymer supports. *Angew. Chemie - Int. Ed. 44*, 2534–2537.

(5) Fumoto, M., Hinou, H., Ohta, T., Ito, T., Yamada, K., Takimoto, A., Kondo, H., Shimizu, H., Inazu, T., Nakahara, Y., and Nishimura, S. I. (2005) Combinatorial synthesis of MUC1 glycopeptides: Polymer blotting facilitates chemical and enzymatic synthesis of highly complicated mucin glycopeptides. *J. Am. Chem. Soc. 127*, 11804–11818.

(6) Matsushita, T., Hinou, H., Fumoto, M., Kurogochi, M., Fujitani, N., Shimizu, H., and Nishimura, S. I. (2006) Construction of highly glycosylated mucin-type glycopeptides based on microwave-assisted solid-phase syntheses and enzymatic modifications. *J. Org. Chem. 71*, 3051–3063.

(7) Naruchi, K., Hamamoto, T., Kurogochi, M., Hinou, H., Shimizu, H., Matsushita, T., Fujitani, N., Kondo, H., and Nishimura, S. I. (2006) Construction and structural characterization of versatile lactosaminoglycan-related compound library for the synthesis of complex glycopeptides and glycosphingolipids. *J. Org. Chem. 71*, 9609–9621.

(8) Matsushita, T., Sadamoto, R., Ohyabu, N., Nakata, H., Fumoto, M., Fujitani, N., Takegawa, Y., Sakamoto, T., Kurogochi, M., Hinou, H., Shimizu, H., Ito, T., Naruchi, K., Togame, H., Takemoto, H., Kondo, H., and Nishimura, S. I. (2009) Functional neoglycopeptides: synthesis and characterization of a new class of MUC1 glycoprotein models having core 2-based O-glycan and complex-type N-glycan chains. *Biochemistry 48*, 11117–11133.

(9) Matsushita, T., Nagashima, I., Fumoto, M., Ohta, T., Yamada, K., Shimizu, H., Hinou, H., Naruchi, K., Ito, T., Kondo, H., and Nishimura, S. I. (2010) Artificial golgi apparatus: Globular protein-like dendrimer facilitates fully automated enzymatic glycan synthesis. *J. Am. Chem. Soc. 132*, 16651–16656.

(10) Naruchi, K., and Nishimura, S. I. (2011) Membrane-bound stable glycosyltransferases: Highly oriented protein immobilization by a C-terminal cationic amphipathic peptide. *Angew. Chemie - Int. Ed. 50*, 1328–1331.

(11) Mathieux, N., Paulsen, H., Meldal, M., and Bock, K. (1997) Synthesis of glycopeptide sequences of repeating units of the mucins MUC 2 and MUC 3 containing oligosaccharide

side-chains with core 1, core 2, core 3, core 4 and core 6 structure. *J. Chem. Soc. Perkin Trans. 1* 2359–2368.

(12) Meinjohanns, E., Meldal, M., Schleyer, A., Paulsen, H., and Bock, K. (1996) No Title 985–993.

(13) Matsushita, T., Hinou, H., Kurogochi, M., Shimizu, H., and Nishimura, S. I. (2005) Rapid microwave-assisted solid-phase glycopeptide synthesis. *Org. Lett. 7*, 877–880.

(14) Garcia-Martin, F., Hinou, H., Matsushita, T., Hayakawa, S., and Nishimura, S.-I. (2012) An efficient protocol for the solid-phase synthesis of glycopeptides under microwave irradiation. *Org. Biomol. Chem. 10*, 1612.

CHAPTER 3

Micro Array Studies

3-1. Introduction

Micro array[1, 2] is a biological technique used for detection of a cell surface antigen by its specific reactivity with an antibody. It is a high throughput method which can used in different levels such as DNA microarrays, Protein microarrays, Peptide microarrays, Tissue microarrays, Cellular microarrays, Chemical compound microarrays, Antibody microarrays, Carbohydrate arrays, Phenotype microarrays, Reverse Phase Protein Microarrays etc. Our laboratory has used this technique to check the interaction levels between series of synthesised MUC1 glycopeptide against mAb [3-5] (KL6[6], DF3[7] and SM3[8]). In this technique synthesised glycopeptides are immobilised onto the glass slide and are treated with afore mentioned mAb's [9, 10]. Such type of 20mer glycopeptides library has not been checked for interaction yet.

In this chapter synthesis of 23 glycopeptides for micro array studies and 6 glycopeptides for NMR studies were undertaken.

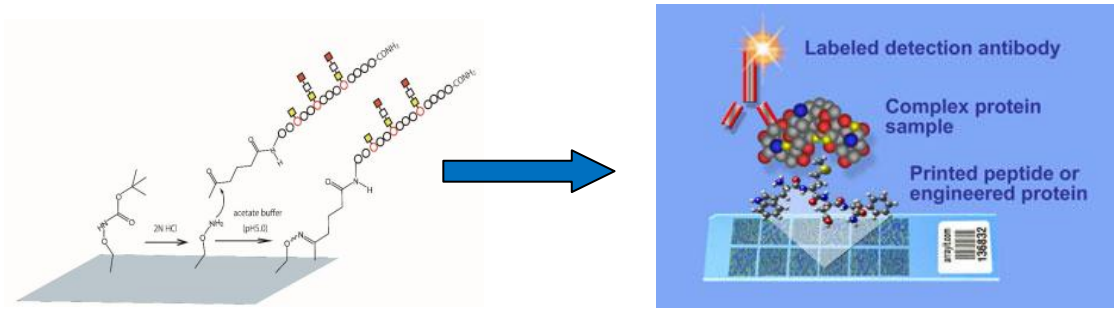


Figure 1: Schematic representation of Micro Array.

3-2. Results and Discussion

The micro array results of epitope screening depending on the basis of signal intensities were determined by clear specific interaction with particular antibodies. Three different monoclonal antibodies (KL6, SM3, and DF3) were used against the microwave synthesised series of MUC1 glycopeptides. Initially 19 variants of MUC1 glycopeptides were synthesised. Three different concentrations of synthesised compounds were printed on the array slides (0.5μM, 5.0μM and 50μM).

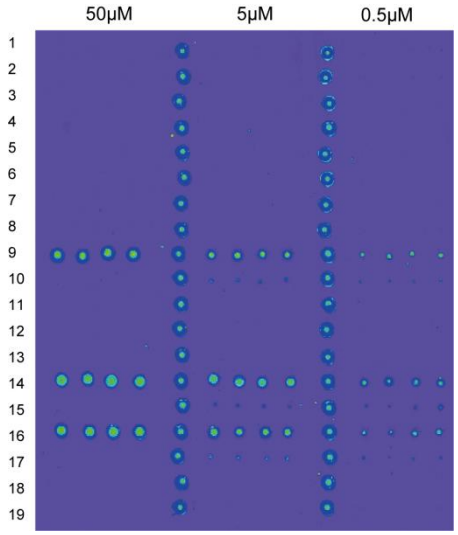
Micro array studies against antibodies:

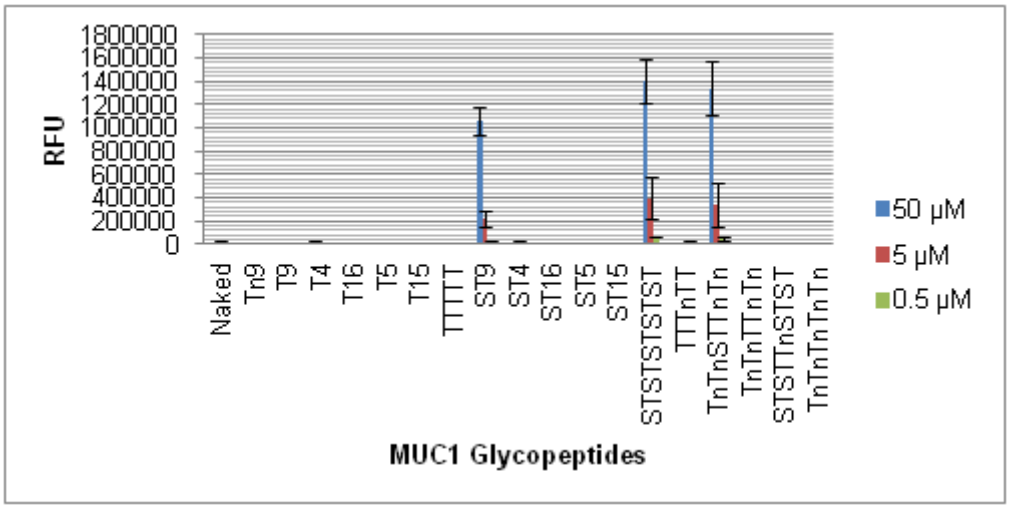
3-2.1. KL6:

Focussing on the KL6 antibody Human lung adenocarcinoma-derived cell line (VMRC-LCR) was used with no dilution was done from the kit. Analysing the graphical results it was clearly seen that the KL6 interaction was highly specific to the MUC1 glycopeptide sequence which comprised of a minimal epitope region i.e. core 1 sialylated T glycans moiety present at the PDTR epitope (glycopeptide ST9).

The MUC1 glycopeptide containing core 1 sialylated T antigen on all the Ser/Thr amino acids present also showed interaction with KL6 monoclonal antibody (glycopeptide STSTSTSTST). Also the glycopeptide bearing core 1 sialylated glycan on the Thr amino acid of the PDTR motif with Tn antigen on the neighbouring glycosylation site showed interaction with KL6 monoclonal antibody. Observing other compounds, no interaction was observed interaction. Even Glycopeptides bearing core 1 sialylated T at all *O* glycosylation sites except

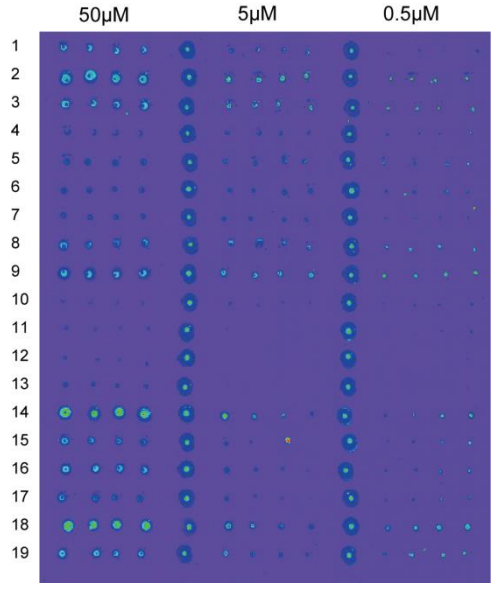
Thr amino acids at the PDTR sequence glycosylated by Tn antigen showed no interaction against the KL6 mAb as observed by micro array image and corresponding data analysis this results indicate the evidence that multiple *O* glycosylation at other ser/thr residues with Tn or ST antigens do not influence the interaction of anti KL6 mAb with the epitope region.

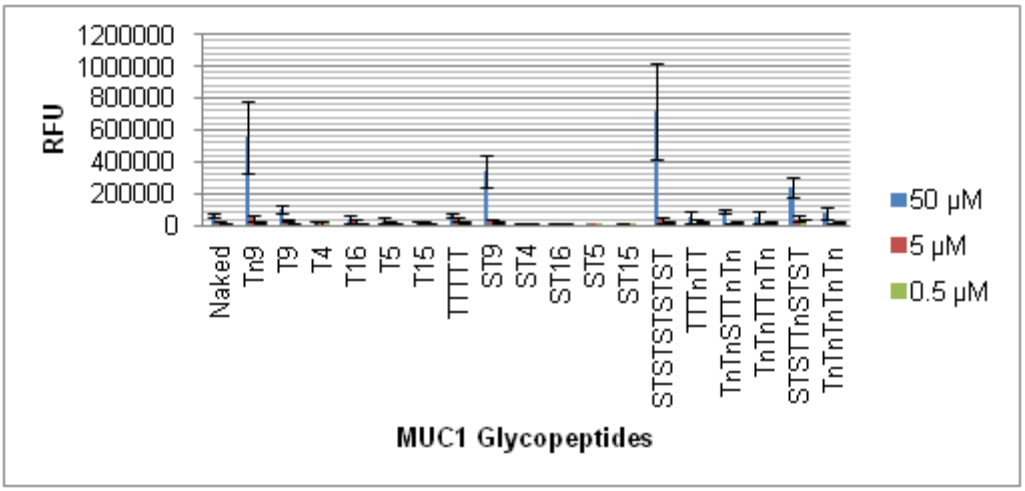




3-2.2. DF3:

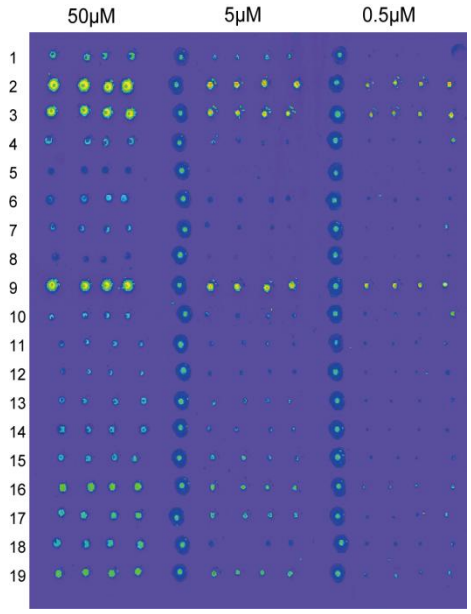
DF3, mAb membrane enriched fractions of human metastatic breast carcinoma. In case of mAb, it was revealed that the broad binding characteristics of DF3 towards PDTR region are affected strongly by *O* glycosylation states at other ser/thr residues. Because as observed the binding affinity of DF3 with most MUC1 fragments used herein were reduced drastically when four other *O*-glycosylation sites are occupied by Tn (glycopeptides 16, 17, and 19) or T antigen (glycopeptides 8 and 15). However, glycopeptides 14 and 18 modified with ST antigen at these four amino acid residues allowed exceptionally for the recognition by DF3.

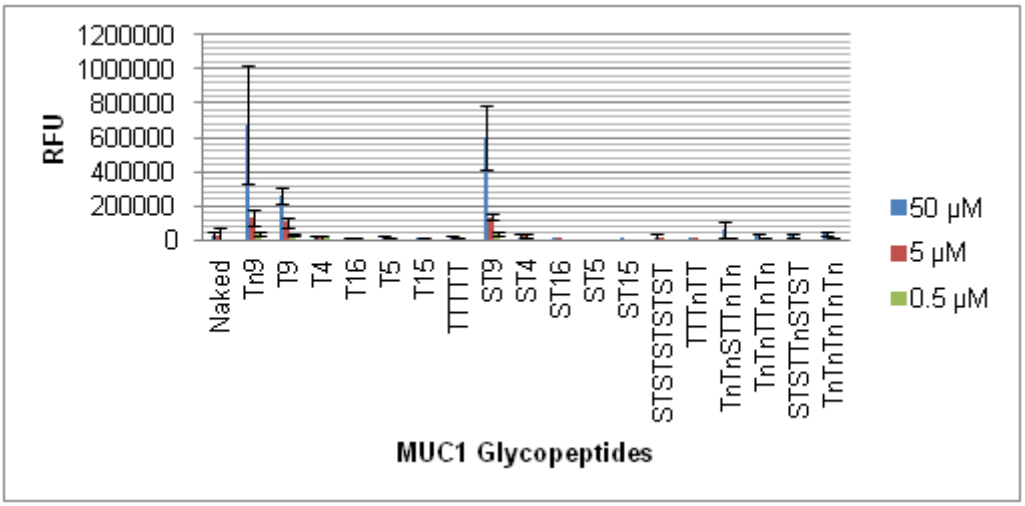




3-2.3. SM3:

Deglycosylated purified milk mucin mAb were also tested against the series of glycopeptides synthesised. SM3 mAb interaction results showed binding affinity against MUC1 glycopeptides that were single glycosylated at the PDTR motif by either Tn antigen, T antigen or Sialyl T antigen. It was observed that multiple *O* glycosylation at adjacent ser/thr residue disturb significantly the interaction between SM3 and glycopeptides 14-19 in a glycoform dependent manner when compared with glycopeptides 2, 3 and 9.





Therefore concluding micro array results with series A synthesised glycopeptides it clearly demonstrated for the first time that the *O*-glycosylation states around the immunodominant PDTR motif in the MUC1 tandem repeats influence strongly the binding potency and profile of the two important anti-MUC1 mAb's, DF3 and SM3. However, it is also important to note that the molecular mechanisms of the glycoside cluster effects on the antibody recognition of the PDTR motif might be different between DF3 and SM3, respectively.

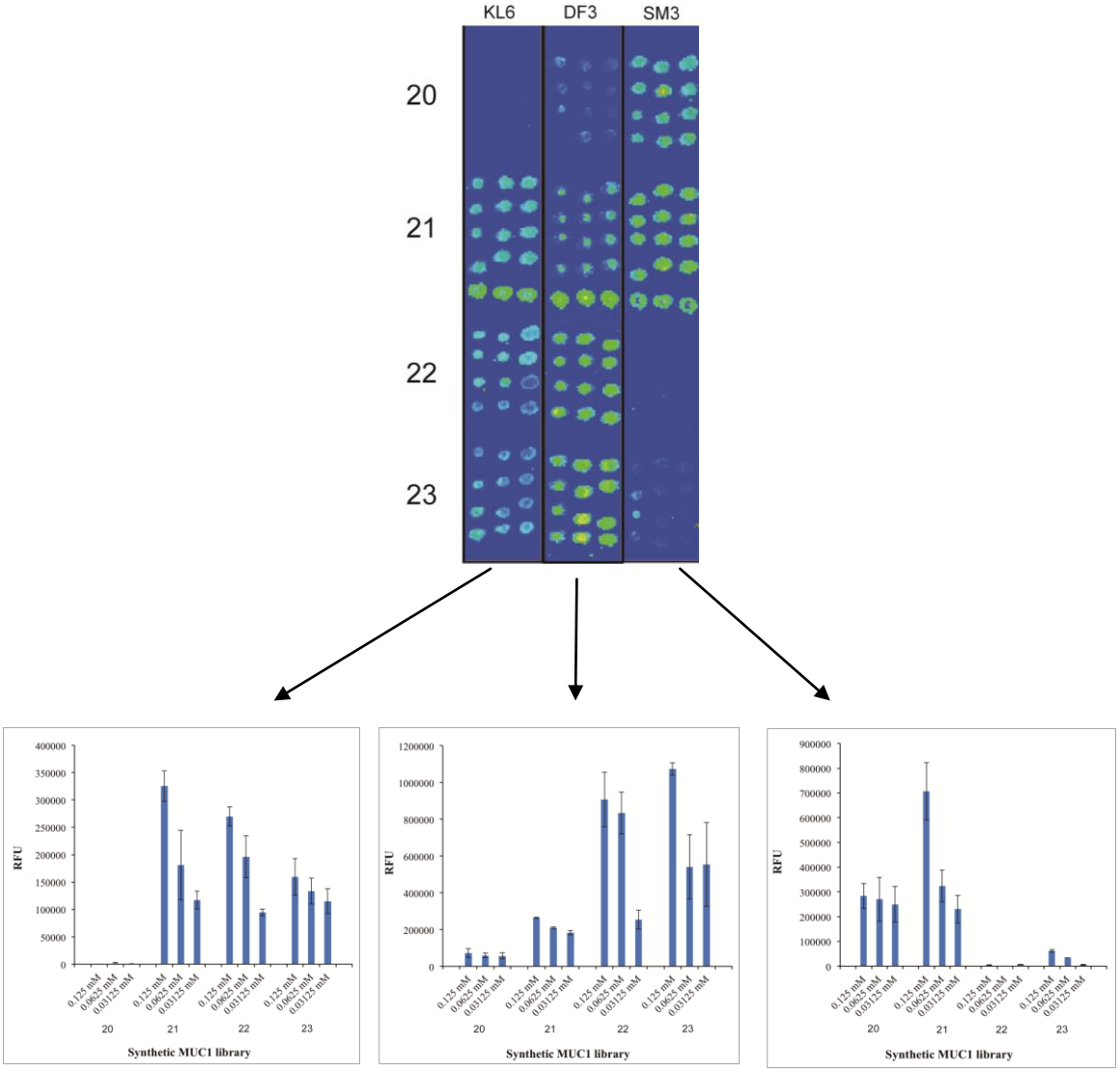
3-2.4 Core 2 MUC1 Glycopeptides Against mAb.

Further study was focussed on the heterogeneously glycosylated glycopeptides comprising of core 2 glycan structures. Here we used Naked peptide, MUC1 peptide comprising of Sialylated T antigen at the PDTR motif and two complex glycopeptides on MUC1 peptide i.e. Pro-Pro-Ala-His-Gal-Val-Thr-Ser-Ala-Pro-Asp-Thr-Arg-Pro-Ala-Pro-Gly-Ser-Thr-Ala-Pro-Pro-Ala sequence was synthesised[11, 12]. Similar micro array experiments were conducted for these 4 glycopeptides to check the binding affinity of KL6, DF3 and SM3. Three concentrations were used such as 0.125mM, 0.625mM and 0.03125mM.

Considering that naturally occurring MUC1 tandem repeating domains express much more complicated *O*-glycosylation states than those of the MUC1 A series tested,26,33 our attention was next directed to the binding features of anti-KL6/MUC1 mAbs with MUC1 B series compounds 20~23 to assess the effects of the *O*-glycosylation states containing more complicated and heterogeneous *O*-glycoforms at the neighbouring Ser/Thr residues on the antibody recognition of the essential epitope structure for the anti-KL6 mAb, the PDTRPAP region carrying ST antigen. As shown in Figure below the binding profiles of antibodies owards glycopeptides 20 and 21 were found to be mostly the same as the results obtained in cases for the compounds 1 and 9 (previous results), indicating that the difference of the location of the immunodominant PDTR motif between the two MUC1 tandem repeating units

did not affect the interaction between these antibodies and this epitope region. Importantly, it was clearly demonstrated that multiple *O*-glycosylations at neighbouring Thr/Ser residues did not disturb this specific epitope recognition by anti-KL6 mAb even when MUC1 fragments were modified with sterically hindered core2 type pentasaccharide moieties (SC2) seen in cases for the glycopeptides 22 and 23. It is not surprising that multiple *O*-modification at Ser/Thr residues of the compounds 22 and 23 abrogated markedly the binding ability of SM3 toward this epitope region detected in the MUC1 fragments 20 and 21. As anticipated, the modifications at the neighbouring *O*-glycosylation sites of the PDTR region with SC2 and Tn antigen can interrupt drastically the interaction of SM3 with these MUC1 fragments. This result is quite similar to the binding profiles of SM3 obtained by means of the microarray displaying MUC1 A series compounds, suggesting that the binding of SM3 against MUC1 extracellular tandem repeats might be influenced significantly by the stearic hindrance rather than the conformational effects of the multiple *O*-glycosylation states in the MUC1 tandem repeats. Surprisingly, the affinity of DF3 directing the essential KL6-epitope region of the compound 21, PDTR having ST antigen, was enhanced markedly in the MUC1 fragments 22 and 23 bearing bulky SC2 19 moiety at two Ser residues within the tandem repeat. These results clearly indicate that the enhanced binding of DF3 to MUC1 fragments involving the KL6-epitope region is mediated only when they are modified with the sialic acid-containing glycoforms, ST (glycopeptides 14 and 18 of MUC1 A series) and SC2 moieties (glycopeptides 22 and 23 of MUC1 B series), respectively. Therefore, it was considered that the conformation of the proximal peptide backbone at the immunodominant PDTR motif may be altered by the *O*-glycosylation at two adjacent Ser residues to afford more preferred structure for the binding with DF3. However, these multiple glycosylation with ST or SC2 moiety did not influence the binding mode of the anti-KL6 mAb with the MUC1 fragments having an essential glycopeptidic epitope, Pro-Asp-

Thr[Neu5Ac・(2→3)Gal・(1→3)GalNAc・・→]-Arg-Pro-Ala-Pro, such as compounds 9, 14, 16, 21, 22, and 23. These results show the evidence that anti-KL6 mAb recognises specifically the glycopeptidic neo-epitope while the binding affinity of DF3 with this immunodominant motif is greatly dependent on the stability of the peptide backbone structure, an extended *trans*-like conformation, rather than the glycoforms linked to the Thr residue in the PDTR motif.



3-3. Experimental Section

Micro Array:

Commercially available reagents and solvents were used without further purification. Microarray slides (75 x 25 x 1 mm) and Hybridization covers (60 x 25 x 0.7 mm) was supplied from Sumitomo Bakelite Co., Ltd. (Tokyo, Japan). Slides were initially immersed in 2N hydrochloric acid (HCl) solution overnight due to the presence of amino-oxy functionalised methacrylic copolymer with phosphoryl choline group. After overnight incubation the slides were washed in H_20 for 1min twice each and dried by centrifugation. Synthesised glycopeptides were dissolved in 25mM ACOH-Pyr, 0.0025%, triton X100 was printed onto the array slides at 3 different concentrations (50mM, 5mM, 0.5mM). The synthesised compounds were printed onto the array slides using Omni grid Micro 9 Digilab Inc., Marlborough, USA) with a 0.8 mm pitch using chip marker TM CMP6 micro spotting pins(200µm spot diameter, Arraylt Corporation, Sunnyvale, CA, USA). Each compound was printed in quadruplicate with 0.8mm distance one with 2.4mm distance between different concentrations having marker in-between different concentrations. Different compounds were printed one below the other with 0.8mm distance. After printing of slides they were incubated for 90min at 80°C to complete the oxime bond formation and later washed with H_20 for 1 min. Next, succinic anhydride (10mg/ml) was used to cap the non reactive amino groups at RT for 4h, followed by washing with H_20 2X 1 min each and dried by centrifugation. Next the slides were used for binding assay with 3 different anti MUC1 MAb (Kl6 labelled with alkaline phosphatase from Lumipulse® KL-6 Eisai was acquired from Sanko Junyaku Co., Ltd. (Tokyo, Japan), DF3 (Covance Research Products, Inc. (Shirley, MA, USA), SM3(Cruz Biotechnology, Inc (Santa Cruz, CA, USA). Same reaction buffer was used for microarray studies. Printed slides were placed in reaction vessel, relatively high humidity was maintained. Initially, hybridisation covers were mounted onto the slides and 70-80µl of anti

MUC1 mAb solution presented was carefully placed so that the gaps between the slide and covered by hybridization cover slip and made sure that no bubbles were formed. After adding the solution the slides were incubated for 2hrs at rt. Hybridization covers were carefully removed and the slides were subjected to washing with washing buffer 3 X 1 min each and the with H2O 1 X 1 min and dried by centrifugation. Secondary antibody used for analysing and identification of antigen antibody binding was performed using FluoroLink™ Cy™3-labeled goat anti-mouse IgG (H+L) from Amersham Biosciences (Buckinghamshire, UK) at 1μg/ml concentration in reaction buffer. Secondary antibody incubation was also performed using similar hybridization covers. Since the Cy3 labelled is fluorescent dye incubation was performed in dark at RT for 1 hour and later washed with washing buffer 3 X 1 min each. Quantification Analysis of binding studies was succeeded by Fluorescence images of microarray slides measured at 10 μm resolution by Typhoon Trio plus variable mode imager (GE Healthcare) with green laser (532 nm) and 580 BP 30 filter at PMT voltage of 600 V and normal sensitivity. Histogram was constructed showing the level of antigen antibody interaction analyzed using ArrayVision™ software version 8.0 (GE Healthcare) [13].

Error bars were included showing standard deviation for each interaction between mAb and glycopeptide. The array slides were stored at 4°C. The used reaction buffer was prepared by a combination of 50mM Tris HCl,100Mm Nacl,1mM CaCl2, MnCl2,MgCl2, 0.05% tween 20,0.1% BSA, at pH7.4, and that of washing buffer was prepared with a combination of 50mM Tris –HCL,100mM NaCl, 1mM CaCl2,MnCl2,MgCl2, 0.05% TRITON x100,pH7.4.

3-4. Conclusion

In this chapter, micro array revealed interesting results which trigger to study the conformational analysis using NMR.

In case of micro array with mAb KL6 it was confirmed that interaction is possible only with glycopeptides bearing minimal sialylated core 1 structure at Thr amino acid at the crucial epitope PDTR motif even though the neighbouring *O* glycosylation sites were glycosylated by any type of glycans.

In case of DF3 mAb it was observed clearly that single glycosylated by Tn, T or sialylated T antigen at the PDTR motif showed binding interaction. Heavily glycosylated Glycopeptide bearing sialylated T antigen at all glycosylated sites also shows strong binding interaction whereas other multiple glycosylated glycopeptides showed decrease binding interaction against DF3mAb.

SM3 mAb showed specific interaction with all single glycosylated glycopeptides at the PDTR epitope and with naked peptide. Whereas no interaction was observed with singly glycosylated glycopeptides at any other *O* glycosylation sites than Thr of PDTR motif and also no interaction with glycopeptides glycosylated at multiple sites.

3-5. References

(1) Tse-Wen Chang. (1983) Binding of cells to matrixes of distinct antibodies coated on solid surface. *J. Immunol. Methods 65*, 217–223.

(2) Wang, D., Carroll, G. T., Turro, N. J., Koberstein, J. T., Kovác, P., Saksena, R., Adamo, R., Herzenberg, L. a, Herzenberg, L. a, and Steinman, L. (2007) Photogenerated glycan arrays identify immunogenic sugar moieties of Bacillus anthracis exosporium. *Proteomics 7*, 180–4.

(3) Matsushita, T., Takada, W., Igarashi, K., Naruchi, K., Miyoshi, R., Garcia-Martin, F., Amano, M., Hinou, H., and Nishimura, S. I. (2014) A straightforward protocol for the preparation of high performance microarray displaying synthetic MUC1 glycopeptides. *Biochim. Biophys. Acta - Gen. Subj. 1840*, 1105–1116.

(4) Garcia-Martin, F., Matsushita, T., Hinou, H., and Nishimura, S.-I. (2014) Fast Epitope Mapping for the Anti-MUC1 Monoclonal Antibody by Combining a One-Bead-One-Glycopeptide Library and a Microarray Platform. *Chem. - A Eur. J. 20*, 15891–15902.

(5) Burchell, J. M., Mungul, a, and Taylor-Papadimitriou, J. (2001) O-linked glycosylation in the mammary gland: changes that occur during malignancy. *J. Mammary Gland Biol. Neoplasia 6*, 355–64.

(6) Ohyabu, N., Hinou, H., Matsushita, T., Izumi, R., Shimizu, H., Kawamoto, K., Numata, Y., Togame, H., Takemoto, H., Kondo, H., and Nishimura, S. I. (2009) An essential epitope of anti-MUC1 monoclonal antibody KL-6 revealed by focused glycopeptide library. *J. Am. Chem. Soc. 131*, 17102–17109.

(7) Friedman, E. L., Hayes, D. F., and Kufe, D. W. (1986) i 5189–5195.

(8) Dokurno, P., Lally, J. M., Bates, P. a, Taylorpapadimitriou, J., Band, H. a, Snary, D., and Freemont, P. S. (1997) Crystallization of an antitumour antibody sm3 complexed with a peptide epitope. *Acta Crystallogr. Sect. D-Biological Crystallogr. 53*, 780–781.

(9) Marcelo, F., Garcia-Martin, F., Matsushita, T., Sardinha, J., Coelho, H., Oude-Vrielink, A., Koller, C., André, S., Cabrita, E. J., Gabius, H.-J., Nishimura, S.-I., Jiménez-Barbero, J., and Cañada, F. J. (2014) Delineating Binding Modes of Gal/GalNAc and Structural Elements of the Molecular Recognition of Tumor-Associated Mucin Glycopeptides by the Human Macrophage Galactose-Type Lectin. *Chem. - A Eur. J. 20*, 16147–16155.

(10) Marcelo, F., Cañada, F. J., André, S., Colombo, C., Doro, F., Gabius, H.-J., Bernardi, A., and Jiménez-Barbero, J. (2012) α-N-Linked glycopeptides: conformational analysis and bioactivity as lectin ligands. *Org. Biomol. Chem. 10*, 5916.

(11) Matsushita, T., Hinou, H., Fumoto, M., Kurogochi, M., Fujitani, N., Shimizu, H., and Nishimura, S. I. (2006) Construction of highly glycosylated mucin-type glycopeptides based on microwave-assisted solid-phase syntheses and enzymatic modifications. *J. Org. Chem. 71*, 3051–3063.

(12) Naruchi, K., Hamamoto, T., Kurogochi, M., Hinou, H., Shimizu, H., Matsushita, T., Fujitani, N., Kondo, H., and Nishimura, S. I. (2006) Construction and structural characterization of versatile lactosaminoglycan-related compound library for the synthesis of complex glycopeptides and glycosphingolipids. *J. Org. Chem. 71*, 9609–9621.

(13) Coelho, H., Matsushita, T., Artigas, G., Hinou, H., Cañada, F. J., Lo-Man, R., Leclerc, C., Cabrita, E. J., Jiménez-Barbero, J., Nishimura, S.-I., Garcia-Martín, F., and Marcelo, F. (2015) The Quest for Anticancer Vaccines: Deciphering the Fine-Epitope Specificity of Cancer-Related Monoclonal Antibodies by Combining Microarray Screening and Saturation Transfer Difference NMR. *J. Am. Chem. Soc.* 150922162735006.

CHAPTER 4

NMR Studies

4-1. Introduction

NMR/ Nuclear Magnetic Resonance spectroscopy[1] is a technique which is mainly used to study the structure of molecules[2]. It is also used to study the interaction of various molecules, the dynamics of molecules and also the composition of mixture of biological or synthetic solutions. There are 2 types of NMR i.e. solid state NMR in which the solid samples are directly measured whose chemical shift is measured in KHz, Liquid phase NMR used a solvent system which is a reference and said to be the most predictable structural analysis method. In our experiments used solution state of 800MHz instrument to predict the conformational aspects of array of glycopeptides[3, 4]. The study included using the parameters like TOCSY (H-H), DQF COSY (C-C), and NOESY (NH^1 –H alpha[2)] with H_20 and D_2O solvent[5].

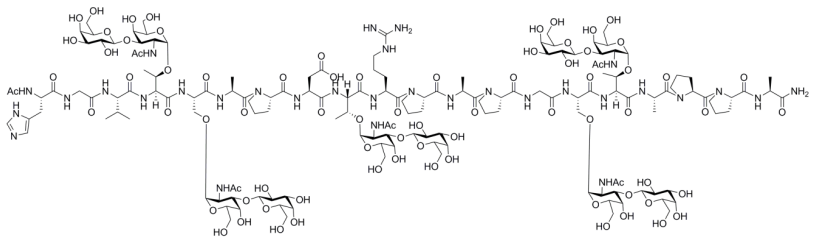
4-2. Results and Discussion

NMR Results:

Glycopeptides structural relationship between each other was determined using 800MHz NMR spectra which was recorded at 273K and at pH 5.5. The MUC1 compounds comprising of 20 amino acids had seen variation between the JαN values. Higher the JαN values show fixed structure in the glycopeptide. The glycopeptides used for structural determination has been explained separately as follows.

Glycopeptide 8

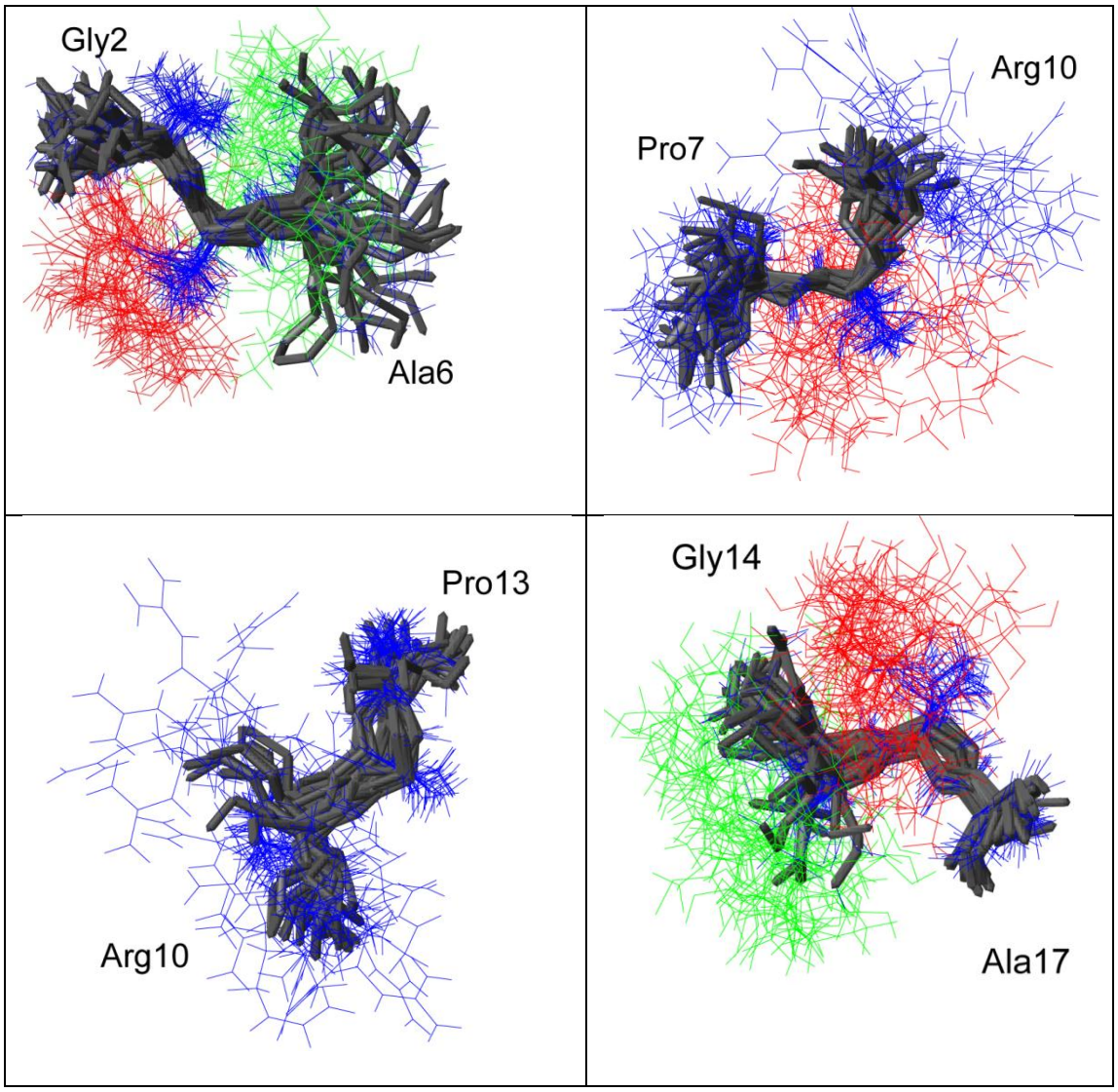
MUC1 20mer amino acid sequence bearing T antigen at all O glycosylation sites was analysed by NMR machine to determine the 3D structure.



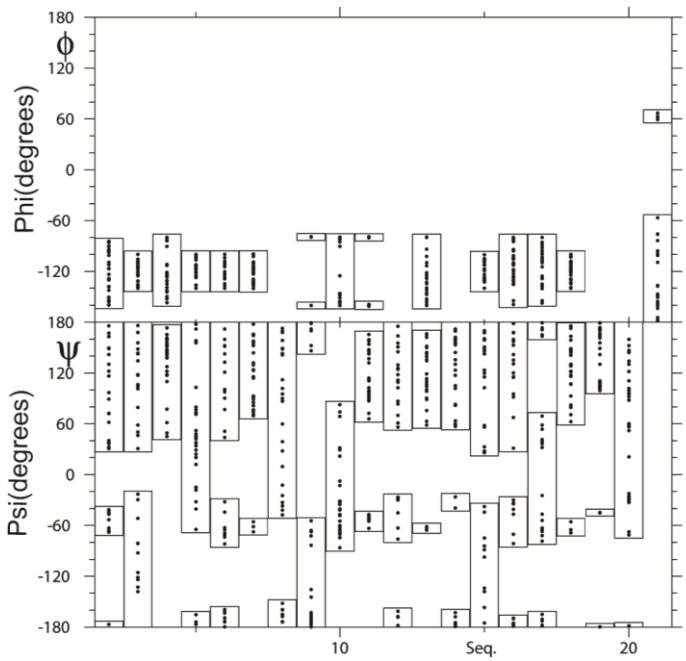
After all the spectra was derived from NMR analysis. Peaks were labelled corresponding using NOESY, COSY, TOCSY, HSQC spectra of each amino acid. Totally 163 restraints were labelled successfully and 84 dihedral angle restraints were labelled as mentioned in the table below which is further used to determine the 3D structure.

Once the labelling of peaks was completed 3D structure was obtained by CNS software. The number of peaks labelled inter residue, intra residue of amino acids and glycans are given as table below.

	Glycopeptide 8
NO of distance restraints	
total	163
intraresidue	
peptide	77
glycan	34
sequential	
peptide	41
glycan	11
medium range	
peptide	0
glycan	0
long range	
peptide	0
glycan	0
peptide to glycan	
within the same glycosylated residue	33
glycan on other peptide residues	12
No. of dihedral restraints	
total	84
peptide backbone	14
glycan	70

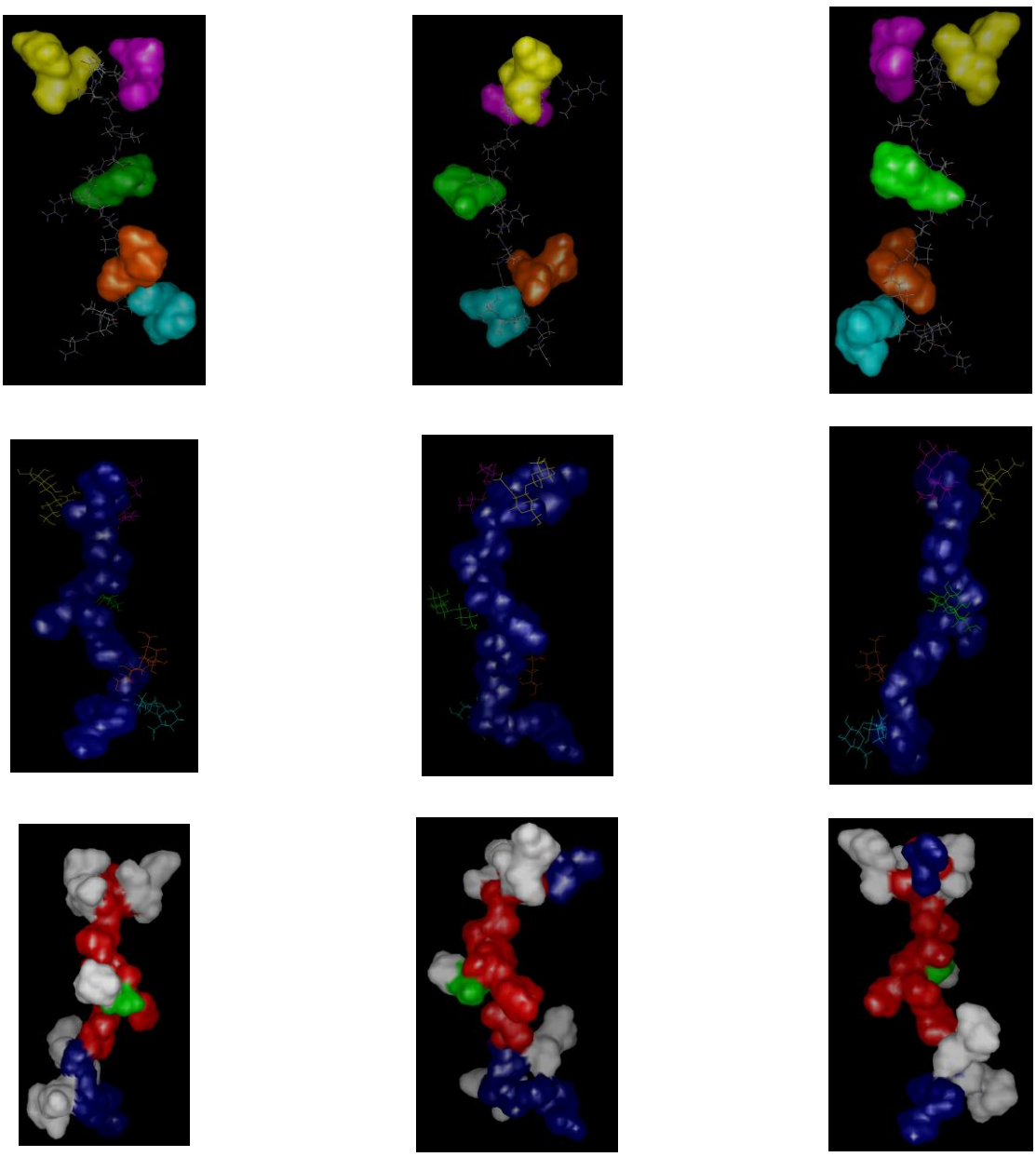


The best 30 calculated structures from NMR data. The ^{2}GVTSA^{6}, ^{7}PDTR^{10} , ^{10}RPAP^{13}, ^{14}GSTA^{17} regions are superimposed on backbone atoms ^{3}VTS^{5},^{8}DT^{9},^{11}PAP^{13}, ^{15}STA^{17} resp. The main chain of the Glycopeptide backbone is colored in Grey,side chain is blue, and GalNAc attached to Thr amino acid is red color, and the GalNAc attached to Ser amino acid is green color.



. Angle bar plots of the 30 obtained structures of MUC1 Glycopeptide 8 (black dots), for all residues prolines φ and ψ are given and for prolines ψ angles are given. Observing the 3D structure and the angle bar data it can be seen whether all 30 structures amino acid psi and phi angles are aligned fitting near to each other or distributed at different degrees.

Discussing the glycopeptide 8 interaction against the 3 antibodies used it was clearly seen that no binding affinity could be observed against KL6 antibody and SM3 mAb, whereas, in case of DF3 interaction weak interaction was seen when compared with KL6 and SM3 interaction.



Lowest energy structure in the 30 calculated models for the glycopeptide8. The left and the right structures are -90° and 90° rotations of the middle structures respectively.

A: Connolly surfaces of the Sialyl core 1 O glycans colour yellow at Thr 4, magenta at Ser5, green at Thr9, orange at Ser 15 and aqua at Thr 16.B:van der Waals Surfaces of the peptide backbone (blue) and five Sialyl core 1 branches indicated by wireframe representation: van der walls surface of the backbone peptide with the GalNAc moiety at the PDTR segment (GalNAc9) and the Connolly surfaces of the Sialyl core 1 based O glycans attached at the other glycosylation sited in which the peptide GVTSAPDTRPAP is red colour and other segments (PPAH and GSTA) are colour blue and the GalNAc 9 moiety is green colour. All images were generated by Mol feat version 5.0(FiatLux)

Analysing the statistical information about the glycopeptide the impact about binding and non binding with antibodies could be explained.

Statistical analysis:

	Glycopeptide 8
average potential energy(kcal/mol)a	
E_{total}	110.83± 10.09
E_{bond}	3.37 ± 0.41
E_{angle}	20.67 ± 2.79
E_{impr}	6.79± 0.71
E_{VDW}	12.3 ± 2.42
E_{NOE}	18.83 ± 1.92
E_{cdih}	0.06 ± 0.03
deviation from idealized geometry	
bond length(A°)	0.0026 ± 0.00021
bond angle (deg)	0.927 ± 0.051
Improper (deg)	0.557 ± 0.029
Average pairwise rmsd (A°)	
Backbone atoms	
Val 3 -Ser5	0.66 ± 0.35
Asp8 -Thr 9	0.36 ± 0.20
Pro11- Pro 13	0.73 ± 0.29
Ser15-Ala 17	0.76 ± 0.29
Heavy atoms	
Val 3 -Ser5	1.45 ± 0.52
Asp8 -Thr 9	1.24 ± 0.34
Pro11- Pro 13	1.40 ± 0.55
Ser15-Ala 17	1.47 ± 0.46
Val 3 -Ser5 and GalNAc 4	1.72 ± 0.60
Val 3 -Ser5 and GalNAc 5	2.45 ± 0.82
Val 3 -Ser5, GalNAc 4, GalNAc 5	2.75 ± 0.83
Asp8 -Thr 9 and GalNAc 9	2.03 ± 0.61
Ser15-Ala 17 and GalNAc 15	2.53 ± 0.58
Ser15-Ala 17 and GalNAc 16	2.29 ± 0.69
Ser15-Ala 17 , GalNAc 15 and GalNAc 16	3.01 ± 0.65

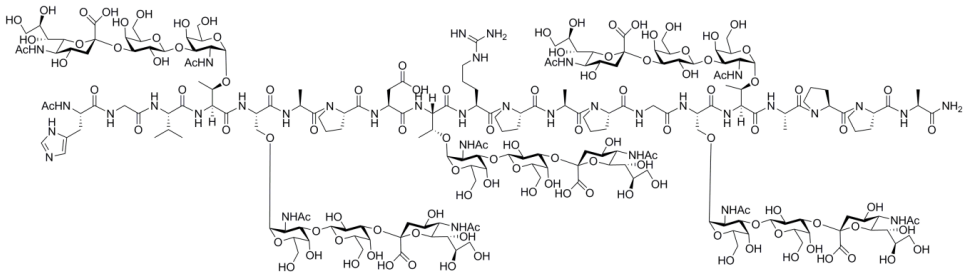
Dihedral angle data:

		Peptide 8
HIS1	ϕ	-127.8 ± 26.1
HIS1	χ1	-161.8 ± 85.8
HIS1	ψ	61.2 ± 84.4
GLY 2	ϕ	-120.4 ± 11.4
GLY 2	ψ	142.9 ± 89.0
VAL 3	ϕ	-113.6 ± 27.1
VAL 3	χ1	147.1 ± 89.1
VAL 3	ψ	137.8 ± 34.5
THR 4	ϕ	-116.0 ± 14.4
THR 4	χ1	14.3 ± 33.7
THR 4	ψ	50.0 ± 75.4
SER 5	ϕ	-120.4 ± 14.8
SER 5	χ1	-30.0 ± 74.4
SER 5	ψ	176.5 ± 81.2
ALA 6	ϕ	-120.3 ± 14.6
ALA 6	χ1	-3.1 ± 95.9
ALA 6	ψ	114.9 ± 66.1
PRO 7	ψ	141.2 ± 93.8
ASP 8	ϕ	-88.0 ± 28.0
ASP 8	χ1	-69.8 ± 73.2
ASP 8	ψ	-158.6 ± 45.8
THR 9	ϕ	-127.1 ± 37.9
THR 9	χ1	-1.9 ± 71.8
THR 9	ψ	-35.0 ± 46.1
ARG 10	ϕ	-132.8 ± 39.6
ARG 10	χ1	-112.4 ± 73.1
ARG 10	ψ	114.0 ± 86.4
PRO 11	ψ	123.4 ± 86.3
ALA 12	ϕ	-130.4 ± 29.17
ALA 12	χ1	33.3 ± 80.5
ALA 12	ψ	121.3 ± 85.0
PRO 13	ψ	136.7 ± 68.6
GLY 14	ϕ	-124.6 ± 13.8
GLY 14	ψ	148.6 ± 86.7
SER 15	ϕ	-111.1 ± 22.3
SER 15	χ1	-38.1 ± 75.9
SER 15	ψ	152.4 ± 87.7
THR 16	ϕ	-113.0 ± 22.0
THR 16	χ1	-15.9 ± 43.0
THR 16	ψ	-93.5 ± 94.0
ALA 17	ϕ	-116.6 ± 13.8
ALA17	χ1	42 ± 80.9
ALA 17	ψ	125.9 ± 62.3
PRO 18	ψ	139.8 ± 55.4
PRO 19	ψ	72.6 ± 75.6
20 ALA	ϕ	-141.6 ± 72.9
20 ALA	χ1	9.7 ± 80.6

The dihedral angle information shows each bond angle orientation due to which the binding studies could be discussed.

Glycopeptide 14:

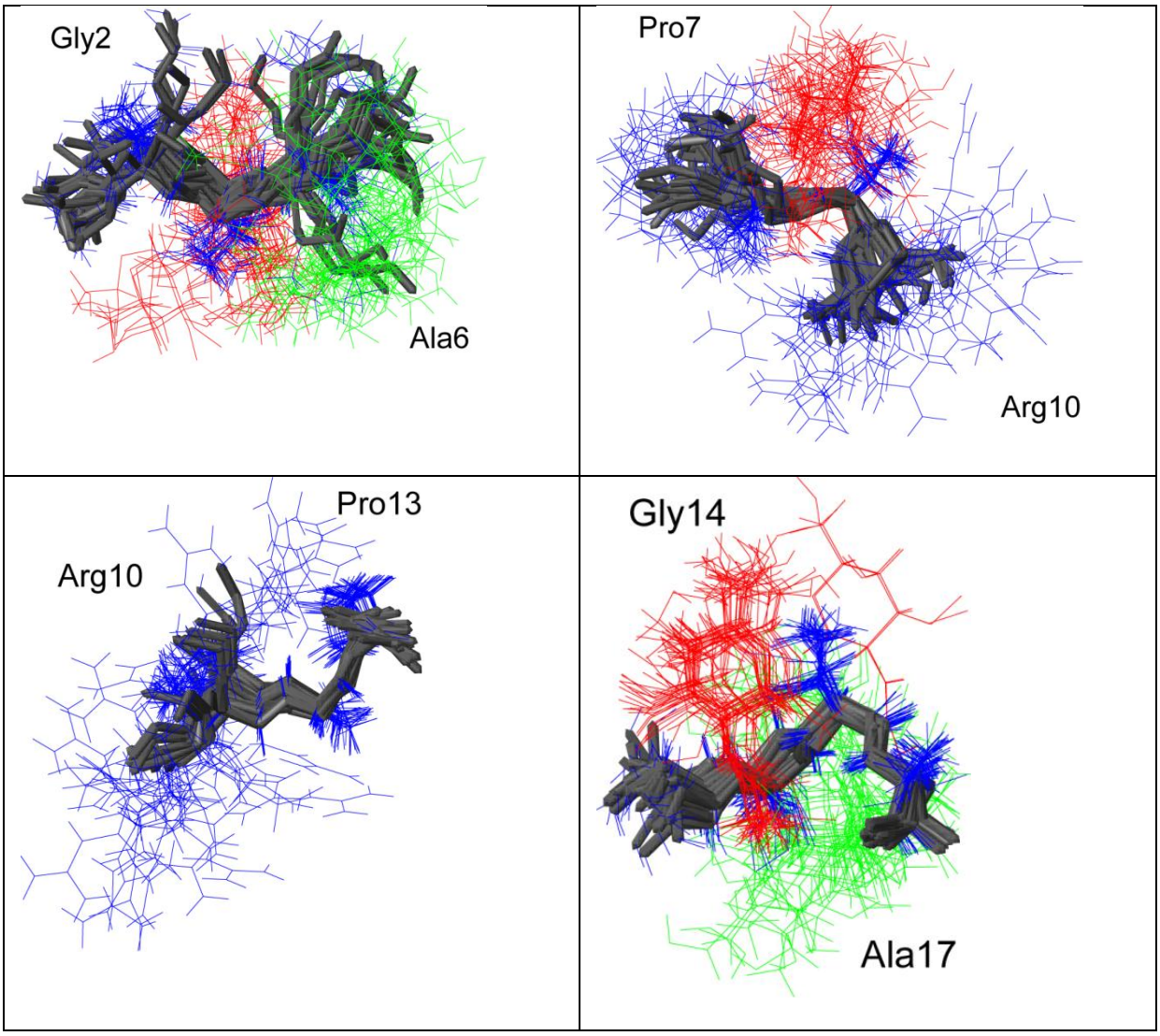
MUC1 20 mer amino acid sequence bearing sialylated T antigen at all O glycosylation sites was analysed by NMR machine to determine the 3D structure.



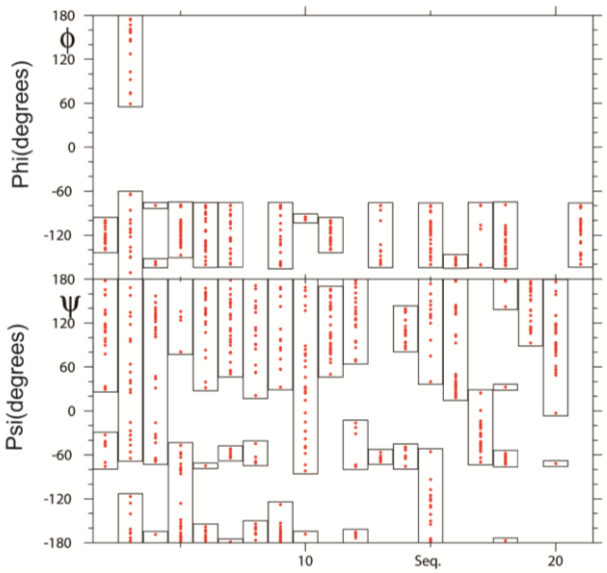
After all the spectra was derived from NMR analysis. Peaks were labelled corresponding using NOESY COSY TOCSY HSQC spectra of each amino acid. Totally 280 restraints were labelled successfully and 135 dihedral angle restraints were labelled as mentioned in the table below which is further used to determine the 3D structure.

Once the labelling of peaks was completed 3D structure was obtained by CNS software. The number of peaks labelled inter residue, intra residue of amino acids and glycans are given as table below.

	Glycopeptide 14
NO of distance restraints	
total	280
intraresidue	
peptide	76
glycan	141
sequential	
peptide	46
glycan	16
medium range	
peptide	1
glycan	0
long range	
peptide	0
glycan	0
peptide to glycan	
within the same glycosylated residue	32
glycan on other peptide residues	19
no of dihedral restraints	
total	
peptide backbone	135
glycan	15
	120

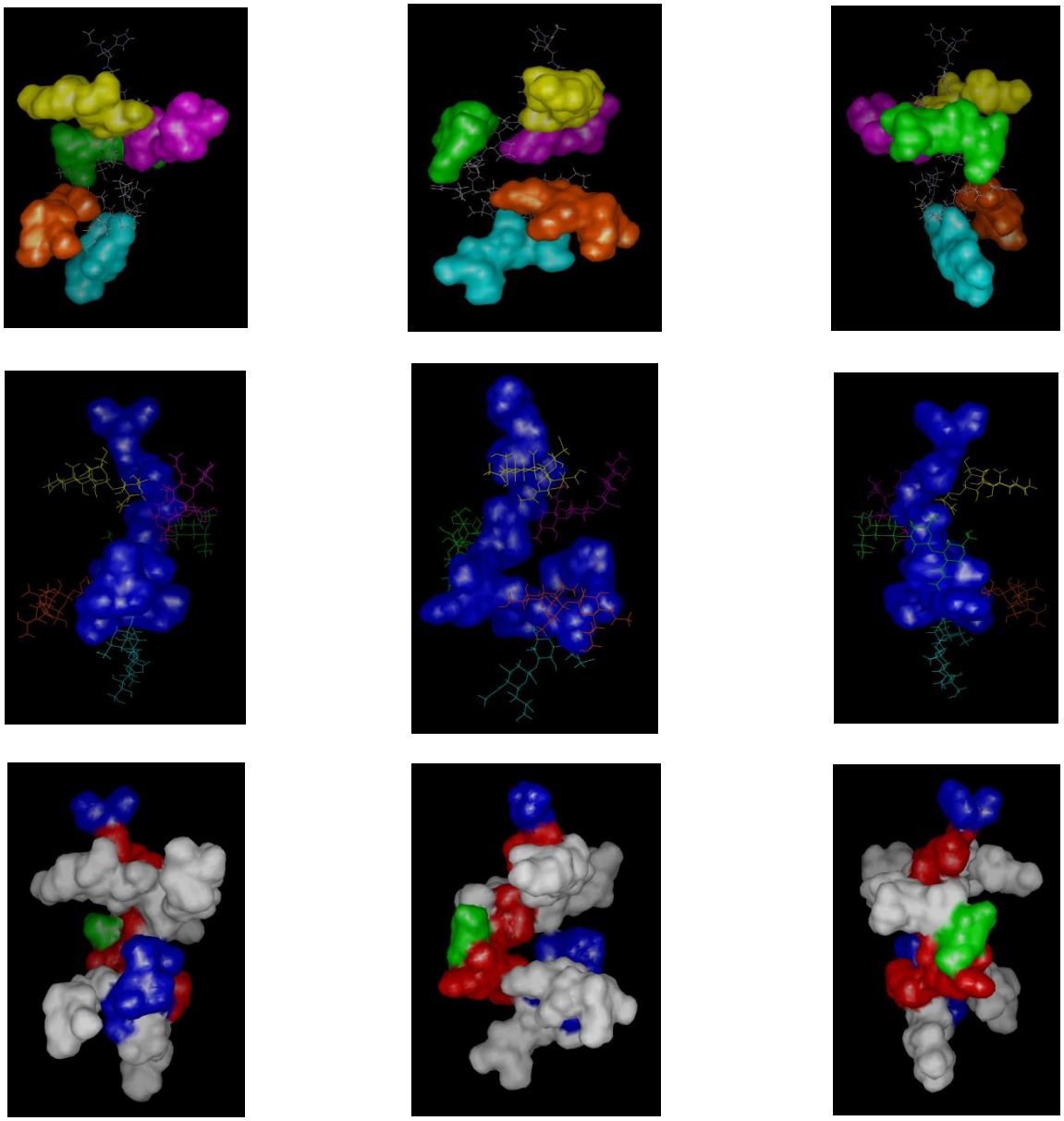


The best 30 calculated structures from NMR data. The ^{2}GVTSA^{6}, ^{7}PDTR^{10} , ^{10}RPAP^{13}, ^{14}GSTA^{17} regions are superimposed on backbone atoms ^{3}VTS^{5},^{8}DT^{9},^{11}PAP^{13}, ^{15}STA^{17} resp. The main chain of the Glycopeptide backbone is colored in Grey,side chain is blue, and GalNAc attached to Thr amino acid is red color, and the GalNAc attached to Ser amino acid is green color.



Angle bar plots of the 30 obtained structures of MUC1 Glycopeptide 1 4 (red dots), for all residues prolines φ and ψ are given and for prolines ψ angles are given. Observing the 3D structure and the angle bar data it can be seen whether all 30 structures amino acid psi and phi angles are aligned fitting near to each other or distributed at different degrees.

Discussing the glycopeptide 14 interaction against the 3 antibodies used it was clearly seen that strong binding affinity could be observed against KL6 antibody and DF3 mAb, whereas, in case of SM3 interaction no interaction was seen



Lowest energy structure in the 30 calculated models for the glycopeptide14. The left and the right structures are -90° and 90° rotations of the middle structures respectively.

A: Connolly surfaces of the Sialyl core 1 O glycans colour yellow at Thr 4, magenta at Ser5, green at Thr9, orange at Ser 15 and aqua at Thr 16.B:van der Waals Surfaces of the peptide backbone (blue) and five Sialyl core 1 branches indicated by wireframe representation: van der walls surface of the backbone peptide with the GalNAc moiety at the PDTR segment (GalNAc9) and the Connolly surfaces of the Sialyl core 1 based O glycans attached at the other glycosylation sited in which the peptide GVTSAPDTRPAP is red colour and other segments (PPAH and GSTA) are colour blue and the GalNAc 9 moiety is green colour. All images were generated by Mol feat version 5.0(FiatLux).

Analysing the statistical information about the glycopeptide the impact about binding and non binding with antibodies could be explained.

Statistical analysis:

	Glycopeptide 14
average potential energy(kcal/mol)a	
E_{total}	140.76± 20.33
E_{bond}	7.58 ± 1.05
E_{angle}	46.57 ± 6.45
E_{impr}	12.74 ± 1.02
E_{VDW}	25.92 ± 7.49
E_{NOE}	30.54 ± 5.89
E_{cdih}	1.4 ± 1.13
deviation from idealized geometry	
bond length(A°)	0.0037 ± 0.0003
bond angle (deg)	1.099 ±0.096
Improper (deg)	0.666 ± 0.022
Average pairwise rmsd (A°)	
Backbone atoms	
Val 3 -Ser5	0.68 ± 0.28
Asp8 -Thr 9	0.35 ± 0.26
Pro11- Pro 13	0.46 ± 0.21
Ser15-Ala 17	0.38 ± 0.18
Heavy atoms	
Val 3 -Ser5	1.46 ± 0.49
Asp8 -Thr 9	1.15 ± 0.32
Pro11- Pro 13	0.87 ± 0.35
Ser15-Ala 17	0.77 ± 0.34
Val 3 -Ser5 and GalNAc 4	1.82 ± 0.89
Val 3 -Ser5 and GalNAc 5	2.06 ± 0.61
Val 3 -Ser5, GalNAc 4, GalNAc 5	2.36 ± 0.78
Asp8 -Thr 9 and GalNAc 9	1.93 ± 0.85
Ser15-Ala 17 and GalNAc 15	1.31 ± 0.52
Ser15-Ala 17 and GalNAc 16	1.07 ± 0.86
Ser15-Ala 17 , GalNAc 15 and GalNAc 16	1.55 ± 0.75

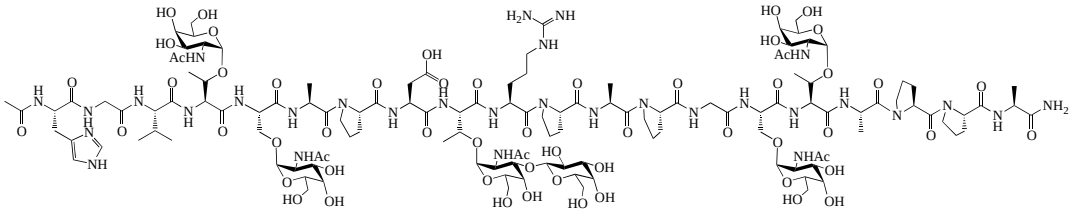
Dihedral angle data:

		Peptide 14
HIS1	ϕ	-123.2 ± 12.9
HIS1	χ1	-120.8 ± 98.3
HIS1	ψ	81.9 ± 82.3
GLY 2	ϕ	-167.1 ± 67.3
GLY 2	ψ	116.9 ± 102.5
VAL 3	ϕ	-113.0 ± 40.0
VAL 3	χ1	145.1 ± 83.2
VAL 3	ψ	111.1 ± 95.1
THR 4	ϕ	-110.7 ± 19.4
THR 4	χ1	42.0 ± 43.3
THR 4	ψ	-141.4 ± 60.0
SER 5	ϕ	-111.9 ± 28.8
SER 5	χ1	116.3 ± 92.0
SER 5	ψ	151.1 ± 51.1
ALA 6	ϕ	-113.6 ± 28.4
ALA 6	χ1	-22.9 ± 103.8
ALA 6	ψ	115.2 ± 86.8
PRO 7	ψ	146.8 ± 77.3
ASP 8	ϕ	-129.4 ± 34.8
ASP 8	χ1	-49.1 ± 84.3
ASP 8	ψ	157.4 ± 57.0
THR 9	ϕ	-97.6 ± 1.4
THR 9	χ1	-14.8 ± 25.7
THR 9	ψ	24.5 ± 79.4
ARG 10	ϕ	-119.4 ± 14.3
ARG 10	χ1	-98.5 ± 65.3
ARG 10	ψ	110.2 ± 34.1
PRO 11	ψ	138.4 ± 73.2
ALA 12	ϕ	-136.0 ± 35.0
ALA 12	χ1	-48.0 ± 93.2
ALA 12	ψ	-65.0 ± 2.3
PRO 13	ψ	109.7 ± 84.0
GLY 14	ϕ	-125.5 ± 25.3
GLY 14	ψ	-176.6 ± 61.8
SER 15	ϕ	-158.8 ± 3.1
SER 15	χ1	142.4 ± 65.3
SER 15	ψ	63.1 ± 53.7
THR 16	ϕ	-106.6 ± 38.1
THR 16	χ1	2.1 ± 42.8
THR 16	ψ	-35.8 ± 22.9
ALA 17	ϕ	-138.1 ±19.9
ALA17	χ1	0.4 ± 100.7
ALA 17	ψ	-81.0 ± 58.0
PRO 18	ψ	138.8 ± 27.6
PRO 19	ψ	94.2 ± 47.8
20 ALA	ϕ	-119.3 ± 24.4
20 ALA	χ1	169.1± 85.8

The dihedral angle information shows each bond angle orientation due to which the binding studies could be discussed.

Glycopeptide 17:

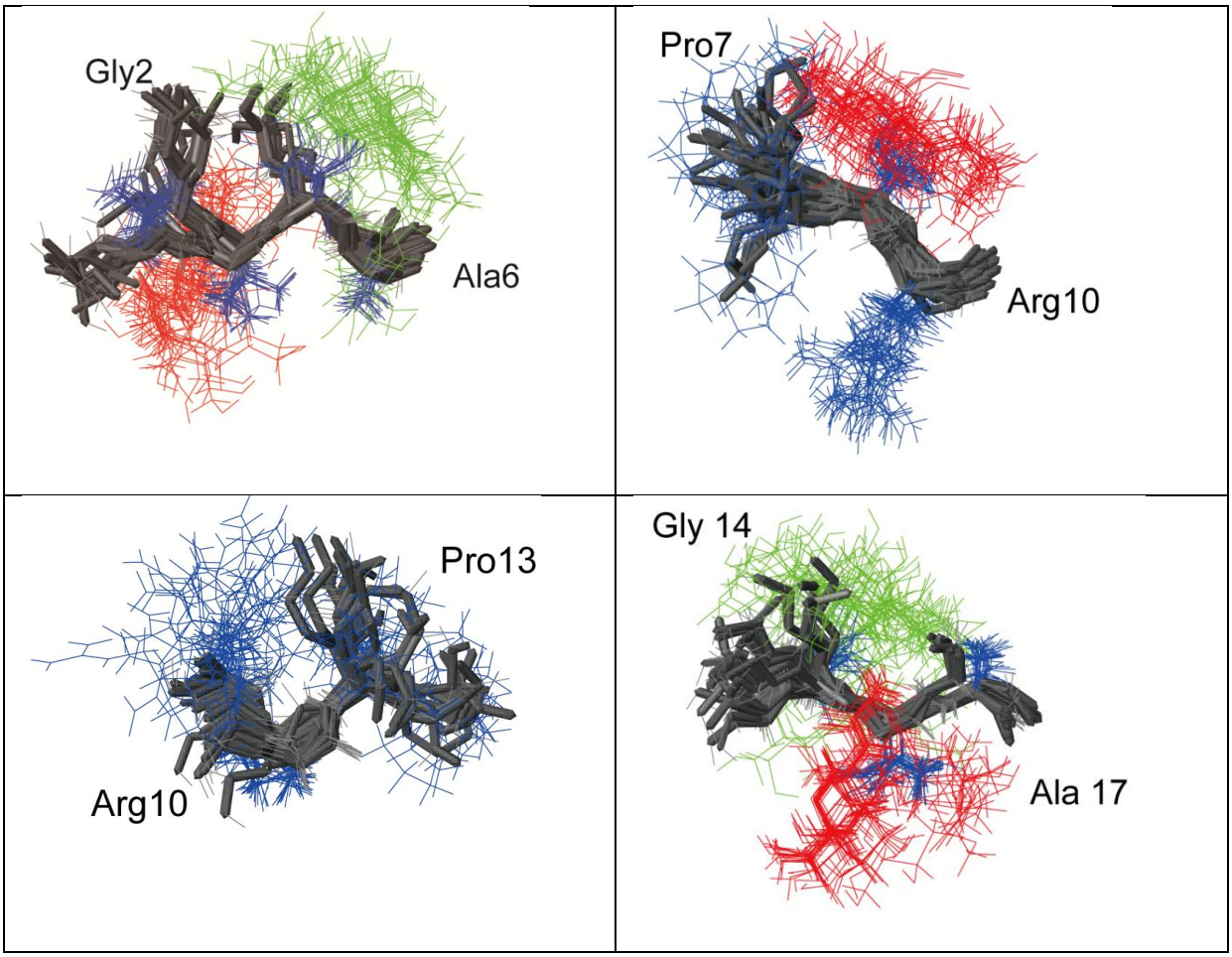
MUC1 20mer amino acid sequence bearing T antigen at thr of PDTR motif whereas neighbouring O glycosylation sites glycosylated by Tn antigen was analysed by NMR machine to determine the 3D structure.



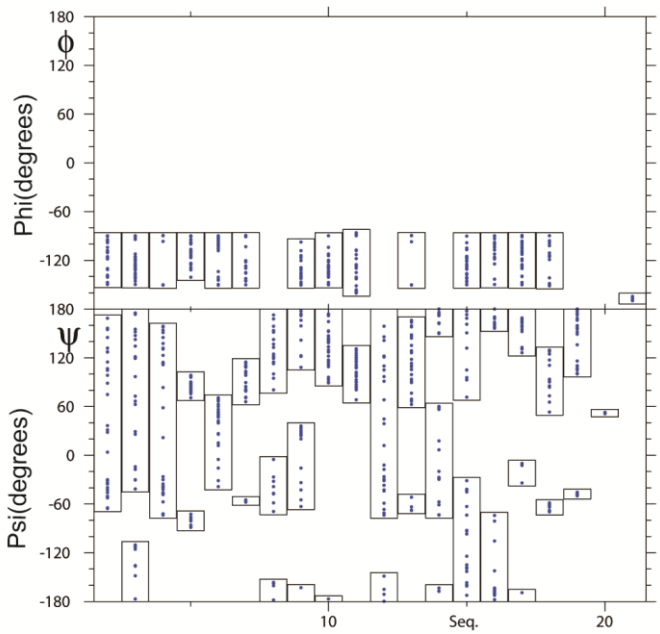
After all the spectra was derived from NMR analysis. Peaks were labelled corresponding using NOESY COSY TOCSY HSQC spectra of each amino acid. Totally 238 restraints were labelled successfully and 62 dihedral angle restraints were labelled as mentioned in the table below which is further used to determine the 3D structure.

Once the labelling of peaks was completed 3D structure was obtained by CNS software. The number of peaks labelled inter residue, intra residue of amino acids and glycans are given as table below.

	Glycopeptide 17
NO of distance restraints	
total	238
intraresidue	
peptide	111
glycan	76
sequential	
peptide	44
glycan	7
medium range	
peptide	0
glycan	0
long range	
peptide	0
glycan	0
peptide to glycan	
within the same glycosylated residue	28
glycan on other peptide residues	10
no of dihedral restraints	
total	
peptide backbone	62
glycan	14
	46

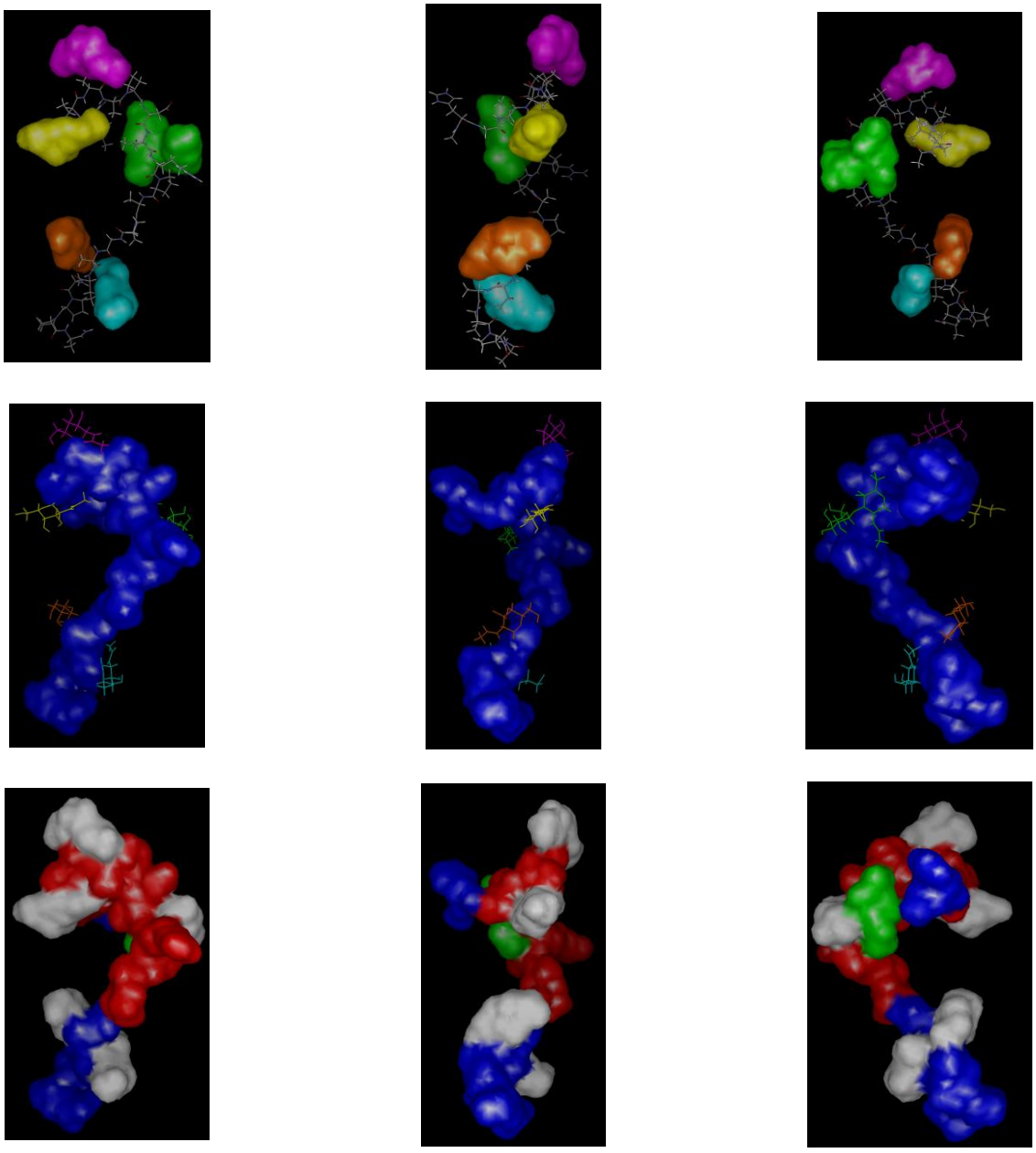


The best 30 calculated structures from NMR data. The ^{2}GVTSA^{6}, ^{7}PDTR^{10} , ^{10}RPAP^{13}, ^{14}GSTA^{17} regions are superimposed on backbone atoms ^{3}VTS^{5},^{8}DT^{9},^{11}PAP^{13}, ^{15}STA^{17} resp. The main chain of the Glycopeptide backbone is colored in Grey,side chain is blue, and GalNAc attached to Thr amino acid is red color, and the GalNAc attached to Ser amino acid is green color.



Angle bar plots of the 30 obtained structures of MUC1 Glycopeptide 17 (blue dots), for all residues prolines φ and ψ are given and for prolines ψ angles are given. Observing the 3D structure and the angle bar data it can be seen whether all 30 structures amino acid psi and phi angles are aligned fitting near to each other or distributed at different degrees.

Discussing the glycopeptide 17 interaction against the 3 antibodies used it was clearly seen no interaction or binding was observed with all the antibodies. Glycopeptide 17 was considered for structural analysis to reveal reason for no interaction with antibodies and understand the difference between the positive interaction and no interaction.



Lowest energy structure in the 30 calculated models for the glycopeptide17. The left and the right structures are -90° and 90° rotations of the middle structures respectively.

A: Connolly surfaces of the Sialyl core 1 O glycans colour yellow at Thr 4, magenta at Ser5, green at Thr9, orange at Ser 15 and aqua at Thr 16.B:van der Waals Surfaces of the peptide backbone (blue) and five Sialyl core 1 branches indicated by wireframe representation: van der walls surface of the backbone peptide with the GalNAc moiety at the PDTR segment (GalNAc9) and the Connolly surfaces of the Sialyl core 1 based O glycans attached at the other glycosylation sited in which the peptide GVTSAPDTRPAP is red colour and other segments (PPAH and GSTA) are colour blue and the GalNAc 9 moiety is green colour. All images were generated by Mol feat version 5.0(FiatLux)

Analysing the statistical information about the glycopeptide the impact about binding and non binding with antibodies could be explained.

Statistical analysis:

	Glycopeptide 17
average potential energy(kcal/mol)a	
E_{total}	99.78±4.24
E_{bond}	7.52±0.33
E_{angle}	31.26±1.31
E_{impr}	4.49±0.5
E_{VDW}	7.82±2.23
E_{NOE}	42.89±1.53
E_{cdih}	0.14±0.1
deviation from idealized geometry	
bond length(A°)	0.0044±0.0001
bond angle (deg)	1.038±0.299
Improper (deg)	0.474±0.029
Average pairwise rmsd (A°)	
Backbone atoms	
Val 3 -Ser5	0.60 ± 0.29
Asp8 -Thr 9	0.44 ± 0.28
Pro11- Pro 13	0.68 ± 0.25
Ser15-Ala 17	0.64 ± 0.46
Heavy atoms	
Val 3 -Ser5	1.39 ± 0.60
Asp8 -Thr 9	1.35± 0.42
Pro11- Pro 13	1.32 ± 0.47
Ser15-Ala 17	1.14 ± 0.56
Val 3 -Ser5 and GalNAc 4	2.07 ± 1.00
Val 3 -Ser5 and GalNAc 5	1.93 ± 0.87
Val 3 -Ser5, GalNAc 4, GalNAc 5	2.59 ± 1.08
Asp8 -Thr 9 and GalNAc 9	1.46 ± 0.49
Ser15-Ala 17 and GalNAc 15	1.70 ±0.57
Ser15-Ala 17 and GalNAc 16	1.54 ± 0.84
Ser15-Ala 17 , GalNAc 15 and GalNAc 16	2.20 ± 0.76

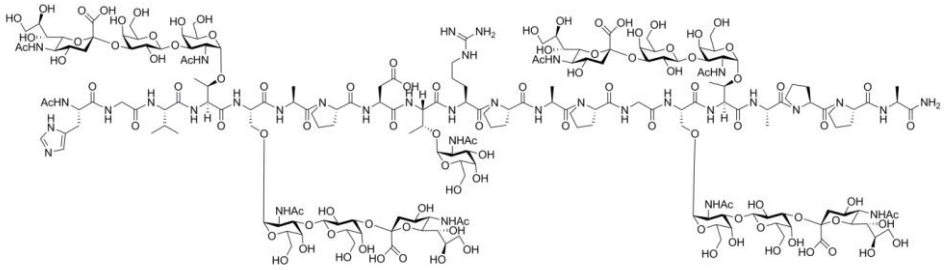
Dihedral angle data:

		Peptide 17
HIS1	ϕ	-117.0 ± 21.2
HIS1	χ1	-103.1 ± 62.3
HIS1	ψ	44.3 ± 80.1
GLY 2	ϕ	-129.0 ± 17.2
GLY 2	ψ	165.8 ± 95.4
VAL 3	ϕ	-122.5 ± 30.5
VAL 3	χ1	116.6 ± 79.4
VAL 3	ψ	25.2 ± 91.8
THR 4	ϕ	-103.5 ± 15.2
THR 4	χ1	8.7 ± 42.1
THR 4	ψ	79.8 ± 79.2
SER 5	ϕ	-106.6 ± 21.9
SER 5	χ1	173.7 ± 42.5
SER 5	ψ	42.5 ± 29.8
ALA 6	ϕ	-117.4 ± 26.8
ALA 6	χ1	-142.3 ± 84.7
ALA 6	ψ	95.4 ± 51.1
PRO 7	ψ	148.7 ± 91.6
ASP 8	ϕ	-137.4 ± 14.1
ASP 8	χ1	-114.4 ± 90.9
ASP 8	ψ	48.2 ± 78.0
THR 9	ϕ	-133.2 ± 14.8
THR 9	χ1	2.9 ± 52.3
THR 9	ψ	134.8 ± 24.2
ARG 10	ϕ	-138.9 ± 25.1
ARG 10	χ1	-119.8 ± 16.5
ARG 10	ψ	102.3 ± 17.4
PRO 11	ψ	8.6 ± 98.0
ALA 12	ϕ	-131.3 ± 28.8
ALA 12	χ1	-66.1 ± 82.4
ALA 12	ψ	121.1 ± 70.9
PRO 13	ψ	170.9 ± 103.4
GLY 14	ϕ	-125.9 ± 18.6
GLY 14	ψ	-149.6 ± 71.3
SER 15	ϕ	-99.2 ± 16.2
SER 15	χ1	43.3 ± 79.3
SER 15	ψ	-165.0 ± 37.9
THR 16	ϕ	-120.5 ± 20.2
THR 16	χ1	-50.1 ± 64.3
THR 16	ψ	159.5 ± 78.4
ALA 17	ϕ	-136.1 ± 21.1
ALA17	χ1	92.6 ± 106.8
ALA 17	ψ	1.5 ± 85.4
PRO 18	ψ	154.7 ± 87.0
PRO 19	ψ	51.7 ± 0.4
20 ALA	ϕ	-166.6 ± 1.3
20 ALA	χ1	42.9 ± 97.0

The dihedral angle information shows each bond angle orientation due to which the binding studies could be discussed.

Glycopeptide 18:

MUC1 20mer amino acid sequence bearing Tn antigen at thr of PDTR motif whereas neighbouring O glycosylation sites glycosylated by sialylated T antigen was analysed by NMR machine to determine the 3D structure.

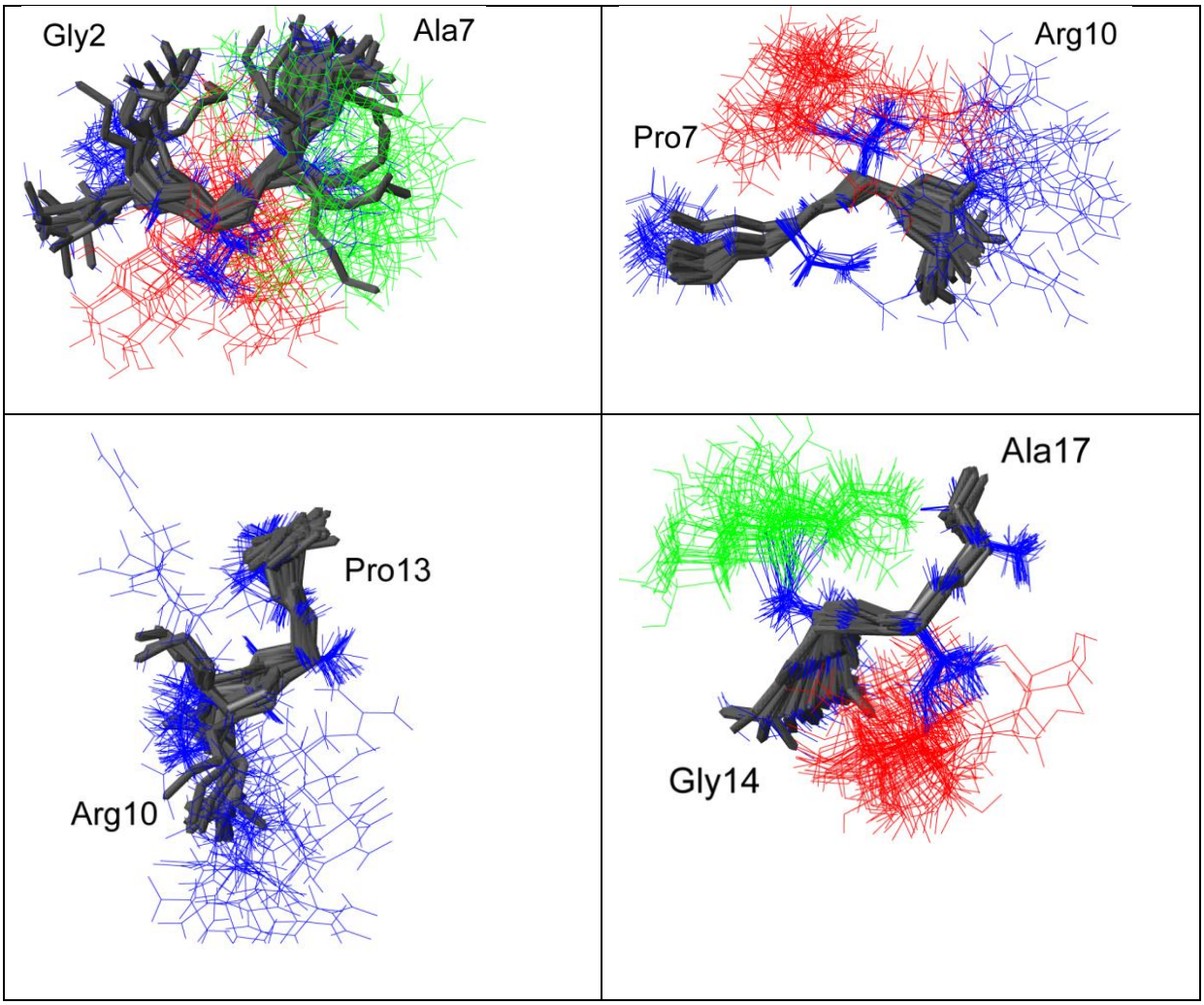


After all the spectra was derived from NMR analysis. Peaks were labelled corresponding using NOESY COSY TOCSY HSQC spectra of each amino acid. Totally 214 restraints were labelled successfully and 114 dihedral angle restraints were labelled as mentioned in the table below which is further used to determine the 3D structure.

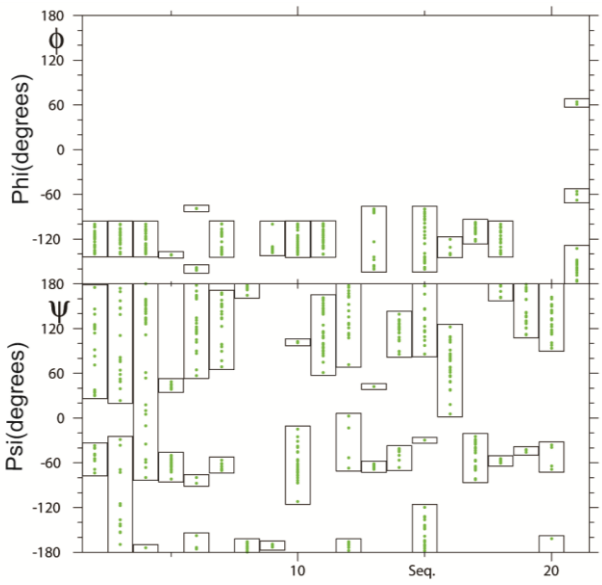
Once the labelling of peaks was completed 3D structure was obtained by CNS software. The number of peaks labelled inter residue, intra residue of amino acids and glycans are given as table below.

	Glycopeptide 18
NO of distance restraints	
total	214
intraresidue	
peptide	75
glycan	79
sequential	
peptide	43
glycan	16
medium range	
peptide	1
glycan	0
long range	
peptide	0
glycan	0
peptide to glycan	
within the same glycosylated residue	27
glycan on other peptide residues	11
no of dihedral restraints	
total	
peptide backbone	114
glycan	14
	100

.

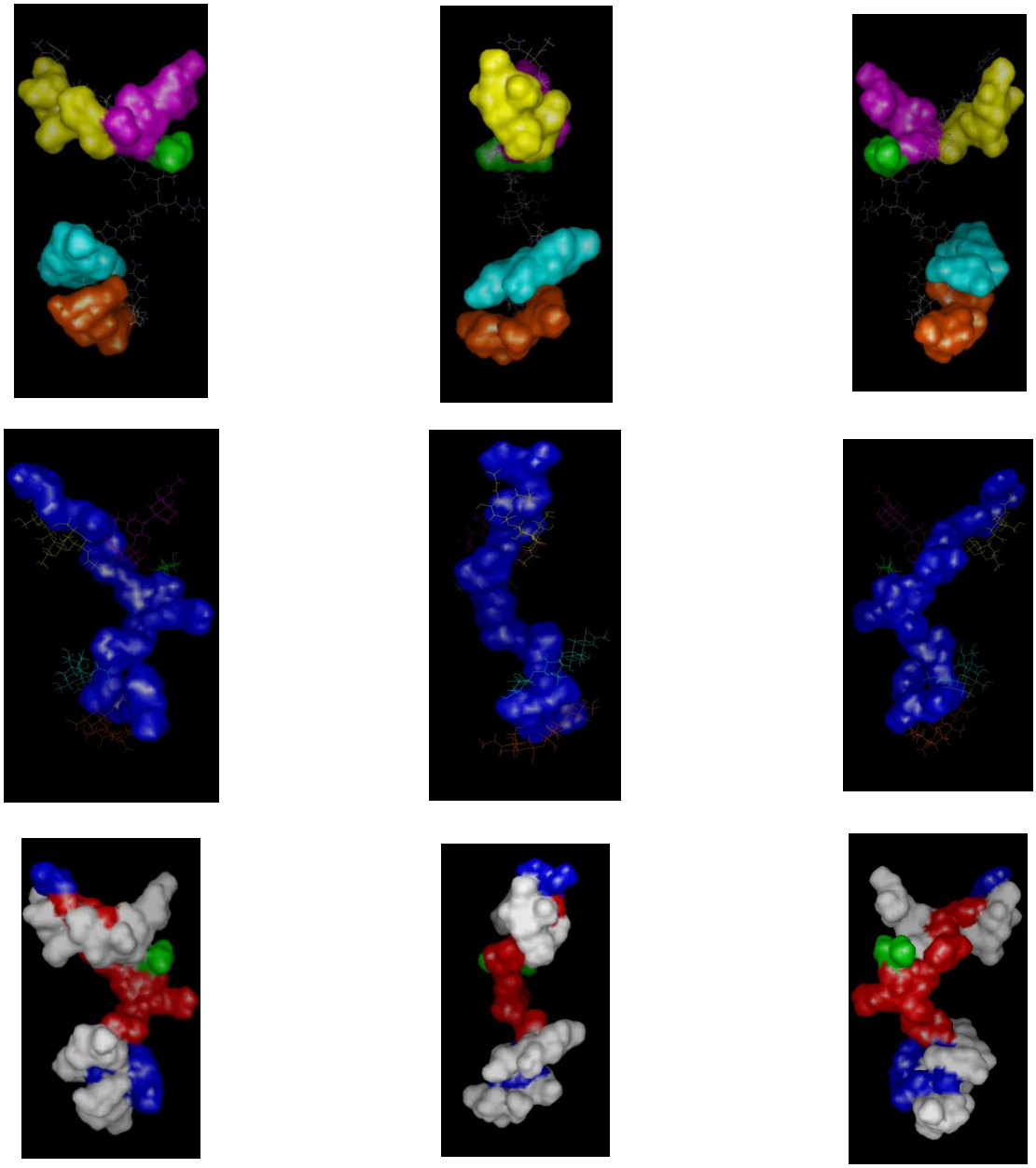


The best 30 calculated structures from NMR data. The ^{2}GVTSA^{6}, ^{7}PDTR^{10} , ^{10}RPAP^{13}, ^{14}GSTA^{17} regions are superimposed on backbone atoms ^{3}VTS^{5},^{8}DT^{9},^{11}PAP^{13}, ^{15}STA^{17} resp. The main chain of the Glycopeptide backbone is colored in Grey,side chain is blue, and GalNAc attached to Thr amino acid is red color, and the GalNAc attached to Ser amino acid is green color.



Angle bar plots of the 30 obtained structures of MUC1 Glycopeptide 18 (green dots), for all residues prolines φ and ψ are given and for prolines ψ angles are given. Observing the 3D structure and the angle bar data it can be seen whether all 30 structures amino acid psi and phi angles are aligned fitting near to each other or distributed at different degrees.

Discussing the glycopeptide 18 interaction against the 3 antibodies used it was clearly seen no interaction or binding observed against KL6 and SM3. GP18 showed good binding affinity against DF3 antibody.



Lowest energy structure in the 30 calculated models for the glycopeptide18. The left and the right structures are -90° and 90° rotations of the middle structures respectively.

A: Connolly surfaces of the Sialyl core 1 O glycans colour yellow at Thr 4, magenta at Ser5, green at Thr9, orange at Ser 15 and aqua at Thr 16.B:van der Waals Surfaces of the peptide backbone (blue) and five Sialyl core 1 branches indicated by wireframe representation: van der walls surface of the backbone peptide with the GalNAc moiety at the PDTR segment (GalNAc9) and the Connolly surfaces of the Sialyl core 1 based O glycans attached at the other glycosylation sited in which the peptide GVTSAPDTRPAP is red colour and other segments (PPAH and GSTA) are colour blue and the GalNAc 9 moiety is green colour. All images were generated by Mol feat version 5.0(FiatLux)

Analysing the statistical information about the glycopeptide the impact about binding and non binding with antibodies could be explained.

Statistical analysis:

	Glycopeptide 18
average potential energy(kcal/mol)a	
E_{total}	140.45 ± 9.99
E_{bond}	7.11 ± 0.46
E_{angle}	46.43 ± 3.02
E_{impr}	13.23 ± 0.48
E_{VDW}	21.09 ± 5.16
E_{NOE}	36.29 ± 2.35
E_{cdih}	0.47 ± 0.24
deviation from idealized geometry	
bond length(A°)	0.0034 ± 0.00013
bond angle (deg)	1.108 ± 0.032
Improper (deg)	0.696 ± 0.012
Average pairwise rmsd (A°)	
Backbone atoms	
Val 3 -Ser5	0.57 ± 0.31
Asp8 -Thr 9	0.10 ± 0.07
Pro11- Pro 13	0.45 ± 0.26
Ser15-Ala 17	0.30 ± 0.14
Heavy atoms	
Val 3 -Ser5	1.39 ± 0.58
Asp8 -Thr 9	0.81 ± 0.34
Pro11- Pro 13	0.89 ± 0.46
Ser15-Ala 17	0.59 ± 0.23
Val 3 -Ser5 and GalNAc 4	1.99 ± 1.04
Val 3 -Ser5 and GalNAc 5	2.30 ± 0.75
Val 3 -Ser5, GalNAc 4, GalNAc 5	2.72 ± 0.95
Asp8 -Thr 9 and GalNAc 9	1.65 ± 0.72
Ser15-Ala 17 and GalNAc 15	1.20 ± 0.53
Ser15-Ala 17 and GalNAc 16	1.04 ± 0.66
Ser15-Ala 17 , GalNAc 15 and GalNAc 16	1.44 ± 0.60

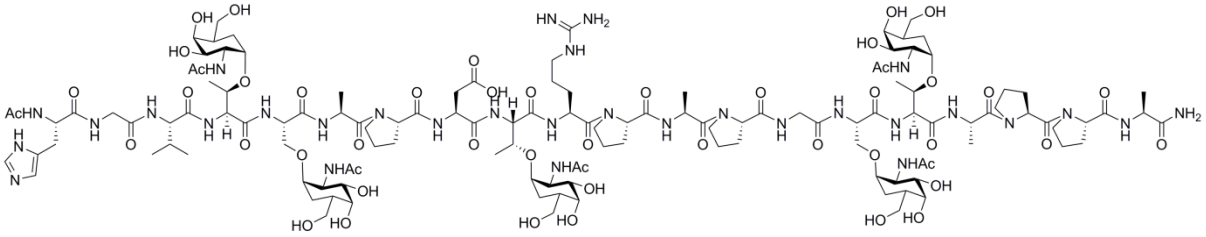
Dihedral angle data:

		Peptide 18
HIS1	ϕ	-121.9 ± 13.7
HIS1	χ1	-81.5 ± 85.9
HIS1	ψ	55.2 ± 81.0
GLY 2	ϕ	-117.0 ± 13.8
GLY 2	ψ	136.2 ± 94.2
VAL 3	ϕ	-124.6 ± 13.8
VAL 3	χ1	-89.3 ± 91.3
VAL 3	ψ	131.1 ± 92.3
THR 4	ϕ	-141.2 ± 0.2
THR 4	χ1	31.6 ± 76.5
THR 4	ψ	-48.3 ± 44.1
SER 5	ϕ	-147.4 ± 33.5
SER 5	χ1	175.1 ± 85.5
SER 5	ψ	142.4 ± 51.6
ALA 6	ϕ	-118.6 ± 16.8
ALA 6	χ1	-133.7 ± 98.6
ALA 6	ψ	150.0 ± 90.4
PRO 7	ψ	-177.9 ± 6.7
ASP 8	ϕ	-133.2 ± 6.6
ASP 8	χ1	111.5 ± 1.5
ASP 8	ψ	-172.0 ± 1.1
THR 9	ϕ	-119.0 ± 14.9
THR 9	χ1	-49.9 ± 80.5
THR 9	ψ	-60.6 ± 47.4
ARG 10	ϕ	-122.3 ± 14.8
ARG 10	χ1	-107.7 ± 75.9
ARG 10	ψ	118.5 ± 29.0
PRO 11	ψ	158.2 ± 65.1
ALA 12	ϕ	-138.7 ± 34.6
ALA 12	χ1	-107.4 ± 90.2
ALA 12	ψ	-63.1 ± 19.7
PRO 13	ψ	99.8 ± 84.2
GLY 14	ϕ	-125.5 ± 30.5
GLY 14	ψ	175.0 ± 53.8
SER 15	ϕ	-139.3 ± 3.9
SER 15	χ1	-160.1 ± 21.3
SER 15	ψ	72.5 ± 26.9
THR 16	ϕ	-105.8 ± 7.6
THR 16	χ1	-34.7 ± 35.3
THR 16	ψ	-49.5 ± 17.8
ALA 17	ϕ	-126.2 ± 15.0
ALA17	χ1	-144 ± 89.6
ALA 17	ψ	-73.5 ± 58.2
PRO 18	ψ	132.3 ± 68.6
PRO 19	ψ	130.5 ± 65.3
20 ALA	ϕ	-153.8 ± 49.0
20 ALA	χ1	-82.0 ± 94.3

The dihedral angle information shows each bond angle orientation due to which the binding studies could be discussed.

Glycopeptides 19

MUC1 20 mer amino acid sequence bearing Tn antigen at all O glycosylation sites was analysed by NMR machine to determine the 3D structure.

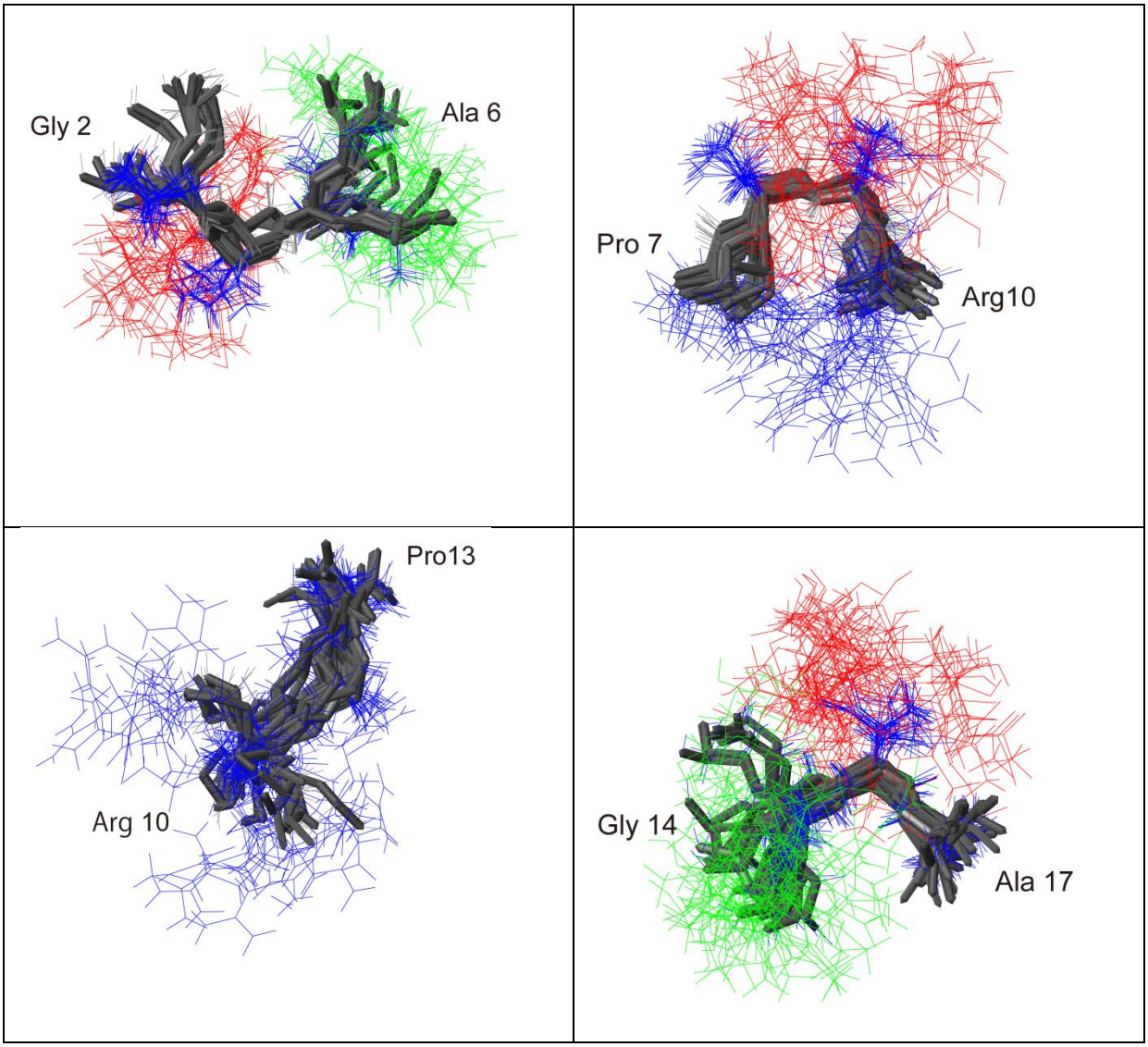


After all the spectra was derived from NMR analysis. Peaks were labelled corresponding using nosey cosy tocsy hsqc spectra of each amino acid. Totally 134 restraints were labelled successfully and 54 dihedral angle restraints were labelled as mentioned in the table below which is further used to determine the 3D structure.

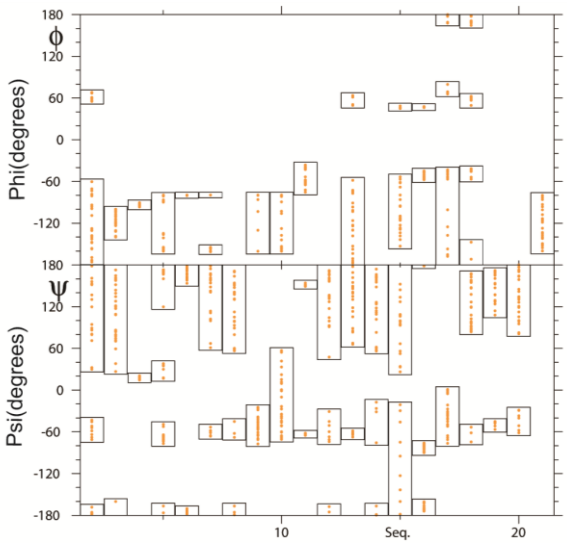
Once the labelling of peaks was completed 3D structure was obtained by CNS software. The number of peaks labelled inter residue, intra residue of amino acids and glycans are given as table below.

	Glycopeptide 19
NO of distance restraints	
total	134
intraresidue	
peptide	67
glycan	35
sequential	
peptide	32
glycan	0
medium range	
peptide	0
glycan	0
long range	
peptide	0
glycan	0
peptide to glycan	
within the same glycosylated residue	16
glycan on other peptide residues	15
no of dihedral restraints	
total	
peptide backbone	54
glycan	14
	40

.

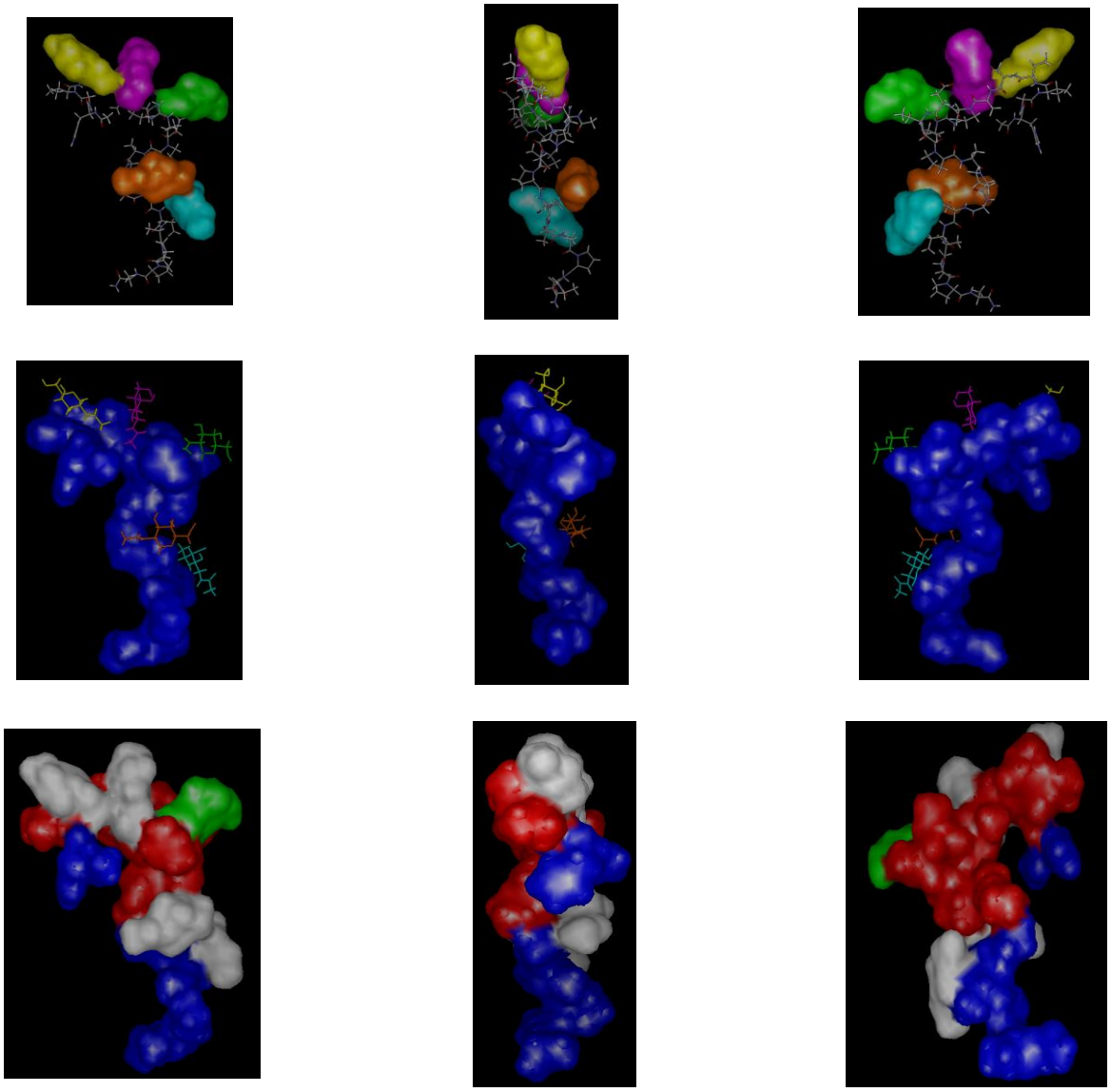


The best 30 calculated structures from NMR data. The ^{2}GVTSA^{6}, ^{7}PDTR^{10} , ^{10}RPAP^{13}, ^{14}GSTA^{17} regions are superimposed on backbone atoms ^{3}VTS^{5},^{8}DT^{9},^{11}PAP^{13}, ^{15}STA^{17} resp. The main chain of the Glycopeptide backbone is colored in Grey,side chain is blue, and GalNAc attached to Thr amino acid is red color, and the GalNAc attached to Ser amino acid is green color



Angle bar plots of the 30 obtained structures of MUC1 Glycopeptide 19(orange dots), for all residues prolines φ and ψ are given and for prolines ψ angles are given. Observing the 3D structure and the angle bar data it can be seen whether all 30 structures amino acid psi and phi angles are aligned fitting near to each other or distributed at different degrees.

Discussing the glycopeptide 19 interaction against the 3 antibodies used it was clearly seen no interaction or binding observed against KL6 and SM3. GP19 showed minimal binding affinity against DF3 antibody.



Lowest energy structure in the 30 calculated models for the glycopeptide19. The left and the right structures are -90° and 90° rotations of the middle structures respectively.

A: Connolly surfaces of the Sialyl core 1 O glycans colour yellow at Thr 4, magenta at Ser5, green at Thr9, orange at Ser 15 and aqua at Thr 16.B:van der Waals Surfaces of the peptide backbone (blue) and five Sialyl core 1 branches indicated by wireframe representation: van der walls surface of the backbone peptide with the GalNAc moiety at the PDTR segment (GalNAc9) and the Connolly surfaces of the Sialyl core 1 based O glycans attached at the other glycosylation sited in which the peptide GVTSAPDTRPAP is red colour and other segments (PPAH and GSTA) are colour blue and the GalNAc 9 moiety is green colour. All images were generated by Mol feat version 5.0(FiatLux)

Analysing the statistical information about the glycopeptide the impact about binding and non binding with antibodies could be explained.

Statistical analysis:

	Glycopeptide19
average potential energy(kcal/mol)a	
E_{total}	79.98 ± 4.03
E_{bond}	6.61±0.41
E_{angle}	15.22± 0.89
E_{impr}	2.51±0.16
E_{VDW}	8.96±1.69
E_{NOE}	42.3±1.96
E_{cdih}	0.06±0.02
deviation from idealized geometry	
bond length(A°)	0.0040±0.0001
bond angle (deg)	0.716±0.015
Improper (deg)	0.372±0.012
Average pairwise rmsd (A°)	
Backbone atoms	
Val 3 -Ser5	0.72 ± 0.35
Asp8 -Thr 9	0.42 ± 0.38
Pro11- Pro 13	1.14 ± 0.63
Ser15-Ala 17	1.14 ± 0.65
Heavy atoms	
Val 3 -Ser5	1.34 ± 0.60
Asp8 -Thr 9	1.28 ± 0.49
Pro11- Pro 13	1.76 ± 0.69
Ser15-Ala 17	1.72 ± 0.69
Val 3 -Ser5 and GalNAc 4	1.33 ± 0.49
Val 3 -Ser5 and GalNAc 5	2.08 ± 0.68
Val 3 -Ser5, GalNAc 4, GalNAc 5	2.38 ± 0.74
Asp8 -Thr 9 and GalNAc 9	2.03 ± 0.83
Ser15-Ala 17 and GalNAc 15	2.39 ± 1.01
Ser15-Ala 17 and GalNAc 16	2.55 ± 0.91
Ser15-Ala 17 , GalNAc 15 and GalNAc 16	3.12 ± 1.08

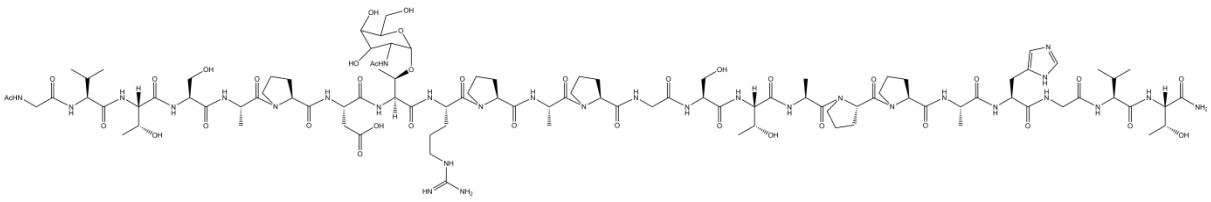
Dihedral angle data:

		Peptide 19
HIS1	ϕ	-128.8 ± 89.0
HIS1	χ1	-151.5 ± 54.4
HIS1	ψ	133.3 ± 98.8
GLY 2	ϕ	-118.2 ± 14.2
GLY 2	ψ	109.1 ± 42.5
VAL 3	ϕ	-94.6 ± 1.3
VAL 3	χ1	-167.8 ± 0.4
VAL 3	ψ	18.5 ± 1.3
THR 4	ϕ	-124.6 ± 36.9
THR 4	χ1	48.5 ± 32.2
THR 4	ψ	-85.0 ± 88.5
SER 5	ϕ	-79.7 ± 0.1
SER 5	χ1	-166.0 ± 5.8
SER 5	ψ	172.5 ± 10.5
ALA 6	ϕ	-144.0 ± 34.7
ALA 6	χ1	102.2 ± 102.6
ALA 6	ψ	156.3 ± 80.9
PRO 7	ψ	133.7 ± 66.5
ASP 8	ϕ	-123.1 ± 39.5
ASP 8	χ1	-126.2 ± 36.6
ASP 8	ψ	-53.4 ±15.3
THR 9	ϕ	-134.1 ± 34.8
THR 9	χ1	11.9 ± 89.4
THR 9	ψ	-24.2 ± 38.7
ARG 10	ϕ	-53.6 ± 12.4
ARG 10	χ1	-175.6 ± 33.3
ARG 10	ψ	-112.4 ± 74.9
PRO 11	ψ	142.2 ± 85.0
ALA 12	ϕ	-119.1 ± 81.3
ALA 12	χ1	15.3 ± 92.6
ALA 12	ψ	132.6 ± 91.6
PRO 13	ψ	126.7 ± 69.1
GLY 14	ϕ	-104.1 ± 55.7
GLY 14	ψ	92.6 ± 83.1
SER 15	ϕ	-41.1 ± 34.0
SER 15	χ1	-53.7 ± 42.0
SER 15	ψ	-130.2 ± 45.1
THR 16	ϕ	-96.6 ± 98.0
THR 16	χ1	-68.7 ± 63.7
THR 16	ψ	-32.3 ± 24.3
ALA 17	ϕ	81.5 ± 94.0
ALA17	χ1	-144.9 ± 100.9
ALA 17	ψ	121.1 ± 69.9
PRO 18	ψ	159.0 ± 77.1
PRO 19	ψ	133.2 ± 87.1
20 ALA	ϕ	-123.8 ± 24.6
20 ALA	χ1	31.5 ± 104.6

The dihedral angle information shows each bond angle orientation due to which the binding studies could be discussed.

Glycopeptide 24:

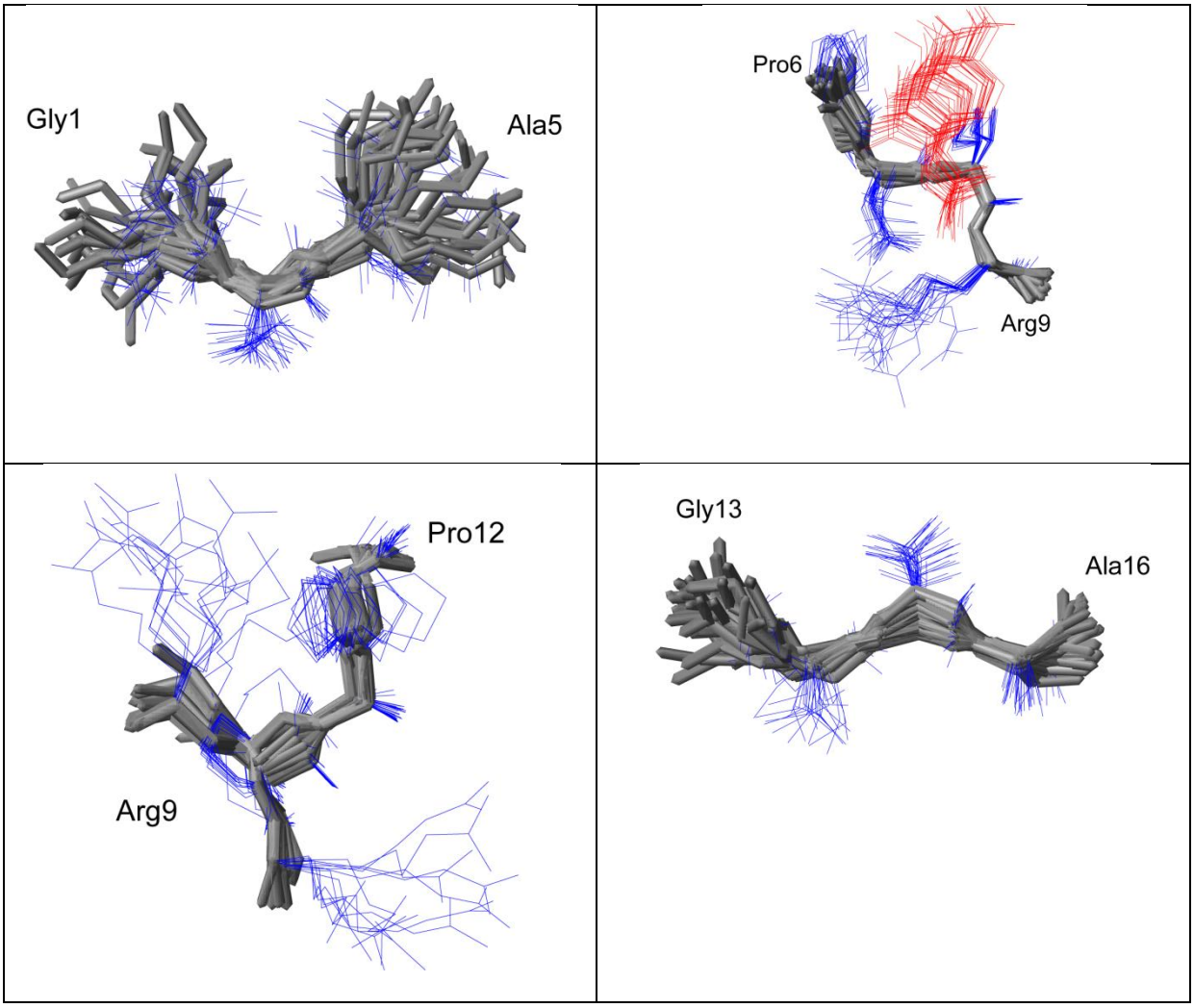
MUC1 24mer amino acid sequence bearing Tn antigen at Thr O glycosylation site of PDTR motif was analysed by NMR machine to determine the 3D structure



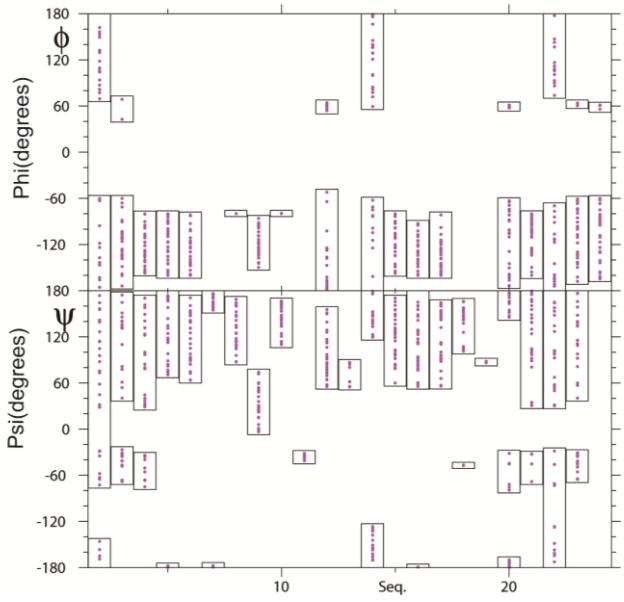
After all the spectra was derived from NMR analysis. Peaks were labelled corresponding using NOESY COSY TOCSY HSQC spectra of each amino acid. Totally 153 restraints were labelled successfully and 19 dihedral angle restraints were labelled as mentioned in the table below which is further used to determine the 3D structure.

Once the labelling of peaks was completed 3D structure was obtained by CNS software. The number of peaks labelled inter residue, intra residue of amino acids and glycans are given as table below.

	Glycopeptide 25
NO of distance restraints	
total	153
intraresidue	
peptide	90
glycan	11
sequential	
peptide	41
glycan	
medium range	
peptide	0
glycan	0
long range	
peptide	0
glycan	0
peptide to glycan	
within the same glycosylated residue	11
glycan on other peptide residues	0
no of dihedral restraints	
total	19
peptide backbone	11
glycan	8



The 30 lowest energy NMR structure of 24-mer MUC1 glycopeptide carrying Tn antigen at Thr8 in ^{6}PDTR^{9} region. The ^{1}GVTSA^{5}, ^{6}PDTR^{9}, ^{9}RPAP^{12}, and ^{13}GSTA^{16} regions were individually superimposed on the backbone atoms of ^{2}VTS^{4}, ^{7}DTR^{9}, ^{10}PAP^{12}, and ^{14}STA^{16}, respectively. The glycopeptide structures represent all heavy atoms, including side chains and GalNAc moiety. The main chain of the peptide backbone is colour gray neon; the side chain is shown as a blue line, the GalNAc moiety attached at Thr residue as a red line.



Angle bar plots of the 30 obtained structures of MUC1 Glycopeptide 25(Maroon dots), for all residues prolines φ and ψ are given and for prolines ψ angles are given. Observing the 3D structure and the angle bar data it can be seen whether all 30 structures amino acid psi and phi angles are aligned fitting near to each other or distributed at different degrees.

Discussing the glycopeptide 19 interaction against the 3 antibodies used it was clearly seen no interaction or binding observed against KL6 and SM3. GP 19 showed minimal binding affinity against DF3 antibody.

Analysing the statistical information about the glycopeptide the impact about binding and non binding with antibodies could be explained.

Statistical analysis

	Tn
average potential energy(kcal/mol)a	
E_{total}	32.35 ± 0.66
E_{bond}	2.56 ± 0.11
E_{angle}	7.66 ± 0.44
E_{impr}	0.71 ± 0.02
E_{VDW}	3.3 ± 0.32
E_{NOE}	17.58 ± 0.53
E_{cdih}	0.03 ± 0.006
deviation from idealized geometry	
bond length(A°)	0.002 ± 6.39
bond angle (deg)	0.41 ± 0.006
Improper (deg)	0.199 ± 0.003
Average pairwise rmsd (A°)	
Backbone atoms	
Val 3 -Ser5	0.77 ± 0.26
Asp8 -Thr 9	0.24 ± 0.14
Pro11- Pro 13	0.61 ± 0.30
Ser15-Ala 17	0.52 ± 0.25
Heavy atoms	
Val 3 -Ser5	1.73 ± 0.44
Asp8 -Thr 9	0.81 ± 0.31
Pro11- Pro 13	1.15 ± 0.58
Ser15-Ala 17	1.07 ± 0.32
Asp8 -Thr 9 and GalNAc 9	0.83 ± 0.31

Dihedral angle data:

		Tn
GLY 1	ϕ	147.9 ± 67.5
GLY 1	ψ	107.8 ± 93.5
VAL 2	ϕ	-117.8 ± 60.5
VAL 2	χ1	-63.2 ± 90.9
VAL 2	ψ	50.9 ± 87.7
THR 3	ϕ	-120.8 ± 22.3
THR 3	χ1	-7.6 ± 83.7
THR 3	ψ	51.2 ± 79.6
SER 4	ϕ	-119.0 ± 25.7
SER 4	χ1	-123.5 ± 73.4
SER 4	ψ	131.0 ± 39.8
ALA 5	ϕ	-128.4 ± 22.9
ALA 5	χ1	168.6 ± 96.5
ALA 5	ψ	116.9 ± 32.3
PRO 6	ψ	164.9 ± 8.0
ASP 7	ϕ	-79.7 ± 0.0
ASP 7	χ1	-170.1 ± 3.3
ASP 7	ψ	133.1 ± 25.6
THR 8	ϕ	-116.2 ± 17.9
THR 8	χ1	33.2 ± 1.5
THR 8	ψ	33.9 ± 20.8
ARG 9	ϕ	-79.6 ± 0.1
ARG 9	χ1	-21.9 ± 50.0
ARG 9	ψ	144.4 ± 19.3
PRO 10	ψ	-35.4 ± 1.8
ALA 11	ϕ	147.0 ± 85.1
ALA 11	χ1	-101.5 ± 103.3
ALA 11	ψ	94.3 ± 29.4
PRO 12	ψ	64.9 ± 9.0
GLY 13	ϕ	158.3 ± 75.4
GLY 13	ψ	172.3 ± 39.7
SER 14	ϕ	-111.8 ± 23.1
SER 14	χ1	-93.1 ± 84.5
SER 14	ψ	120.4 ± 29.9
THR 15	ϕ	-134.2 ± 20.6
THR 15	χ1	-67.2 ± 88.7
THR 15	ψ	115.1 ± 33.9
ALA 16	ϕ	-134.5 ± 19.3
ALA16	χ1	116.3 ± 101.0
ALA 16	ψ	118.6 ± 35.0
PRO 17	ψ	135.7 ± 76.7
PRO 18	ψ	87.4 ± 0.5
ALA 19	ϕ	-102.8 ± 86.4
ALA19	χ1	7.3 ± 102.3
ALA19	ψ	-172.7 ± 60.7
HIS20	ϕ	-112.9 ± 24.9
HIS20	χ1	-93.9 ± 77.1
HIS20	ψ	114.4 ± 81.2
GLY 21	ϕ	167.0 ± 68.6
GLY 21	ψ	137.8 ± 84.7
VAL 22	ϕ	-108.9 ± 54.5
VAL 22	χ1	160.7 ± 89.7
VAL 22	ψ	70.0 ± 88.9
THR 23	ϕ	-106.3 ± 68.3
THR23	χ1	-18.6 ± 95.2

The dihedral angle information shows each bond angle orientation due to which the binding studies could be discussed

Structural Conformation at each Motif of MUC1 Glycopeptide:

Structural calculations were performed using CNS for the analysis of the 3D structure. Few other NMR experiments were conducted to determine the distance between each binding and dihedral restraints. Around 200 ensembles were calculated and considered and a family of 30 accepted structures were selected having the lowest potential energies. The statistical analysis of NMR restraints and structural statistics were listed down for better understanding. The average dihedral angles and the angle bar plot of Phi(φ), Chi (χ1), Psi(ψ) angles of the accepted 30 structures were summarised. Due to the statistical analysis of restraints, structural statistics and various angle information structural alterations in different immunodominant motif could be compared. Number of NOE signal observed were listed (Figure 6a)

Understanding GVTSA motif and analysing the statistical values, it was observed that the back bone of heavy atoms of Val3, Thr4, and Ser5 exhibited a tendency (1.45±0.52, 1.46±0.49, 1.39±0.58, 1.39±0.60, 1.34 ± 0.60) suggesting that glycopeptide 8, 18, 14,17 and 19 has a similar structural orientation. When Val3, Thr4 and Ser5 bearing GalNAc residues on both Thr4 and Ser5 amino acids of all the 5 glycopeptides were observed, it was detected to have similar statistical value hence bringing a stabilised and rigid conformation structure.

Statistical analysis of the φ, χ1, ψ angles were observed. The phi angle of the Thr4 amino acid demonstrated the stability effect due to the presence of glycosylation extended on the peptide backbone structure in glycopeptide 18 than Thr 4 on glycopeptide 14, glycopeptide 17, glycopeptide 8 and glycopeptide 19. The phi angle in glycopeptide 18 shows higher stability compared to glycopeptide 8, 19 and 14 even though glycopeptide 14 and glycopeptide 18 both being sialylated amino acid. Chi angle, the important bond angle

relating to the attachment of the glycan moiety it was observed that glycopeptide 14 show more stability in the chi angle than glycopeptide 18 because glycopeptide 18 had (31.6 ±76.5) glycopeptide 14 (42.0 ±43.3), regarding the chi angle glycopeptide 8 and glycopeptide 17 show similar orientation and stability.

The psi angles of the Thr 4 showed similar stability but change in bond conformation could be visible. The psi angle of the Ser5 residue determined that the glycopeptide 19 has a comparatively stable amino acid structure when compared with other glycopeptides. Similarly in the Val3 of the GVTSA motif where psi angle was observed that in the glycopeptide 19 was comparably more stable than same Val3 in glycopeptide 8, glycopeptide 18, glycopeptide14 and glycopeptide 17.Further analysis of the glycopeptide comprising of crucial epitope i.e. Pro-Asp-Thr-Arg motif was conducted in the similar manner as explained before. Analysis of the 3D structure with the statistical value showed evidences of site specific alteration during core 1 glycosylation and enzymatic sialylation in the MUC1 tandem repeating peptide.

In the PDTR sequence segment, statistical analysis of the backbone heavy atoms pair wise r.m.s.d values before and after glycosylation could be seen to be elevated. The glycopeptide 8 with the absence of GalNAc showed (1.24 ± 0.34) where as in the presence of GalNAc showed (2.03 ± 0.61). The glycopeptide 18 with the absence of GalNAc showed (0.81 ± 0.34) where as in the presence of GalNAc showed (1.65 ± 0.72). The glycopeptide 14 with the absence of GalNAc showed (1.15 ± 0.32) where as in the presence of GalNAc showed (1.93 ± 0.85). The glycopeptide 17 with the absence of GalNAc showed (1.35 ± 0.42) where as in the presence of GalNAc showed (1.46 ± 0.49) .The glycopeptide 19 with the absence of GalNAc showed (1.28 ± 0.49) where as in the presence of GalNAc showed (2.03 ± 0.83) This showed that even minimum glycosylation in the peptides could change r.m.s.d value.

Clearly determining the difference in the PDTR motif was required to understand each amino acid angles between each glycopeptide. The psi angle of Pro7 and Asp8 has more stability in the glycopeptide 18 comprising of Tn antigen than the glycopeptide 8, 14, 17 and 19. Whereas, Thr9 residue of glycopeptide 14 showed that the phi angle's high stability comparing the glycopeptide 8 and glycopeptide 17, glycopeptide 18, glycopeptide 19. The psi angle of the Thr9 showed conformational change and was also unstable in the bond angle in glycopeptide 14 than the peptide 8, 17, 18 and 19. These conformational alterations have shown the dispersed or non fitted Arg10 the PDTR motif which could also be observed in the 3D structure. Further observation in the PDTR motif showed that the chi angle which played major role in relating the back bone to the glycan moiety showed glycopeptide 8(-1.9 ± 71.8), glycopeptide 18 (-49.9 ±80.5), glycopeptide 14 (-14.8 ±25.7), glycopeptide 17(2.9±52.3) , glycopeptide 19(11.9 ± 89.4) values. By these values it could be clearly seen that the core 1 sialylated peptide 14 was highly stable when compared to the non sialylated core 1 glycopeptide 8 and the Tn antigen glycopeptide 18 and glycopeptide 17 comprising of T antigen and glycopeptide19 comprising of completely Tn glycosylated peptide. These results clearly have shown that the fully sialylated MUC1 glycopeptide provided required stability and suitable angle conformation against KL6 mAb.

In the MUC1 20mer peptide the segment of amino acid region which comprises of no glycosylation sites was the RPAP sequence. RPAP being the naked sequence was also a part of crucial heptamer epitope region (PDTRPAP). Understanding the rmsd values of backbone heavy atoms of this segment glycopeptide 8 (1.40 ± 0.55), glycopeptide 18 (0.89 ± 0.46), glycopeptide 14(0.87 ± 0.35), glycopeptide 17 (1.32±0.47), glycopeptide19 (1.76 ± 0.69) showed no significant difference.

The Pro 11 and Pro 13 of the RPAP segment showed similar stability and conformation of bond angles between glycopeptide 8, 18, 14 and 19 where as in glycopeptide 17 there was

change in bond angle orientation comparing with other peptides at Pro 11. The Ala12 amino acid psi angles glycopeptide 8 (121.3 ±85.0), glycopeptide 18 (-63.1 ±19.7), glycopeptide 14(-65.0 ±2.3) and glycopeptide 17(121.1±70.9), glycopeptide 19(132.6 ± 91.6) revealed that fully sialylated glycopeptide 14 was stable compared to the other 4 glycopeptide. This shows the centre Ala 12 residue of the RPAP segment have shown good stability structure variation which also related due to the presence of sialylated Thr9 of the PDTR segment.

Analysis of the final segment of the MUC1 peptide was the GSTA segment, the pair wise backbone heavy atoms of the Ser15,Thr16 and Ala17 with no glycans in glycopeptide 8,18,17,14 and 19 were observed to be (1.47 ± 0.46, 0.59 ± 0.23, 1.14 ± 0.56 , 0.77 ± 0.34, 1.72 ± 0.69) respectively. When either one of Ser15/Thr16 residue bearing the GalNAc residue rmsd value were compared i.e. rmsd during presence of GalNAc at only Ser15 of glycopeptide 8 (2.53 ± 0.58), glycopeptide 18 (1.20 ± 0.53), glycopeptide 14 (1.31 ± 0.52), glycopeptide (1.70 ± 0.57), glycopeptide 19 (2.39 ± 1.01) or during the presence of GalNAc at only Thr16 of glycopeptide 8 (2.29 ± 0.69), glycopeptide 18(1.04 ± 0.66), glycopeptide 14 (1.07 ± 0.86), glycopeptide 17(1.54±0.84), glycopeptide19(2.55 ± 0.91) minimal decrease in rmsd value could be observed in all glycopeptides.

Further, the pair wise backbone rmsd value including the GalNAc in both Ser15 and Thr16 residue of the GSTA motif was glycopeptide 8 (3.01 ± 0.65), glycopeptide 18 (1.44 ± 0.60), glycopeptide 14 (1.55 ± 0.75), glycopeptide 17(2.20±0.76), glycopeptide 19(3.12 ± 1.08). It was observed that sialylated structures glycopeptide 14, glycopeptide 18 values was similar suggesting that core1 sialylation at both Ser15 and Thr16 residues contributed strongly for the stabilization of the GSTA segment.

Dihedral angle of the GSTA segment showed striking r.m.s.d value changes in the Ser15 phi angle of glycopeptide 8 (-111.1 ±22.3), glycopeptide 18(-139.3 ± 3.9), glycopeptide 14

(-158.8 ± 3.1), glycopeptide 17(-99±16.2), glycopeptide 19(-41.1±34.0) showing that sialylated glycopeptides 18 and 14 and highly stable conformation than glycopeptide 8, glycopeptide 17 and glycopeptide 19. Similar stability conformation was observed with the psi angles. In the Thr16 residue of the psi angle it was observed that the sialylated glycopeptide 18 and 14, 19 showed better conformational stability when compared with glycopeptide 8 and glycopeptide 17. The chi angle of the ser residue in glycopeptide 8(-38.1 ± 75.9), glycopeptide 18 (-160.1 ±21.3), glycopeptide 14 (142.4 ± 65.3) glycopeptide 17(43.3± 79.3), glycopeptide 19 (-53.7 ± 42.0), was different to each other thus, signifying that sialylation contributed to the bond orientation in the Ser15 of GSTA segment.

3D Structural Analysis:

Totally 30 super imposed structures determined by MOLMOL software were calculated to represent the characteristic features of the PDTRPAP heptapeptide region when compared with the three dimensional structures of the neighbouring GVTSA and GSTA motif. For better understanding the hepta peptide region was calculated as PDTR and RPAP to get the best fit structure and for better understanding. Determining the gamma turn, and beta turn certain criteria should be identified given that the psi and phi angles of central residue in the inverse gamma turn adopt a φ (i+1) of -79 ± 40 and ψ (i+1) of 69.0 ± 40, however the criteria could not be satisfied by any of the five glycopeptides, at the PDT or DTR region. Recent studies revealed that Inverse gamma turn structure could be clearly identified by Hα(i+1)-NH(i+2) had a distance of 2.4A°, the NH(i+1)-NH(i+2) distance of 3.8A° and Hα(i)-NH(i+2) distance of 4.3A°.

Analysing the lowest energy ensemble as a representative model at the PDT region showed a small difference i.e. < 2 A° from the above definition of the distance criteria for the ideal inverse gamma turn, Asp9 Hα(i+1)-Thr10 NH(i+2) distance of 2.7 A°,2.66A°,2.77A° and

2.36 A° whereas glycopeptide 19 showed 3,54 A°, Asp9NH(i+1)-Thr Nh(i+2) distance of 4.45A°,4,31A°, 4.43A° , 4.49 A° and 2.64 A° resp, ProHα(i)-Thr NH(i+2) distance of 6.66A°, 6.51A° , 6.68 A° , 6.65 A° and 4.38 A° of peptides 8, 18, 14 and 17,19 respectively. Whereas for DTR region indicated a much smaller difference <1A° to that of the ideal inverse gamma turn Thr10 Hα(i+1)-Arg11 NH(i+2) distance of 3.56 A°,3.56 A°,3.47A° , 2.14 A°,3.56 A°, Thr10NH(i+1)-ArgNH(i+2) distance of 2.42°,2.37A°, 2.15A° , 4.09 A° and 2.62 A°, Asp9Hα(i)-Arg11 NH(i+2) distance of 4.83A°, 5.0A°, 4.54 A° , 6.12 A° and 5.66 of glycopeptides 8,18 14 ,17 and 19 respectively. Glycopeptide 17 shows greater distance between Thr10 NH(i+1)- ArgNH(i+2) and Asp9Hα(i)-Arg11 NH(i+2) due to which angle change at the Arg residue could be observed when compared with other structures in the 3D model.

Criteria of the beta turn which had been reported earlier showed that any tetra peptide i.e. in our case PDTR should have a distance between Cα (i) and Cα (i+3) is <7A° reference plane. The central residues i.e. the Asp and Thr residues were not in the helical conformation. According to our lowest 30 calculated structures none of our residues show <7A° hence not fulfilling the beta turn criteria except glycopeptide 19. But comparing the distance of Cα(i) and Cα(i+3) is 10.1A°,9.93A°, 8.59A° , 10.5 A° and 5.66 A° whereas for CO (i)and Arg10 NH(i+3) distance are 6.66 A°,7.30 A°, 6.41 A° , 8.17 A° and 5.42 of glycopeptides 8,18, 14, 17 and 19 resp. This distance interestingly showed that the glycopeptide14 had the lower distance and visual images of the 3D structure showed that the converged structure which played the major role for the KL6 mAb interaction even though glycopeptide 19 shows lower value but due to absence of sialylation and also glycopeptide 8 shows the similar orientation structure the absence of the Sialyl group also could be determined as the driving force for the KL6 mAb interaction.

Effects of Multiple *O*-Glycosylation States on the Conformation of Immunodominant PDTR Motif in the MUC1 Tandem Repeats

The dihedral angle bar plots of the 20 amino acid residues and the 30 superimposed calculated structures focusing on the PDTR motif of the MUC1 glycopeptides 8, 14, 17, 18, and 19 have been explained to show its importance. The NMR structures elaborated from the results of the average dihedral angles and pairwise rmsd values clearly showed the characteristic features of the PDTR region of these glycopeptides. It was revealed for the first time that the conformation of the immunodominant PDTR motif is restricted not only by the glycosylation at the Thr residue of this region but by the glycosylation states of the neighbouring Ser/Thr residues in the tandem repeats. For example, the difference in the conformation of the PDTR region between the glycopeptides 8 and 17 is evident, even though both MUC1 fragments have the same T antigen at Thr residue in the PDTR region. Similarly, glycopeptides 18 and 19 having the same Tn antigen in this area exhibited obviously different structural features as shown in its respective figures. It is interesting to note that the well-extended *trans*-like peptide backbone structure and the fixed glycan orientation of the PDTR motif in the compound 18 might be due to the multiple ST antigens attached at the adjacent Ser/Thr residues when compared with the structure of glycopeptide 19. The multiple modifications with Tn antigens at five Ser/Thr residues in the glycopeptide 19 can convert an extended *trans*-like peptide backbone of this epitope region into *cis*- (β-turn) like conformation, in which amino acid side chains and carbohydrate moiety appeared to be very flexible. Since compound 8 having five T antigens at all potential *O*-glycosylation sites also exhibited a quite similar *cis*- (β-turn-) like conformation and the flexibility in the carbohydrate moiety attached at the PDTR motif, an extended *trans*-like conformation at the PDTR region with a restricted glycan directionality might be critical for the recognition by common anti-MUC1 mAbs such as DF3 and SM3. Given the results of the epitope mapping analysis that DF3 showed significantly higher binding affinity toward glycopeptides 14 and

18 than compound 16 and 19, the multiple modification with ST antigens may provide the immunodominant PDTR motif of the MUC1 tandem repeats with a *trans*- like extended conformation as well as the fixed carbohydrate directionalities to be more preferred antigenic structure for DF3 recognition. In contrast, it is noteworthy that multiple *O*-glycosylation states in the MUC1 tandem repeats did not affect entirely the binding mode of anti-KL6 mAb when the MUC1 fragment contains an essential KL6-epitope (PDTRPAP with ST antigen) such as compound 14, 16, 22, and 23 as shown in its respective data.

Experimental Section

NMR Spectroscopy. The glycopeptides were dissolved at concentration of 3.0mM in 300µl of a 90% H2O/ 10%D2O mixture at pH5.5 using NaOH and HCl. NMR spectra were observed by Bruker Avance 800Hz.Compounds used for NMR structural analysis were observed by Bruker Avance with a proton frequency of 800MHz at 278K temperature. Samples prepared with final concentration 3.0mM in 300µl of 90% H_20/10%D_2O. The samples were maintained at pH 5.0 using HCl or NaOH. The NMR peak labelling and structural determination was completed using two dimensional DQF-COSY, TOCSY with MLEV-17 sequence and NOESY spectra were recorded in the indirect dimension using States-TPPI phase cycling[6]. Also the two dimensional homonuclear 13C-edited HSQC and HSQC TOCSY measurements were taken in echo-anti echo mode for sensitivity enhancement. Spin lock time of TOCSY experiments were done at 60ms, whereas NOESY spectra were recorded with mixing times of 100,200 and 400ms. Water signals was suppressed by pre saturation at 2s relaxation delay and by 3-9-10 WATERGATE pulse sequence including a field gradient[7,8]. Recorded NOESY and TOCSY spectra were acquired with 2048X512 frequency data points and were zero filled to yield 2048X2048 data matrices respectively. DQF-COSY spectra having 16384X512 frequency data points was recorded and zero filled yield of 16384X16384 matrix was used to measure the coupling constants. Applied Sweep width was noted to be 9578.544 Hz. Sine bell window function with 90°phase shift prior were multiplied prior to Fourier transformation by time domain data in both the dimensions. Topspin3.2 version was used to process the derived data and analysed by Sparky software [9, 10].

NMR Data:

The compounds spectra were determined by 800MHz NMR machine. The 3D structure was calculated using CNS 1.1[11].Distance restraints for calculations were estimated from the cross peak intensities in the NOESY spectra having mixing time of 400ms for all compounds used for structural studies. For the structural analysis the restraints estimated was classified as strong, medium and weak signals and the upper limit was noted to be 2.6,3.5,5.0 A°. 5.0A° was set as the upper bound for the NOEs found only in the NOESY spectra having the mixing time of 400ms. Dihedral angle restraints were based on JHNα coupling constants which was measured by high resolution DQF-COSY. For JHNα calculated more than 8.0Hz, dihedral angle was constricted to -120 ± 30°. Complete RMSD values of secondary and tertiary structures were performed using MOLMOL software[12].

4-4. References

(1) Ernst, R. R. (1992) Nuclear magnetic resonance Fourier transform spectroscopy. *Biosci. Rep. 12*, 143–187.

(2) Alonso, D. E., and Warren, S. E. (2005) NMR Analysis of Unknowns: An Introduction to 2D NMR Spectroscopy. *J. Chem. Educ. 82*, 1385.

(3) Imberty, a, and Perez, S. (2000) Structure, conformation, and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations. *Chem Rev 100*, 4567–4588.

(4) Pérez, S., and Tvaroška, I. (2014) Carbohydrate-protein interactions: molecular modeling insights. *Adv. Carbohydr. Chem. Biochem.*

(5) a, S., S, D., a, R., and S, P. (2015) Databases of Conformations and NMR Structures of Glycan Determinants *25*, 1480–1490.

(6) Bothner-By, A. a., Stephens, R. L., Lee, J., Warren, C. D., and Jeanloz, R. W. (1984) Structure determination of a tetrasaccharide: transient nuclear Overhauser effects in the rotating frame. *J. Am. Chem. Soc. 106*, 811–813.

(7) Bax, A., and Davis, D. G. (1985) Practical aspects of two-dimensional transverse NOE spectroscopy. *J. Magn. Reson. 63*, 207–213.

(8) Piotto, M., Saudek, V., and Sklenář, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR 2*, 661–665.

(9) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277–293.

(10) Lee, W., Hu, K., Tonelli, M., Bahrami, A., Neuhardt, E., Glass, K. C., and Markley, J. L. (2013) Fast automated protein NMR data collection and assignment by ADAPT-NMR on Bruker spectrometers. *J. Magn. Reson. 236*, 83–88.

(11) 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., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D Biol. Crystallogr. 54*, 905–921.

(12) Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph. 14*, 51–55.

CHAPTER 5

Docking Studies

5-1. Introduction

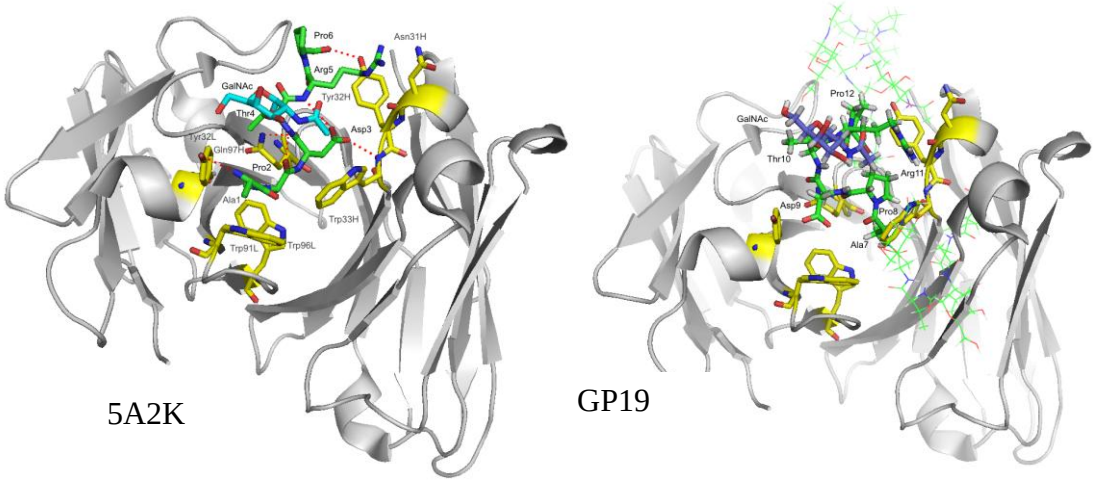
Molecular docking is a method used to predict the orientation between one molecule to another and determine whether they form a strong or weak complex[1]. The binding affinity can be determined using this method. Docking of small molecules into a known binding socket i.e. receptor, can be designed which can be helpful for drug designing [2-7].

PyMol is software used to perform docking studies; it is a structure based prediction tool highly efficient to understand the binding alignment[8-10]. The method involves compilation of an extensive library of structural motif templates based on well characterized enzyme structures with active sites that are defined.

In our laboratory we used this software to check the interaction and alignment between SM3 antibodies against the MUC1 glycopeptides synthesised for NMR studies. In this docking study we used the crystallographic structure of SM3 antibody docked against a hexa peptide sequence of MUC1 peptide, but we replaced our structure with the hexa peptide and analysed the alignment between each other[11] .

5-2. Results and Discussion

To conduct the docking studies reported crystallographic structure derived from PDB i.e. (5A2K) of SM3 mAb docked with a hexa peptide sequence APDTRP bearing GalNAc residue at the Thr amino acid. To this 5A2K our glycopeptides focussed for NMR studies were used for understanding the interaction. Lowest energy structure of each glycopeptide was selected for studying the binding orientation pattern with each other.



5A2K GP19

GP 19 is MUC1 glycoprotein bearing Tn antigen (GalNAcα) at all 5 O glycosylation sites. The RMS value of this docking data was observed to be 4.591 with 46 atoms being aligned; this value was observed to be having the highest value comparing the other glycopeptides docked with 5A2K. Therefore the shape complimentary between receptor-ligand interactions is said to be not a good hit.

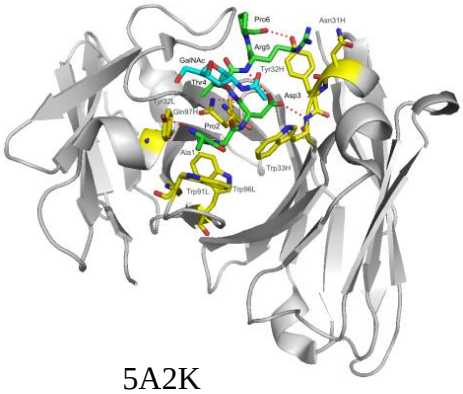
GP 8 comprises of Gaβ1-3GalNAcα (T antigen) at all 5 O glycosylation sites. The RMS value of GP8 with 5A2K was observed to be 2.072 with 39 atoms aligned. This shows that interaction with the SM3 crystallographic structure shows better fitted interaction compared with other structures.

Docking studies of GP 14 comprising of Neu5Acα2,3 Gaβ1-3GalNAcα (Sialyl T) glycan on all 5 O glycosylation sites shows RMS value of 2.500 with 43 aligned atoms. Similarly GP 17 comprising of Gaβ1-3GalNAcα residue at the Thr of PDTR motif whereas the neighbouring 4 O glycosylation sites with GalNAcα residue shows 2.698 RMS value with 42 atoms aligned.

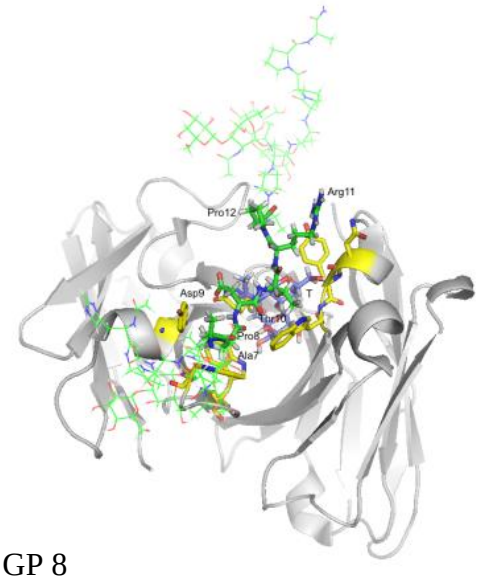
GP 18 comprising of Tn antigen (GalNAcα) at the PDTR Thr amino acid and neighbouring *O* glycosylation sites with Neu5Acα2, 3 Gaβ1-3GalNAcα (Sialyl T) showed 3.148 as the RMS value with 44 atoms being aligned. Therefore not being the best fitted interaction compared to GP 8 GP14 and GP17.

GP24 comprising of a single glycosylated amino acid i.e. Tn antigen (GalNAcα) at Thr amino acid residue of the PDTR motif showed the best hit interaction between the receptor and ligand with 2.0005 as the RMS value with 42 aligned atoms which can be observed in the docking structures.

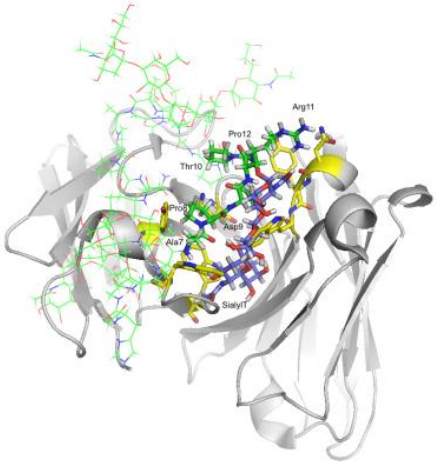
S3: PyMol



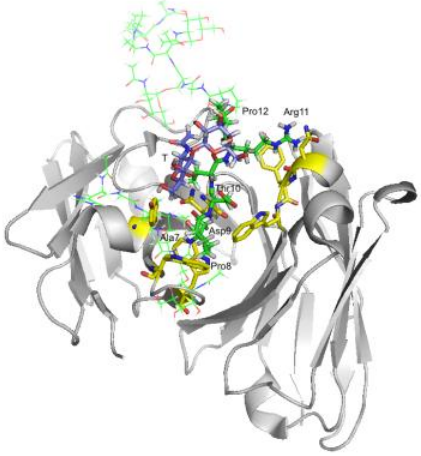
5A2K



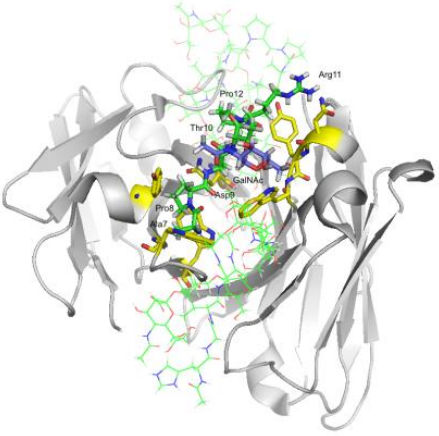
GP 8



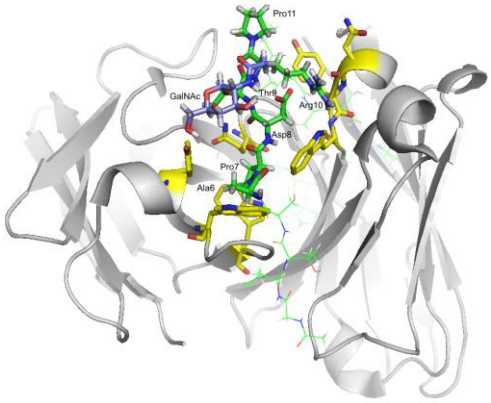
GP 14



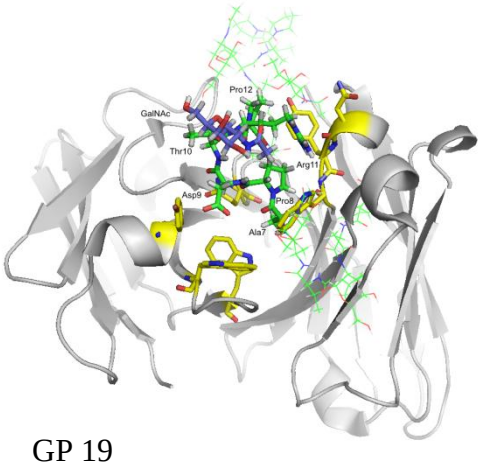
GP 17



GP 18



GP 24



GP 19

Key Binding interaction of peptides 1 (a) glycopeptide 8(GP8), (b) glycopeptide 14 (GP14), (c) glycopeptide 17(GP17), (d) glycopeptide 18(GP18), 5A2K is the SM3 structure obtained from PDB data which comprises of binding with hexpeptide ligand APDTRP having GalNAc residue at the Thr amino acid. Peptide carbon atoms are shown in green. Glycan residues are purple. Carbon atoms of key residues of SM3 are colour yellow surface and Red dashed lines indicate hydrogen bonds between peptide backbones and SM3 antibody.

GP 8: r.m.s.d 2.072, 39 atoms

GP 14: r.m.s.d 2.500, 43 atoms

GP 17:r.m.s.d 2.698, 42 atoms

GP 18: r.m.s.d 3.148, 44 atoms.

GP 19: r.m.s.d 4.591, 45 atoms.

GP 24: r.m.s.d 2.005, 42 atoms.

When, our attention was directed to the results that SM3 lost drastically the binding potency with most MUC1 fragments containing its immunodominant APDTRD motif after the

multiple *O*-glycosylation at the neighbouring potential *O*-glycosylation sites (Micro array results). It was thought that the epitope recognition mode of SM3 antibody might be different from those of other anti-MUC1 mAb's, DF3 or anti-KL5 mAb. Considering that the abovementioned inhibitory effects observed in the multiply *O*-glycosylated MUC1 derivatives on the SM3 recognition were independent from the glycoforms it was hypothesised that the size and flexibility of the binding pocket of the SM3 may be crucial for the recognition of the glycosylated MUC1 fragments. To test this hypothesis, we examined the docking experiments using the NMR structures of the synthetic MUC1 fragments, the monovalent and multivalent MUC1 glycopeptides as the positive and negative models, with the binding groove of the 3D structure of the recombinant SM3 antibody obtained by the X-ray crystallography. Docking with Glycopeptide 24 it revealed that the NMR structure of the monovalent MUC1 fragment having a Tn antigen at the Thr residue in the PDTR region [GVTSAPDT(GalNAcα□→)RPAPGSTAPPAHGVT]16 can be docked into and nicely fit with the surface groove of the recombinant SM3 antibody in a quite similar conformation of the peptide backbone and the geometry of the glycoside linkage to that found in the crystal Structure reported for the simplified MUC1 glycopeptide [APDT(GalNAcα□→)RP] as represented x ray crystallographic structure of SM3(5A2K) Surprisingly, this superimposed view clearly showed that the NMR structure of the overlaid monovalent MUC1 glycopeptide may allow considerably the formation of some of the key intramolecular interactions within the APDTRP region indicated i.e.. (a) the stacking of the Pro with Trp91L, Trp96L, and Tyr32L, (b) the hydrophobic contacts of the side chains of Asp and Arg with Trp33H and Tyr32H, and (c) the hydrogen bonds between the hydroxymethyl group of GalNAc and the side chain of Tyr32L on the SM3 antibody. It is also important to note that two non-glycosylated peptide regions of this longer MUC1 glycopeptide, notably *N*-terminal GVTS

and *C*-terminal APGSTAPPAHGVT moieties, can reside outside of the binding groove of the SM3 and did not disturb the interaction of the APDTRP region with SM3 antibody.

On the contrary, it was indicated that the surface groove of the SM3 antibody cannot fit any other NMR solution structures of the multivalent MUC1 fragments tested herein (shown above). These results demonstrate that SM3 antibody can interact more tightly with the glycosylated MUC1 fragments than naked non-glycosylated MUC1 peptides only when glycosylated in a site-specific manner at the PDTR motif. It is clear that the first *O*-glycosylation with α-GalNAc residue at the Thr renders this area more converged and an extended *Trans*-like structure as shown with GP24. This conformational stabilisation of the immunodominant PDTR motif might allow for the formation of the key intramolecular interactions between the monovalent MUC1 glycopeptide and the surface groove of the SM3 antibody as represented in GP24.

The multiple *O*-glycosylation with Tn antigens made the interaction between the APDTRP moiety of GP19 and the SM3 difficult. It is clear that the unstable α-turn-like conformation at PDTR motif and clustered GalNAc directionalities in the GVTSA segment of the GP19 cannot adopt the surface groove of the SM3 antibody. Since the PDTR region of the GP 14 and GP18 maintains *trans*-like extended structure even after the multiple *O*-glycosylation at other Ser/Thr residues , stearic hindrance due to these multiple *O*-glycosylation may simply block the access of the SM3 antibody to this epitope area as observed in GP14 and GP18 namely negative glycoside cluster effect on the antibody recognition. Although characterization for the surface groove of DF3 and anti-KL6 mAb remains to be elucidated, it is clear that their molecular mechanisms in the interaction with multiply glycosylated MUC1 fragments are entirely different from that of the SM3 antibody.

5-3. Experimental Section

PyMol.

PyMol (TM) Molecular Graphics System Version 0.99rc6 software was used to align glycopeptides against established SM3 structure derived from the PDB. The SM3 structure (5A2K) comprised having binding socket against a hexa peptide ligand APDTRP bearing a GalNAc residue at the Thr amino acid. Least energy structure of each of five MUC1 glycopeptide derived by NMR data was used to for alignment studies using PyMol [6, 12] .

- 5A2K: Crystal structure of scFv-SM3 in complex with APD-GalNAc-RP

  DOI: 10.2210/pdb5a2k/pdb

5-4. Conclusion

In this chapter, docking studies were performed to check the interaction and orientation between the ligand and the protein. All 6 MUC1 glycopeptides were used for docking studies using the PyMOl software. GP 24 showed best hit orientation and interaction between the Glycopeptide and SM3 structure having the least RMS value. Where was GP19 and GP18 showed least hit orientation and interaction having 4.591 , 3.148 as RMS value, and GP17 GP14, GP8 showed average hit structures with 2.698, 2.500 and 2.072 as the RMS value. Hence the fitting orientation and level of interaction of Glycopeptides and SM3 antibody could be understood by this study.

5-5. References

(1) Kitchen, D. B., Decornez, H., Furr, J. R., and Bajorath, J. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications : Abstract : Nature Reviews Drug Discovery. *Nat. Rev. Drug Discov. 3*, 935–49.

(2) Taylor-Papadimitriou, J., Burchell, J., Miles, D. W., and Dalziel, M. (1999) MUC1 and cancer. *Biochim. Biophys. Acta - Mol. Basis Dis. 1455*, 301–313.

(3) Kanekura, T., Sakuraba, H., Matsuzawa, F., Aikawa, S., Doi, H., Hirabayashi, Y., Yoshii, N., Fukushige, T., and Kanzaki, T. (2005) Three dimensional structural studies of alpha-N-acetylgalactosaminidase (alpha-NAGA) in alpha-NAGA deficiency (Kanzaki disease): different gene mutations cause peculiar structural changes in alpha-NAGAs resulting in different substrate specificities and c. *J. Dermatol. Sci. 37*, 15–20.

(4) Corzana, F., Busto, J. H., Jiménez-Osés, G., De Luis, M. G., Asensio, J. L., Jiménez-Barbero, J., Peregrina, J. M., and Avenoza, A. (2007) 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. 129*, 9458–9467.

(5) Corzana, F., Busto, J. H., Jiménez-Osés, G., Asensio, J. L., Jiménez-Barbero, J., Peregrina, J. M., and Avenoza, A. (2006) Supporting information: New insights into α-GalNAc-Ser motif: Influence of hydrogen bonding versus solvent interactions on the preferred conformation. *J. Am. Chem. Soc. 128*, 14640–14648.

(6) Brooks, C. L., Schietinger, A., Borisova, S. N., Kufer, P., Okon, M., Hirama, T., Mackenzie, C. R., Wang, L.-X., Schreiber, H., and Evans, S. V. (2010) Antibody recognition of a unique tumor-specific glycopeptide antigen. *Proc. Natl. Acad. Sci. U. S. A. 107*, 10056–61.

(7) Brister, M. a., Pandey, A. K., Bielska, A. a., and Zondlo, N. J. (2014) OGlcNAcylation and phosphorylation have opposing structural effects in tau: Phosphothreonine induces particular conformational order. *J. Am. Chem. Soc. 136*, 3803–3816.

(8) Osipovitch, M., Lambrecht, M., Baker, C., Madha, S., Mills, J. L., Craig, P. a., and Bernstein, H. J. (2015) Automated protein motif generation in the structure-based protein function prediction tool ProMOL. *J. Struct. Funct. Genomics 16*, 101–111.

(9) Seeliger, D., and De Groot, B. L. (2010) Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J. Comput. Aided. Mol. Des. 24*, 417–422.

(10) Lerner, M. G., Meagher, K. L., and Carlson, H. a. (2008) Automated clustering of probe molecules from solvent mapping of protein surfaces: new algorithms applied to hot-spot mapping and structure-based drug design. *J. Comput. Aided. Mol. Des. 22*, 727–36.

(11) Dokurno, P., Bates, P. a, Band, H. a, Stewart, L. M., Lally, J. M., Burchell, J. M., Taylor-Papadimitriou, J., Snary, D., Sternberg, M. J., and Freemont, P. S. (1998) Crystal structure at 1.95 A resolution of the breast tumour-specific antibody SM3 complexed with its peptide epitope reveals novel hypervariable loop recognition. *J. Mol. Biol. 284*, 713–728.

(12) Martínez-Sáez, N., Castro-López, J., Valero-González, J., Madariaga, D., Compañón, I., Somovilla, V. J., Salvadó, M., Asensio, J. L., Jiménez-Barbero, J., Avenoza, A., Busto, J. H., Bernardes, G. J. L., Peregrina, J. M., Hurtado-Guerrero, R., and Corzana, F. (2015) 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. 54*, 9830–9834.

CHAPTER 6

Concluding remarks

MUC1 is one of the most common mucins present in the human body which undergoes various alterations at the O glycosylation sites. The alteration of glycan plays a vital role for the change in structural diversity and its functions. To study and understand the change in structural pattern and binding affinity against different antibodies due to different glycosylation at the epitope PDTR motif and it's surrounding O glycosylation sites provides detailed information and helps us to develop an effective binding target associated with certain diseases.

In **Chapter 2**, we synthesised a library of 23 MUC1 glycopeptides which were used to study the binding interaction against various monoclonal antibodies and 6 interesting MUC1 glycopeptides which was used for the NMR based structural studies. All compounds were synthesised by micro wave assisted solid phase synthesis for rapid synthesis of compounds, sialylation was done using enzymatic sialylation method. All compounds were purified by reverse phase HPLC method.

In **chapter 3**, we focussed on understanding the level of interaction studies between the synthesised MUC1 compound library and 3 mAb's i.e. KL6, SM3, DF3 by micro array experiment. Glycopeptide samples showed remarkable interaction against antibodies, which were mainly responsible due to the glycosylation at PDTR motif or the surrounding O glycosylation sites.

In **chapter 4**, after analysing the micro array data we focussed on 6 interesting MUC1 glycopeptide to study the structural changes due to change in glycosylation pattern. Overall analysis of structure showed that glycosylation patter changes at not only the PDTR motif bring out the change in structure but also the neighbouring Ser/Thr O glycosylation sites also bring about effective changes in the peptide backbone due to which the interaction mechanism can be clearly understood.

Chapter 5, Docking studies by PyMol showed better information about the alignment of the glycopeptides with SM3 mAb. SM3 being one of the known existing crystallographic structure also used in micro array studies was used against the synthesised glycopeptides used for NMR studies. Alignment and orientation of the glycopeptides and SM3 antibody were clearly observed. These docking studies provided an insight into the 3D dimensional studies which supported the 2D NMR structures.

This work accelerates and influences for further mucin based MUC1 glycopeptide studies.

AKNOLEDGEMENTS:

The present dissertation is described author's study from 2012-2016 at Graduate school of Life Science, Hokkaido University.

Immeasurable appreciation and deepest gratitude for the help and support extended to the following persons who is in one way or another have contributed in making this study possible.

Professor Shin-Ichiro Nishimura, principle investigator for his support, advices, guidance, valuable comments, suggestion and provision that benefited me much in the completion and success of this study.

Professor Kenjo Monde for his time and effort in checking and giving suggestions on the dissertation manuscript.

Assistant Professor Hiroshi Hinou for encouraging me and sharing his knowledge in doing research and helped in analysis of data and its statistical computations.

Ms. S. Oka and T. Hirose at the centre of instrumental analysis, Hokkaido University, for ESI-MS measurement and amino acid analysis.

Dr. Shun Hayakawa for helping me during NMR experiments and sharing his knowledge.

To all Nishimura's Laboratory colleagues for cooperation daily supports and encouragements.

My family Prof. K.S Rangappa, Ms. Poornima Devi Rangappa, Mr. Shreyas Rangappa and Dr. Namrata Shreyas for their moral support and warm encouragements.