Ana Sofia de Campos Grosso

Licenciatura em Bioquímica

Molecular recognition of tumor-associated antigens by lectins and antibodies

Dissertação para obtenção do Grau de Mestre em Química Bioorgânica

Orientador: Doutora Filipa Marcelo, Investigadora Auxiliar, Faculdade de Ciências e Tecnologia- Universidade Nova de Lisboa

Júri:

Presidente: Prof. Doutora Paula Cristina de Sério Branco
Arguente: Doutora Maria Angelina de Sá Palma
Vogal: Doutora Filipa Margarida Barradas de Morais Marcelo

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro 2017

2017

Molecular recognition of tumor-associated antigens by lectins and antibodies
Ana Grosso

Ana Sofia de Campos Grosso

Licenciatura em Bioquímica

Molecular recognition of tumor-associated antigens by lectins and antibodies

Dissertação para obtenção do Grau de Mestre em Química Bioorgânica

Orientador: Doutora Filipa Marcelo, Investigadora Auxiliar, Faculdade de Ciências e Tecnologia- Universidade Nova de Lisboa

Júri:

Presidente: Prof. Doutora Paula Cristina de Sério Branco
Arguente: Doutora Maria Angelina de Sá Palma
Vogal: Doutora Filipa Margarida Barradas de Morais Marcelo

Molecular recognition of MUC1 tumor-associated antigens by lectins and antibodies

Copyright © Ana Sofia de Campos Grosso, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa
A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Agradecimentos

Este foi um ano maravilhoso, onde aprendi imenso sobre coisas que não teria aprendido de outra forma, especialmente sobre açúcares. Os açúcares são biomoléculas fantásticas, mas raramente mencionados durante as unidades curriculares de Mestrado ou Licenciatura. Devido a isto, a esta oportunidade de aprender não só sobre os açúcares, mas também sobre técnicas experimentais que nunca tinha feito, RMN e ter-me dado a oportunidade de ter conhecido o seu grupo de investigação, gostaria de agradecer à Investigadora Filipa Marcelo, por ter aceite orientar-me durante esta minha viagem no conhecimento. Obrigada por toda a ajuda ao longo deste ano, as boleias e o computador que me emprestou.

Outra pessoa muito importante neste ano e nesta maravilhosa experiência foi a bolseira Ana Diniz. A Ana ajudou-me a vários níveis, quer no laboratório, onde estava sempre um passo à frente e me ensinou imensas coisas, como também facilitou o meu processo de integração no grupo e laboratório. Ela é realmente o espírito deste laboratório e está sempre disposta a ajudar todos e a facilitar a vida de todos no laboratório. Para além da Ana, também tenho que agradecer à aluna de Doutoramento Helena Coelho, por me ter ajudado na última parte da tese. Não conheci muito bem a Helena, mas ela esteve sempre disposta em ajudar-me em tudo o que precisei. Por isso, obrigada pelo vosso esforço em me ajudar ao longo do meu percurso durante a minha tese.

Para além destas pessoas, devo dizer que todo o grupo de investigação foi muito acolhedor, simpático e prestativo sempre que tive problemas. O Doutor Jorge Dias ajudou sempre que tinha dúvidas em relação à parte de purificação e expressão de proteínas e HPLC; a Doutora Ana Sofia Ferreira ajudou quando tive problemas informáticos; o Professor Eurico Cabrita ensinou imensa coisa sobre RMN; o aluno de Doutoramento Micael Silva tornou o ambiente do laboratório mais cheio de vida; os restantes alunos de Doutoramento com quem não interagi muito também foram sempre muito simpáticos e dispostos a ajudar.

Fora do grupo tenho de agradecer à minha Tia Catarina por me disponibilizar um sítio para dormir e me ouvir e aconselhar sempre que havia algo que me preocupava e aos meus pais por sempre terem se interessado pela minha educação. Às minhas maninhas que serviram para me distrair e esquecer da tese.

A todos um sincero obrigado!!

Abstract

Every living cell on Earth is covered by glycans. They are inserted in proteins and lipids by a posttranslational modification called glycosylation. Their recognition by specific receptors is translated into distinct biological signals.

In cancer cells, a misregulation in expression and/or activity of glycosyltransferases, alters the mechanism of glycosylation, creating new glycan epitopes dubbed tumor-associated carbohydrate antigens (TACAs). These are recognized by various receptors, playing a major role in tumor immune responses and metastasis. To target cancer-associated glycan phenotype is crucial to disentangle the molecular recognition process that involves TACAs recognition and biosynthesis.

Therefore, NMR techniques were employed to investigate distinct glycan-protein systems: i) the molecular interactions between a mucin-1 (MUC1) related Tn-glycopeptide mimetic containing a non-natural amino acid and distinct antibodies by saturation transfer-difference (STD-NMR); ii) the molecular interactions between galectin-3 (Gal-3) and TF-antigen (TF-Thr and TF-peptide), by heteronuclear single quantum coherence $^{1}H,^{15}N$-HSQC titrations, STD-NMR and line broadening analysis and iii) the glycosylation of MUC1 tandem repeated protein $(G^{1}VT^{3}S^{4}APDT^{8}RPAPGS^{14}T^{15}APPAH^{20})_4$ by GalNAc-T3 using $^{1}H,^{15}N$-HSQC and STD-NMR.

In i), the STD-NMR binding experiments show that all antibodies under study recognize the Tn-glycopeptide mimetic and point out structural differences that explain antibodies' binding preferences.

In ii), the $^{1}H,^{15}N$-HSQC titrations experiments indicate that Gal-3 binds both TF-derivatives. The dissociation constant K_D estimated for both through chemical shift analysis also shows the same range of affinity (275 μM and 413 μM for TF-antigen and TF-peptide, respectively). STD-NMR results demonstrate that the protons from galactose in the TF-moiety govern the recognition process of Gal-3.

In iii), the $^{1}H,^{15}N$-HSQC experiments of MUC1 in presence of GalNAc-T3 show that the enzyme has preference to glycosylate first the Thr at –GVTS-, followed by the residue Thr at –GSTA-. STD-NMR confirms the cooperative mechanism between the lectin and catalytic domain of GalNAc-T3.

Keywords: Carbohydrate-protein interactions; NMR Spectroscopy; Galectins; Antibodies; GalNAc-transferases; Mucin-1.

Resumo

Os açúcares cobrem todas as células vivas do planeta. São inseridos em proteínas e lípidos por uma modificação pós-translacional, a glicosilação. A sua tradução em sinais biológicos ocorre através do reconhecimento por certos receptores.

No cancro, uma disfunção na regulação na expressão e/ou actividade de glicosiltransferases altera o mecanismo de glicosilação, criando novos epitopos de açúcares, os antigénios de carbohidratos-associados a tumor (*TACAs*). Estes são reconhecidos por vários receptores, influenciando a resposta imune e promovendo metástases. Por isso, é importante perceber o processo de reconhecimento molecular que envolve o reconhecimento e biossíntese dos *TACAs*.

Para tal, técnicas de RMN foram utilizadas para investigar vários sistemas açúcar-proteína: i) as interações moleculares entre um Tn-glicopéptido mucina-1 (MUC1) com um aminoácido não natural e vários anticorpos, por *Saturation Transfer-Difference* (STD-RMN), ii) as interações moleculares entre a Galectina-3 (Gal-3) e o antigénio-TF (TF-Thr e péptido-TF), por titulações *Heteronuclear Single Quantum Coherence* ^{1}H,^{15}N-HSQC, STD-RMN e análise de largura de linha e iii) a glicosilação da proteína MUC1 ($G^{1}VT^{3}S^{4}APDT^{8}RPAPGS^{14}T^{15}APPAH^{20})_{4}$ pela GalNAc-T3 usando ^{1}H,^{15}N-HSQC e STD-RMN.

Em i), as experiências de STD-RMN mostraram que os anticorpos estudados conseguem reconhecer o Tn-glicopéptido e que as diferenças estruturais explicam as diferentes preferências de interação.

Em ii), as experiências de titulação de ^{1}H,^{15}N-HSQC mostraram que a Gal-3 interage com ambos os derivados-TF. A constante de dissociação (K_D) estimada através da análise dos desvios químicos mostrou a mesma gama de afinidade para o antigénio-TF e o péptido-TF de 275 μM e 413 μM, respectivamente. STD-RMN mostrou que os protões da Galactose do antigénio-TF governam o processo de reconhecimento da Gal-3.

Em iii), as experiências de ^{1}H,^{15}N-HSQC da proteína MUC1 na presença de GalNAc-T3, mostraram que a enzima tem preferência na glicosilação do resíduo Thr de -GVTS-, seguido pelo resíduo Thr de -GSTA-. STD-RMN provou que o domínio lectina e catalítico da GalNAc-T3 têm um mecanismo cooperativo.

Palavras-chave: Interações açúcar-proteína; Espectroscopia de RMN; Galectinas; Anticorpos; GalNAc-Transferases; Mucina-1.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Ala (A)- Alanine

Arg (R)- Arginine

Asn (N)- Asparagine

Asp (D)- Aspartic acid

BLI- Bio-layer Interferometry technique

BMRB- Biological Magnetic Resonance Bank

CBP- Carbohydrate Binding Protein

Cys (C)- Cysteine

CRD- Carbohydrate Recognition Domain

CSP- Chemical Shift Perturbation

Da- Dalton

DTT- Ditriothreitol

EDTA- Ethylenediaminetetraacetic Acid

ER- Endoplasmatic Reticulum

FL- Full Length

Gal- Galactose

Gal-3- Galectin-3

GalNAc- Acetylgalactosamine

GalNAc-Ts- Uridine Diphosphate *N*-Acetylgalactosamine polypeptide *N*-Acetylgalactosaminyl-Transferases

Gln (Q)- Glutamine

Glu (E)- Glutamic acid

Gly (G)- Glycine

His (H)- Histidine

Hnv- Hydroxy-Norvaline

HPLC- High-Performance Liquid Chromatography

HSQC- Heteronuclear Single Quantum Coherence

Ile (I)- Isoleucine

IPTG- Isopropyl β-D-1-Thiogalactopyranoside

ITC- Isothermal Titration Calorimetry

K- Kelvin

K_a- Association Constant

K_D- Dissociation Constant

LB- Luria-Bertani Medium

Leu (L)- Leucine

Lys (K)- Lysine

mAb- Monoclonal Antibody

MALDI-TOF- Matrix-Assisted Laser Ionization- Time-of-flight Mass Spectrometer

MGL- Macrophage Galactose-type Lectin

MUC1- Mucin-1

MUC1-4TR- A mucin-1 construct with 4 Tandem Repeats of 20 amino acids

Neu5Ac- *N*-Acetylneuraminic Acid

NHAc- *N*-Acetyl group

NMR- Nuclear Magnetic Resonance

NOE- Nuclear Overhauser Effect

NOESY- Nuclear Overhauser Effect Spectroscopy

PBS- Phosphate Buffer Saline

PDB- Protein Data Bank

Phe (F)- Phenylalanine

ppm- part per million

Pro (P)- Proline

PTS region- Proline, Threonine and Serine region

rpm- revolutions per minute

SDS-PAGE- Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SEA domain- sea urchin sperm protein, enterokinase and agrin domain of the glycoprotein MUC1

Ser (S)- Serine

STD- Saturation Transfer Difference

STn- Sialyl Tn

STF- Sialyl TF

T1- Longitudinal Relaxation Time

T2- Transversal Relaxation Time

TACAs- Tumor-Associated Carbohydrate Antigens

TF or T-antigen- Thomsen-Friedenreich antigen

Thr (T)- Threonine

Tn-antigen- Thomsen nouvelle antigen

TOCSY-Total Correlation Spectroscopy

TRIS-Tris(hydroxymethyl)aminomethane

Trp (W)- Tryptophan

TSP- Trimethylsilylpropanoic acid

Tyr (Y)- Tyrosine

UDP- GalNAc- Uridine Diphosphate *N*-Acetylgalactosamine

Val (V)- Valine

1. Introduction

1.1. Carbohydrate-protein interactions

Carbohydrates coat every living cell on Earth. Furthermore, carbohydrate interactions with extracellular receptors, like lectins, antibodies, enzymes (Figure 1.1) play a key role in several biological processes, such as, cellular transport and adhesion, cell signaling processes, cell-cell communication, immune response, hyperacute rejection of tissue transplants of nonhuman sources, fertilization, tissue maturation, apoptosis, blood clotting, infection by bacterial and viral pathogens, tumor growth and metastasis [1]. Therefore, determination of the structural and conformational features that govern the molecular recognition of these biomolecules is of paramount importance [2-7]. So, understanding the structure and function of carbohydrates is crucial to realize their function in health and disease, as well as, to develop new glycan-based therapeutics.

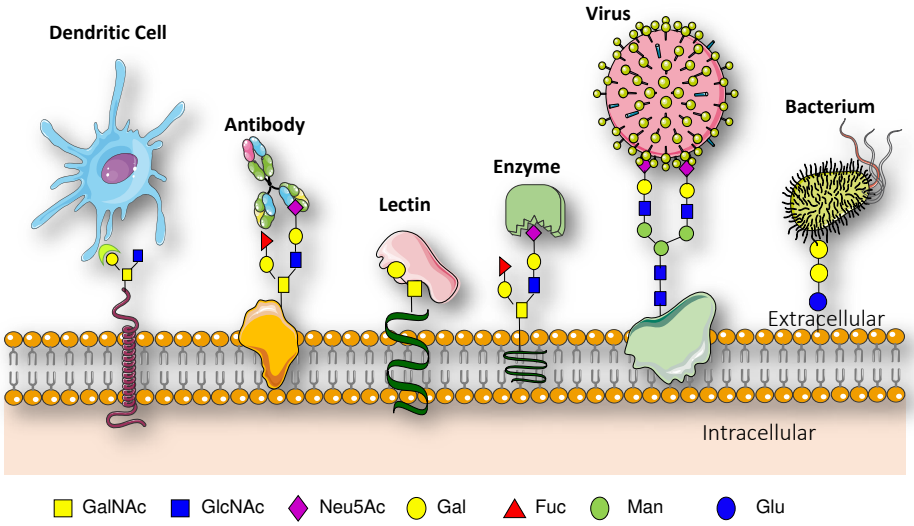


Figure 1.1: Examples of carbohydrate-protein interactions. GalNAc represents *N*-acetylgalactosamine; GlcNAc: *N*-acetylglucosamine; Neu5Ac: *N*-Acetylneuraminic Acid; Gal: Galactose; Fuc: Fucose; Man: Manose and Glu: Glucose [1].

Carbohydrate-protein complexes can be achieved by different kinds of forces, due to the amphipathic character of oligosaccharides, in the recognition process [2,3]. Carbohydrate-protein interactions are more dynamic than other protein-ligand complexes and their affinity arises from several weak interactions. Carbohydrate-binding specificity results in the balance of electrostatic, hydrogen bonding and hydrophobic interactions between the protein, solvent and the carbohydrate, resulting in changes of enthalpy and entropy after binding [8].

The hydroxyl groups of oligosaccharides can make intermolecular hydrogen bonds to side chains of polar amino acids. The residues more frequently found in intermolecular hydrogen bonds with carbohydrates are Asp, Asn>Glu>Arg, His, Trp, Lys>Tyr, Gln>Ser, Thr [9]. Hydrogen bonds play a major role in protein-carbohydrate interactions. They convey stability, give specificity and control

dynamics. They are stable enough to contribute to the affinity and are highly directional, which gives them specificity. Furthermore, the characteristic stereochemical arrangement of hydroxyl groups plays a key role in protein specificity towards a given sugar type [10,11]. In addition, there are various characteristic phenomena in carbohydrate-protein interactions: i) cooperative hydrogen bonding, where the same hydroxyl group can act as the donor and acceptor of hydrogen bonds; ii) bidentate hydrogen bonds established between two adjacent hydroxyl groups of a sugar and both carboxylate oxygens of either aspartic or glutamic acids; iii) salt bonds between charged residues of some sugar units and protein residues of opposite charge; iv) some proteins require coordination of a divalent cation to connect certain sugar hydroxyls and negatively charged aspartates or glutamates [11].

In protein-carbohydrate complexes, it is also possible to observe van der Waals interactions between the carbohydrate and nonpolar aromatic or aliphatic residues, like Trp, Phe, Tyr, Leu, Val and Ala [9]. This interaction is very common in recognition sites of proteins and mediates the interaction between carbohydrates and the aromatic residues of the side chains of Trp, Tyr, Phe and His of the receptors. They have origin in dispersion forces, which have an impact on the enthalpy of the process. Depending on the stereochemistry of the saccharide the nonpolar C-H patch can interact with the aromatic residues of the protein side chains, by van der Waals, CH-π and hydrophobic interactions. These interactions have an entropic contribution, by protecting both apolar surfaces from the bulk water and an enthalpic contribution by formation of three non-conventional hydrogen bonds between the sugar hydrogens and the aromatic ring. The nature and structure of the solvent molecules and the way they behave (individual or bulk) are important to determine the specificity or lack of interaction between two different entities or between any entity and the solvent [2,3,11,12].

Besides, intermolecular interactions between carbohydrates and proteins, interactions with water molecules at the binding site also provide additional intra and intermolecular connections important to stabilize the complex. The water molecules occupy the sugar-binding site of free proteins, orienting the key residues, thus the binding site can be preorganized to accommodate the ligand, in a way that the entropy loss is minimized in the formation of the complex [9].

However, the level of importance of carbohydrate-protein interactions depends on the architecture of the binding site and chemical nature of the interacting sugar. The patterns of recognition for neutral, positively charged and negatively charged sugars are very distinct. Another factor is the number of valency used in the formation of these interactions [9]. Normally, carbohydrate-protein interactions are very selective, but with very weak affinity. This affinity can be enhanced by multivalent interactions. These interactions are characterized by the binding of multiple ligands through noncovalent interactions on multiple receptors, simultaneously [13,14]. Multivalent interactions convert weak monovalent interactions into strong, highly specific and thermodynamic and kinetic stable recognition events [14]. A multivalent ligand has multiple copies of a recognition element in a central scaffold, in this case, multiple carbohydrates. (Figure 1.2) These types of interactions, with multivalent ligands can access receptor-binding modes inaccessible to monovalent compounds [15]. Therefore, are extremely important in various recognition processes, like the interactions between carbohydrate and lectins [16].

Receptors can also adopt multivalent presentation. In fact, multivalency of receptors is as important as, the multivalency of ligands. Multivalent recognition of substrates requires a large contact area and enough complementary recognition sites, which can be achieved by multiple interactions through a

combination of amino acids and nucleotides [17]. Multivalency of receptors is important in many biological phenomena, such as agglutination of carbohydrates by lectins and antigen-antibodies interactions [17,11].

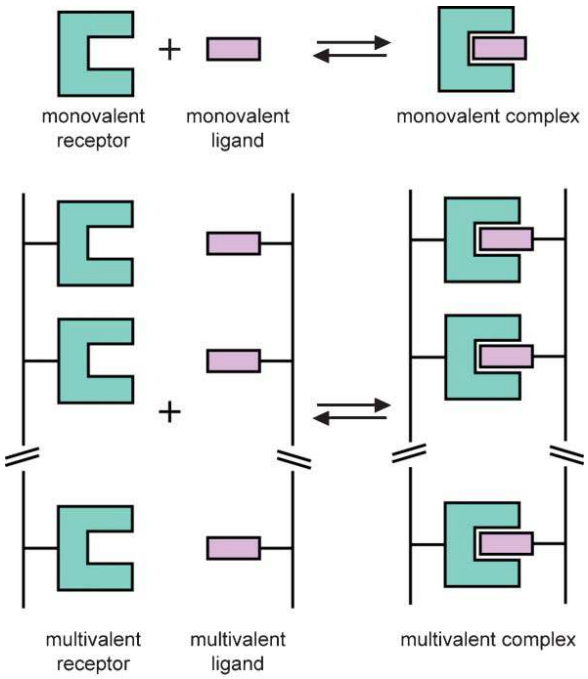


Figure 1.2: Differences in the binding interactions between a monovalent interaction and a multivalent interaction. [14].

In comparison to proteins, carbohydrates are flexible molecules containing several bounds with free rotation, so they often populate multiple conformational families, requiring both temporal and spatial descriptors to quantify their conformational properties [18,6].

X-ray crystallography is used to extract structural data (conformational features and intermolecular interactions) from crystalline molecules, but depends on the size of the molecule and its ability to form crystals. For small disaccharides or oligosaccharides crystallization in the unbound state may be very difficult. However, oligosaccharides can be crystallized covalently linked to the peptide chain or as ligands complexed with the macromolecular receptor. Noteworthy, the glycan moiety is often poorly resolved, with only one or two carbohydrate residues seen [19]. Furthermore, the K_a values of carbohydrate complexes are typically weak, ranging from 10^2 to 10^6 M^{-1}, making the crystallization of the complex sometimes very difficult [4].

Therefore, a multidisciplinary approach that combines distinct structural techniques like X-ray crystallography, NMR spectroscopy, molecular modeling protocols complemented with biophysical binding methods, are the best choice to achieve the binding and dynamics of the carbohydrate-protein recognition process.

1.2.Glycosylation on proteins

The addition of glycans can modulate the structure and function of the proteins where they are attached, by altering their 3D structure, making modifications on their surface and extending as large molecular masses away from them. By directly interacting with protein surfaces, they partially occlude regions of the protein surface and reduce the protein dynamics, due to their large mass and inertial resistance [8,20]. Hence, carbohydrates linked to the proteins can alter their properties and biological activity, by making them more soluble, influencing their stability and protecting them from proteolysis. They are inserted by a co or post-translational modification of proteins, called glycosylation. This modification is the most common and complex post-translational modification in proteins and is classified by the way the carbohydrate is bonded to the protein. The most common forms of glycosylation are the *N*-glycosylation and the *O*-glycosylation [7, 21].

While, both types of glycosylation have the same purpose, the covalent addition of carbohydrates to different biomolecules, they have different mechanisms. The *N*-glycosylation is initiated in the endoplasmatic reticulum (ER) with the transfer of preassembled blocks of lipid-linked carbohydrates, onto the *N*-atom of the side chain of an asparagine in forming proteins and terminates in the Golgi apparatus. It requires a sequon, a consensus sequence motif of Asn-Xaa-Ser/Thr, where Xaa can be any amino acid except Pro; whereas in *O*-glycosylation, there is no defined consensus sequence, but generally the residues glycosylated are Ser/Thr located in Pro-rich sequences, especially in the position -1 and +3 of the glycosylation site. These residues must be accessible and exposed, in finished and folded proteins. It's initiated in the Golgi apparatus and starts with the addition of a GalNAc residue to the hydroxyl oxygen of a Ser/Thr. After this step, other monosaccharides are added in a complex and controlled stepwise enzymatic manner to build linear or branched structures, forming *O*-glycans smaller and less branched than *N*-glycans [22, 23, 20, 24].

Most glycoproteins exhibit both types of glycans, since both have distinct effects in their structure and properties. *N*-glycans are important during the folding process and in the detection of incorrectly folded proteins. They also decrease the conformational mobility of the peptide backbone. *O*-glycans increase the stability in helices [7].

1.3.Glycosylation and cancer

Alteration in the glycosylation pattern of proteins and lipids has a strong impact in the biology of the cell with strong impact in many neoplastic transformations, like cancer [25].

In cancer, the altered glycosylation of glycoproteins and glycolipids, affect both *N*- and *O*-glycans, during cancer progression. Some of the most common alterations are excessive sialylation and fucosylation of glycans, increased branching of *N*-glycans and incomplete biosynthesis, resulting in truncated glycans [25]. *N*-glycans, in cancer cells are associated with invasion and metastization. *O*-glycans are also highly overexpressed. The major carriers of *O*-glycans in cancer are mucins, glycoproteins with 50 to 90 % of their molecular mass as *O*-linked glycans [26,25], due to their repetitive sequences rich in serine and threonine (tandem repeats), known as the PTS region. Normally, they contribute to the protective mucous gel through *O*-glycosylated tandem repeats that form rod-like structures, extended from the cell surface [27]. However, when aberrant glycosylated, mucins facilitate cell adhesion during tumor metastasis and alter the function of proteins interacting with their carbohydrate

moieties [28]. In cancer cells, mucins are overexpressed and aberrantly expressed, in contrast to their restricted and tissue-specific expression in normal cells. These modifications of mucin expression sometimes are linked to modified glycosylation. Their overexpression amplifies cancer cells' surface alterations and elevated concentrations are associated with elevated tumorigenesis and poor prognosis, making them good cancer biomarkers [25].

1.3.1. Mucin-1 (MUC1)

Mucin-1 (MUC1) was the first mucin to be cloned from mammary carcinomas and is also the best characterized to date [29]. MUC1 is a transmembrane glycoprotein of 500-1000-kDa heavily *O*-glycosylated expressed in the apical surfaces of ductal and glandular epithelial cells [30]. The full-length MUC1 encodes the N- and C-terminal subunits (Figure 1.3). The N-terminal subunit of MUC1 is composed by most of the extracellular domain and contains a variable number of 20-120 tandem repeats (VNTR) of a polymorphic sequence of 20 amino acids, with five potential sites for *O*-glycosylation in serine and threonine residues (HGV**TS**APD**T**RPAPG**ST**APPA, underlined the 5 potential sites of glycosylation) [31, 32, 33, 28]. The MUC1 extracellular domain extends 200-500 nm above the plasma membrane, beyond the 10 nm of glycocalyx [33]. The VNTR of MUC1 has three relevant regions: the GVTSA, a good substrate for GalNAc transferases; the PDTR the most immunogenic domain of MUC1 and a well-known epitope recognized by several anti-MUC1 antibodies and the GSTA which is recognized by different antibodies and represents a potential tool for diagnosis and therapeutic applications [34].

The C-terminal subunit is composed by 58 amino acid residues of an extracellular domain, a single hydrophobic spanning transmembrane domain (28 amino acids) and a cytoplasmatic tail of 72 amino acids with various tyrosine, serine and threonine phosphorylation sites that can bind to various signaling motifs (kinases and growth factor receptors) [31]. The later interactions affect and regulate several cancer processes, like the proliferation, apoptosis and transcription of various genes [32, 33, 35, 28].

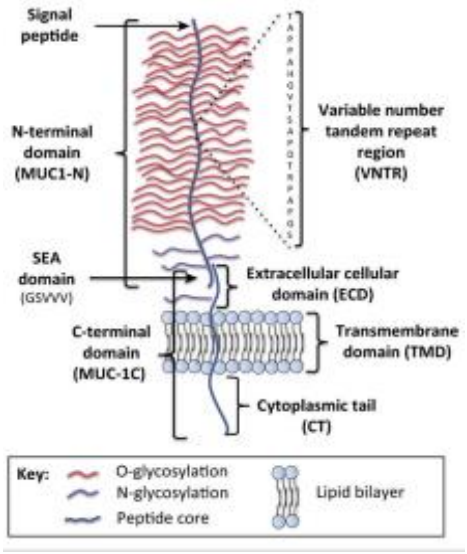


Figure 1.3: Schematic representation of the structure of MUC1. The N-terminal constituted by the VNTR domain and the SEA domain, while the C-terminal composed by the Transmembrane domain, Cytoplasmic domain and some residues of the Extracellular domain [31].

In normal cells (Figure 1.4 A), MUC1 is restricted to the luminal side of the cells [33]. The protein backbone is extensively O-glycosylated and characterized by complex sugar chains, usually branched core 2 O-glycans extending from an α-O-GalNAc unit directly linked to the hydroxyl group of either serine or threonine. Typically, core-2 can be elongated by several lactosamine units and usually terminate with fucose and/or sialic acid [30]. The clustering of O-linked negatively charged glycans leads to an extended protein core, long and with a rod-like molecule far above the plasma membrane. It has a protective role in the modulation and retention of secreted mucins, by providing a scaffold for the presentation of glycans for recognition of bacteria and viruses. MUC1 also lubricates the cell, keeps it hydrated and protects it from pathogen invasion [32, 36, 35, 30]. The peptide core is masked by the sugar moieties, which protects it from proteolytic cleavage by environmental enzymes and stabilizes mucins at the cell surface [31].

In tumor cells (Figure 1.4 B), the cell polarity is lost, resulting in MUC1 expression in the entire cell surface [33]. This overexpression contributes to an aggressive tumor phenotype where the extended peptide core inhibits the normal cell-cell and cell-matrix interactions. Furthermore short, truncated and prematurely *O*-glycans are now expressed acting as ligands for protein receptors on endothelial cells and strongly contributing to immune invasion and metastasis [35, 36, 31, 30].

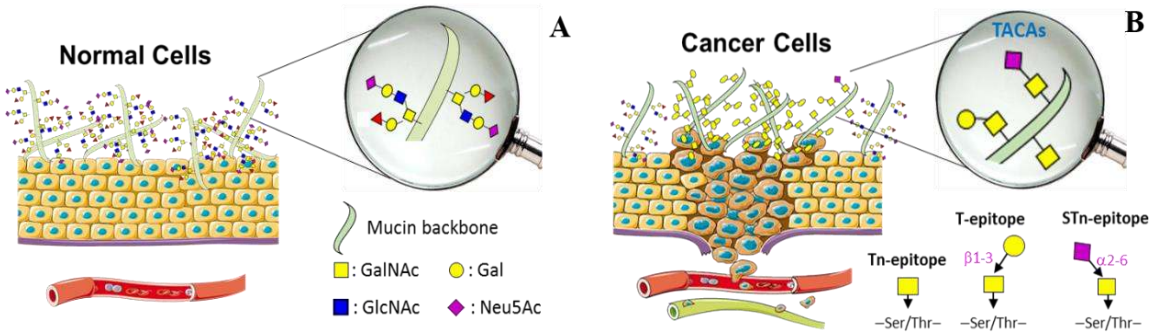


Figure 1.4: **A**- MUC1 in normal cells, normally glycosylated and expressed only at the cell surface; **B**- MUC1 in cancer cells, where it is overexpressed and aberrantly glycosylated. [37]

1.3.1.1. MUC1 and tumor-associated carbohydrate antigens (TACAs)

These *O*-glycans dubbed tumor-associated carbohydrate antigens (TACAs) are absent in normal cells and commonly present in 90 % of the human carcinomas. The more common TACAs are the Tn (α-*O*-GalNAc-Ser/Thr), sialyl Tn (STn) (Neu5Acα2-6GalNAcα-*O*-Ser/Thr), TF (Galβ1-3GalNAc) and sialyl TF (STF) (NeuAcα2-3Galβ1-3GaNAc-*O*-Ser/Thr)-antigens (Figure 1.5). They are present not only in MUC1, but also in various secreted and membrane mucin proteins. Aberrant glycosylation of mucin glycoproteins is associated to a misregulation in the expression levels of glycosyltransferases, which may be caused by mutation, inactivation or lack of their functional chaperone proteins, the overexpression of sialyltransferases and disorganization of secretory pathway organelles in cancer cells [30, 38, 25].

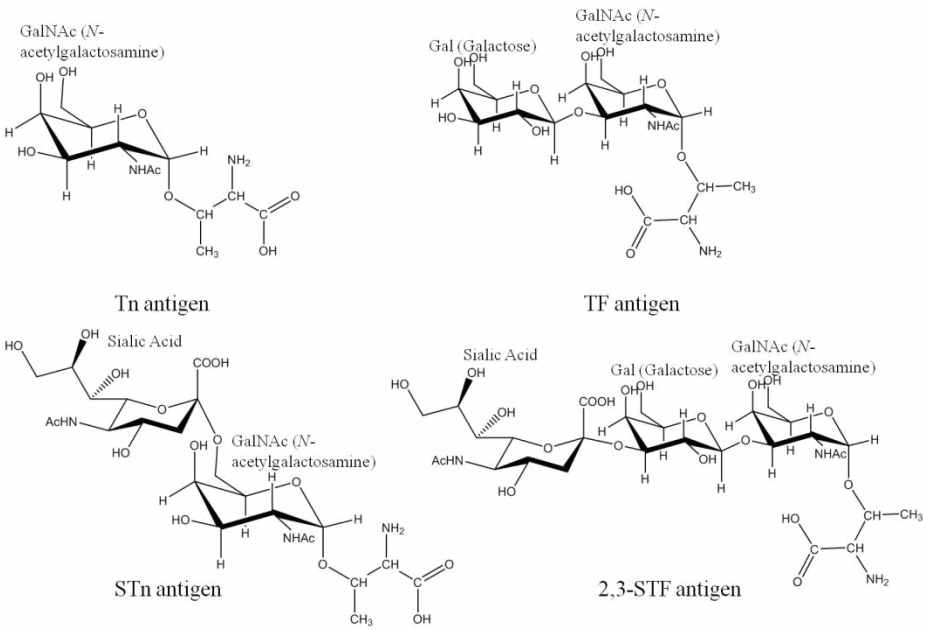


Figure 1.5: Structure of TACAs.

1.3.1.2. Interaction of TACAs with antibodies and lectins

TACAs are exposed to the immune system and have been used to design glycan-based cancer vaccines, including MUC1-based antitumor vaccines, with the main goal to educate the immune system to create antibodies [39]. In addition, several structural and recognition studies have been carried out to understand the minimal features that modulate glycan-antibody recognition [36, 30, 40]. In previous studies, this research group has already studied the interactions of Tn-antigen and two families of antibodies: anti-MUC1 and anti-Tn. The anti-MUC1 antibodies bind the MUC1 peptides in a strict peptide-sequence-dependent manner, with improved binding affinity, after the introduction of the GalNAc residue at the more immunogenic region of MUC1 sequence (PDTRP) [41]. The anti-Tn monoclonal antibodies only recognize Tn-glycopeptides. However, the type of residue glycosylated (Ser/Thr) modulates the binding. The anti-Tn antibodies used on the study have preference to the Tn linked to the serine residue [41].

TACAs can also be sensed by endogenous lectins, which translate the appearance of the new sugar signal into cellular activities [42, 43]. In this context, this research group studied the interactions between the Tn antigen and human macrophage galactose-type lectin (MGL) by STD-RMN experiments in tandem with molecular dynamic simulations [44]. MGL is a C-type lectin expressed in the surface of monocyte-derived immature dendritic cells and macrophages and is proposed to act as an immune-modulatory receptor. This lectin binds the Tn antigen carried by MUC1 in colon cancer cells and activates dendritic cells to uptake these antigens in a MGL-mediated way [45, 46]. Structural studies demonstrate that MGL binds preferentially to Tn and sialyl-Tn antigens and recognition occurs mainly through the GalNAc moiety [44, 47, 48]. However, some protons of close amino acids also received saturation, which means that MGL also contacts with amino acids. MGL also binds galactose, however has a high degree of selectivity towards GalNAc. This is explained by the fact that the NHAc group of the GalNAc residue forms additional hydrogen bonds and CH-π interactions with MGL residues [44].

Another very studied and important lectin in cancer progression is galectin-3 (Gal-3), which is an animal lectin that belongs to the family of carbohydrate-binding proteins (CBPs) with affinity for β-galactosides commonly present in glycoproteins [49-51]. Gal-3 has approximately 31 kDa and is composed by a conserved sequence of approximately 130 amino acids, responsible for the carbohydrate-binding activity, the carbohydrate region domain (CRD) and a flexible non-lectin domain made of 7-14 tandem repeats of a short nine proline/glycine/tyrosine-rich consensus sequence with a total of 120 amino acid residues [52, 53, 49, 54]. Gal-3 is the only galectin classified as Chimera-type with the ability to form pentamers after binding to multivalent carbohydrates by self-association of the *N*-terminal non-lectin region (Figure 1.6) [50, 55].

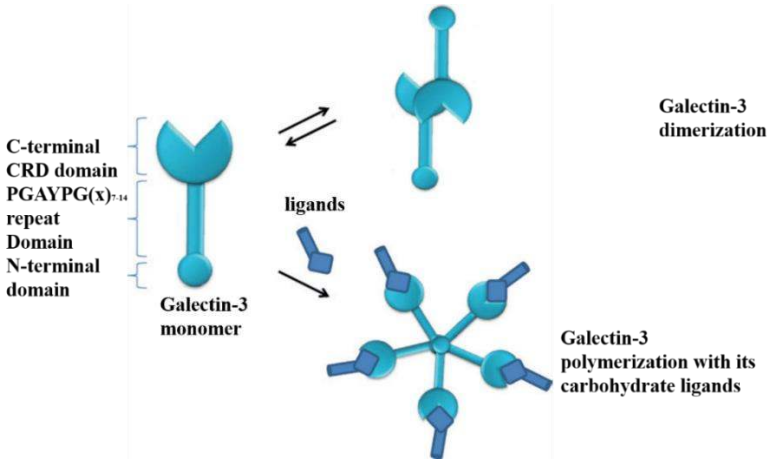


Figure 1.6: Structure of galectin-3 monomer and its pentamer form. [50]

In cancer, Gal-3 is widely expressed by epithelial and immune cells and depending in the place, where it is expressed, its functions will be different [51]. Gal-3 exhibits specific affinity for the Thomsen-Friedenreich (T or TF) antigen. This antigen contains a galactose residue β1-3 linked to an α-*N*-acetylgalactosamine linked to a serine or threonine residue of a glycoprotein. As Tn-antigen, TF-antigen is normally occulted by further glycosylation, however in cancer cells is exposed to interact with Gal-3. These interactions are responsible for the enhancement of cancer cell-endothelial adhesion and promote metastasis [28]. The binding of TF-antigen by Gal-3 was previously explored [56, 57]. X-ray crystallography of the complex showed that the glycans bind in a concave surface of β-strands, especially the S4-S6 strands. The crystal structure showed that the Gal-moiety interacted with the residues His158, Asn160, Arg162, Asn174, Trp181 and Glu184 by hydrogen bonds or van der Waals contacts, while the GalNAc residue uses a hydrogen bond network of two water molecules, where one of them interacts with the residue Arg186 and the oxygen in the carbohydrate ring of GalNAc and the other interacts with the residue Glu165 and the NHAc group of GalNAc (Figure 1.7) [56, 57]. Recently isothermal titration calorimetry (ITC) studies showed that the presentation of the carbohydrate by the natural peptide backbone contributes to the enhanced affinity for Gal-3 [57].

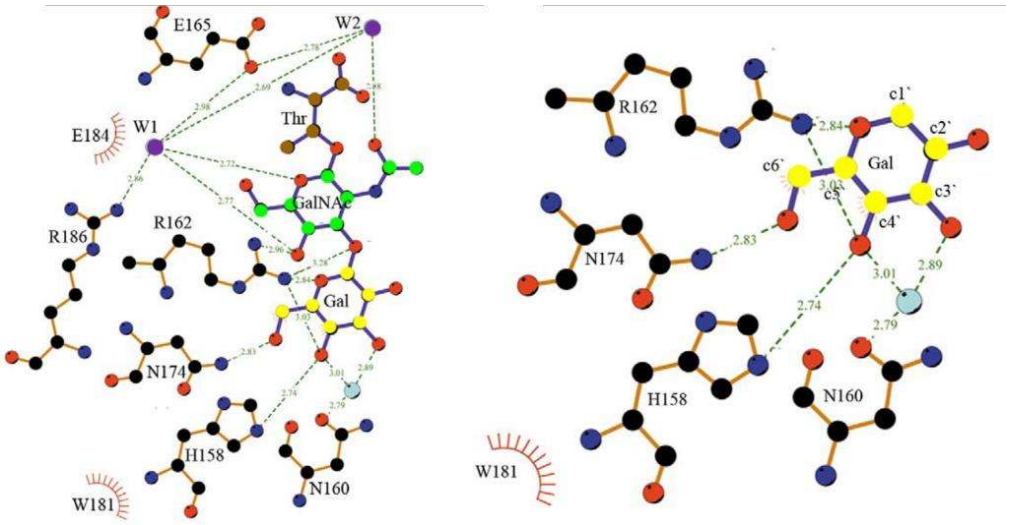


Figure 1.7: Representation of the X-Ray Crystallography structure obtained with the different types of interactions between Gal-3 and TF-antigen. [56]

1.4.GalNAc-Ts and the biosynthesis of O-glycans

Mucin-type *O*-linked glycans constitute 80 % of all mammalian cancer antigens. The glycosylation pathway of mucins is initiated with the formation of the Tn antigen, by the transfer of GalNAc from UDP-GalNAc, a sugar donor, to the hydroxyl group in the side chain of a serine or threonine residue and is controlled by the large family of UDP-GalNAc-polypeptide *N*-acetylgalactosaminyl-transferases (GalNAc-Ts), localized in the Golgi apparatus [25, 58].

GalNAc-Ts family encodes 20 different isoforms that present different tissue expression and acceptor substrate specificities which provide them unique functions depending on the cell type and organ where they are expressed. GalNAc-Ts control the sites and density of *O*-glycan occupancy of the mucin tandem repeats. There are sites of proteins that can be glycosylated by more than one GalNAc-T enzyme (redundant sites). However, there are others that are restricted for one specific GalNAc-T.

From a structural perspective, GalNAc-Ts hold a catalytic domain attached via a flexible linker to a ricin-like lectin domain [59]. Both domains are essential for efficient *O*-GalNAc glycosylation. The lectin domain on the GalNAc-Ts may be mobile and its location relative to the catalytic domain varies among isoforms [60]. The recognition of *O*-glycosylation sites by GalNAc-Ts depends whether these enzymes interact with naked or previous glycosylated regions of the substrate. While naked peptides appear to be exclusively recognized by the catalytic domain, glycopeptides recognition relies on the cooperative mechanism between the catalytic and lectin domain [61]. Lectin domain mediates the GalNAc-peptide substrate specificity also increasing the efficiency of the enzymatic action [62].

Although some GalNAc-Ts structures have been reported, the structural features of the peptide substrates recognition by distinct GalNAc-Ts, as well as, the dynamic landscape between the lectin and catalytic domains upon glycosylation remain uncertain [59].

Misregulation in expression and activity of specific GalNAc-Ts has also a significant influence in cancer biology by affecting cell differentiation, adhesion, invasion and metastasis [63-65]. The altered

expression and activity of distinct GalNAc-Ts, namely GalNAc-T2, T3, T6 and T12, may be one of the mechanisms involved in changes in mucin *O*-glycosylation during malignant transformation [25].

In particular, GalNAc-T3 exhibits a high expression in several human cancers [63-65]. Overexpression of this enzyme strongly correlates with shorter survival rates, more aggressiveness, probability of metastasis and reoccurrence after operation [66, 63]. Inhibition of GalNAc-T3 expression in ovarian [67], renal [64], early stage squamous gastric [68, 69], esophageal [69], oral squamous [63] and pancreatic cancer cells inhibit their invasive capacities arguing GalNAc-T3 a potential target for developing new cancer therapies [66, 67, 63, 64, 58]. While, in colorectal [70], non-small lung and lung [71, 72], gallbladder [69], extrahepatic bile duct [69], gastric [73], hepatocellular [74], pancreatic cancer cells [65], the loss or decrease of expression of GalNAc-T3 shows the opposite effect, in poorly differentiated cancer cells. This shows that the correlation between GalNAc-T3 expression and prognosis depends on specific type of tumors [65] and that GalNAc-T3 can be a useful parameter for clinical management, especially in an early post-operative phase [63].

GalNAc-T3 glycosylates preferentially peptide sequences with valine in position -1 [24]. In the particular MUC1 sequence PD**T**RPAPG**ST**APPAHGV**TS**A, the threonine residues in -TS- and -ST- are glycosylated and to a lesser degree the S in -ST-, while the isolated threonine in -DTR- is not. GalNAc-T3 shows preference in initiating MUC1 glycosylation at the T in -TS- and only after finishing, does it proceed to -ST- sites, the opposite of GalNAc-T2 [75].

Rationalizing at the molecular level the interactions between the lectin and catalytic domain of GalNAc-T3 and MUC1 peptides and glycopeptides will be of great value for the rational design of inhibitors that target the lectin or/and catalytic domain to regulate GalNAc-T3 activity in cancer.

1.5.Methodology

The study of carbohydrate-protein interactions has impact on medical research, namely in the field of cancer. Several antibodies are under clinical trials that target glycan-binding proteins or glycan-antigens expressed in cancer cells [76-80].

From this perspective is of paramount importance to understand the recognition process at a molecular and atomic level of carbohydrates by lectins, antibodies and enzymes. Structural, conformational and dynamic knowledge of the molecular recognition processes between carbohydrates and their macromolecular targets has the potential to explain binding and selectivity, as well as, to assist in the design of new molecular probes for diagnosis and novel sugar-based drugs and vaccines.

Structural information can be obtained by X-ray Crystallography or NMR Spectroscopy combined with computational methods [9]. X-ray Crystallography has been used to characterize various biomolecules complexes, however, the resulting structures are static pictures of the stabilized complexes. Indeed, the characteristic flexibility of carbohydrates can be critical in the recognition process and in most of the cases cannot be reflected in the crystal structure. In addition, the K_a values of carbohydrate-protein complexes are weak, only improved by multivalence effects, making its crystallization difficult. Therefore, to complement X-Ray results, it can be also used a combined approach of NMR and computational studies, where each technique offers different features of the binding process [81, 4, 11, 9, 82].

NMR techniques can have a very wide range of utilities, which can go from identification of new ligands for carbohydrates-binding proteins, the determination of the ligands epitopes and protein groups which are involved in the recognition to the description of the dynamics and conformational features of the complexes. The application of a multidisciplinary strategy, combining the experimental data with molecular simulations and modelling is the best to characterize the main structural features that govern the recognition of saccharides by receptors [81, 4, 11].

In NMR, two different approaches can be employed to investigate ligand-receptor interactions: the receptor or ligand-based approach. The choice of the method depends on the problem and the required structural information, especially the off-rate of the dissociation process and the receptor's size. If the exchange conditions between free and bound state are fast and the binding transient to weak, it's more suitable to use ligand-based methods and follow the ligand's parameters. However, if the receptor's size allows, chemical shift and relaxation of protein resonances can be analyzed in free and bound states, using receptor-based methods. In these cases, the binding regime is slow to intermediate, meaning the binding is weak or irreversible [4, 82].

1.5.1. Protein-based Methods

In protein-based methods, the interaction studies between small CBPs and carbohydrates can be done, by monitoring the chemical shift or line-width variations of specific proton resonances of the CBP, during the binding event [11]. These methods require selective isotopically labelled receptors and specific resonance assignment of the protein NMR spectra [82].

However, these methods depend on the observation of the protein signal and are limited by the protein molecular weight, usually below 30 kDa, but this limit can be extended to 100 kDa by using methods involving deuteration of the protein or selective residue labelling. Furthermore, the advances in methodology and the access to high field magnets are increasing the limits of NMR spectroscopy and study of biomolecules with large molecular size and complexity is now possible [82].

Receptor-based methods give information on the residues directly involved in the recognition. However, for large proteins, it is essential isotope labeling with ^{13}C or ^{15}N, unless the binding process is monitored by the ligand perspective. By adding another dimension with the isotope labeling it reduces the spectral overlap and helps in the resonance assignment. The most used spectrum for protein assignment is the heteronuclear correlation spectrum $^{1}H,^{15}N$-HSQC. This spectrum is also considered as the fingerprint of the protein, since it correlates the N-H pair, showing in the spectrum every backbone amide hydrogen peak. The dispersion of the peaks is a consequence of the different chemical environments generated by the three-dimensional structure of the protein [82].

Since the chemical shift is very sensitive to structural changes in the local environment binding of a ligand will produce a chemical shift perturbation (CSP) in the N-H pair. Experimentally successive $^{1}H,^{15}N$-HSQC spectra are recorded, in a titration of the protein with the ligand. When ligand binds, the chemical environment of the protein backbone changes due to the presence of the ligand and their chemical shift is altered. This happens usually to residues directly involved in the binding. The alteration in the $^{1}H,^{15}N$-HSQC spectra of the protein will depend on the type of chemical exchange regime between the free and bound protein and on the strength of the interaction with the ligand. In some of the cases it is also possible to estimate the dissociation constant, K_D [82].

When a ligand binds reversibly to a receptor, this binding has an association constant, which normally range from 10^2 to 10^{12} M^{-1}. If the exchange between free and bound state is slow, in the spectrum it is observed two sets of signals, one corresponding to the resonances of the free protein and another for the complex. With the increase of the concentration of the ligand, the intensity of the signals corresponding to the free protein will be reduced with simultaneous increase in the intensity of the corresponding signals of the complex [82]. (Figure 1.8 A) The complex is formed instantly, so in those cases it is needed low ligand-to-protein ratios to make the identification of the perturbed residues in the interaction. From this type of experiment, it is possible to determine the apparent K_D based on the intensities of the free and bound signals of the protein. This method was used to study the binding of hexaacetyl chitohexaose to MoCVNH-LysM [83].

However, for complexes with carbohydrates these are weaker and range from 10^2 to 10^6 M^{-1}, due to the association rate being rapid, limited only to molecular diffusion. The dissociation rates are also rapid, allowing a fast exchange between the free and bound states of the ligand [5]. In fast exchange, it's only possible to observe one resonance, in the HSQC spectrum, with its position determined by the population of free and bound protein. The shift in the chemical shift observed in this type of experiment, is due to the decrease of the free population of the protein and the increase of the population in bound protein, with the increase of ligand concentration [82]. (Figure 1.8 B) In those cases, it is needed a large ligand to protein ratio to observe the shifts and it is also possible to estimate an apparent K_D value. This method was used to identify the binding cleft of the family 11 carbohydrate binding module from *Clostridium thermocellum* through protein titrations with cellotetraose and cellohexaose [82].

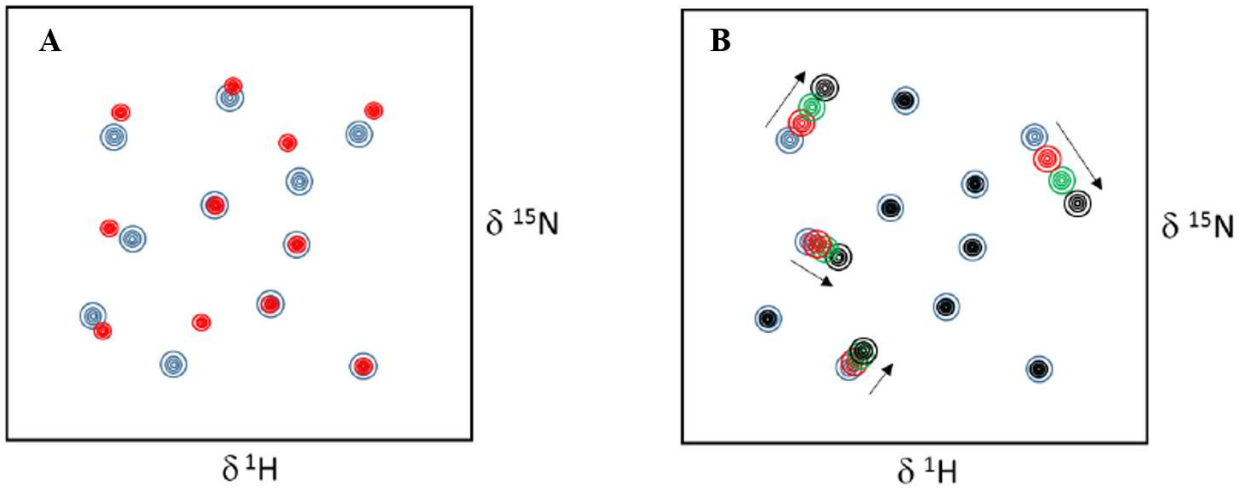


Figure 1.8: Representation of the chemical shift perturbation induced by interaction of an unlabelled small ligand and a ^{15}N-labelled protein detected in ^{1}H,^{15}N-HSQC spectrum. **A-** Example of slow exchange interactions; **B-** Example of fast exchange interactions. [82]

1.5.2. Ligand-based methods

Ligand-based methods can also be used to deduce the ligand-protein binding, through changes in motion, orientation and diffusion properties of the ligands, when passing from the free state to being recognized by a large receptor [82]. They require very small amounts of receptor and do not require stable isotope labelling [81].

In these methods, the changes in the NMR parameters of the ligand are followed to monitor the binding process and determine the ligand epitope mapping. However, they only study changes in ligands, thus it is not possible to extract directly the protein binding site. Therefore, it is necessary additional NMR experiments, to determine the binding site, namely binding experiments using a competitive ligand.

One of the first NMR carbohydrate-based methods for determination of ligand-protein complexes was based in the measurement of T_1/T_2 relaxation rates of ligand protons, which change drastically upon binding to a macromolecular receptor. This leads to the observation of selective broadening of certain resonances of the ligand. The close contact with the protein results in a shortening of their T_1/T_2 relaxation time. This selective line broadening of certain resonances of the ligand, may occur due to enhanced T_2 relaxation, since the large size of the ligand-protein complex and its slow tumbling in solution, will increase the rate of T_2 relaxation, giving rise to greater line widths for the ligand resonances. Hence, from proton ligand line width analysis it is possible to obtain the epitope mapping, since the protons of the ligand closer to the protein will have larger signal line width [11, 5]. This method was applied to monitor the interaction of MUC1 glycopeptides with the fragment antigen unity of the anti-MUC1 mAb B27.29 [4].

Nowadays, Nuclear Overhauser effect (NOE) based techniques are the most powerful and robust to detect biomolecular interactions from a ligand perspective. One of the most used to obtain key aspects of carbohydrate-protein interactions is the experiment saturation transfer-difference spectroscopy (STD-NMR) [84]. This technique permits the deduction of interactions between a small molecule and the target receptor, by determining the protons in close contact with the protein and thus defining the ligand epitope mapping [81, 82]. It is based in intermolecular NOE transfer from the large receptor to an interacting ligand.

In this method, two ^{1}H-NMR spectra are acquired: on-resonance and off-resonance. In the on-resonance spectrum, the receptor protons are selectively saturated, with a selective radiofrequency pulse, at frequencies where there are no ligand resonances at least within 1-2 ppm. Usually, the frequencies chosen are in the aromatic and aliphatic regions. This results in a spreading of magnetization throughout the protein by spin diffusion, which will reach the binding site and affect the nuclei of any interacting ligand. The off-resonance spectrum is recorded without saturation of the protein. If the ligand interacts with the protein, there is a decrease of the ligand resonance intensities in the on-resonance spectrum, due to the intermolecular transfer of magnetization from the receptor to the bound ligand, via ^{1}H-^{1}H cross-relaxation. This can be seen in the STD spectrum, which is obtained from the subtraction of the on-resonance spectrum from the off-resonance. STD spectrum only displays the ligand signals in direct contact with the saturated protein resonances. (Figure 1.9) If a ligand is not interacting with the receptor it will not receive saturation from the protein and its intensities will not be affected, therefore its signals will not appear in the STD spectrum. The same happens with a strong binder ligand. The amount of magnetization transference from the receptor to the ligand is proportional to the inverse of the sixth power of the distance between the ligand nuclei and the receptor protons and depends on the residence time of the ligand in the protein pocket, relating the signal intensities of the ligand protons with their proximity to the protons in the receptor binding site. The dissociation of the ligand will transfer this saturation into solution where the free ligand gives rise to resonance signals with narrow line widths. In this way it is possible to quantify the individual contributions of each ligand's proton to the recognition event and obtain the ligand epitope. To determinate the STD-derived epitope mapping, the STD intensities are

expressed as relative STD percentages, after normalizing to 100% the most intense signal [82, 11, 4, 81, 85].

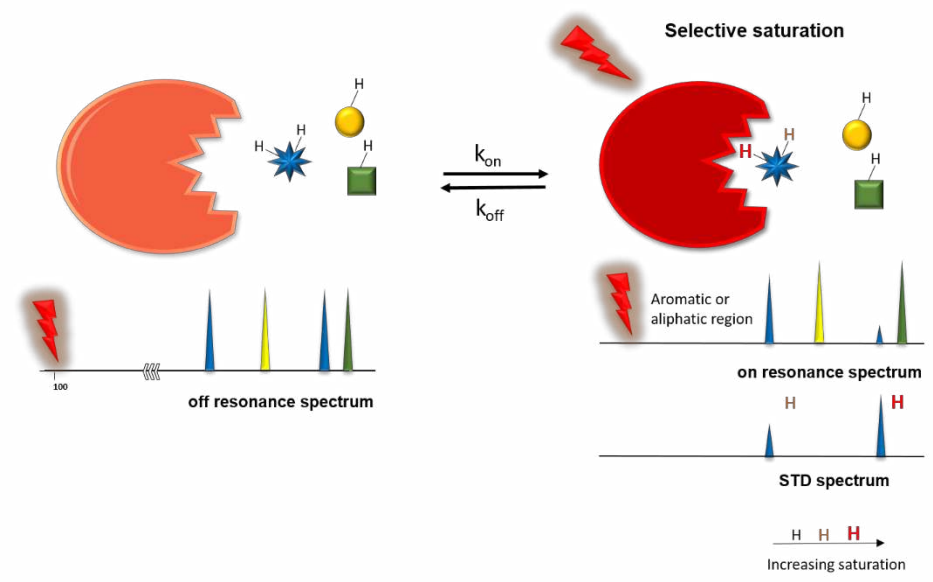


Figure 1.9: Schematic representation of STD-NMR experiment.

In NOE-based experiments, like STD-NMR there is a strong requirement in the kinetics of the dissociation process, where the off-rates should be fast on the relaxation time scale. This means that they are more appropriate to be employed in the case of moderate or weak binding processes, typically in the range of mM or µM [4, 81].

1.6. Objectives

Carbohydrates are present in every tissue and cell type playing a key role in several molecular recognition processes in health and disease. They are inserted through glycosylation, which is the most abundant and complex posttranslational modification in proteins and lipids.

An alteration in glycosylation is a well-known hallmark in various diseases, especially cancer, one of the main causes of mortality worldwide. Altered glycans or glycoproteins are useful biomarkers for diagnosis and prognosis in cancer. Moreover, they are implicated in tumor development, progression and metastasis due to altered cellular interactions with other biomolecules, namely lectins involved in immune response [86].

Understanding the structural bases beyond the molecular mechanism involved in the biosynthesis of glycans and their recognition by specific receptors will certainly help the rational design of vaccines or therapeutic drugs for cancer and other diseases. For structural information, X-Ray Crystallography, NMR spectroscopy and molecular dynamics simulations might be the best choice.

In this context, this thesis covers distinct systems in the context of carbohydrate-protein recognition processes using different NMR binding spectroscopy techniques.

As mentioned before, MUC1 antigens can be recognized by specific antibodies, thus the first chapter of this thesis describes the molecular recognition study of a MUC1 Tn-antigen peptidomimetic containing a non-natural amino acid and different antibodies, using STD-RMN. Understanding the molecular mechanism of these interactions will give important notions for the rational design of anti-MUC1 antigen vaccines.

Chapter two is focused in the interactions between Gal-3 and MUC1-based TF-antigen using $^{1}H,^{15}N$-HSQC, STD-NMR and line broadening analysis. The protocol established and optimized during this thesis will allow to study in an atomic level, future synthetic mimetics of TF-antigen binders designed as potential inhibitors of Gal-3.

Lastly, the third chapter relies on the study of MUC1 *O*-glycosylation using the enzyme GalNAc-T3 employing $^{1}H,^{15}N$-HSQC experiments to follow the glycosylation process of a MUC1 construct containing four tandem repeat domains. STD-NMR of MUC1 glycopeptide fragments were also accomplished to determine the epitope mapping and to help rationalize the glycosylation process. The obtained data are relevant to understand GalNAc-T3 specificity towards MUC1 and prompt the development of inhibitors towards GalNAc-T3 regulation.

2. Materials and Methods

All NMR spectroscopy experiments were acquired in a 600 MHz Bruker Avance III spectrometer equipped with 5 mm inverse detection triple-resonance z-gradient cryogenic probe, head (CP TCI) operating at 600.13 MHz for hydrogen and 60.82 MHz for nitrogen. All NMR data was processed in Bruker TopSpin 3.5.

2.1. Chapter i): Carbohydrate-antibodies interactions

The anti-MUC1 monoclonal antibody clone VU-3C6 was purchased from GeneTex, Inc. provided at 1 mg/mL concentration in phosphate buffer saline and 0.09% NaN_3.

The anti-MUC1 monoclonal antibody SM3 was synthetized and obtained through a collaboration with Dr. Ramón Hurtado-Guerrero from Institute of Biocomputation and Physics of Complex Systems, University of Zaragoza, Instituto de Quimica Fisica Rocasolano, Consejo Superior de Investigaciones Científicas Joint Unit, Zaragoza, Spain.

The anti-Tn monoclonal antibody 14D6 was synthetized by Dr. Claude Leclerc from Unité de Régulation Immunitaire et Vaccinologie, Equipe Labellisée Ligue Contre le Cancer, Paris, Institut Pasteur, France.

The Tn-glycopeptide mimetic APD(Hnv)*RP (* indicates the site of Tn-glycosylation) was synthetized and obtained from a collaboration with Dr. Francisco Corzana from Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química, Logroño, Spain.

The NMR characterization of Tn-glycopeptide with sequence APD(Hnv)*RP (* indicates the site of Tn-glycosylation) was accomplished using 1D and 2D experiments: ^{1}H-NMR, ^{1}H,^{1}H-TOCSY and ^{1}H,^{1}H-NOESY at 278 K, with spectral width of 9001.9 to 9014.4 Hz and O1=2797.3 Hz. In TOCSY experiments mixing times of 30 ms and 80 ms were used, while the in the NOESY experiment a 400 ms of mixing time was employed. The glyco peptidemimetic was characterized in buffer containing PBS 20 mM, NaCl 20 mM, NaN_3 (0.09%) in 90:10 H_2O:D_2O at pH=7.1. The concentration of Tn-glycopeptide mimetic was set at 1 mM. To the sample was added 0.05 mM of TSP as internal reference.

For the study of interactions between the Tn-antigen peptidomimetic and antibodies VU3-C6 and 14D6, a 100% deuterated buffer was used, containing PBS 20 mM, NaCl 20 mM, NaN_3 0.09% at pH=7.1, except in the case of antibody SM3, where the buffer was composed by Tris-d11 25 mM, NaN_3 0.09% at pH= 6.5. In all samples it was added 0.05 mM of TSP as internal standard. The NMR experiments were recorded in 3 mm NMR tubes.

STD-NMR experiments were acquired at 310 K, except for the STD-NMR experiment in presence of mAb 14D6 that was recorded at 298 K. For the STD-NMR experiments an antibody/ligand molar ratio of 1:40 was used. In the case of SM3 this molar ratio was set to 1:4.

The STD-NMR spectra were acquired with 4160 transients for VU-3C6 and SM3 and 1600 transients for 14D6 in a matrix with 65k data points in t2, in a spectral window of 12335.53 Hz centered at 2822.89 Hz for mAb VU-3C6, 2822.51 Hz for mAb SM3 and 2817.23 Hz for mAb 14D6. Selective saturation of the protein resonances (on resonance spectrum) was performed by irradiating at -0.5 ppm, using a series of Eburp2.1000-shaped 90º pulses (50 ms, 1 ms delay between pulses) for a total saturation time of 2.0 s. For the reference spectrum (off resonance) the samples were irradiated at 100 ppm.

To obtain the STD-NMR-derived epitope mapping the STD-NMR total intensities were normalized with respect of the highest STD-NMR response.

2.2. Chapter ii): Gal-3/TF interactions

2.2.1. Expression and purification of $^{1}H,^{15}N$-labelled Gal-3 CRD

To study the interaction between Gal-3 and TF-antigen the expression and purification of ^{15}N-isotopically labelled carbohydrate recognition domain of lectin Gal-3 (Gal-3 CRD) was carried out. The procedure used was based in previously reported work [87].

The Gal-3 CRD region corresponds to the residues 114-250 and was subcloned in the expression vector pET21, produced by NZYTech (appendix 6.3). To transform the cells 1 µL of the expression vector was added to 50 µL of *E. coli* BL21 competent cells and then incubated for 15 min on ice. The incubate was then submitted to a thermal shock for 40 s, at 42 ºC, in a heating plate AccuBlock Digital Dry Bath and transferred to ice for 15 min. Then, 500 µL of sterile Luria-Bertani medium (LB) (Appendix 6.4 Table 6.2) were added to the cells and this culture was incubated for 1 hour, at 37 ºC, under agitation using an Orbital Shaker ES-20 incubator. Then the cells were concentrated by centrifugation at 14000 rpm, for 30 s, in a Mikro120 Hettich Zentrifugen centrifuge. The supernatant was discarded and in a Petri dish with LB-agar medium containing ampicillin, 50 µL of the transformed *E. coli* cells were inoculated and incubated, at 37 ºC overnight.

One colony of the transformed *E. coli* cells was inoculated and incubated in 25 mL of LB medium, at 37 ºC overnight. Afterwards, 20 mL of the pre-inoculum were inoculated and incubated in 1 L of M9 minimal medium containing $^{15}NH_4Cl$ as the sole nitrogen source (Appendix 6.4 Table 6.3) at 37 ºC for 3h and 45 min, in an Optic Ivymen System incubator with agitation of 220 rpm. When an optical density value of 0.601 at 600 nm was reached 1 mM of IPTG was added to the culture, to induce Gal-3 CRD expression for 4 hours at 37 ºC with agitation of 220 rpm. The cells were recovered at 6000 rpm for 12 min, at 4ºC in an Avanti J-26 XPI Beckman Coulter with the rotor JA-10. Cells from before the induction, 2 and 4 hours after the induction were saved.

The cells were resuspended in 30 mL of buffer 25 mM PBS, 50 mM NaCl, 1 mM DTT and 0.1% sodium azide, pH 6.8 and lysed by sonication with 10 cycles of 1 min with 80% of amplitude, in a Hielscher Ultrasound Technology (UP100H). The lysate was centrifuged at 16000 rpm, 4 ºC for 30 min, in an Avanti J-26 XPI Beckman Coulter, rotor JA-10 and the supernatant was retained.

For purification, an affinity chromatography with an α-lactose agarose column was performed (AKTA Start). The washing buffer was 25 mM PBS, 50 mM NaCl, 1 mM DTT and 0.1% sodium azide,

pH 6.8 and the elution buffer was 25 mM PBS, 50 mM NaCl, 1 mM DTT, 0.1% sodium azide and 150 mM lactose, pH 6.8. During the washing step, some fractions of Gal-3 CRD were eluted without lactose, lyophilized and stored. During the elution with lactose there were fractions of the Gal-3 CRD with lactose, which were dialyzed against the washing buffer to remove all the lactose and then lyophilized and stored.

The expression and purification process was evaluated by Polyacrylamide and sodium dodecyl sulfate Gel Electrophoresis (SDS-PAGE), with a 45 min run, in a 10% acrylamide gel, constant voltage of 180 V and variable amperage (Bio-Rad).

The quantification of the pure Gal-3 CRD was measured by Ultraviolet Spectroscopy, in an Ultraspec 2100 pro Amersham Biosciences spectrophotometer. The absorbance was read at 280 and 320 nm. The concentration was determined using the Lambert-Beer Law, with an extinction molar coefficient and molecular weight determined using the ExPASy tool [88].

2.2.2. NMR TF-antigen assignment

The TF-glycopeptide PDT*RP (* indicates the site of TF-glycosylation) was obtained through a collaboration with Dr. Ulrika Westerlind from ISAS–Leibniz Institute for Analytical Sciences, Dortmund, Germany.

Two TF antigens were characterized by means of NMR spectroscopy: TF-Thr and TF-glycopeptide. TF-antigen corresponds to the structure Galβ1-3GalNAcα1 linked to Thr/Ser amino acid.

TF-antigen glycopeptide used on our study encodes the peptide sequence PDT*RP, where the * indicates the glycosylation position.

The NMR assignment of the structures was accomplished combining 1D and 2D experiments: ^{1}H, ^{1}H,^{1}H-TOCSY with mixing time of 30 and 80 ms and ^{1}H,^{1}H-NOESY with mixing time of 400 ms. The experiments were acquired at 278 K, with spectral width of 9001.9 to 9014.4 Hz and O1 of 2798.1 Hz for the TF-glycopeptide and 2821.4 Hz for the TF-antigen.

The samples containing TF-antigen (1mM) and TF-glycopeptide (0.5mM) were prepared in a buffer containing PBS 25 mM, NaCl 50 mM, DTT 1 mM, NaN_3 0.1%, pH= 6.5, with a ratio of 90:10, in H_2O and D_2O, respectively. The samples also contained TSP (0.05 mM) as internal reference.

The characterization was performed using CARA software (Computer Aided Resonance Assignment) [89]. The ^{1}H-NMR characterization for both TF-antigen and TF-glycopeptide is displayed in appendix 6.5 and appendix 6.6, respectively.

2.2.3. Gal-3/TF-antigen interactions monitored by ^{1}H,^{15}N-HSQC titrations

For the titration of Gal-3 CRD with the TF-glycopeptide (containing the PDT*RP sequence, * indicates the site of glycosylation) the samples were prepared with constant concentration of Gal-3 CRD (50 μM), 0.05 mM of TSP in a ratio of 90:10 H_2O/D_2O of a buffer containing PBS 25 mM, NaCl 50 mM, DTT 1 mM, NaN_3 0.1%, pH= 6.5 and a variable concentration of TF-glycopeptide. The protein/ligand molar ratios used were 1:0; 1:0.5; 1:1; 1:5; 1:9; 1:45, where 1:0 corresponds to absence of ligand and 1:45

corresponds to 2.25 mM of TF-glycopeptide. The pH was corrected to 6.5 and controlled to be maintained constant during the titration. The $^{1}H,^{15}N$-HSQC were acquired at 298 K, with spectral width of 2189.4 to 9615.4 Hz and O1 of 2793 Hz.

For the titration of Gal-3 CRD with the TF-antigen glycosylated in Thr, the protein/ligand molar ratios used were 1:0; 1:0.5; 1:1; 1:5; 1:15; 1:30, where 1:0 corresponds to the sample of Gal-3 CRD without ligand, and 1:30 corresponds to the sample of Gal-3 CRD with 1.5 mM of TF-antigen. The conditions used were the same as the previous titration described, except the O1, which was set at 2798 Hz. The titrations were accomplished in 5 mm NMR tubes at 298 K.

$^{1}H,^{15}N$-HSQC titrations were analyzed using the software CcpNmr Analysis 2.4.2 [90]. The assignment of the Gal-3 CRD signals was based in the sequential assignment existent in the data base bank BMRB with the code 4909 [91]. The K_D of the interaction was calculated and the amino acids interacting with the TF-glycopeptide and the TF-antigen were identified.

2.2.4. Gal-3/TF-antigen interactions studied by STD-RMN

Besides the titrations, STD-NMR experiments of TF-antigens in presence of Gal-3 were also carried out. For that purpose, it was used Gal-3 full length domain (Gal-3 FL). This domain was expressed by Ana Diniz on the group.

STD-NMR experiment was acquired at 298 K with 1600 transients in a matrix with 65k data points in t2 in a spectral window of 12335.5 Hz centered at 2821.4 Hz. Selective saturation of the protein resonances (on resonance spectrum) was performed by irradiating at -0.5 ppm using a series of Eburp2.1000-shaped 90º pulses (50 ms, 1 ms delay between pulses) for a total saturation time of 2.0 s. For the reference spectrum (off resonance) the samples were irradiated at 100 ppm. The sample was prepared with 45 µM of Gal-3 FL and 1800 µM of TF-antigen, in a ratio of 1:40, 0.05 mM of TSP and PBS buffer in D_2O.

To obtain the STD-NMR-derived epitope mapping the STD-NMR total intensities were normalized with respect of the highest STD-NMR response.

2.2.5. Gal-3/TF-antigen interactions studied by line broadening analysis

Additional line broadening analysis was performed to complement the binding data extracted from STD-NMR experiment. For the line broadening analysis, a comparison between the ^{1}H-NMR spectra of both TF-antigens (TF-antigen at 500 µM and TF-glycopeptide at 400 µM) in absence and presence of Gal-3 CRD at 50 µM was accomplished.

2.3.Chapter iii): MUC1 *O*-Glycosylation by GalNAc-T3

2.3.1. Expression and purification of $^{1}H,^{15}N$-labelled MUC1-4TR

A construct of the N-terminal domain of MUC1 containing 4 tandem repeats with 20 amino acids of the conserved sequence GVTSAPDTRPAPGSTAPPAH (MUC1-4TR) was designed. MUC1-4TR was subcloned in the expression vector pHTP-KSI, produced by NZYTech to yield MUC1-4TR construct (appendix 6.7). The MUC1-4TR construct concomitantly contains a fusion protein KSI tag, a His tag and

a TEV protease recognition sequence (ENLYFQ). To transform the cells 1 μL of the expression vector was added to 50 μL of *E. coli* BL21 competent cells and then incubated for 45 min on ice. The incubate was then submitted to a thermal shock for 40 s, at 42 ºC, in a heating plate AccuBlock Digital Dry Bath and transferred to ice for 15 min. Then, 500 μL of LB medium without antibiotic was added and this culture was incubated for 1 hour, at 37 ºC, under agitation using an Orbital Shaker ES-20 incubator. Then the cells were concentrated by centrifugation at 14000 rpm, for 30 s, in a Mikro120 Hettich Zentrifugen centrifuge. The pellet was retained and re-suspended on the remaining supernatant. The culture was spread on LB plates with kanamycin (50 μg/ml) and incubated, at 37 ºC overnight.

One colony of the transformed *E. coli* cells was inoculated and incubated in 10 mL of sterile LB medium with 50 μg/mL of Kanamycin (Appendix 6.8 Table 6.4) and left to grow overnight at 37ºC, in an Optic Ivymen System incubator with agitation of 220 rpm (pre-inoculum).

The pre-inoculum (10 mL) were inoculated in 500 mL of ^{15}N M9 minimal medium with 50 μg/mL of Kanamycin (Appendix 6.8 Table 6.5) and incubated at 37 ºC, in an optic Ivymen System incubator with agitation of 220 rpm, until the optical density at 600 nm reached 0.600. Then, 1 mM of IPTG was added to induce MUC1-4TR expression and the solution was incubated overnight at 25 ºC, with agitation of 220 rpm. The cells were centrifuged at 6000 rpm for 15 min, at 4 ºC in an Avanti J-26 XPI Beckman Coulter with the rotor JA-10. The supernatant was discarded, and the pellet was retained.

The pellet was re-suspended in 40 mL of the lysis buffer containing PBS 10 mM, NaCl 0.5 M and β-mercaptoethanol 1 mM and sonicated with 10 cycles of 1 min with 80% of amplitude, in a Hielscher Ultrasound Technology (UP100H). After sonication, the lysate was centrifuged at 6000 rpm, 4ºC for 15 min, in an Eppendorf Centrifuge 5804-R. The supernatant was discarded and the pellet was re-suspended in 50 mL of the solubilization buffer containing PBS 10 mM, NaCl 150 mM, Urea 8 M and β-mercaptoethanol 1 mM and left overnight in agitation.

Before the purification procedure, the solution was centrifuged at 10000 rpm, 4ºC for 15 min, in an Eppendorf Centrifuge 5804-R and the pellet was discarded.

The purification of MUC1-4TR construct was carried by affinity chromatography using 5 HisTrap™ FF crude columns (GE HealthCare), in an AKTA Start apparatus and following the protocol: 1) the system was washed and equilibrated in the solubilization buffer (solution A), composed by PBS 10 mM, NaCl 150 mM, Urea 8 M, β-mercaptoethanol 1 mM and imidazole 10 mM; 2) the supernatant was injected; 3) The column was washed with the solubilization buffer and to create a gradient of 4%, an elution buffer containing PBS 10 mM, NaCl 150 mM, imidazole 1 M and β-mercaptoethanol 1 mM (solution B) was used; 4) Fractions were collected; 5) Finished the collection, the gradient was changed to 100%; 6) At the end, the column was washed in 100% of the elution buffer.

Desalting chromatography was recorded using 5 HiTrap™ desalting columns (GE HealthCare), in an AKTA Start apparatus. The buffer containing PBS 10 mM, NaCl 150 mM and β-mercaptoethanol 1 mM was prepared and used as elution buffer. This step was made in cycles of 8 mL per injection, until all fractions, except fraction 2, 6, 7 and 8, were injected and new fractions were collected. A polyacrylamide and sodium dodecyl sulfate gel electrophoresis (SDS-PAGE) was recorded, with a 45 min run, in a 10% acrylamide gel, with constant voltage of 180 V and amperage variable, to select the fractions with the MUC1-4TR construct, for the digestion with TEV protease.

For digestion with TEV protease, 500 μM of EDTA and TEV protease were added to the fractions 4 and 5 and incubated at 4 ºC overnight. The fractions were centrifuged at 10000 rpm, 4 ºC for 15 min. SDS-PAGE was recorded to assess if the digestion with TEV protease was successful.

For the reverse-phase chromatography protocol it was used the HPLC apparatus AKTA Prime Plus (GE HealthCare) and a Purospher® STAR RP-18 endcapped 5μm column (HPLC-Cartridge). The column and system were first equilibrated in a buffer containing water and 0.1% TFA, then the sample was injected and the elution recorded using a gradient from 10-40% of acetonitrile, with a flow of 2 mL/min. All peaks with absorbance at 220 nm were collected. At the end the column was washed with 100% of acetonitrile. (Figure 3.40)

2.3.2. NMR spectroscopy studies of MUC1-4TR glycosylation by GalNAc-T3

The GalNAc-T3 enzyme was obtained through a collaboration with Dr. Ramón Hurtado-Guerrero from Institute of Biocomputation and Physics of Complex Systems, University of Zaragoza, Instituto de Quimica Fisica Rocasolano, Consejo Superior de Investigaciones Científicas Joint Unit, Zaragoza, Spain.

The glycopeptide T3-Tn was synthetized though a collaboration with Dr. Francisco Corzana from Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química, Logroño, Spain.

The $^{1}H,^{15}N$-HSQC, employed to follow the *O*-glycosylation of MUC1-4TR by GalNAc-T3, was acquired at 298 K with the spectral width of 1946.2 to 9615.4 Hz and O1=2800.7 Hz for ^{1}H and 7175.7 Hz for ^{15}N. The samples were prepared in a buffer containing TRIS 25 mM, DTT 1 mM, NaN_3 0.1% at a pH 6.3, in 90:10 of H_2O: D_2O. The concentration of MUC1-4TR was calculated by ^{1}H-NMR experiments and using TSP as internal reference. The concentration of MUC1-4TR was estimated using the aliphatic chain of valine amino acid. Glycosylation process by GalNAc-T3 enzyme was recorded with MUC1-4TR at 155 μM. To monitor the glycosylation event initiated by GalNAc-T3, 11.5 μM of GalNAc-T3 and 250 μM of $MnCl_2$ was added. Further, to this sample, control amounts of the donor substrate UDP-GalNAc were added: i) 38.8 μM of UDP-GalNAc, amount of UDP-GalNAc necessary for glycosylate only ¼ of one position; ii) followed by 38.8 μM, the concentration required to glycosylate ½ of one glycosylation position and then iii) 1mM of UDP-GalNAc.

The ^{1}H-NMR assignment of the glycopeptide T3-Tn was accomplished by Helena Coelho (Appendix 6.9). Sample for STD-experiment was prepared in perdeuterated 25 mM TRIS-d11 in deuterated water, 7.5 mM NaCl and 1 mM DTT, uncorrected pH 7.4. STD-NMR experiment of T3-Tn (875 μM) was performed at 298 K in the presence of 150 μM UDP, 150 μM MnCl2 with 23 μM of GalNAc-T3. The STD-NMR spectra were acquired with 1920 transients in a matrix with 65k data points in t2 in a spectral window of 12335.5 Hz centered at 2820.6 Hz. Selective saturation of the protein resonances (on resonance spectrum) was performed by irradiating at −0.5 ppm using a series of Eburp2.1000-shaped 90° pulses (50 ms, 1 ms delay between pulses) for a total saturation time of 2.0 s. For the reference spectrum (off resonance) the samples were irradiated at 100 ppm.

To obtain the STD-NMR-derived epitope mapping the STD-NMR total intensities were normalized with respect of the highest STD-NMR response.

3. Results and Discussion

3.1. Carbohydrate-antibodies interactions

One of the most immunogenic regions in MUC1 protein is the –PDTRP- region [34]. This sequence is often recognized by monoclonal antibodies anti-MUC1, especially if containing the Tn-antigen linked to the threonine (Thr) residue. Abnormal MUC1-derived glycan antigens are specific of malignant and pre-malignant cells. This observation makes this region a suitable starting point for the rational design of anti-MUC1 based cancer vaccines [92]. However, vaccines using natural amino acids are easily hydrolyzed by proteases. Therefore, it is essential to design (glyco)peptide-based vaccines resistant to proteases. One of the chemical approaches to increase the bioavailability and stability of these structures is to replace the natural amino acids of the peptide sequence by non-natural amino acids.

In this perspective and in collaboration with a synthetic group at Universidad de la Rioja a Tn-glycopeptide mimetic was designed (Figure 3.1). In this compound, the threonine at the –PDTRP- region was replaced by hydroxy-norvaline (Hnv), a non-natural isomer of valine. The peptide sequence is APD(Hnv)*RP, where * indicates that Hnv is glycosylated by α-acetylgalactosamine (GalNAc).

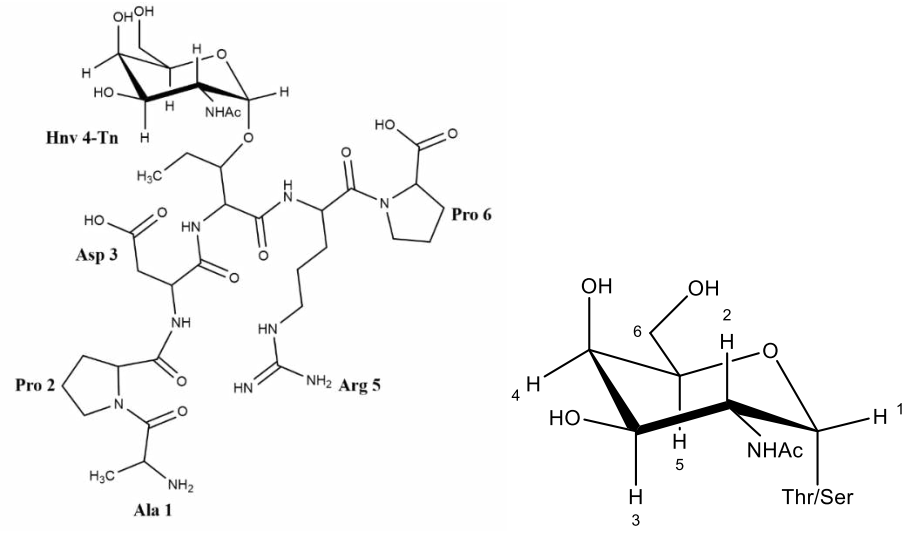


Figure 3.1: Structure of the Tn-glycopeptide mimetic with sequence APD(Hnv)*RP and structure of the Tn-antigen numbered.

In order to design efficient anti-MUC1 vaccines it is essential to study the interactions that govern the molecular recognition event between non-natural antigens and antibodies, at an atomic level. In this context, it was investigated the interaction between the Tn-glycopeptide mimetic APD(Hnv)*RP and distinct commercial antibodies. For the interaction studies, antibodies of the anti-MUC1 family, VU-3C6 and SM3, as well as, the anti-Tn family, 14D6 were used. Moreover, it is also relevant to compare the recognition of the non-natural antigen to the corresponding natural antigen. In this context, the interaction studies of the natural Tn-glycopeptide antigen by anti-MUC1 VU-3C6 and anti-Tn 14D6 was previously investigated and reported by the group and will be used herein for comparison [41].

3.1.1. Characterization of Tn-glycopeptide mimetic APD(Hnv)*RP by NMR Spectroscopy

All the (glyco)peptides in this thesis were characterized using the approach described in this section.

For the assignment of the Tn-glycopeptide, the 2D ^{1}H,^{1}H-TOCSY was used to identify each amino acid spin-system. The 2D ^{1}H,^{1}H-TOCSY shows correlations through-bonds between the ^{1}H atoms in the side chains of a residue. The ^{1}H chemical shifts are different to each amino acid, providing them a fingerprint region. While, the 2D ^{1}H,^{1}H-NOESY shows correlations between protons through the space. Two protons at a distance up to 5 Å will present a NOE cross-peak in the NOESY spectrum.

Hence, from 2D ^{1}H,^{1}H-NOESY analysis the correlations through-space between the NH proton of the amide bond of the amino acid i and the Hα of the lateral chain of the amino acid i-1 permits to do the sequential assignment [93].

For glycopeptides, like the Tn-glycopeptide APD(Hnv)*RP, it is easier to assign first the ^{1}H resonances of the amino acids and only after the resonances of the carbohydrate. Amino acids have specific patterns in ^{1}H,^{1}H-TOCSY spectrum, especially in the NH region (region in the 8 ppm). In contrast, carbohydrate resonances are much more overlapped at the 3-4 ppm region in the spectrum.

The assignment of the amino acids is initiated by identifying the specific patterns in the NH region, commonly named as spin system (Figure 3.2). The only amino acid that doesn't have a NH pattern is the proline, thus cannot be found at the NH region. The amino acid at the N-terminus is not involved in the amide bond and therefore cannot also be identified at this region of the spectrum. This is the case of Ala amino acid in the Tn-glycopeptide.

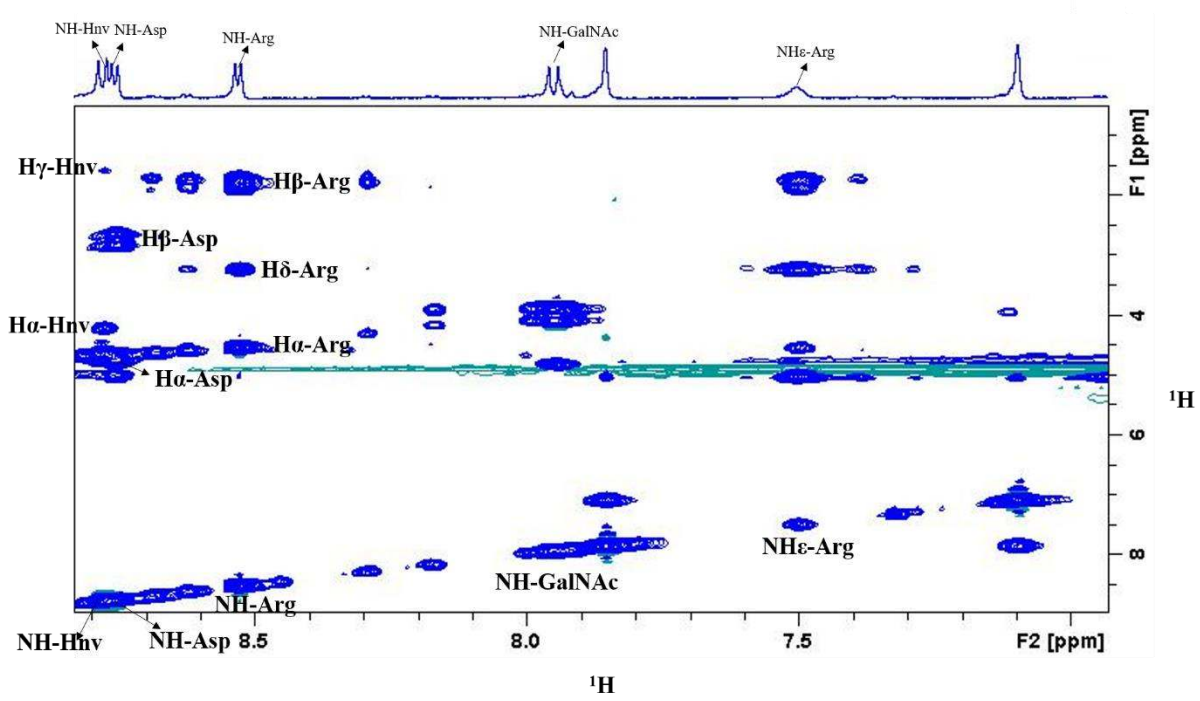


Figure 3.2: NH region of the TOCSY spectrum (80 ms of mixing time) of the Tn-glycopeptide APD(Hnv)*RP (* indicates the site of glycosylation).

In the spectrum above, it can be observed three spin systems corresponding to amino acids Hnv, Asp and Arg, one corresponding to the NH of the α-GalNAc residue and another corresponding to the spin system of the NH_{ε} side chain of the arginine residue. Each amino acid has a specific pattern in TOCSY, because each one has a distinct chemical structure. Comparing the obtained patterns with those reported in the literature it was possible to identify to each amino acid they corresponded (Appendix 6.1).

However, with only the NH region of the spectrum it is not possible to assign all the proton resonances of the amino acids. This is the case of prolines, the N-terminal alanine and some protons of hydroxy-norvaline. To assign the rest of the resonances, it is needed to analyze the whole TOCSY spectrum.

From the protons not yet identified, the easiest to attribute are the aliphatic protons from the side chain of alanine and hydroxy-norvaline, since these protons are more shielded and therefore their signals will appear in a low ppm region (around 1.5 to 0.5 ppm). After assigning these protons it is easy to complete the assignment of Ala and Hnv spin system. In the spin system of alanine, it is possible to observe that it overlaps with one of the prolines (green rectangle). The other spin system unidentified and alike to the proline, will be the other proline (orange rectangle) (Figure 3.3).

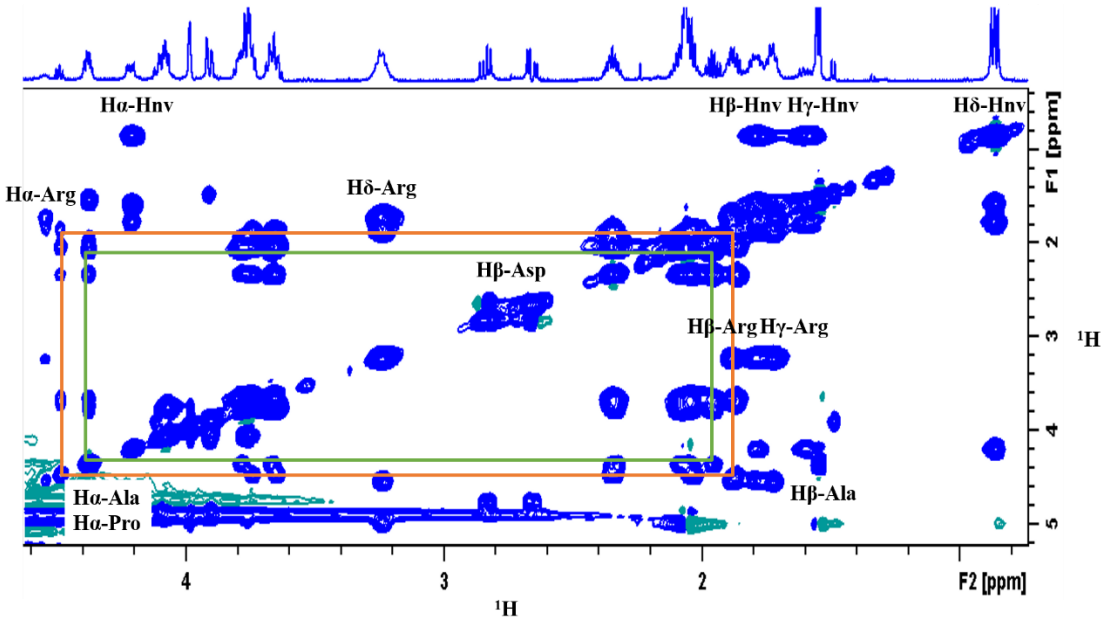


Figure 3.3: Aliphatic region of TOCSY spectrum (80 ms of mixing time) of Tn-glycopeptide APD(Hnv)*RP (* indicates the site of glycosylation) with the attribution of the spin system of the amino acids. The green rectangle is the spin system of one of the prolines and the orange is the spin system of the other proline.

After assignment of all spin systems in ^{1}H-NMR spectrum, it is necessary to do the sequential assignment of the amino acids. For that purpose, it is used the NOESY spectrum. In the case of Tn APD(Hnv)*RP sequence, this step is important to distinguish the two prolines. From superposition of the TOCSY with the NOESY spectra it is possible to distinguish the two prolines (Figure 3.4).

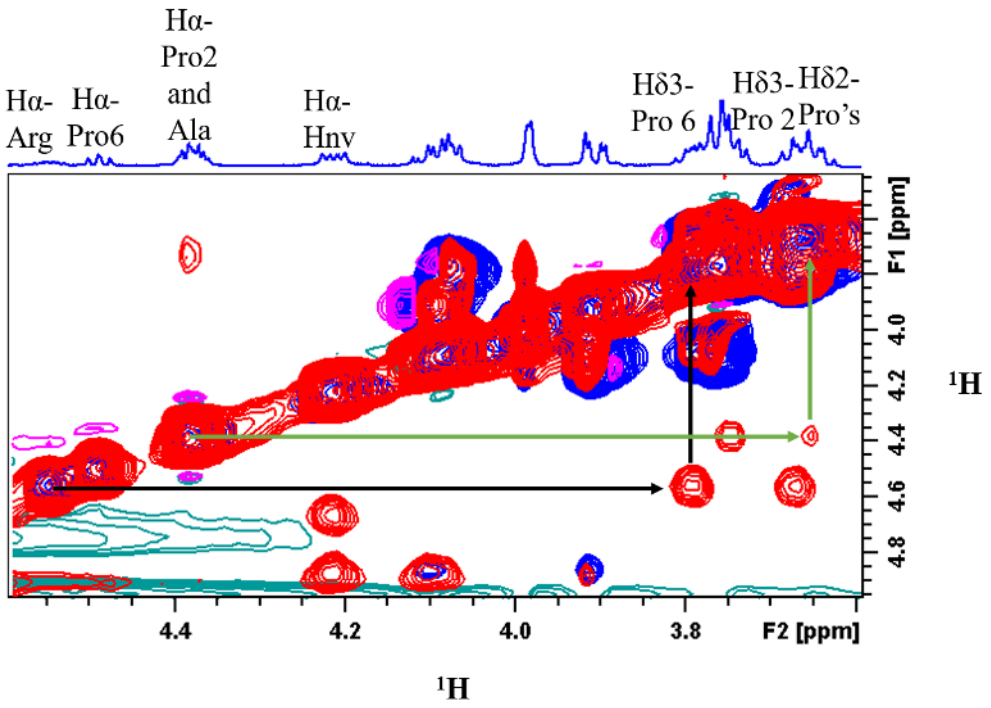


Figure 3.4: Superposition of the aliphatic region of TOCSY (blue) and NOESY (red) spectra of the Tn-glycopeptide APD(Hnv)*RP (* indicates the site of glycosylation). The black arrows represent the NOESY between the Hα of arginine with Hδ3 of proline at position 6 and the green arrows represent the NOESY between the Hα of alanine with Hδ2 of proline at position 2.

From superposition of TOCSY and NOESY spectra it can be noticed that the Hα of arginine has a NOESY peak with the Hδ3 of a proline residue, which according to the peptide sequence must be the proline 6. In contrast, the Hα of alanine also has a NOESY peak with the Hδ2 of a proline that can only be assigned as the proline at position 2.

Now that all the amino acids are identified, it is time to assign the carbohydrate resonances. In the NH region of the spectrum, the NHAc group of GalNAc was identified, see figure 3.2. In GalNAc spin system, it is possible to see two peaks that can be assigned to H2 and H3 of GalNAc. It is possible to discriminate H2 of H3, since H2 presents the strongest correlation with NH of NHAc (Figure 3.5).

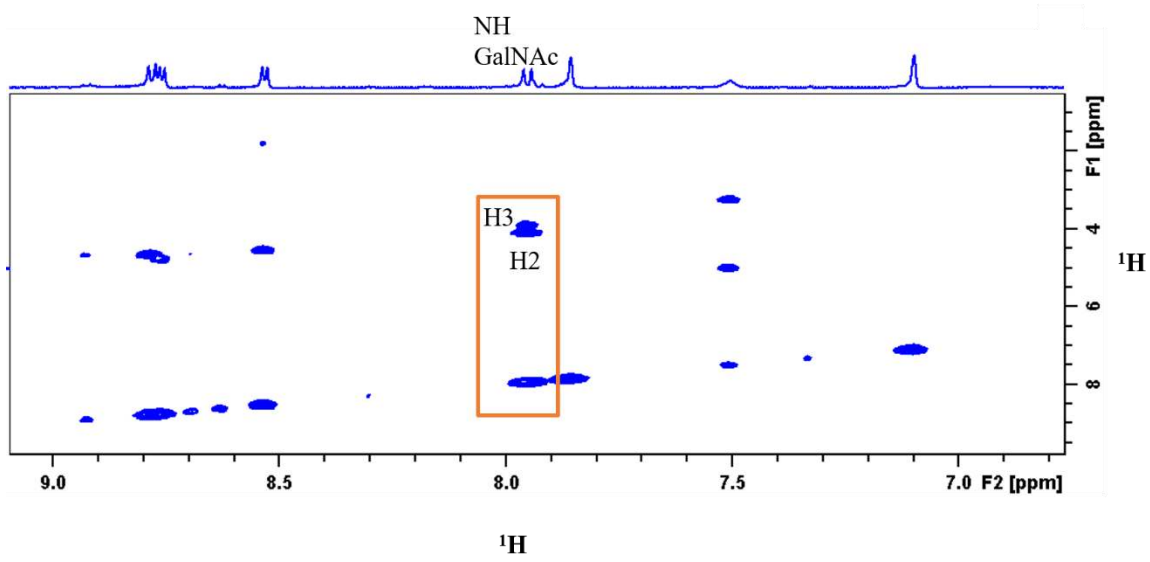


Figure 3.5: NH region TOCSY spectrum (30 ms of mixing time) of the Tn-glycopeptide APD(Hnv)*RP (* indicates the site of glycosylation), with the identification of the carbohydrate's spin system (orange rectangle) and the attribution of the H2 and H3.

The H1 anomeric proton holds a resonance closer to 5 ppm and due to the water signal cannot be detected in the ^{1}H-NMR spectrum. However, it is possible to detect correlations involving the anomeric proton H1 in TOCSY.

The NOESY spectrum can be used to assign the rest of the protons of GalNAc. From GalNAc structure it is expected that H3 correlates with H5. A NOE cross peak can be detected between H3 and H5 since these two protons are closer (less than 4Å). Assigned the H5 proton, it is possible to assign H6s, by spin-spin coupling in the TOCSY spectrum. The signal of H4 in GalNAc is a doublet with a small constant coupling (Figure 3.6).

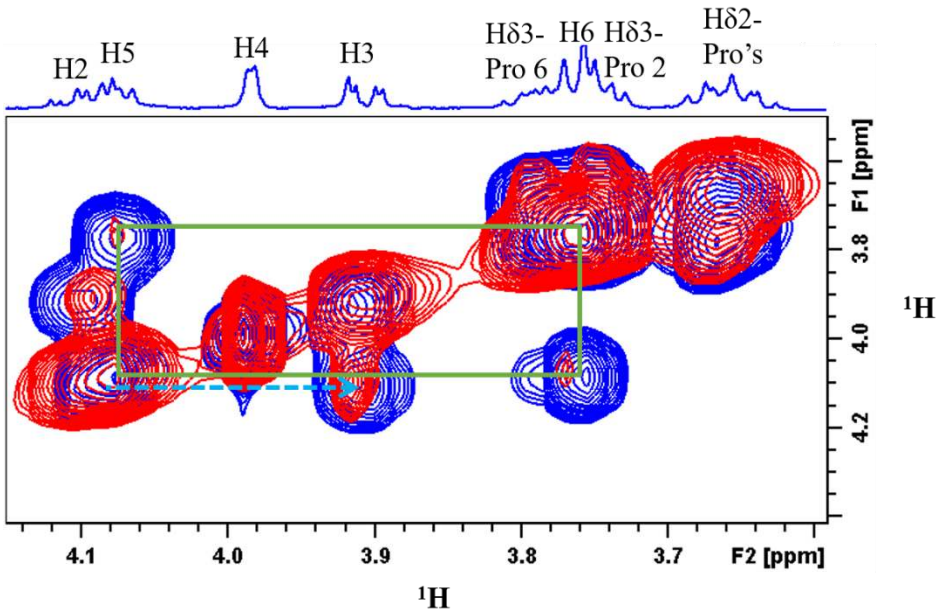


Figure 3.6: Superposition of the aliphatic region of TOCSY (blue) and NOESY (red) spectra of the Tn-glycopeptide APD(Hnv)*RP (* indicates the site of glycosylation). The green rectangle corresponds to the correlation between H5 and H6, while the blue dots arrow shows the spatial correlation between H5 and H3.

The methyl group of NHAc can be assigned through the spatial correlation with the NH of NHAc itself (Figure 3.7).

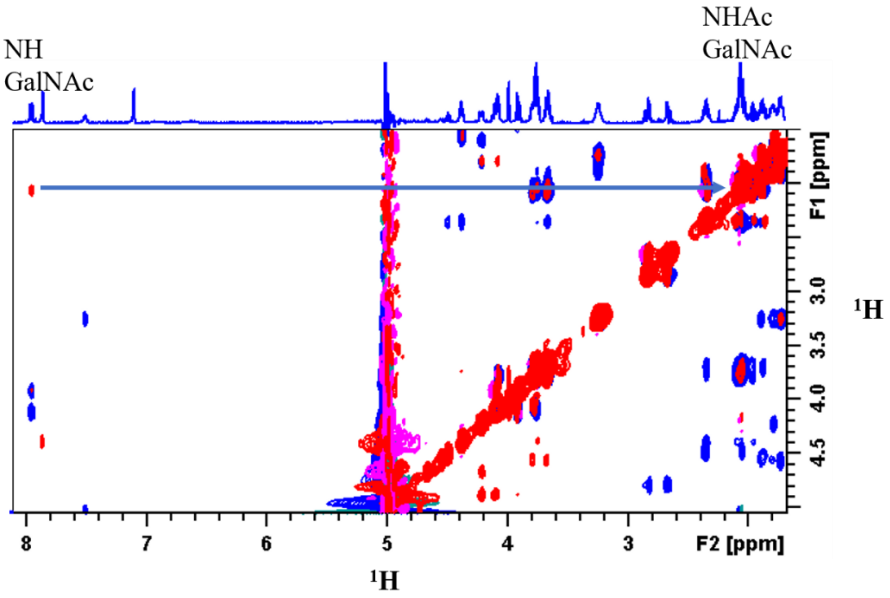


Figure 3.7: Superposition of the TOCSY (blue) and NOESY (red) spectra of the Tn-glycopeptide APD(Hnv)*RP (* indicates the site of glycosylation). The blue arrow represents the spatial correlation between the NH of the GalNAc and the -CH_3 group of the GalNAc residue.

The total assignment of the Tn-glycopeptide mimetic is displayed in appendix 6.2.

3.1.2. STD-NMR binding studies of Tn-glycopeptide mimetic APD(Hnv)*RP

The interactions of the Tn-glycopeptide mimetic APD(Hnv)*RP in presence of distinct antibodies were scrutinized by STD-NMR binding studies. In particular, the anti-MUC1 antibodies VU-3C6 and SM3 were employed (Figure 3.8 and Figure 3.12, respectively) together with the anti-Tn mAb 14D6 (Figure 3.16).

In the case of VU3-C6 antibody the STD-derived epitope binding shows that the most important amino acid residues for the binding interactions with VU-3C6 antibody are the Hnv and Arg (Figure 3.9). Replacement of the Thr in the natural antigen by the non-natural Hnv does not preclude the binding with VU-3C6 and that is a positive result towards the rational design of more stable glycan-based vaccines. In addition, it is clear that the Hδ protons of Hnv receives the highest % of STD response, indicating that the terminal $–CH_3$ group of Hnv amino acid is in closer contact with the VU-3C6 binding site (Figure 3.8 and 3.9). Indeed, the side chain of Hnv is longer than the Thr and may perform additional interactions with the antibody binding site.

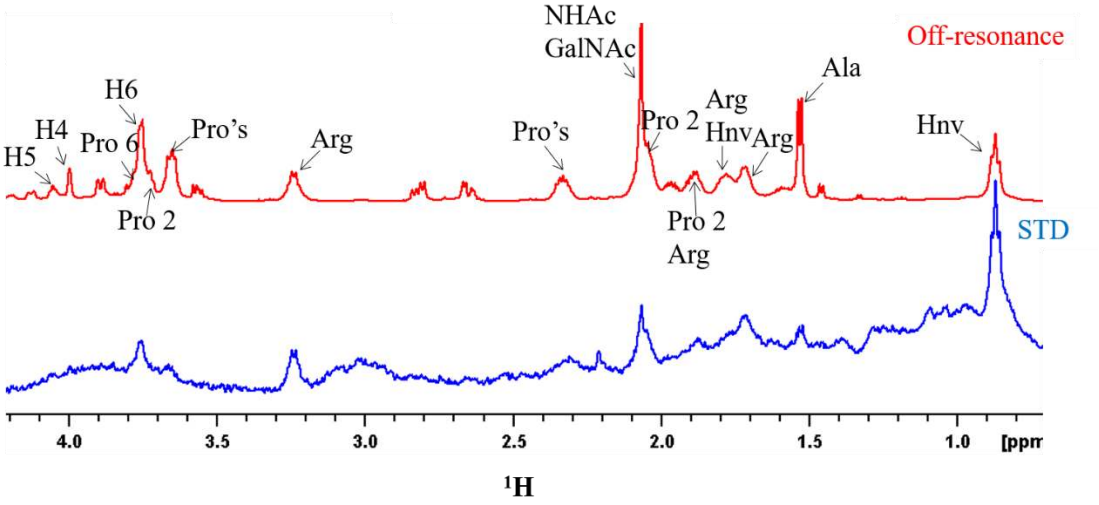


Figure 3.8: STD spectrum (blue) and the off-resonance spectrum (red) for the STD-NMR experiment of the Tn-glycopeptide APD(Hnv)*RP (* indicates the site of glycosylation) in presence of VU-3C6.

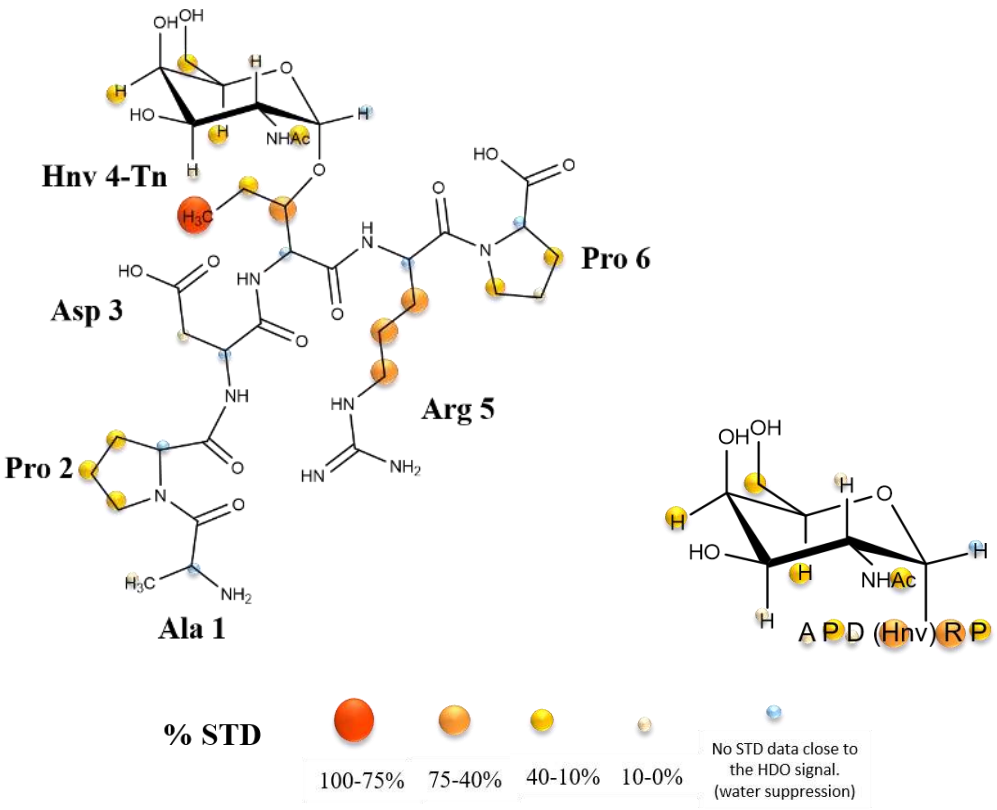


Figure 3.9: STD-derived epitope mapping of Tn-glycopeptide mimetic APD(Hnv)*RP (* indicates the site of glycoslation) in presence of mAb VU-3C6.

However, STD results point out differences in the recognition between the Tn-glycopeptide mimetic APD(Hnv)*RP and the corresponding natural Tn-glycopeptide fragment (Figure 3.10) which seems indicate differences in the binding mode of these two glycopeptides. The interaction of the natural Tn-glycopeptide was previously reported and was used for comparison [41].

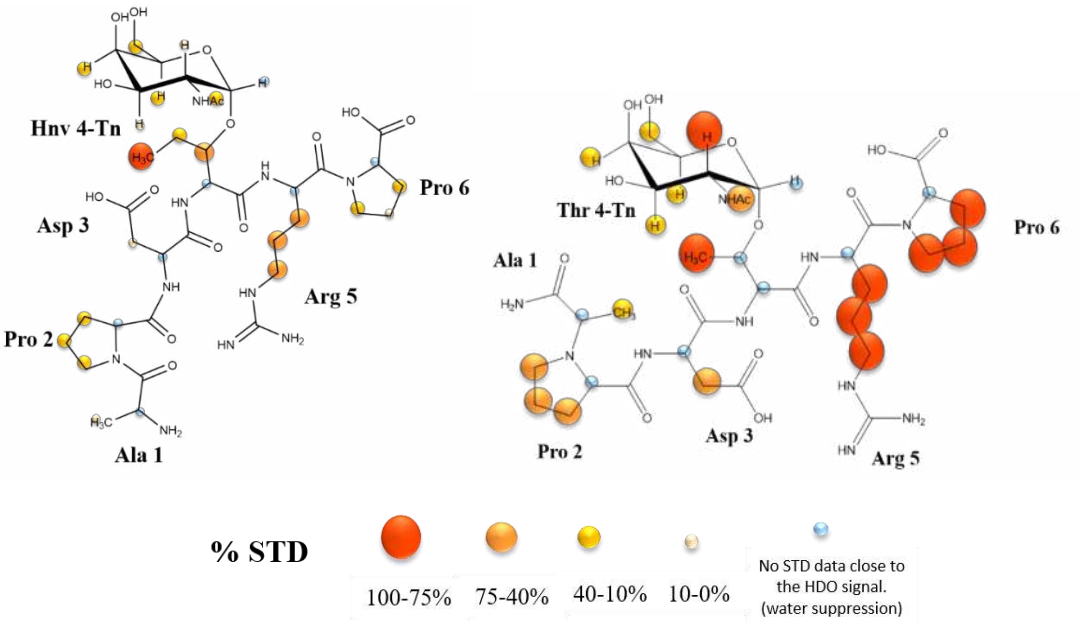


Figure 3.10: Comparison between the STD-derived epitope mapping of Tn-glycopeptide mimetic APD(Hnv)*RP with the one obtained for the Tn-glycopeptide APDT*RP (where * indicates the site of glycosylation) in presence of mAb VU-3C6.

Glycan microarrays also demonstrate binding of Tn-glycopeptide mimetic and VU-3C6. Glycan microarrays were carried out through a project collaboration with the researchers Fayna Garcia-Martin (Hokkaido University, Japan) and Francisco Corzana (Universidad de la Rioja).

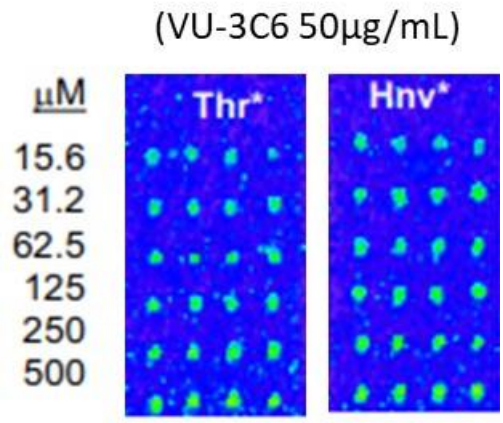


Figure 3.11: Epitope mapping analysis of anti-MUC1 VU-3C6 mAb (50 µg/mL). Fluorescent image scan of natural Tn-glycopeptide (Thr*) and Tn-glycopeptide mimetic (Hnv*). Glycopeptides with natural Thr and non-natural Hnv were printed at 6 different concentrations (15.6, 31.2, 62.5, 125, 250, and 500 µM) onto an aminooxy-functionalized microarray in quadruplicate.

An STD-NMR binding experiment of Tn-glycopeptide mimetic was also accomplished with SM3 Fc fragment antibody (Figure 3.12). The SM3 Fc was prepared in collaboration with Ramón Hurtado-Guerrero (BIFI institute and Universidad de Zaragoza) and Francisco Corzana (Universidad de la Rioja) and following the protocol previously reported [40].

Noteworthy, the STD response obtained for the Tn-glycopeptide mimetic is very weak (Figure 3.12 and 3.13). SM3 Fc is much smaller than VU-3C6 and a lot of interference in the spectrum baseline is detected. Moreover, the transfer of saturation in the case of SM3 Fc fragment will be poor since the spin diffusion through all the receptor will be less efficient which will contribute to the poor quality of the STD spectrum. Despite this it was possible by STD-NMR detect interaction with the residues of Ala 1, Pro 2 and Asp 3 and with less intensity the methyl of the NHAc group of GalNAc residue.

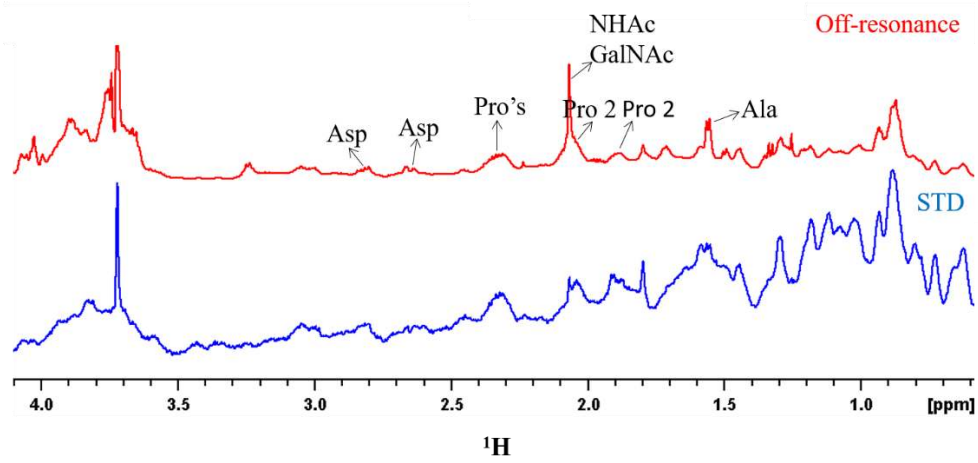


Figure 3.12: STD spectrum (blue) and the off-resonance spectrum (red) for the STD-NMR experiment of the Tn-glycopeptide APD(Hnv)*RP (* indicates the site of glycosylation) in presence of SM3 Fc.

Accordingly to this data, SM3 seems to recognize better the N-terminus of the peptide –APD(Hnv)*RP- sequence.

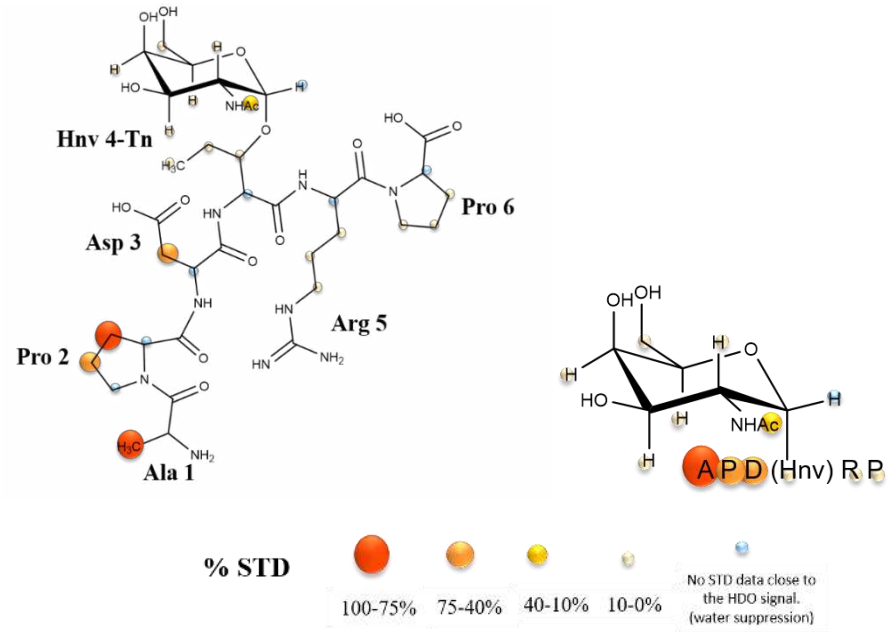


Figure 3.13: STD-derived epitope mapping of Tn-glycopeptide mimetic APD(Hnv)*RP (* indicates the site of glycoslation) in presence of SM3 Fc.

The interaction of the glycopeptide mimetic with SM3 was also investigated by glycan microarrays and bio-layer interferometry technique (BLI). Glycan microarrays clearly demonstrate binding of Tn-glycopeptide mimetic and SM3 antibody (Figure 3.14 Panel A). From BLI data it was possible to extract the affinity constant of the glycopeptide mimetic to the antibody and the results are shown in Figure 3.14 Panel B. In fact, BLI results clearly demonstrate a very high affinity for the non-natural Hnv-containing Tn-glycopeptide (assigned as Hnv* in Figure 3.14 Panel B). The high affinity value obtained can explain the weak STD response in the STD-NMR experiment. In fact, tight ligand binders are not the most appropriate ligands to be studied by STD-NMR spectroscopy.

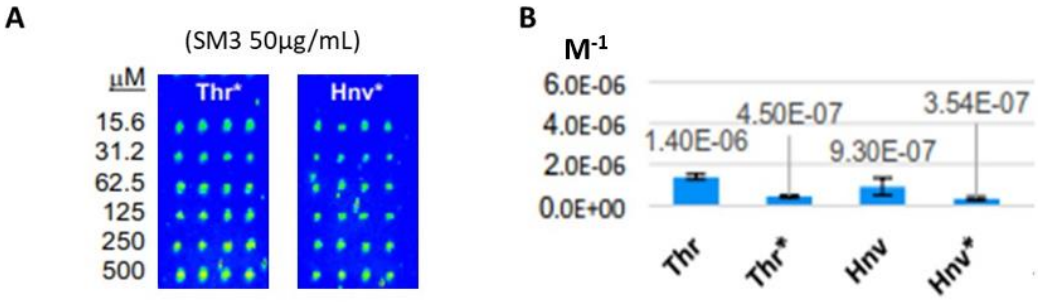


Figure 3.14: A. Epitope mapping analysis of anti-MUC1 SM3 mAb (50 µg/mL). Fluorescent image scan of natural Tn-glycopeptide (Thr*) and Tn-glycopeptide mimetic (Hnv*). Glycopeptides with natural Thr and non-natural Hnv were printed at 6 different concentrations (15.6, 31.2, 62.5, 125, 250, and 500 µM) onto an aminooxy-functionalized microarray in quadruplicate. **B.** Graphical presentation of the K_a values (M^{-1}) determined by BLI technique for peptides containing the natural Thr and non-natural Hnv and their corresponding Tn-derivatives, Thr* and Hnv*.

Structural information of Tn-glycopeptide/SM3 complex was obtained using X-ray crystallography. Crystals were successfully obtained by Hurtado-Guerrero's group in Universidad de Zaragoza. The structure of the complex is displayed in the Figure 3.15.

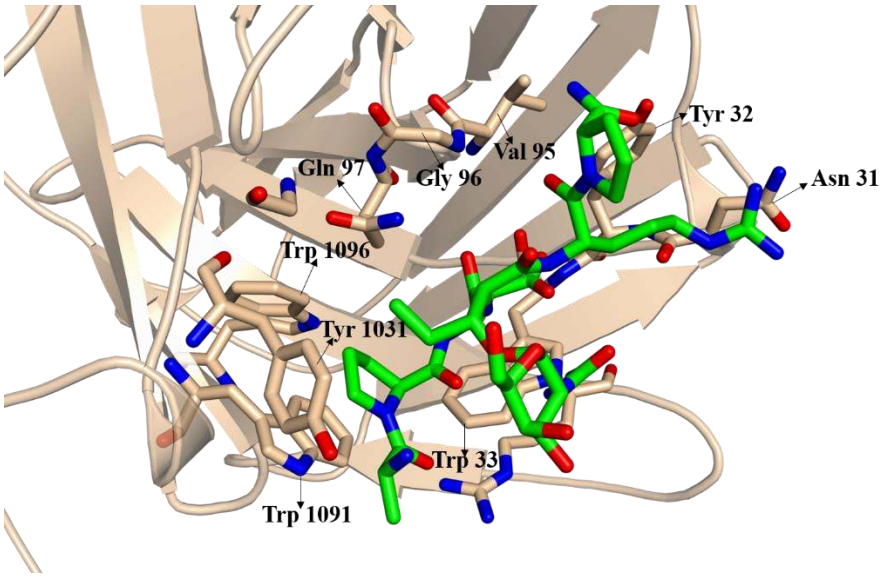


Figure 3.15: X-ray Crystallography structure of SM3 with the peptide APD(Hnv)*RP in the binding site. The ligand is colored according the atom, carbon is green, oxygen red and nitrogen is blue. The residues important for the binding interaction are identified. Image obtained using the software PyMOL [94].

X-ray crystallography structure of the complex shows the residues of Ala 1, Pro 2 and Asp 3 inside the binding site while the rest of the peptide, including the sugar, is turned outside. These results validate our STD-NMR binding experiment highlighting that besides the weak STD response, the experiment of STD was sensitive to extract the binding mode displayed in the X-ray structure. By the X-ray structure, it is possible to appreciate that the binding site of SM3 is mainly composed by aromatic amino acids. This feature adds a nonpolar character to the binding site, turning the recognition process rich in van der Waals and CH-π type of interactions. The Asp can still establish interactions in solution mediated by hydrogen bounds or electrostatic interactions. However, the rest of the peptide, namely Hnv 4, Arg 5 and Pro 6, are solvent exposed. Once again, this observation matches with the STD response deduced in the STD-NMR experiment (Figure 3.13).

Interactions between the natural Tn-glycopeptide and SM3 antibody were previously investigated by STD-NMR [36]. This study demonstrates that Pro 1 in PDT*RP, that corresponds to Pro 2 in APD(Hnv)*RP, receives the highest % of saturation together with the methyl of NHAc group of GalNAc. This result resembles our STD NMR results in the case of the Tn-glycopeptide mimetic. However, in the natural Tn-glycopeptide the STD-NMR spectrum also shows medium STD responses for the peptide amino acids Thr 3, Arg 4 and Pro 5 in –PDTRP- sequence. Therefore, the replacement of the natural Thr by Hnv slightly affects the binding mode towards SM3. This may be explained by the longer side chain of Hnv, which may hinder partially the contacts between the C-terminal part of peptide and the SM3 binding pocket.

At the end, it was studied the interactions of Tn-glycopeptide mimetic and 14D6 mAb. The antibody 14D6 belongs to anti-Tn family. In contrast to SM3 and VU-3C6, which present more specificity to MUC1 peptide sequence, this type of antibody is specific for the Tn-antigen [41]. This antibody binds exclusively Tn-glycopeptides and not non-glycosylated peptides.

STD-NMR spectrum and STD-derived epitope mapping of Tn-glycopeptide in presence of 14D6 is displayed in Figure 3.16 and Figure 3.17, respectively.

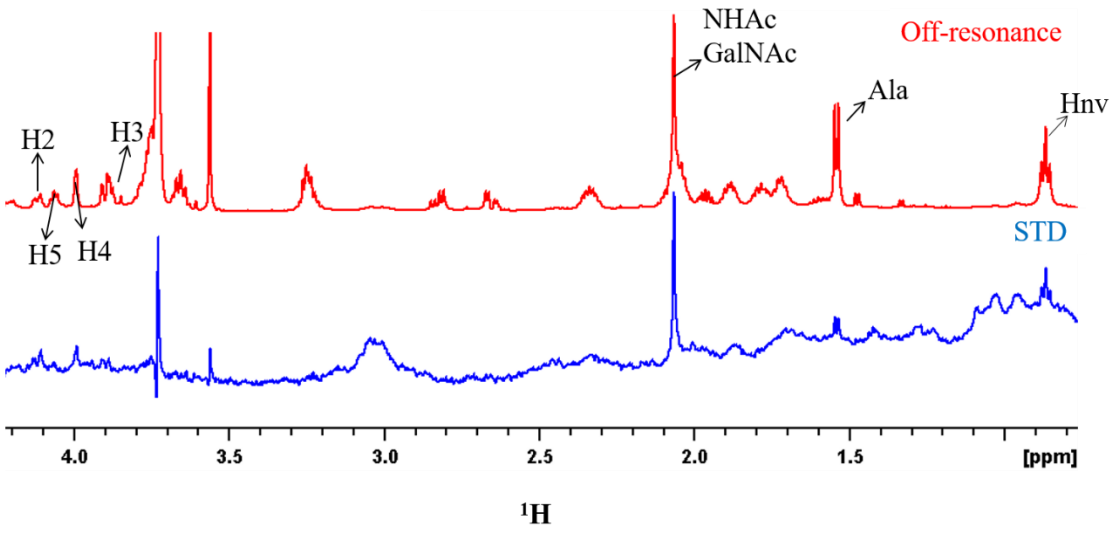


Figure 3.16: STD spectrum (blue) and the off-resonance spectrum (red) for the STD-NMR experiment of the Tn-glycopeptide APD(Hnv)*RP (* indicates the site of glycosylation) in presence of 14D6.

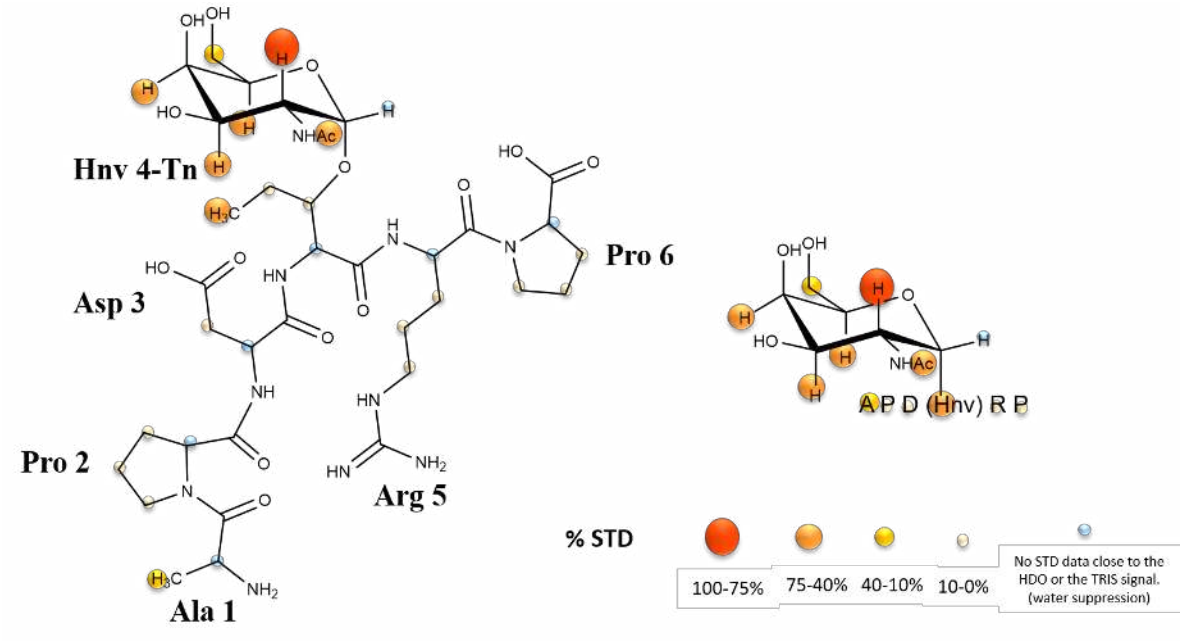


Figure 3.17: STD-derived epitope mapping of Tn-glycopeptide mimetic APD(Hnv)*RP (* indicates the site of glycoslation) in presence of 14D6.

The interaction of the natural Tn-glycopeptide was previously reported and was used for comparison [41]. In particular, for the natural Tn-glycopeptides located at Ser and Thr of MUC1 sequence employed in Helena Coelho et. al. study, it was observed that mainly the GalNAc residue received saturation from the mAb. However, the binding mode of 14D6 seems to be modulated by the type of the amino acid (Ser

vs. Thr) that is glycosylated [41]. Furthermore, this type of antibody recognizes better Ser-conjugated Tn-glycopeptides than Thr-conjugates counterparts.

Herein, for Tn-glycopeptide mimetic it is also evident that GalNAc fragment receives much more saturation from the protein in comparison to the peptide backbone. Nevertheless, Hnv amino acid has much more STD response than that Ser and Thr amino acids. This result indicates that Hnv has a favorable contribution in the recognition of GalNAc by 14D6.

3.1.3. Principal conclusions and perspectives

Overall, with this study it was possible to determine the binding epitope of a glycopeptide mimetic to three distinct antibodies, two from anti-MUC1 family (VU-3C6, SM3) and one from anti-Tn family (14D6). The Tn glycopeptide mimetic APD(Hnv)*RP (* indicates GalNAc glycosylation at Hnv) has a non-natural amino acid, the Hnv (a non-natural isomer of valine), instead of Thr at the APDTRP region of MUC1. Introduction of this non-natural amino acid should increase the bioavailability of this type of fragments and be useful in the design of stable glycan-based cancer vaccines.

Herein, NMR binding experiments were synergistically complemented to other techniques glycan microarrays, BLI measurements and X-ray crystallography, with the main goal to decipher the recognition process of Tn-glycopeptide mimetic.

The anti-MUC1 antibodies recognize the Tn-glycopeptide in a sequence-dependent manner. Our studies show that VU-3C6 and SM3 present different binding preferences towards the same peptide sequence. The 14D6 mAb can recognize the Tn-glycopeptide mimetic APD(Hnv)*RP mainly through the GalNAc and the methyl group of the Hnv amino acid.

In general, these studies can contribute for a structure-guided design of more stable glycan-based cancer vaccines.

In the future, conformational studies should be carried out to understand the impact of non-natural amino acids in the structure presentation of MUC1 antigens to antibodies.

3.2.Chapter ii): Gal-3/TF interactions

3.2.1. Expression and purification of $^{1}H,^{15}N$-labelled Gal-3 CRD

Gal-3 CRD was expressed in *E. coli* cells, in a M9 minimal medium rich in $^{15}NH_4Cl$, in order to express Gal-3 CRD isotopically labelled with ^{15}N for NMR purposes (see material and methods).

To purify the protein, first the cells were centrifuged, the pellet was resuspended and later sonicated to erupt the cellular wall and membrane and free the intracellular medium. Lastly, the protein was extracted via centrifugation, saving the supernatant containing the Gal-3 CRD. The supernatant was then purified by affinity chromatography, using an α-lactose agarose column, resulting in the chromatogram below.

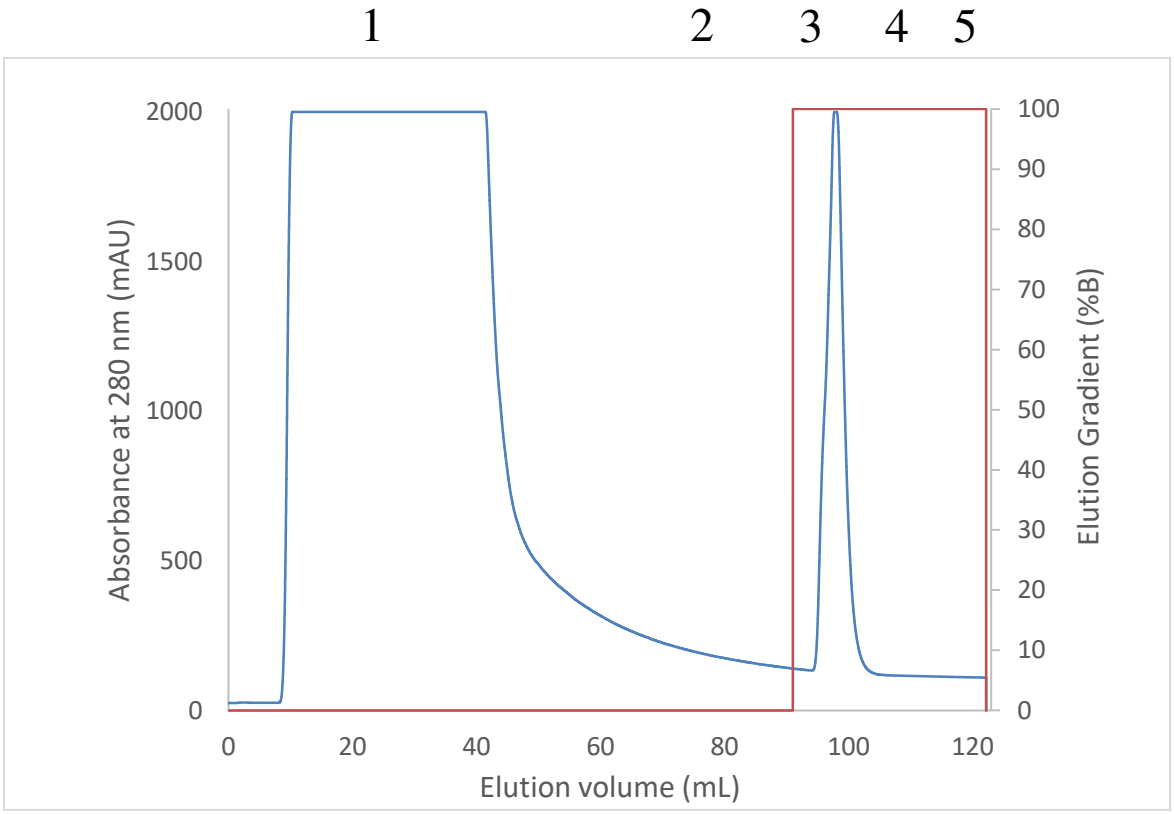


Figure 3.18: Chromatogram obtained in the purification step of Gal-3 CRD, by α-lactose-agarose column affinity chromatography. The blue line corresponds to the absorbance at 280 nm and the red line corresponds to the gradient % B, where B is the elution buffer. The washing buffer is composed by 25 mM PBS, 50 mM NaCl, 1 mM DTT and 0.1% sodium azide, pH 6.8 and the elution buffer contained 25 mM PBS, 50 mM NaCl, 1 mM DTT, 0.1% sodium azide and 150 mM lactose, pH 6.8. The numbers correspond to the fractions collected.

The α-lactose agarose column was chosen for the affinity chromatography, because Gal-3 has high affinity to α-galactosides and α-lactose is a disaccharide form that encodes glucose β1-4 linked to a galactose residue. Hence, during the washing step, Gal-3 will have affinity to the column and will bind and adsorb to it, while the rest of the contaminants will be eluted as demonstrated in fraction 1 of the chromatogram. Before changing to the elution buffer, another fraction was collected, the fraction 2. This fraction contained purified Gal-3 CRD. Then, the column was washed with the elution buffer. This

elution buffer contains lactose, which will compete with the column for Gal-3 allowing to obtain the protein linked with lactose. The increase in absorbance corresponds to the elution of Gal-3 CRD in fraction 3 (Figure 3.18).

To evaluate the process of expression and purification a SDS-PAGE Gel Electrophoresis was performed using all the fractions collected during the expression process and the fractions obtained during the affinity chromatography. The gel is displayed in figure 3.19.

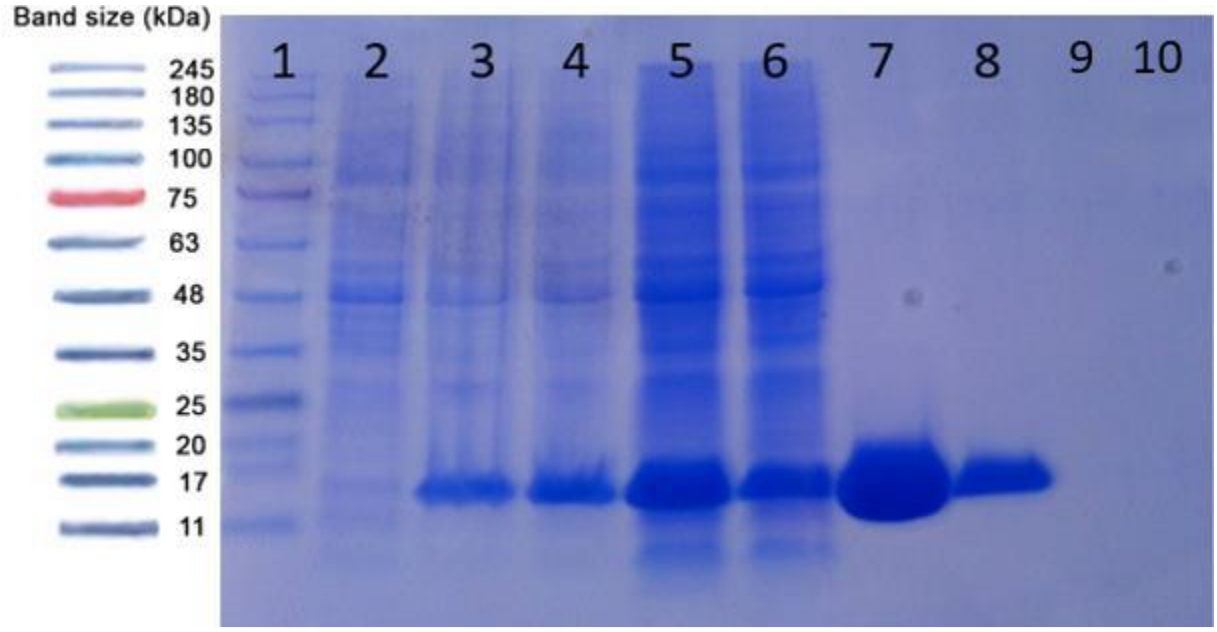


Figure 3.19: 10% polyacrylamide Gel Electrophoresis (SDS-PAGE). **Well 1-** NZYColour Protein Marker II; **Well 2-** Sample before induction; **Well 3-** Sample 2 h after induction; **Well 4-** Sample 4 h after induction; **Well 5-** Supernatant after sonication; **Well 6-** Affinity Chromatography fraction 1; **Well 7-** Affinity Chromatography fraction 3; **Well 8-** Affinity Chromatography fraction 2; **Well 9-** Affinity Chromatography fraction 4; **Well 10-** Affinity Chromatography fraction 5.

By analysis of the SDS-PAGE gel it was possible to follow the expression process of Gal-3 CRD. Well 2 corresponds to a fraction before the induction while wells 3 and 4, correspond to fractions 2 and 4 hours after induction, respectively. In wells 3 and 4 a band corresponding to Gal-3 CRD, at 15 kDa (the value of mass was calculated using ExPASy [88]) is detected. The well 5 shows the fraction after sonication.

From the SDS-PAGE analysis the wells 7 and 8 have pure Gal-3 CRD since the only band visible is the one correspondent to the protein. The well 6 corresponds to the fraction 1 of the affinity chromatography. In this well it is possible to observe all the contaminants of the solution. These contaminants do not have affinity for the column and were eluted before the protein. The Gal-3 CRD collected in fraction 2, corresponds to well 8 and was collected before the elution buffer, therefore, this fraction does not have lactose or any contaminant. Gal-3 CRD collected in well 7 was eluted with lactose. It's possible to observe that the protein was collected without contaminants, however it was required to dialyze this fraction for 3 days to remove lactose.

After the dialysis of fraction 3, the concentration of fraction 2 and 3 were determined by UV-Vis spectroscopy. The concentration was calculated using the Lambert-Beer Law, following equation 1:

$$Abs = c \times l \times \varepsilon, \qquad \text{(Equation 1)}$$

where *Abs* is the absorbance measured at 280 nm to which the absorbance at 320 nm was subtracted, c is the concentration, l is the optical path, in this case was 1 cm and ε is the extinction molar coefficient, whose value was 9970 M^{-1} cm^{-1} obtained using ExPASy [88]. The results obtained are in table 3.1.

Table 3.1: Values obtained of absorbance at 280 nm and 320 nm and concentration.

Sample	***Abs* 280 nm**	***Abs* 320 nm**	***Abs* 280 – 320 nm**	**Concentration (µM)**
Fraction 2	1.869	0.397	1.472	147.64
Fraction 3	2.995	1.146	1.849	185.46

With the volumes of each fraction, a yield of 83.9 mg/L of culture was obtained.

3.2.2. Gal-3/TF-antigen interactions monitored by $^{1}H,^{15}N$-HSQC titrations

The interactions of Gal-3 CRD with the TF-antigen (Figure 3.20) were studied by a $^{1}H,^{15}N$-HSQC titration with different ligand concentrations.

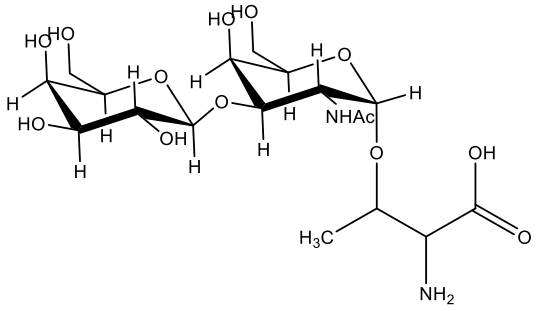


Figure 3.20: Structure of the TF-antigen.

The $^{1}H,^{15}N$-HSQC titration allows us to monitor the chemical shift perturbation for the amino acids involved in the recognition process. A total of six $^{1}H,^{15}N$-HSQC were recorded to perform the titration of Gal-3/TF-antigen. In all points, the concentration of Gal-3 CRD was maintained constant at 50 µM and the ligand concentration was increased during the titration from 0 mM to 1.5 mM (more details in material and methods).

Figure 3.21 shows the superposition of the $^{1}H,^{15}N$-HSQC spectrum of Gal-3 with distinct concentrations of the ligand. Figure 3.22 shows the chemical shift perturbation for the residues His 158 and Thr 175.

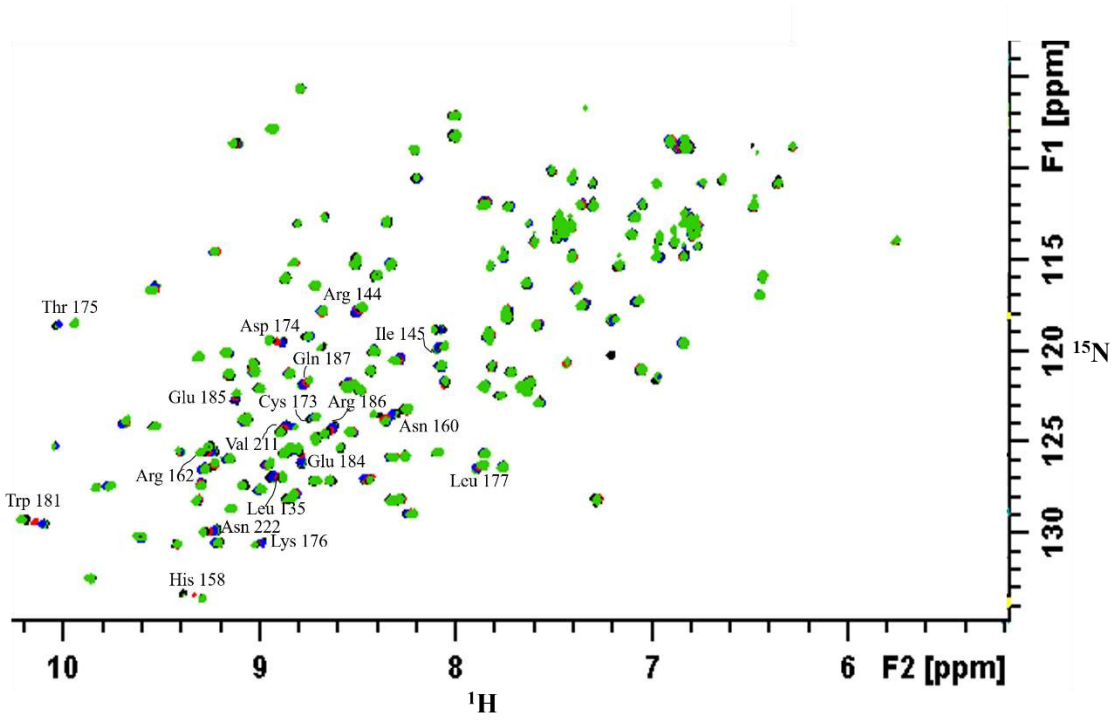


Figure 3.21: Overlap of the six ^{1}H, ^{15}N-HSQC spectra from the titration of Gal-3 CRD with the TF-antigen, with the identification of the peaks with more chemical shift. From left to right: the first black signal corresponds to 1:0, the first green signal corresponds to 1:0.5, the blue signal corresponds to 1:1, the red signal corresponds to the ratio 1:5, the second black signal corresponds to the ratio 1:15 and the second green signal corresponds to the ratio 1:30.

The amino acids with higher chemical shift perturbation upon ligand addition correspond to the amino acids involved in the recognition process with the ligand. Upon interaction with the ligand, the chemical environment of the amino acids at the binding site of the protein will change. If the regime between the free and bound state is in fast chemical exchange in the chemical shift NMR time scale, only one peak of the NH bond is detected. Moreover, during the titration this NH will experience a chemical shift perturbation with the increase of ligand concentration until saturation. By this technique it is possible to determine the protein binding site.

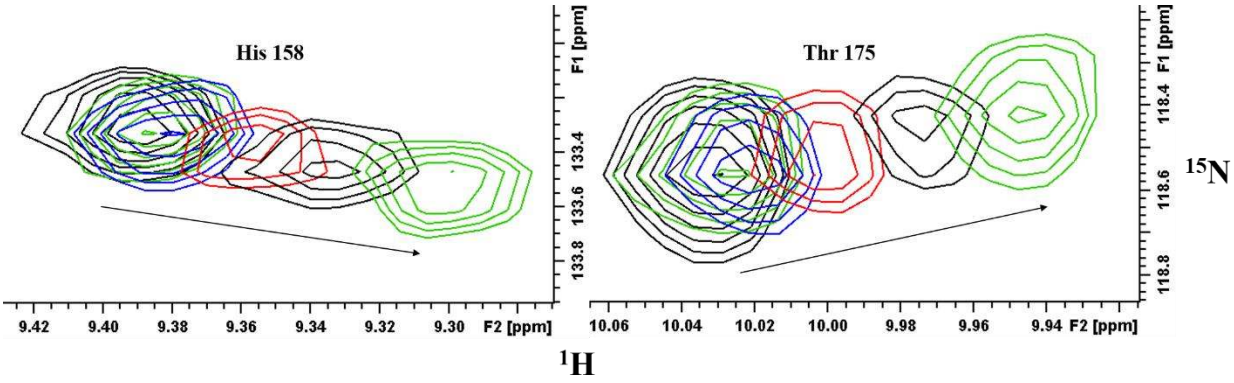


Figure 3.22: Overlap of the titration spectra for the residues His 158 and Thr 175. From left to right: the first black signal corresponds to 1:0, the first green signal corresponds to 1:0.5, the blue signal corresponds to 1:1, the red

signal corresponds to the ratio 1:5, the second black signal corresponds to the ratio 1:15 and the second green signal corresponds to the ratio 1:30.

To determine the residues that have a significant chemical shift due to the interaction with the ligand, the value of the combined chemical shift (Δδ*comb*) was calculated. This value correlates the chemical shift of each titration spectra with the chemical shift of the spectrum without ligand and employs equation 2:

$$\Delta\delta comb = \sqrt{(\Delta\delta H)^2 + (\frac{\omega H}{\omega N}\Delta\delta N)^2}, \qquad \text{(Equation 2)}$$

where $\Delta\delta H$ is the chemical shift difference for the signal obtained in one of the concentrations of the titration with the chemical shift of the proton in the spectrum without ligand, $\Delta\delta N$ is the chemical shift difference between the signal obtained in one of the concentrations of the titration with the chemical shift of the nitrogen atom in the spectrum without ligand and the $\frac{\omega H}{\omega N}$ is the quotient between the gyromagnetic constant of the proton and the gyromagnetic constant of the nitrogen atom.

With the values of $\Delta\delta comb$ obtained for each amino acid for the highest ligand concentration, a column chart was performed. This chart is depicted in figure 3.23.

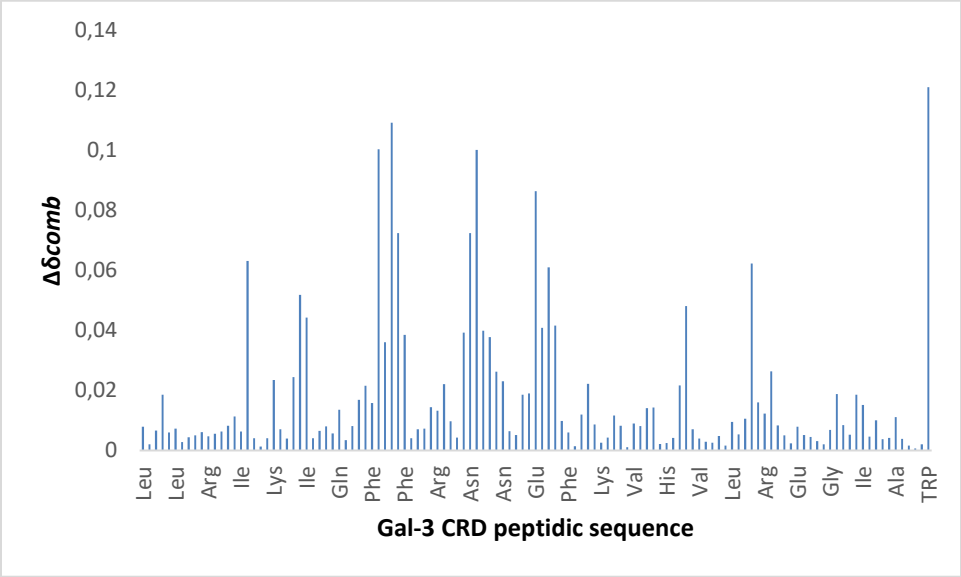


Figure 3.23: Chart with the values of Δδcomb obtained for each amino acid of Gal-3 CRD.

To distinguish which residues, have a higher participation in the interaction process, the standard deviation of the values obtained for Δδ*comb* and a cut-off of three times that value were considered. The calculation of the cut-off is an iterative process, since the standard deviation is recalculated every time a residue has a value of Δδ*comb* higher than the cut-off. These residues are removed from the standard deviation calculation, until all Δδ*comb* values are below the cut-off. (Figure 3.24)

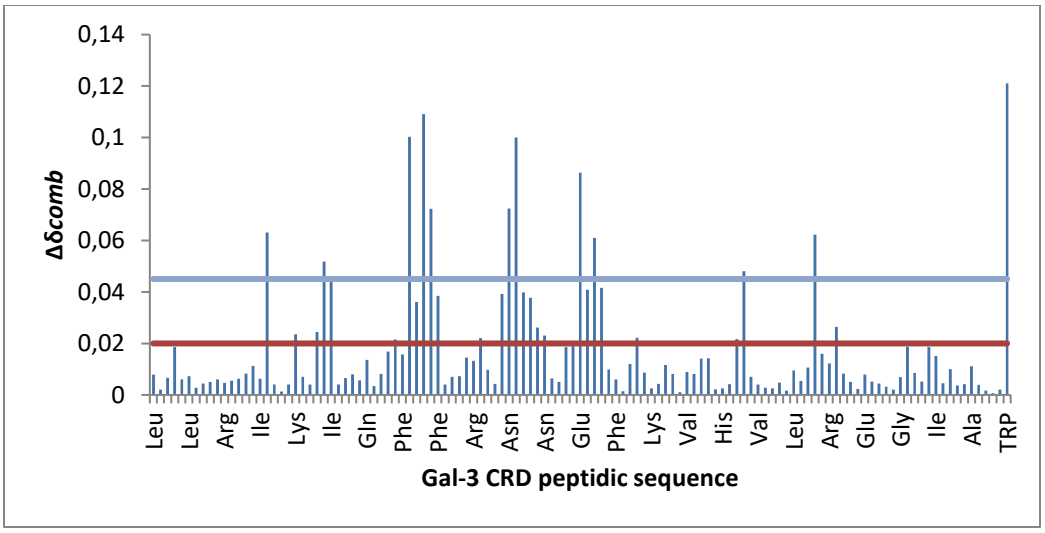


Figure 3.24: Chart with the values of Δδcomb obtained for each amino acid of Gal-3 CRD. The blue line is the second cut-off, while the dark red line is the cut-off used for the determination of the residues participating in the interaction with the TF-antigen.

A total of 29 residues of Gal-3 suffered chemical shift perturbation considering the cut-off of 0.02. These residues were mapped into the X-ray structure of Gal-3 (PDB code 3AYA [56]) and depicted in figure 3.25.

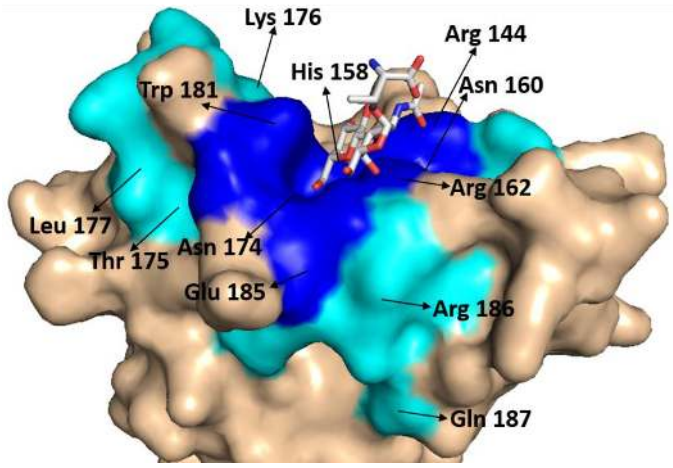


Figure 3.25: Representation of the binding site of the Gal-3 CRD for the TF-antigen (PDB code 3AYA [56]). The ligand is colored according the atom, carbon is white, oxygen red and nitrogen is blue, the binding site at dark blue and the amino acids adjacent to the binding site at cyan. The principal amino acids for the Gal-3/TF-antigen interaction were identified. Image obtained using the software PyMOL [94].

From the NMR results, it was possible to determinate that the principal residues of the binding site were His 158, Arg 162, Arg 144, Asp 160, Asp 174, Glu 184 and the side chain of the Trp 181. These results are in agreement with the X-Ray Crystallography structure [56]. Accordingly to the X-Ray structure the side chain of the Trp 181 is closer to the binding site than the NH of the amide bound. These residues are involved in distinct type of interactions with the residue of galactose in TF-antigen structure, namely such as hydrogen bounds, van der Waals interactions and CH-π stacking.

Besides these residues, there are others that also suffered chemical shift perturbations, namely Thr 175, Lys 176, Leu 177, Gln 187, Arg 186 and Asp 222. These residues are not placed in the primary binding site, but are adjacent and therefore, may suffer chemical shift perturbations due to the conformational changes in the protein upon binding to the ligand.

From NMR titration it was also possible to estimate the K_D of the binding interaction between Gal-3 CRD and the TF-antigen, due to the equilibrium between the bound form and the free form of the complex. The K_D value was obtained following equation 3:

$$\Delta\delta comb = \Delta\delta max \frac{(K_D+[L]_0+[R]_0)-\sqrt{(K_D+[L]_0+[R]_0)^2-4\times([L]_0\times[R]_0)}}{2\times[R]_0}, \quad \text{(Equation 3)}$$

where the K_D is calculated relating the $[L]_0$, which is the concentration of the ligand, the $[R]_0$, the concentration of the receptor, the $\Delta\delta comb$ being the combined chemical shift and $\Delta\delta max$, the maximum value of chemical shift for each residue. Estimation was performed by adjusting the experimental values obtained to the values calculated by the equation 3, using the Solver tool, from Excel. An example of the calculation of K_D is shown below for the residue of Arg 144 (Figures 3.26 and 3.27)

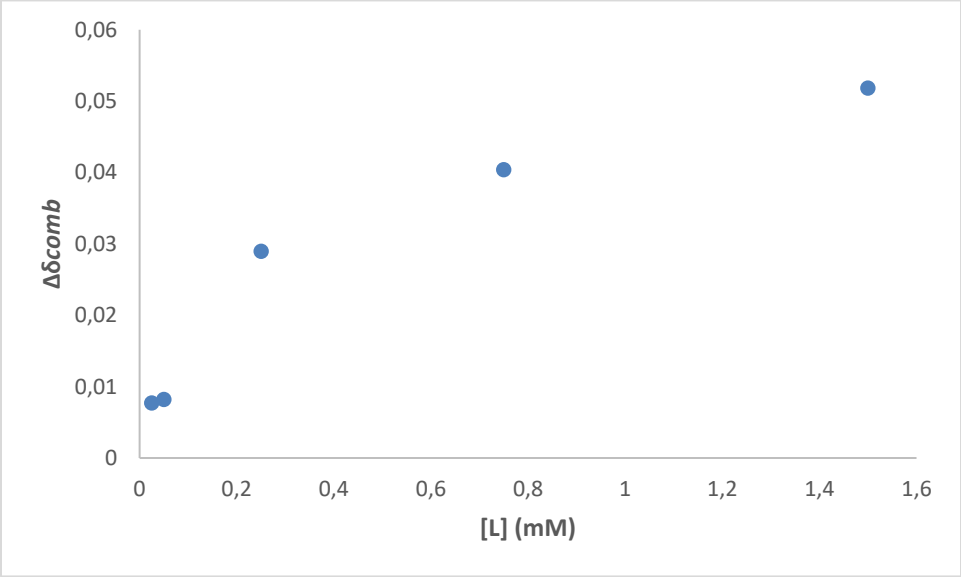


Figure 3.26: Chart of the experimental values of Δδcomb for the residue of Arg 144.

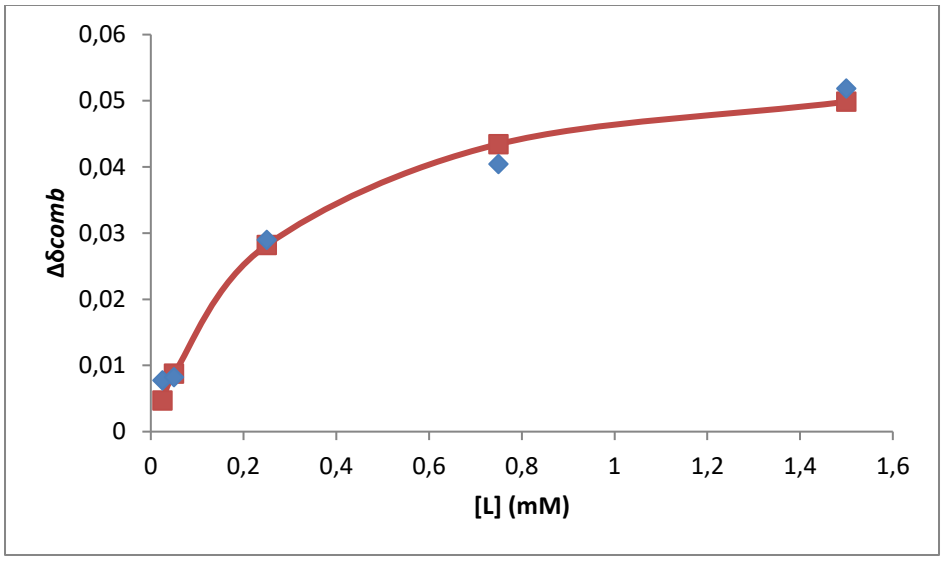


Figure 3.27: Chart with the values of the experimental Δδcomb (blue) and the adjusted calculated Δδcomb (dark red) for Arg 144.

The K_D was estimated for each residue considering their involvement in the binding site and binding interaction and an average of the values was obtained. That average was considered the apparent K_D. The average value obtained for the interaction between Gal-3 CRD and TF-antigen, for a total of 9 residues was 275 μM. This result is in agreement with the K_D obtained by ITC in previous studies [57].

3.2.3. Gal-3/TF-glycopeptide interactions monitored by $^{1}H,^{15}N$-HSQC titrations

The titration of Gal-3 CRD with TF-glycopeptide PDT*RP (where * indicates the site of glycosylation) was also recorded. The structure of the TF-glycopeptide is displayed in the Figure 3.28. The main objective of this study was to understand if the peptide backbone somehow plays a role in the TF-antigen recognition process.

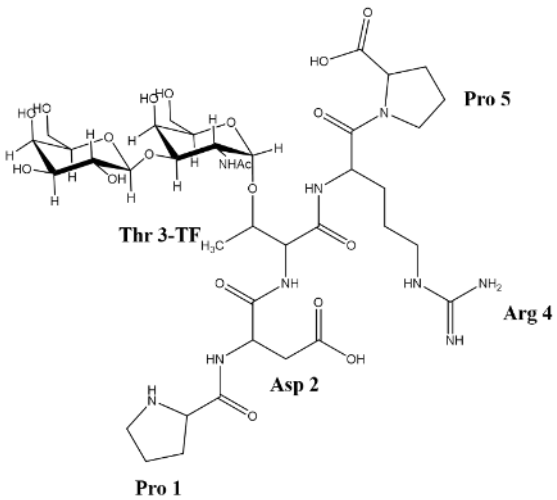


Figure 3.28: Structure of the TF-glycopeptide (PDT*RP where * indicates the site of glycosylation).

A total of six $^{1}H,^{15}N$-HSQC were also recorded in the case of Gal-3/TF-glycopeptide titration. In all points the concentration of Gal-3 CRD was maintained constant at 50 μM and the ligand concentration was increased during the titration from 0 mM to 2.25 mM (more details in material and methods). The superposition of the $^{1}H,^{15}N$-HSQC spectra at distinct conditions is shown in figure 3.29.

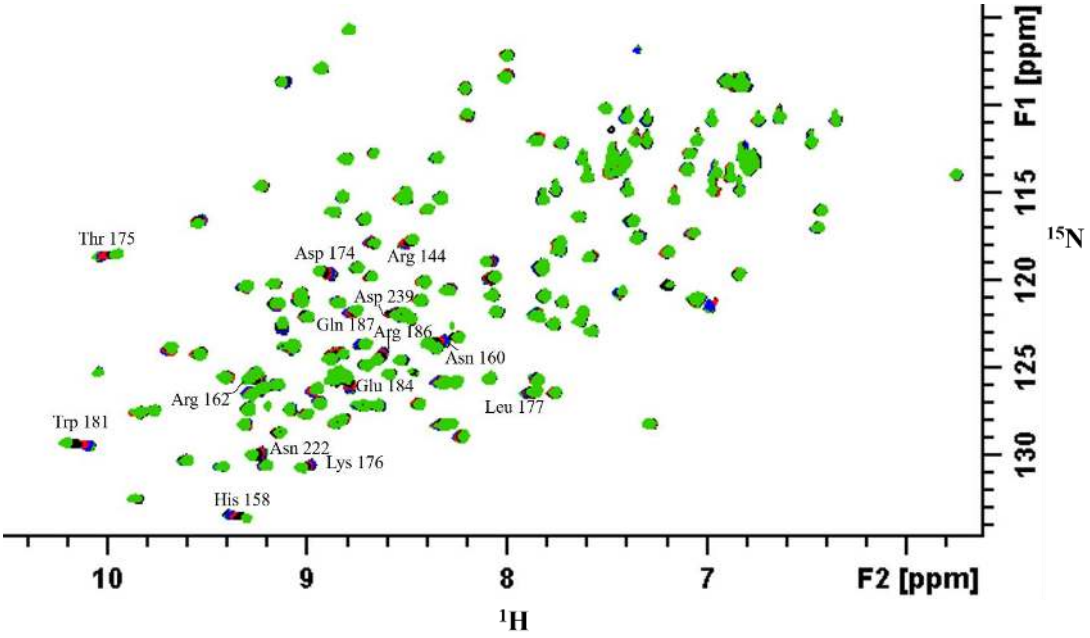


Figure 3.29: Overlap of the six $^{1}H,^{15}N$-HSQC spectra from the titration of Gal-3 CRD with TF-glycopeptide (PDT*RP, where * indicates the site of glycosylation) with the identification of the peaks with more chemical shift. From left to right: the first black signal corresponds to 1:0, the first green signal corresponds to 1:0.5, the blue signal corresponds to 1:1, the red signal corresponds to the ratio 1:5, the second black signal corresponds to the ratio 1:8 and the second green signal corresponds to the ratio 1:45.

The Δδ*comb* was calculated to each residue using the equation 2. A column chart was created with the Δδ*comb* values and corresponding standard deviation. Once again, the cut-off was calculated to identify the amino acids with higher participation in ligand binding (Figure 3.30).

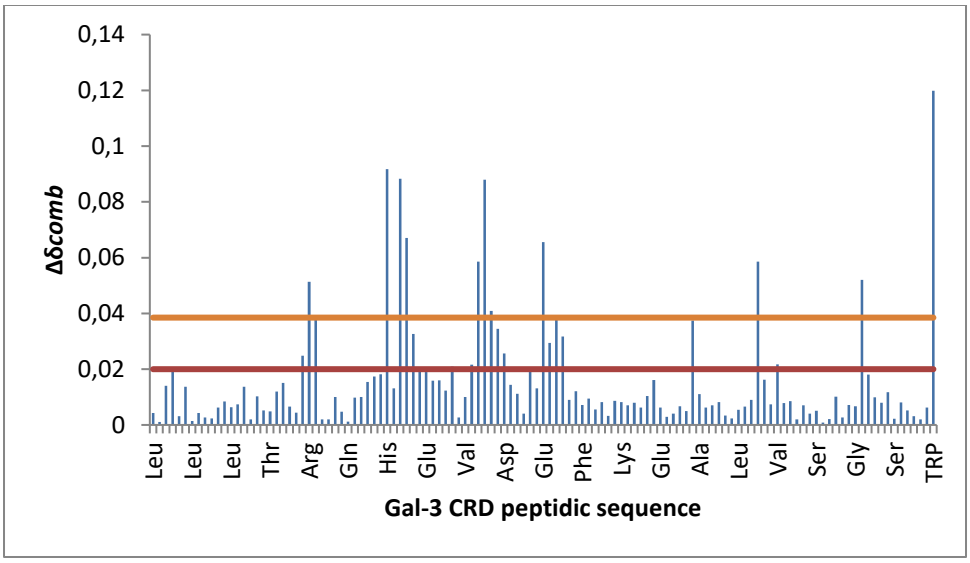


Figure 3.30: Chart with the values of Δδcomb obtained for each amino acid of Gal-3 CRD. The orange line is the second cut-off, while the dark red line is the cut-off used for the determination of the residues participating in the interaction with the TF-glycopeptide (PDT*RP, where * indicates the site of glycosylation).

The residues with values higher than the cut-off were removed from the calculation for the standard deviation and were considered as the ones with a higher participation in the recognition process between the Gal-3 CRD and the TF-peptide PDT*RP (where * indicates the site of glycosylation). A total of 23 residues were identified in these conditions. Considering this information, it was possible to determine the binding site of Gal-3 CRD for this ligand and mapped in the X-ray structure (PDB code 3AYA [56]) (Figure 3.31).

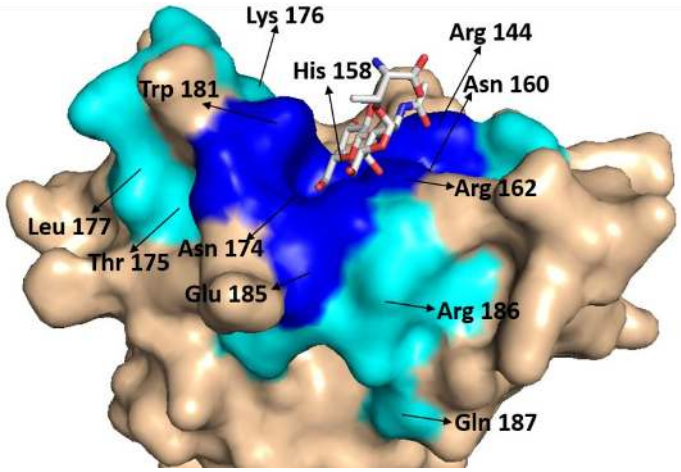


Figure 3.31: Representation of the binding site of the Gal-3 CRD for the TF-glycopeptide (PDT*RP) using X-ray crystallography structure of the complex (PDB code 3AYA [56]). The ligand is colored according the atom, carbon is white, oxygen red and nitrogen is blue, the binding site is represented with dark blue and the amino acids adjacent to the binding site with cyan. The principal amino acids in the Gal-3/TF-peptide interaction were identified in the figure. Image obtained using the software PyMOL [94].

The results obtained from the titration shows that the binding mode TF-glycopeptide to Gal-3 CRD is similar to that obtained for the TF-antigen. This result points out that the sugar moiety is the main feature for Gal-3 recognition.

The apparent K_D was also determined for this interaction, using the same methodology already described and the value obtained was of 413 μM from the average of 9 residues. The value estimated for the K_D of TF-glycopeptide is in the same range of that obtained in the case of TF-antigen, indicating that the PDTRP peptide sequence does not participate in the recognition of TF-antigen by Gal-3. Since, TF-antigen is presented to Gal-3 at the surface of mucins, longer peptides can eventually lead to adjacent interactions with the Gal-3 protein surface and possibly playing a more relevant role in the recognition event. Indeed, ITC results of a TF-glycopeptide encoding a longer peptide sequence show a slight improve in the K_D values from 272 μM to 103 μM due to a switch of the thermodynamics properties [57].

3.2.4. Gal-3/TF-antigen interactions studied by STD-NMR

To better understand the interactions between Gal-3 and TF-antigen, a STD-NMR experiment was performed (see material and methods for more details in experimental conditions). STD-NMR allows mapping interaction from the ligand perspective. With this technique it is possible to determine which ligand protons are in closer contact with the protein. For the STD-NMR experiments, we used the Gal-3 FL expressed and purified by Ana Diniz. Gal-3 FL encodes all Gal-3 domain thus is a larger receptor than Gal-3 CRD. Larger receptors are more appropriate for STD-NMR experiments.

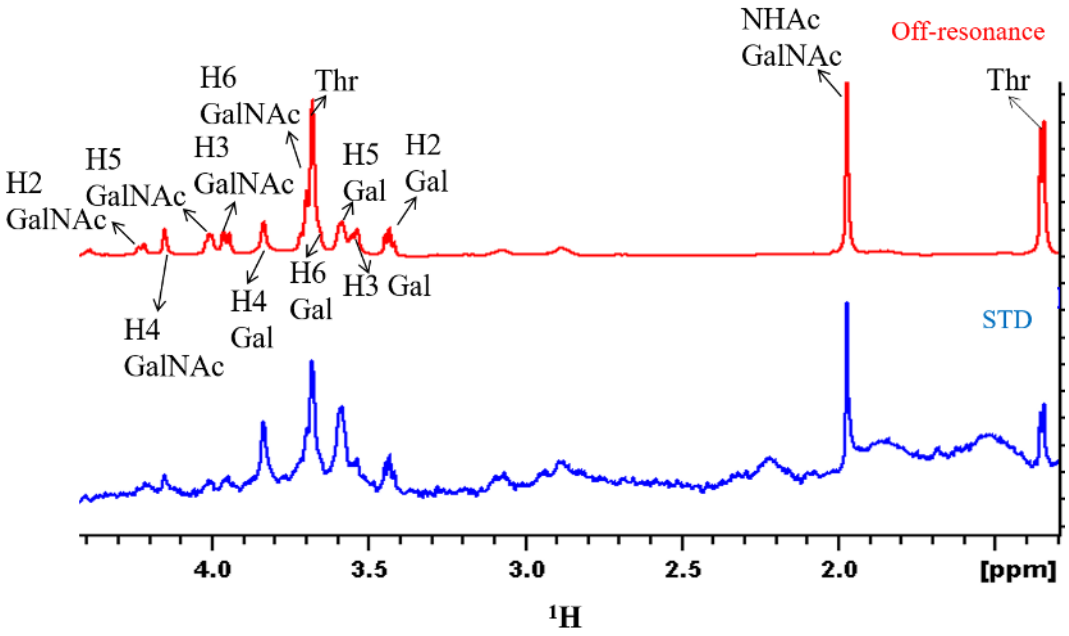


Figure 3.32: STD spectrum (blue) and the off-resonance spectrum (red) of TF-antigen in presence of Gal-3 FL

The STD-derived epitope mapping of TF-antigen is shown in Figure 3.33. From STD analysis it is clear that Gal-3 has a preference to bind the Gal residue of the TF-antigen than GalNAc. The protons H4 and H5 of Gal display more intensity in the STD spectrum. Gal-3 also strongly recognizes the protons H2, H3 and H6s of Gal, as well as, the H2 and protons H6s from GalNAc.

The STD-derived epitope map is totally in agreement with the X-ray crystallography structure (PDB code 3AYA [56]) displayed in Figures 3.25 and 3.31. From X-Ray structure, Gal residue is closer to the Gal-3 binding site, while the GalNAc residue and the Thr are more exposed to the solvent. It is also observed from X-ray structure that protons H2 and H6 of GalNAc moiety are pointing towards Gal-3 surface and thus in closer contact with the protein than the rest of the GalNAc protons. This evidence explains our STD-derived epitope mapping that clearly show H2 and H6s of GalNAc moiety receive higher % of saturation.

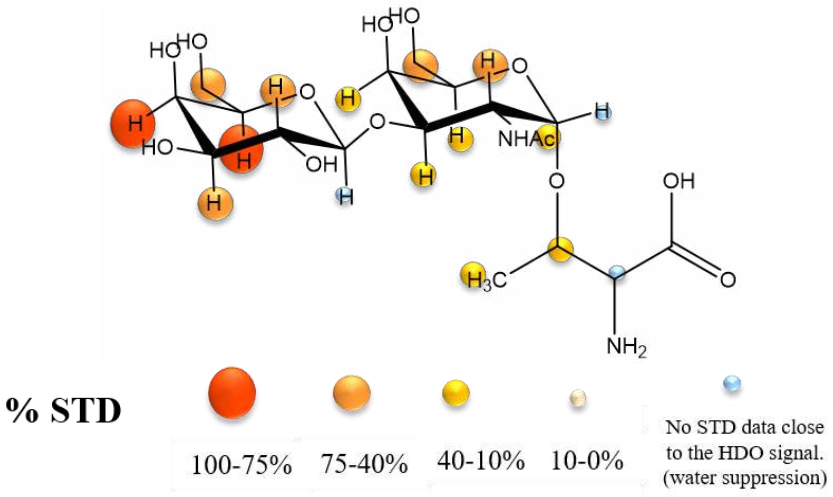


Figure 3.33: STD-derived epitope of TF-antigen in presence of Gal-3 FL.

3.2.5. Gal-3/TF-antigen interactions studied by line broadening analysis

In order to perform a line broadening analysis, two ^{1}H-NMR spectra of TF-antigen in absence and presence of Gal-3 were recorded. These two spectra were superimposed and compared (Figure 3.34).

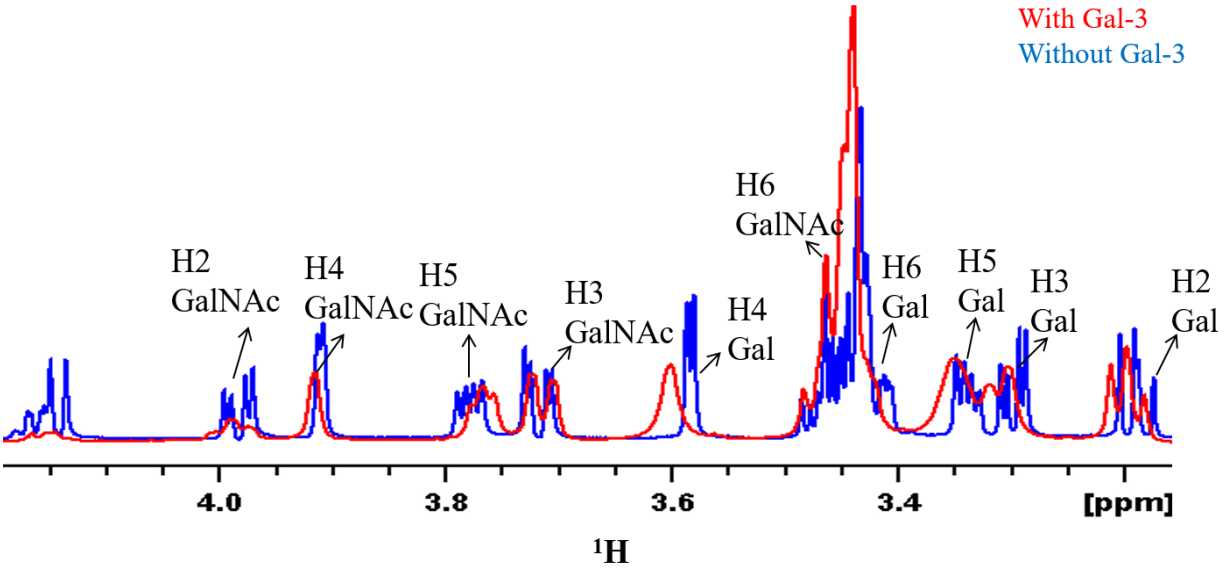


Figure 3.34: Superimposition of TF-antigen ^{1}H–NMR spectrum at 500 µM (selected the sugar proton region in ^{1}H-NMR) in presence of Gal-3 CRD (red) with 50 µM and in absence of Gal-3 CRD (blue).

A strong broadening effect for all sugar protons can be appreciated in presence of Gal-3. However, the increase in the line width is larger for Gal than GalNAc in TF-moiety. This result is totally in agreement with the fact that Gal-3 is a carbohydrate binding protein with affinity for galactosides. Line broadening analysis also supports the STD-NMR results (Figure 3.33). In fact, the signals with a more severe line broadening effect where those that also display more % of saturation in STD-NMR experiment.

The line broadening effect in TF-glycopeptide proton signals in presence of Gal-3 was also studied (Figure 3.35). Herein, once again the Gal residue is the main point of interaction with Gal-3. Indeed, protons H4, H6s, H2 and H5 experienced an intense raise of their line widths, which indicates a strong interaction. The raise in line width is less pronounced in the case of the GalNAc residue, nevertheless it is clear that protons H6s, H2, H4 and H3 are still strongly affected upon binding. Interestingly, the line width of proton signals of the amino acids remained almost unperturbed. This result highlights that the peptide PDTRP does not interact with the protein explaining the similar values of K_D for both molecules.

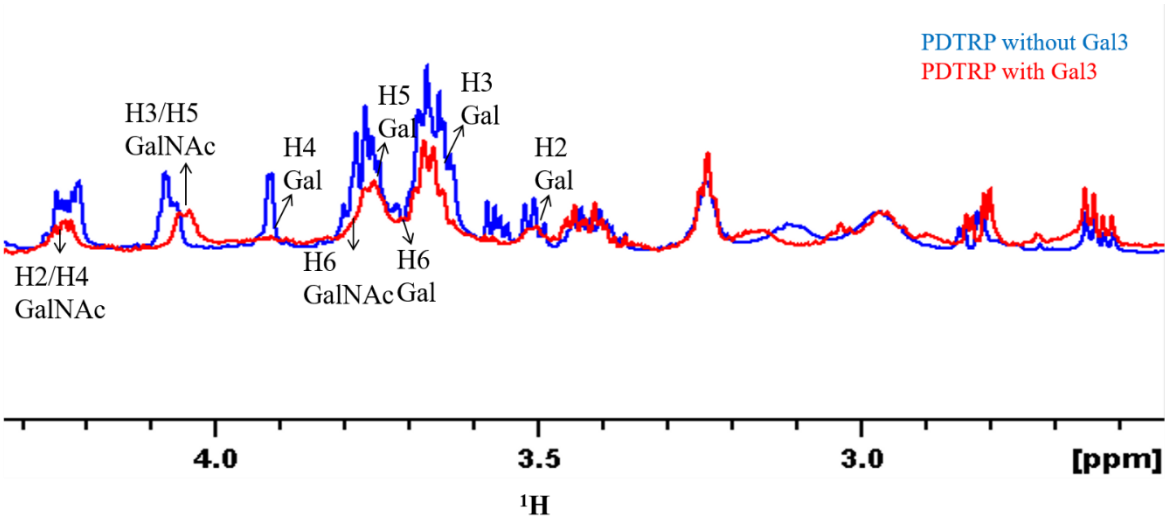


Figure 3.35: Superimposition of TF-glycopeptide (PDT*RP where * indicates the site of glycosylation) ^{1}H–NMR spectrum (from 2.5 to 4.5 ppm) at 400 µM in presence of Gal-3 CRD (red) with 50 µM and in absence of Gal-3 CRD (blue).

3.2.6. Principal conclusions and perspectives

With this study it was possible to investigate the molecular interactions that govern the recognition process of TF-antigen by Gal-3 by combining distinct NMR binding experiments from protein and ligand perspective.

^{1}H,^{15}N-HSQC titrations of Gal-3 CRD with TF-antigen and the TF-glycopeptide (PDT*RP) with Gal-3 allow us to identify the binding site and the apparent K_D for the interaction. Accordingly to ^{1}H,^{15}N-HSQC data, there is not much difference in the recognition between the TF-antigen glycosylated in Thr and TF-glycopeptide linked to the PDTRP peptide sequence upon binding with Gal-3. In addition, the K_D estimated values seem to corroborate this conclusion.

STD-NMR clearly shows that the main key feature for the recognition process is the Gal residue in TF-moiety that is pointing towards the binding site. In contrast, the GalNAc residue and the Thr amino acid are more solvent exposed. Our STD-derived epitope mapping totally matches with X-Ray crystallography data.

Line broadening analysis of TF-antigen matches with the STD-NMR results for this compound. Additionally, this approach allows us to compare the effects on the line width between the disaccharide TF and peptide sequence PDTRP in TF-glycopeptide. The results clearly point out that the 5-amino acids peptide sequence PDTRP should be solvent exposed in the TF-glycopeptide/Gal-3 complex.

The protocol to study TF-based ligands towards Gal-3 was established during this thesis and can now be used to perform a rational structure based design of potential glycomimetics inhibitors of Gal-3. This is especially important, because the interactions of Gal-3 with TF-antigen present in cancer cells are extremely important in the metastasis and tumor invasion process.

3.3. Chapter iii): MUC1 *O*-Glycosylation by GalNAc-T3

3.3.1. Expression and Purification of $^{1}H,^{15}N$-labelled MUC1-4TR

The expression of MUC1-4TR construct was accomplished in *E. coli* cells, in a M9 minimal medium rich in $^{15}NH_4Cl$ to selective isotopically label the protein with ^{15}N (see material and methods).

The isolation of the pure MUC1-4TR suitable for NMR purposes was achieved after three purification steps: 1) affinity chromatography towards his-tag; 2) desalting chromatography and 3) reverse-phase chromatography after digestion with TEV protease. More details are presented in material and methods.

In the first step, the protein construct MUC1-4TR was purified by affinity chromatography using a nickel column since the construct encodes a His-tag with affinity to Ni^{2+} ions. Therefore, while the protein is adsorbed to the column, the other impurities will be eluted first as observed in fractions 1 and 2 (Figure 3.36). Initiating the gradient with buffer B (containing imidazole), the protein starts to renature and slowly starts to lose affinity for the column and elute (fraction 3, Figure 3.36), because the imidazole competes with his-tag for Ni^{2+} ions in the column. Thus, with the increase of the gradient of buffer B and total renaturation of the protein, the protein lost totally the affinity for the column and is finally collected in fractions 4 and 5 (Figure 3.36).

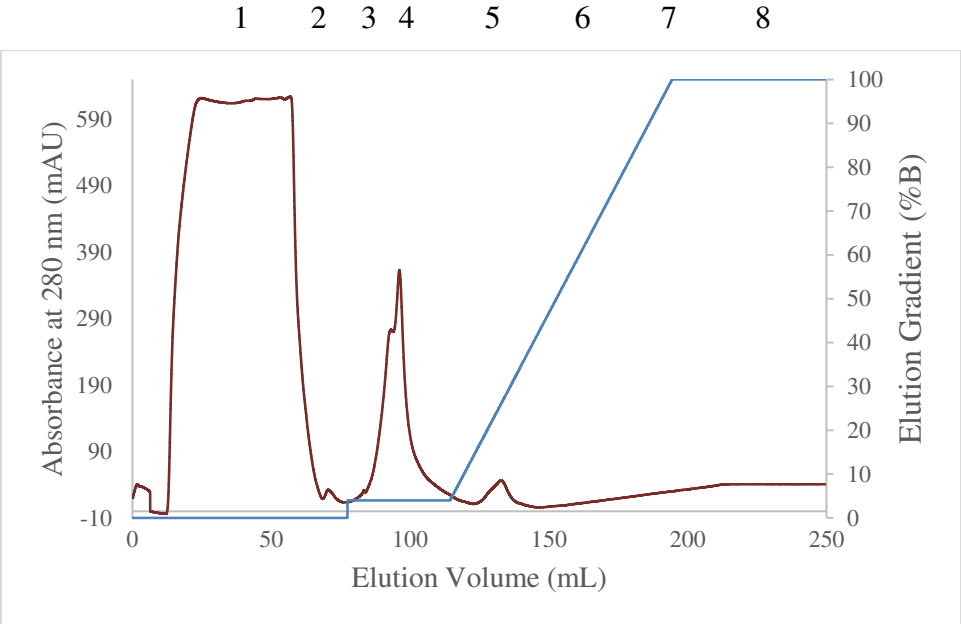


Figure 3.36: Chromatogram obtained from the affinity chromatography, using a Ni column. The brown line corresponds to the absorbance at 280 nm and the blue line corresponds to the gradient of buffer B. Buffer B is the elution buffer containing PBS 10 mM, NaCl 150 mM, imidazole 1 M and β-mercaptoethanol 1 mM. The numbers on top of the chromatogram correspond to the fractions collected.

After the first step of purification, the collected fractions contain high amounts of imidazole. Thus, to remove the imidazole a desalting chromatography was employed (Figure 3.37). In the chromatogram it is possible to observe the different injections from the different fractions collected in the previous

chromatographic step. After each injection three fractions were collected (see material and methods for the details).

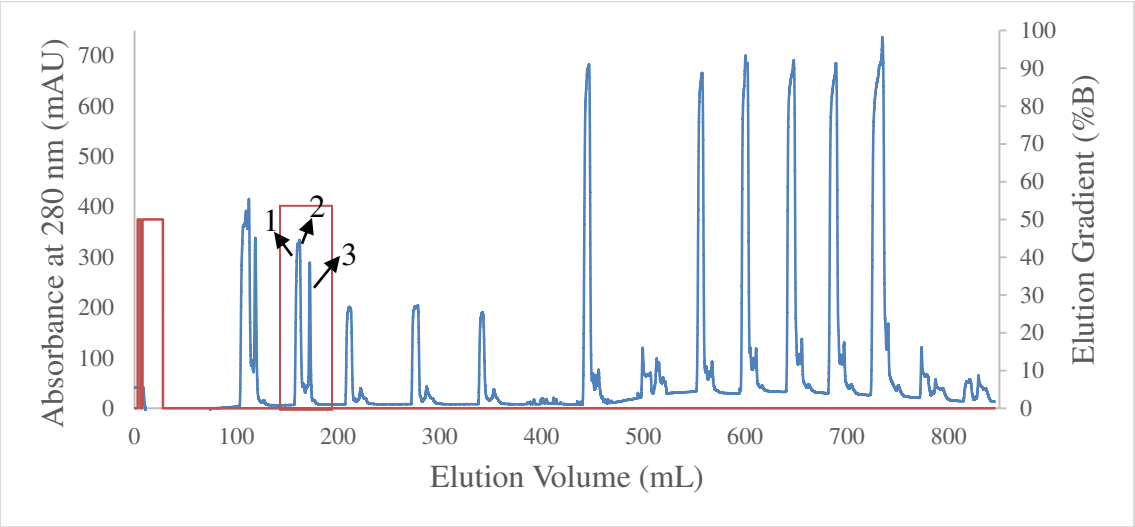


Figure 3.37: Chromatogram obtained from the desalting chromatography. The blue line corresponds to the absorbance at 280 nm and the red line to the elution gradient (%B). Buffer B is the elution buffer composed by PBS 10 mM, NaCl 150 mM and β-mercaptoethanol 1 mM. Each peak is cycle (red rectangle) and each cycle has 3 fractions, like the example showed.

To evaluate if the fractions contain the MUC1-4TR construct, a polyacrylamide and sodium dodecyl sulfate gel electrophoresis (SDS-PAGE) was recorded (material and methods for SDS-PAGE details). In the SDS-PAGE, it was used the fraction 2 of each cycle, from each fraction of the affinity chromatography. Accordingly to SDS-PAGE, well 2 does not contain the protein construct of MUC1-4TR construct; well 3, contains a small amount of MUC1-4TR construct together with an impurity. While, wells 4 and 5 contain the MUC1-4TR construct almost pure (Figure 3.38).

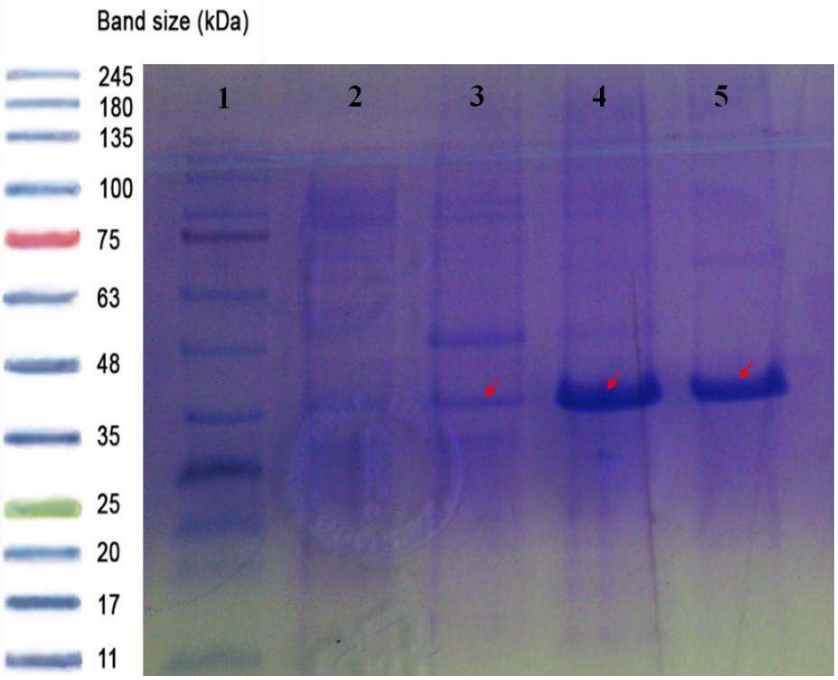


Figure 3.38*:* 10% Polyacrylamide Gel Electrophoresis. **Well 1-** NZYColour Protein Marker II; **Well 2-** Fraction 1 after desalting; **Well 3-** Fraction 3 after desalting; **Well 4- F**raction 4 after desalting; **Well 5-** Fraction 5 after desalting. The red arrow indicates the MUC1-4TR construct.

Fractions 4 and 5 were further subject to the action of TEV protease and SDS-PAGE was accomplished to verify if the digestion with TEV protease was successful (Figure 3.39). The protocol for TEV protease digestion is described in material and methods.

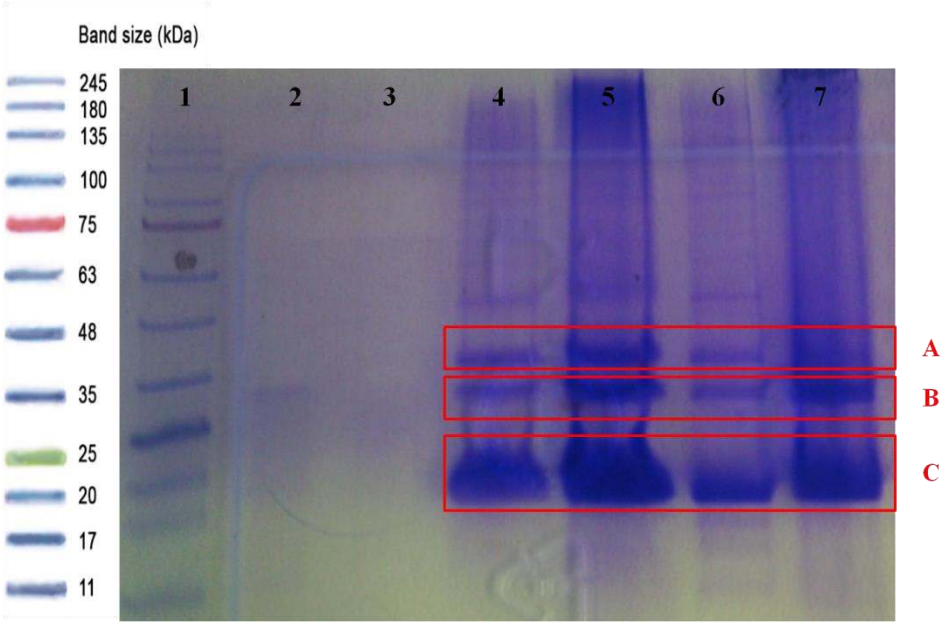


Figure 3.39: 10% Polyacrylamide Gel Electrophoresis of TEV reaction. **Well 1-** NZYColour Protein Marker II; **Well 2-** Fraction 4 after centrifuging; **Well 3-** Fraction 5 after centrifuging; **Well 4-** Pellet diluted of fraction 4; **Well 5-** Pellet diluted of fraction 5; **Well 6-** Pellet fraction 4; **Well 7-** Pellet fraction 5. The bands assigned in the gel with

A correspond to MUC1-4TR that remained undigested (MUC1-4TR+KSI). The bands assigned with **B** correspond to the TEV protease and the bands assigned with **C** correspond to the fusion protein KSI.

SDS-PAGE was used to monitor the digestion with TEV protease (Figure 3.39). Well 2 and 3 contains the supernatant after TEV digestion (fractions 4 and 5, respectively). In these wells no bands are observed. This result indicates that the digestion by TEV protease was successful. Taken into account the MUC1-4TR sequence, this protein should be invisible in the SDS-PAGE upon revelation either by coomassie and silver nitrate stains. The expression vector pHTP-KSI, produced by NZYTech, was deliberately designed, so that after TEV digestion, only the interest protein MUC1-4TR remains soluble (KSI fusion protein precipitates). Therefore, the wells 4, 5, 6 and 7 corresponds to the pellets of fraction 4 and 5 (in different concentration) and can be attributed to TEV protease (Figure 3.39, rectangle **B**), the protein fusion (Figure 3.39, rectangle **C**), and a small amount of the complex MUC1-4TR+KSI that remained undigested (Figure 3.39, rectangle **A**). In addition, MUC1 sequence also does not have aromatic amino acids. Therefore, cannot be detected by absorbance at 280 nm. Thus, the last step to obtain the pure MUC1-4TR protein employs a reversed-phase chromatography in a HPLC apparatus and using a C18 column with the detector placed at 220 nm, 230 nm and 254 nm (Figure 3.40). The first 20 mL (with 10% of acetonitrile and flow 2 mL/min) were used to remove some impurities (Figure 3.40).

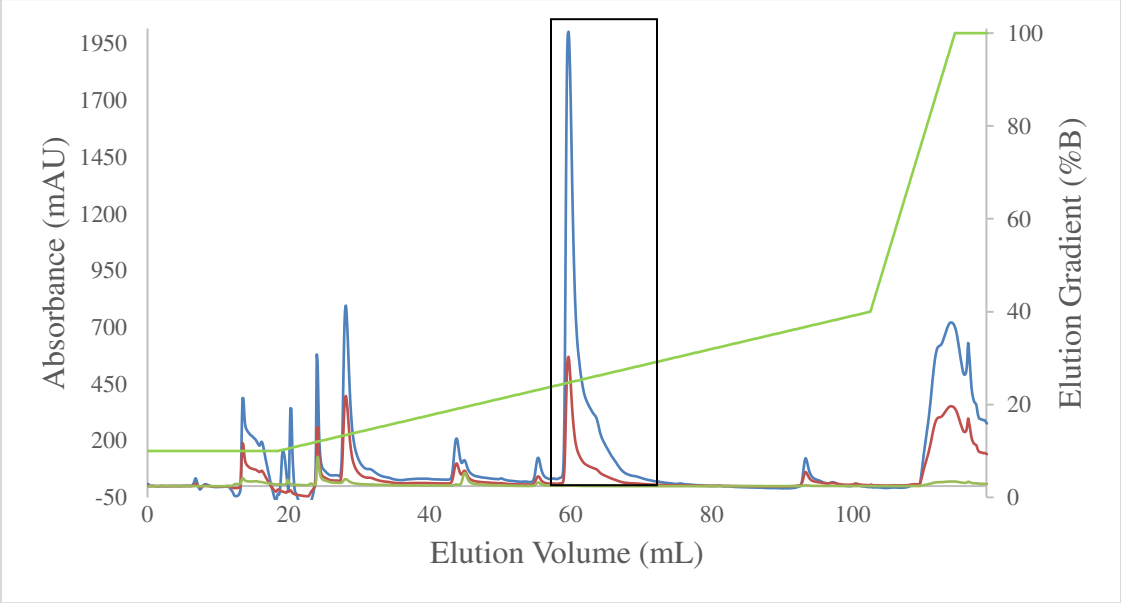


Figure 3.40: Chromatogram obtained from the reversed-phase chromatography. The peak marked with the black rectangle corresponds to the signal of MUC1-4TR. The blue line corresponds to the absorbance at 220 nm, the red line corresponds to the absorbance at 230 nm, the green line corresponds to the absorbance at 254 nm and the green line is the elution gradient (%B). Buffer B is the elution buffer containing 100% acetonitrile.

After analyzing all the peaks collected during the reversed-phase chromatography by $^{1}H,^{15}N$-HSQC and MALDI-TOF it was possible to assign the peak collected approximately at 60 mL, isolated from impurities, as MUC1-4TR (Figure 3.40).

The $^{1}H,^{15}N$-HSQC NMR spectrum of MUC1-4TR was assigned by the PhD student Helena Coelho in the group (Figure 3.41, Panel B). Figure 3.41 also shows the sequence of MUC1-4TR with the

glycosylation sites identified (Panel A), as well as, the mass spectrum that confirms the mass of MUC1-4TR (Panel C).

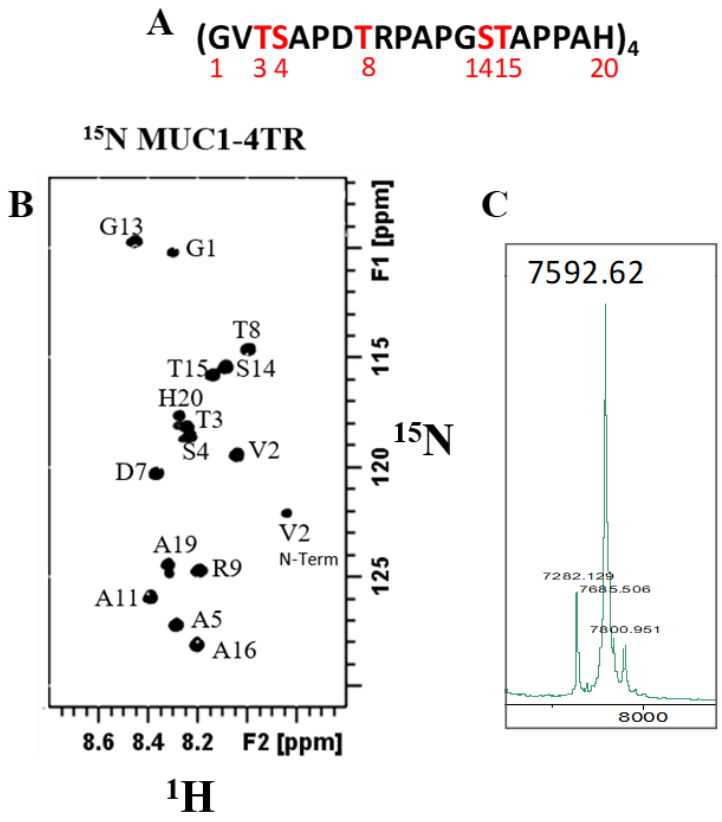


Figure 3.41: **A** MUC1-4TR sequence; **B** ^{1}H,^{15}N HSQC of MUC1-4TR with assignment of peptide sequence; **C** MALDI-TOF spectrum of ^{15}N-MUC1-4TR.

To note that the assignment is in agreement with the previously reported data in the literature for a His-tag MUC1-5TR construct [95] and confirms that ^{15}N-isotopically labelled MUC1-4TR was successfully expressed with high degree of purity.

3.3.2. Study of MUC1-4TR *O*-Glycosylation by GalNAc-T3 using NMR spectroscopy

From structure viewpoint, GalNAc-Ts hold a lectin and catalytic domain. The specificity of *O*-glycosylation sites by GalNAc-Ts depends whether these enzymes interact with naked or previous glycosylated regions of the substrate. Naked peptides are recognized by the catalytic domain while glycopeptides recognition relies on the cooperative mechanism between the catalytic and lectin domain [61, 62].

GalNAc-T3 enzyme is able to glycosylate three of the five positions of MUC1 sequence, namely T3, S14 and T15 of MUC1 sequence (Panel A, Figure 3.41) [75]. The S4 and T8 of MUC1 sequence are glycosylated by another enzyme GalNAc-T, the GalNAc-T4 [96]. Herein, the main objective was to

follow this event using NMR spectroscopy by detecting chemical shift perturbations in the $^{1}H,^{15}N$-HQSC spectrum of MUC1-4TR during the glycosylation event.

Glycosylation of the OH group of the lateral chain of Ser and Thr amino acids induce chemical shift of the amide bond of Ser and Thr directly glycosylated, as well as, in the neighboring amino acids of the glycosylation site. To assign which amino acid is directly glycosylated it is required assigning each glycosylation products.

The $^{1}H,^{15}N$-HSQC of the final product of GalNAc-T3 was previously assigned by the PhD student Helena Coelho and confirmed by mass spectrometry (Figure 3.42, Panel A and B). This tri-glycosylated product resulting from GalNAc-T3 catalysis was obtained after addition of an excess of the UDP-GalNAc donor substrate. MUC1-4TR has 4 tandems with 5 possible sites of glycosylation, thus there is a total of 20 glycosylation positions for glycosylation.

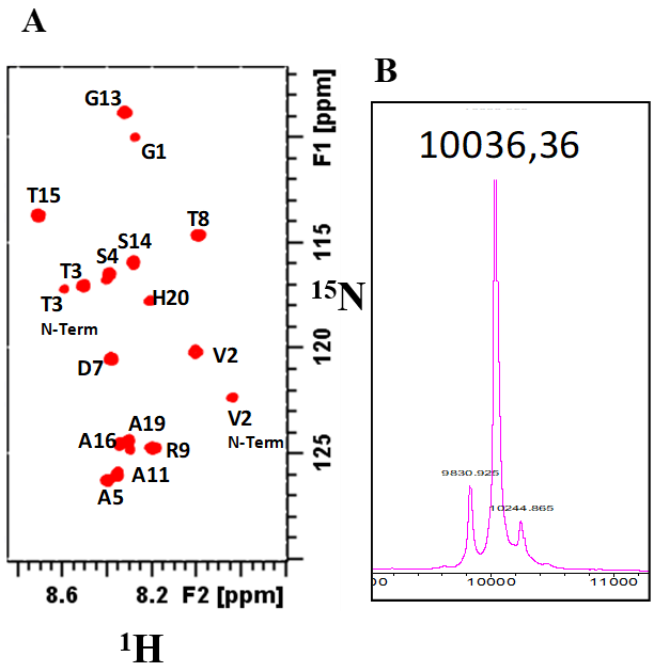


Figure 3.42: **A**.$^{1}H,^{15}N$-HSQC of the final product of GalNAc-T3 with corresponding assignment of peptide sequence. **B**. MALDI-TOF spectrum of the final product of GalNAc-T3.

Herein, we attempted to monitor step-by-step the *O*-glycosylation process of MUC1-4TR by GalNAc-T3 by adding control amounts of the donor substrate (UDP-GalNAc).

The $^{1}H,^{15}N$-HSQC of MUC1-4TR spectrum in absence and presence of GalNAc-T3 and $MnCl_2$ is identical. To this sample the amount of UDP-GalNAc necessary to glycosylate only ¼ of one position was added. A third $^{1}H,^{15}N$-HSQC was acquired with the objective to detect which amino acid was first glycosylated by GalNAc-T3. However, with this concentration of UDP-GalNAc it was not possible to observe glycosylation (data not shown, the spectrum is identical as the non-glycosylated MUC1-4TR). Additional amount of UDP-GalNAc was added to reach the amount required to glycosylate ½ of one glycosylation position. A fourth $^{1}H,^{15}N$-HSQC spectrum was measured. In this spectrum it was possible to detect new signals indicated in green in Figure 3.43. These new signals are assigned in the final tri-

glycosylated product to T3, S4 and V2 of MUC1 sequence. Besides the appearance of these new signals, the original signals of T3, S4 and V2 (unglycosylated MUC1-4TR) remain intense. Accordingly to the relative intensities of the signals T3/S4/V2 glycosylated to T3/S4/V2 non-glycosylated it is possible to conclude that only one tandem repeat of MUC1-4TR was glycosylated and this glycosylation is not indeed complete.

Accordingly to the assignment of the final glycosylation product of GalNAc-T3, S4 is not glycosylated therefore the chemical shift detected for Ser 4 arises from the glycosylation in T3 (Figure 3.43). The appearance of new signals for V2 and S4 are due to the perturbation of the chemical environment induced by the glycosylation of T3 (Figure 3.43). This result together with the previously analyzed of the final compound allows us to conclude that T3 is the first residue to be glycosylated by GalNAc-T3. Glycosylation of T3 at MUC1-4TR is exclusively dependent of the affinity of the catalytic domain towards MUC1 sequence. This result is in agreement with the literature where it is described that GalNAc-T3 shows preference to glycosylate T in -TS- instead of T in –ST– in MUC1 sequences [75].

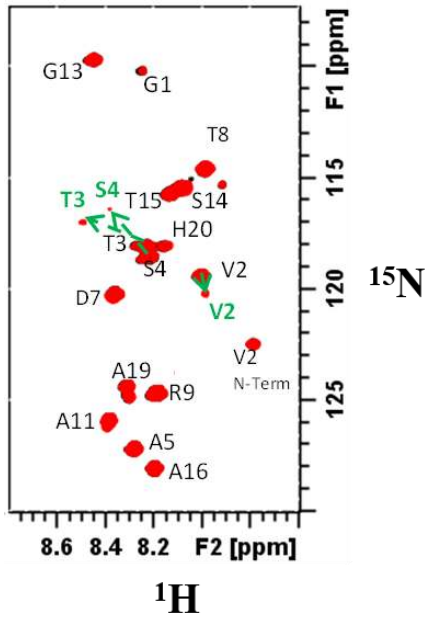


Figure 3.43: Overlap of the $^{1}H,^{15}N$-HSQC spectrum of MUC1-4TR naked in black and the $^{1}H,^{15}N$-HSQC spectrum of MUC1 with T3 glycosylated in red. Arrows in green highlight new V2/T3/S4 signals due to glycosylation.

Continuing the glycosylation process and taking into account the final assignment of the product of glycosylation by GalNAc-T3, it was also possible to conclude that the second residue to be glycosylated is the T15. Once again, the glycosylation only occurred in one tandem repeat, since the original signal remains much more intense than the new one. Besides the glycosylation of T15, it is also possible to observe a new signal for A16, due to the perturbation of the chemical environment caused by the glycosylation of T15. This is represented by blue arrows in Figure 3.44. Furthermore, the intensity of the T3/S4/V2 corresponding to a glycosylated product increases. Glycosylation of T15 starts before GalNAc-T3 finishes glycosylation in T3, indicating that GalNAc-T3 is glycosylating T15 with assistance of the lectin domain.

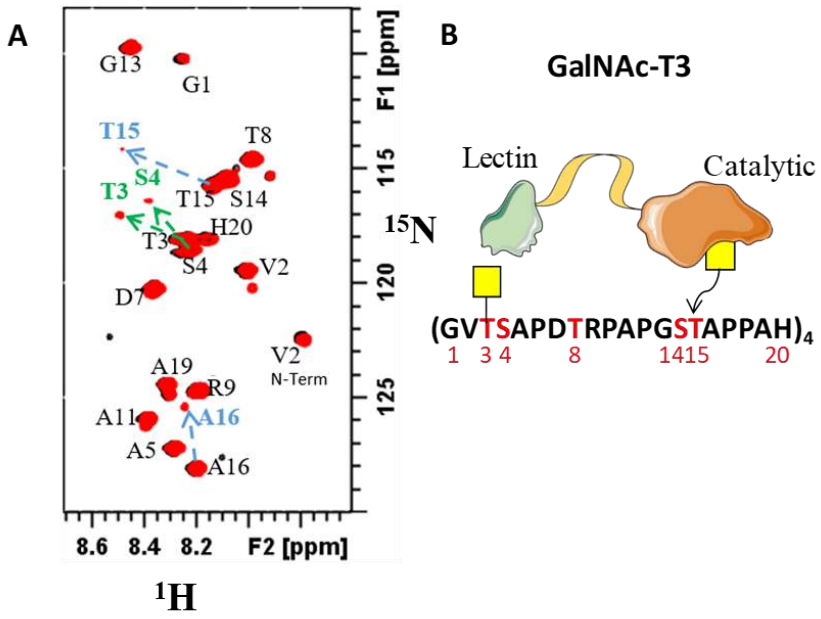


Figure 3.44: A. Overlap of the $^{1}H,^{15}N$-HSQC spectrum of MUC1-4TR naked in black with the $^{1}H,^{15}N$-HSQC spectrum of MUC1-4TR with T3 and T15 glycosylated in red. The shift of the signals is represented by a green arrow (glycosylation of T3) and blue arrow (glycosylation of T15); **B.** Scheme of GalNAc-T3 orientation upon T15 glycosylation.

An excess of UDP-GalNAc was added to yield the final product of the GalNAc-T3 glycosylation and another $^{1}H,^{15}N$-HSQC spectrum was acquired. The final product of GalNAc-T3 glycosylation has T3/S14/T15 glycosylated and can be observed in green in Figure 3.45.

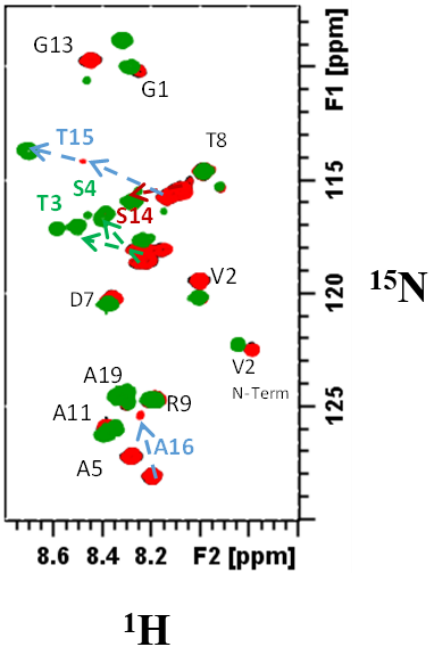


Figure 3.45: Overlap of the $^{1}H,^{15}N$-HSQC spectrum of MUC1-4TR with T3 and T15 glycosylated in red and the $^{1}H,^{15}N$-HSQC spectrum of the final product tri-glycosylation product in green. The shift of the signals is represented by green arrows (glycosylation T3), blue arrows (glycosylation of T15) and dark red arrow (glycosylation of S14).

Glycosylation of S14, represented by the dark red arrow in Figure 3.45, induces chemical shift perturbation of T15 residue. As well, it can also be observed that T3 has two signals in the final product (accordingly to the assignment of MUC1-4TR established by Helena Coelho). This happens, because the T3 of the N-terminal of MUC1-4TR has a different chemical environment than T3 in the other 3 tandem repeats (Figure 3.45).

3.3.3. Interaction studies of glycopeptide T3-Tn by GalNAc-T3

In order to better understand the mechanism of glycosylation, STD-NMR experiments were also acquired using a glycopeptide. The glycopeptide has the peptide sequence GVTSAPDTRPAPGSTAPPAHGVT and it is glycosylated at the T3 residue (T3-Tn).

The objective was to study the molecular recognition event of the second step of glycosylation by GalNAc-T3. In this second glycosylation event, the lectin domain of GalNAc-T3 is expected to bind to GalNAc moiety at –GVTS- in the glycopeptide structure to direct the catalytic domain towards the peptide region –GSTA–. The STD spectrum and STD-derived epitope mapping is shown in Figure 3.46.

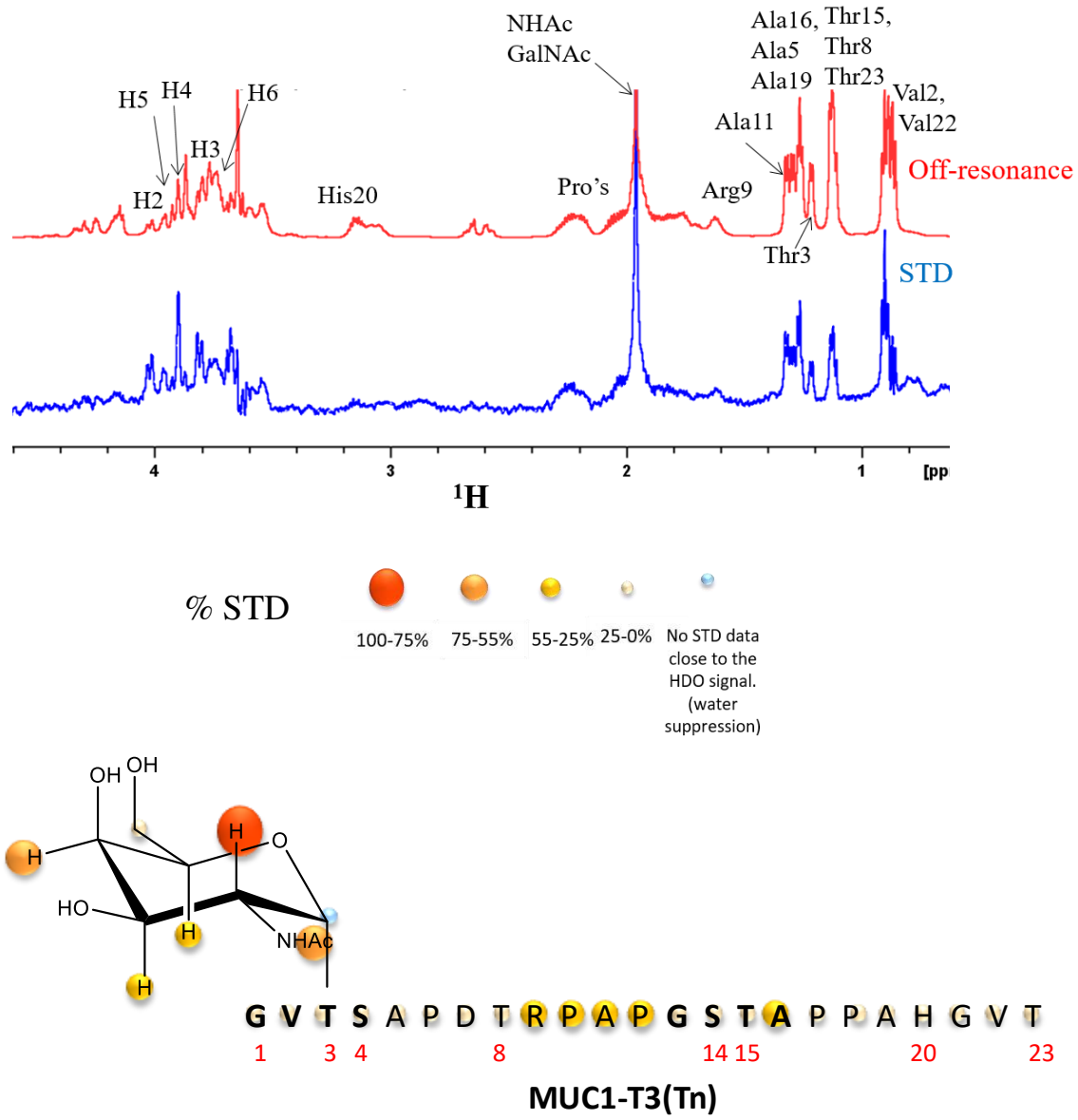


Figure 3.46: Overlap of the STD spectrum (blue) and Off-resonance (red) of the spectrum T3-Tn with GalNAc-T3. The epitope obtained and the STD percentage scale used.

In fact, according to the STD-derived epitope mapping GalNAc-T3 recognizes the sugar protons of GalNAc moiety together with amino acids of PAP region. This region is located near to the glycosylation site T15 (second position that is glycosylated by GalNAc-T3). To note that the PxP region in peptides/proteins is a common feature previously identified to be associated to potential glycosylation sites for GalNAc-Ts [97].

3.3.4. Principal conclusions and perspectives

NMR spectroscopy using ^{1}H,^{15}N-HSQC chemical shift analysis shows to be suitable to follow the glycosylation process of MUC1-4TR construct by GalNAc-T3.

Accordingly to $^{1}H,^{15}N$-HSQC data, GalNAc-T3 enzyme has preference to glycosylate the Thr at –GVTSA– sequence (T3) in agreement with the previously reported data [75]. This result shows that the catalytic domain of GalNAc-T3 has more affinity to the –GVTSA- region of MUC1. The second event of glycosylation by GalNAc-T3 seems to be a cooperative mechanism between the lectin and catalytic domain. Thus, the lectin domain of GalNAc-T3 binds GalNAc at T3 of the MUC1 sequence to direct the catalytic domain to the next residue to be glycosylated, the Thr residue at –GSTA– sequence (T15).

STD-derived epitope map of a glycopeptide of 23 amino acids, containing GalNAc moiety at T3 shows clearly STD signals for GalNAc residue. STD response of GalNAc moiety in the glycopeptide can be assigned to the interaction with the lectin domain. In addition, amino acids at PAP region of the peptide sequence also receive STD response. Here STD signals can be explained by the transient interaction with the catalytic domain.

At the end, our study provides new structural insights into the mechanism of glycosylation of MUC1-4TR by GalNAc-T3 that will be essential to the rational design of selective inhibitors towards this specific enzyme. However, the study of the process of glycosylation of GalNAc-T3 is far from being totally uncovered. Further experiments, in particular, $^{1}H,^{15}N$-HSQC of MUC1-4TR, will be recorded to unveil if glycosylation at S14 occurs before GalNAc-T3 finishes to glycosylate T3 and T15. STD-NMR binding experiments will be also carried out using a di-glycopeptide with GalNAc moiety at T3 and T15 positions. Using relative intensities of signals in the $^{1}H,^{15}N$-HSQC we will try to estimate the kinetic information (km) of the glycosylation process.

4. Conclusions

Carbohydrate-protein interactions are incredibly important in molecular recognition processes. These interactions are key factors in many biological functions, such as cellular transport, signaling and cell-cell communication. Furthermore, they play a key role in several diseases, from infection, autoimmune and neurological diseases to cancer. Therefore, understanding at an atomic level the interactions that govern carbohydrate recognition is of upmost importance.

In this context, in this thesis distinct biological systems that involve carbohydrate-protein recognition processes were investigated by employing NMR techniques.

In the first case, the interactions between a Tn-glycopeptide mimetic APD(Hnv)*RP (* means the site of Tn-glycosylation), where the Thr was replaced by a non-natural isomer of valine, the hydroxy-norvaline (Hnv) and diverse antibodies, two of the anti-MUC1 family (VU-3C6 and SM3) and one of the anti-Tn family (14D6), were scrutinized. STD-NMR binding technique was used to determine the binding epitope of the Tn-mimetic and STD results were further complemented by other biochemical techniques. At the end, it was shown that the anti-MUC1 antibodies recognize the Tn-glycopeptide in a sequence-dependent manner, however with different binding preferences towards the same peptide sequence. In contrast, the anti-Tn antibody 14D6 recognizes the Tn-glycopeptide mimetic APD(Hnv)*RP mainly through the GalNAc and the methyl group of the Hnv amino acid residue.

These studies and results are useful and relevant in the rational design of more resistant and stable glycan-based anti-cancer vaccines.

In the second case, the recognition process of galectin-3 (Gal-3) and the tumor-associated ligands TF-antigen and a TF-peptide PDT*RP (* means the site of TF-glycosylation) was studied. This system was studied from the receptor and ligand viewpoint. $^{1}H,^{15}N$-HSQC titrations of the isotopic labelled Gal-3 in presence of both ligands were performed. The results from $^{1}H,^{15}N$-HSQC titrations of Gal-3 carbohydrate recognition domain does not show significant differences in the recognition mode of TF-antigen linked to Thr or the TF-antigen linked to the peptide. STD-NMR and line broadening analysis determined that the Gal residue is the main key unit that establishes interactions with the protein and that the peptide PDTRP should be solvent exposed and therefore not strongly involved in the binding. These ligand-based binding experiments were also supported by the similar values of K_D obtained from the titrations.

This study allows us to establish a NMR binding protocol, which can now be used to study Gal-3 interactions with TF-glycomimetics helping to perform a rational structure based design of potential glycomimetics inhibitors of Gal-3.

Understanding the molecular features that dictate how the carbohydrates are introduced to proteins by glycosylation is also crucial. This is especially important if we consider that alteration of this process is a hallmark of cancer. To study the glycosylation process of MUC1 glycoprotein, a GalNAc-Transferase was used, the GalNAc-T3. The glycosylation mechanism of GalNAc-T3 was followed by $^{1}H,^{15}N$-HSQC and STD-NMR. By $^{1}H,^{15}N$-HSQC, it was revealed that GalNAc-T3 has preference to glycosylate the Thr residue present in -GVTSA- (T3), followed by the Thr -GSTA- and then Ser at –GSTA-. It was also

possible to prove that the second step of glycosylation, the glycosylation of T15 is a cooperative mechanism between the lectin and catalytic domain of GalNAc-T3, by STD-NMR. The STD showed that the lectin domain binds to the GalNAc residue present in -GVTSA-, which directs the catalytic domain to the next residue to be glycosylated, the Thr residue at –GSTA– sequence (T15).

The results obtained herein allow us to uncover important structural insights into the mechanism of glycosylation by GalNAc-T3 essential to the rational design of selective inhibitors for this enzyme.

Overall, this thesis shows the importance of NMR spectroscopy in molecular recognition studies and the importance of an interdisciplinary approach to investigate carbohydrate-protein recognition processes. By NMR spectroscopy techniques, it was possible to determine the main points for the recognition process of a Tn-glycopeptide mimetic with a non-natural amino acid by three different antibodies, as well as, of TF-antigen and the TF-peptide PDTRP by Gal-3 lectin. Furthermore, NMR also uncovered valuable aspects of the glycosylation process by GalNAc-T3.

5. References

[1] Varki, A.; Cummings, R.D.; Esko, J.D. *Essentials of Glycobiology*. 3rd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015.

[2] Chávez, M.; Andreu, C.; Vidal, P.; Aboitiz, N.; Freire, F.; Groves, P.; Asensio, J.L.; Asensio, G.; Muraki, M.; Cañada, F. J.; Jiménez-Barbero, J. *Chem. Eur. J.* **2005**, 11, 7060-7074.

[3] Fernández-Alonso, M. C.; Cañada, F. J.; Jiménez-Barbero, J.; Cuevas, G. *J. Am. Chem. Soc.* **2005**, 127, 7379-7386.

[4] Marcelo, F.; Cañada. F. J.; Jiménez-Barbero, J. In *Anticarbohydrate Antibodies: from molecular basis to clinical application*; Kosma, P., Müller-Leonnies, S., Eds.; Springer-Verlag: Vienna, 2012; pp 385-402

[5] Johnson, M. A.; Pinto, B. M. *Carbohydrate Research* **2004**, 339, 907-928.

[6] Kirschner, K. N.; Woods, R. J. *PNAS* **2001**, 98, 10541-10545.

[7] Lütteke, T. *Acta Cryst.* **2009**, 65, 156-168.

[8] Fadda, E.; Woods, R. J. *Drug Discovery Today* **2010**, 15, 596-609.

[9] Roldós, V.; Cañada, F. J.; Jiménez-Barbero, J. *ChemBioChem* **2011**, 12, 990-1005.

[10] Quiocho, F. A. *Pure & Appl. Chem.* **1989**, 7, 1293-1306.

[11] Fernández-Alonso, M. C.; Díaz, D.; Berbis, M. Á.; Marcelo, F.; Cañada, J.; Jiménez-Barbero, J. *Current Proteinand Peptide Science* **2012**, 13, 816-830.

[12] Ramírez-Gualito, K.; Alonso-Ríos, R.; Quiroz-García, B.; Rojas-Aguilar, A; Díaz, D.; Jiménez-Barbero, J.; Cuevas, G. *J. Am. Chem. Soc.* **2009**, 131, 18129-18138.

[13] Lim, Y.-b.; Lee, M. *Org. Biomol. Chem.* **2007**, 5, 401-405.

[14] Baldini, L.; Casnati, A.; Sansone, F.; Ungaro, R. *Chem. Soc. Rev.* **2007**, 36, 254-266.

[15] Kiessling, L. L.; Gestwicki, J. E.; Strong, L. E. *Angew. Chem. Int. Ed. Engl.* **2006**, 45, 2348-2368.

[16] Deniaud, D.; Julienne, K.; Gouin, S. G. *Org. Biomol. Chem.* **2011**, 9, 966-979.

[17] ten Cate, M. G. J.; Reinhoudt, D. N.; Crego-Calama, M. *J. Org. Chem.* **2005**, 70, 8443-8453.

[18] Imberty, A.; Pérez, S. *Chem. Rev.* **2000**, 100, 4567-4588.

[19] Bush, C. A.; Martin-Pastor, M. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, 28, 269-293.

[20] Stowell, S. R.; Ju, T.; Cummings, R. D. *Annu. Rev. Pathol. Mech. Dis.* **2015**, 10, 473-510.

[21] Christiansen, M. N.; Chik, J.; Lee, L.; Anugraham, M.; Abrahams, J. L.; Packer, N. H. *Proteomics* **2014**, 14, 525-546.

[22] Brooks, S. A.; Carter, T. M.; Royle, L.; Harvey, D. J.; Fry, S. A.; Kinch, C.; Dwek, R. A.; Rudd, P. M. *Anti-Cancer Agents in Medicinal Chemistry* **2008**, 8, 2-21.

[23] Munkley, J.; Elliott, D. J. *Oncotarget* **2016**, 7, 35478-35489.

[24] Schjoldager, K. T.-B. G.; Clausen, H. *Biochimica et Biophysica Acta* **2012**, 1820, 2079-2094.

[25] Reis, C. A.; Osorio, H.; Silva, L.; Gomes, C.; David, L. *J. Clin. Pathol.* **2010**, 63, 322-329.

[26] Burchell, J. M.; Mungul, A.; Taylor-Papadimitriou, J. *Journal of Mammary Gland Biology and Neoplasia* **2001**, 6, 355-364.

[27] Kufe, D. W. *Nature Reviews Cancer* **2009**, 9, 874-885.

[28] Bafna, S.; Kaur, S.; Batra, S. K. *Oncogene* **2010**, 29, 2893-2904.

[29] Tarp, M. A.; Clausen, H. *Biochimica et Biophysica Acta* **2008**, 1780, 546-563.

[30] Song, W.; Delyria, E. S.; Chen, J.; Huang, W.; Lee, J. S.; Mittendorf, E. A.; Ibrahim, N.; Radvanyi, L. G.; Li, Y.; Lu, H.; Xu, H.; Shi, Y.; Wang, L.-X.; Ross, J. A.; Rodrigues, S. P.; Almeida, I. C.; Yang, X.; Qu, J.; Schocker, N. S.; Michael, K.; Zhou, D. *INTERNATIONAL JOURNAL OF ONCOLOGY* **2012**, 41, 1977-1984.

[31] Nath, S.; Mukherjee, P. *Trends in Molecular Medicine* **2014**, 24, 332-342.

[32] Gendler, S. J. *Journal of Mammary Gland Biology and Neoplasia* **2001**, 6, 339-353.

[33] Lakshmanan, I.; Ponnusamy, M. P.; Macha, M. A.; Haridas, D.; Majhi, P. D.; Kaur, S.; Jain, M.; Batra, S. K.; Ganti, A. K. *Journal of Thoracic Oncology* **2015**, 10, 19-27.

[34] Madariaga, D.; Martínez-Sáez, N.; Somovilla, V. J.; Coelho, H.; Valero-González; Castro-López, J.; Asensio, J. L.; Jiménez-Barbero, J.; Busto, J. H.; Avenoza, A.; Marcelo, F.; Hurtado-Guerrero, R.; Corzana, F.; Peregrina, J. M. *ASC Chem. Biol.* **2015**, 10, 747-756.

[35] Kinlough, C. L.; McMahan, R. J.; Poland, P. A.; Bruns, J. B.; Harkleroad, K. L.; Stremple, R. J.; Kashian, O. B.; Weixel, K. M.; Weisz, O. A.; Hughey, R. P. *THE JOURNAL OF BIOLOGICAL CHEMISTRY* **2006**, 281, 12112-12122.

[36] Möller, H.; Serttas, N.; Paulsen, H.; Burchell, J. M.; Taylor-Papadimitriou, J.; Meyer, B. *Eur. J. Biochem.* **2002**, 269, 1444-1455.

[37] Pinho, S. S.; Reis, C. A. *Nat. Rev. Cancer* **2015**, 15, 540-555.

[38] Ju, T.; Lanneau, G. S.; Gautam, T.; Wang, Y.; Xia, B.; Stowell, S. R.; Willard, M. T.; Wang, W.; Xia, J. Y.; Zuna, R. E.; Laszik, Z.; Benbrook, D. M.; Hanigan, M. H.; Cummings R. D. *Cancer Res* **2008**, 68, 1636-1646.

[39] Slovin, S. F.; Keding, S. J.; Ragupathi, G. *Immunology and Cell Biology* **2005**, 83, 418-428.

[40] 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.; Corzana, F. *Argew. Chem. Int. Ed.* **2015**, 54, 9830-9834.

[41] Coelho, H.; Matsushita, T.; Artigas, G.; Hinou, H.; Cañda, F. J.; Lo-Man, R.; Leclerc, C.; Cabrita, E. J.; Jiménez-Barbero, J.; Nishimura, S.-I.; Garcia-Martín, F.; Marcelo, F. *J. Am. Chem. Soc.* **2015**, 137, 12438-12441.

[42] Murphy, P. V.; André, S.; Gabius, H.-J. *Molecules* **2013**, 18, 4026-4053.

[43] Nangia-Makker, P.; Conklin, J.; Hogan, V.; Raz, A. *TRENDS In Molecular Medicine* **2002**, 8, 187-192.

[44] 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.; Cañada, F. J. *Chem. Eur. J.* **2014**, 20, 16147-16155.

[45] Saeland, E.; van Vliet, S. J.; Bäckström, M.; Broks-van den Berg, V. C. M.; Geijtenbeek, T. B. H.; Meijer, G. A.; van Kooyk, Y. *Cancer Immunol Immunother* **2007**, 56, 1225-1236.

[46] Cazet, A.; Julien, S.; Bobowski, M.; Burchell, J.; Delannoy, P. *Breast Cancer Research* **2010**, 12, 204-217.

[47] Mortezai, N.; Behnken, H. N.; Kurze, A.-K.; Ludewig, P.; Buck, F.; Meyer, B.; Wagener, C. *Glycobiology* **2013**, 23, 844-852.

[48] Beatson, R.; Maurstad, G.; Picco, G.; Arulappu, A.; Coleman, J.; Wandell, H. H.; Clausen, H.; Mandel, U.; Taylor-Papadimitriou, J.; Sletmoen, M.; Burchell, J. M. *PLoS ONE* **2015**, 10, e0125994.

[49] Fukumori, T.; Kanayama, H.-i.; Raz, A. *Drug Resistance Updates* **2007**, 10, 101-108.

[50] Larsen, L.; Chen, H.-Y.; Saegusa, J.; Liu, F.-T. *Journal of Dermatological Science* **2011**, 64, 85-91.

[51] Argüeso, P.; Panjwani, N. *Exp. Eye Res.* **2011**, 92, 2-3.

[52] Dumic, J.; Dabelic, S.; Flögel, M. *Biochimica et Biophysica Acta* **2006**, 1760, 616-635.

[53] Hughes, R. C. *Biochimica et Biophysica Acta* **1999**, 1473, 172-185.

[54] Newlaczyl, A. U.; Yu, L.-G. *Cancer Letters* **2011**, 313, 123-128.

[55] Rabinovich, G. A.; Baum, L. G.; Tinari, N.; Paganelli, R.; Natoli, C.; Liu, F.-T.; Iacobelli, S. *TRENDS In Immunology* **2002**, 23, 313-320.

[56] Bian, C.-F.; Zhang, Y.; Sun, H.; Li, D.-F.; Wang, D.-C. *PLoS ONE* **2011**, 6, e25007.

[57] Rodriguez, M. C.; Yegorova, S.; Pitteloud, J.-P.; Chavaroche, A. E.; André, S.; Ardá, A.; Minond, D.; Jiménez-Barbero, J.; Gabius, H.-J.; Cudic, M. *Biochemistry* **2015**, 54, 4462-4474.

[58] Taniuchi, K.; Cerny, R. L.; Tanouchi, A.; Kohno, K.; Kotani, N.; Honke, K.; Saibara, T; Hollingsworth, M. A. *Oncogene* **2011**, 30, 4843-4854.

[59] Hurtado-Guerrero, R. *Biochem. Soc. Trans.* **2016**, 44, 61-67, and references cited herein.

[60] Gerken, T. A.; Revoredo, L.; Thome, J. J. C.; Tabak, L. A.; Vester-Christensen, M. B.; Clausen; H.; Gahlay, G. K.; Jarvis, D. L.; Johnson, R. W.; Moniz, H. A.; Moremen, K. *J. Biol. Chem.* **2013**, 288, 19900-19914.

[61] Wandall, H. H.; Hassan, H.; Mirgorodskaya, E.; Kristensen, A. K.; Roepstorff, P.; Bennett, E. P.; Nielsen, P. A.; Hollingsworth, M. A.; Burchell, J.; Taylor-Papadimitriou, J.; Clausen, H. *J. Biol. Chem.* **1997**, 272, 23503-23514.

[62] Wandall, H. H.; Irazoqui, F.; Tarp, M. A.; Bennett, E. P.; Mandel, U.; Takeuchi, H.; Kato, K.; Irimura, T.; Suryanarayanan, G.; Hollingsworth, M. A.; Clausen, H. *Glycobiology* **2007**, 17, 374-387.

[63] Harada, Y.; Izumi, H.; Noguchi, H.; Kuma, A.; Kawatsu, Y.; Kimura, T.; Kitada, S.; Uramoto, H.; Wang, K.-Y.; Sasaguri, Y.; Hijioka, H.; Miyawaki, A.; Oya, R.; Nakayama, T.; Kohno, K; Yamada, S. *Tumor Biol.* **2016**, 37, 1357-1368.

[64] Kitada, S.; Yamada, S.; Kuma, A.; Ouchi, S.; Tasaki, T.; Nabeshima, A.; Noguchi, H.; Wang, K.-Y.; Shimajiri, S.; Nakano, R.; Izumi, H.; Kohno, K.; Matsumoto, T.; Sasaguri, Y. *British Journal of Cancer* **2013**, 109, 472-481.

[65] Chugh, S.; Meza, J.; Sheinin, Y. M.; Ponnusamy, P.; Batra, S. K. *British Journal of Cancer* **2016**, 114, 1376-1386.

[66] Song, L.; Linstedt, A. D. *eLife* **2017**, 6, e24051.

[67] Wang, Z.-Q.; Bachvarova, M.; Morin, C.; Plante, M.; Gregoire, J.; Renaud, M.-C.; Sebastianelli, A.; Bachvarov, D. *Oncotarget* **2014**, 5, 544-560.

[68] Ishikawa, M.; Kitayama, J.; Nariko, H.; Kohno, K.; Nagawa, H. *Journal of Surgical Oncology* **2004**, 86, 28-33.

[69] Inoue, T.; Eguchi, T.; Oda, Y.; Nishiyama, K.; Fujii, K.; Izumi, H.; Kohno, K.; Yamaguchi, K.; Tanaka, M.; Tsuneyoshi, M. *Modern Pathology* **2007**, 20, 267-276.

[70] Shibao, K.; Izumi, H.; Nakayama, Y.; Ohta, R.; Nagata, N.; Nomoto, M.; Matsuo, K.-I.; Yamada, Y.; Kitazato, K.; Itoh, H.; Kohno, K. *Cancer* **2002**, 94, 1939-1946.

[71] Gu, C.; Oyama, T.; Osaki, T.; Li, J.; Takenoyama, M.; Izumi, H.M Sugio, K.; Kohno, K.; Yasumoto, K. *British Journal of Cancer* **2004**, 90, 436-442.

[72] Dosaka-Akita, H.; Kinoshita, I.; Yamazaki, K.; Izumi, H.; Itoh, T.; Katoh, H.; Nishimura, M.; Matsuo, K.; Yamada, Y.; Kohno, K. *British Journal of Cancer* **2002**, 87, 751-755.

[73] Onitsuka, K.; Shibao, K.; Nakayama, Y.; Minagawa, N.; Hirata, K.; Izumi, H.; Matsuo K.-I.; Nagata, N.; Kitazato, K.; Kohno, K.; Itoh, H. *Cancer Sci* **2003**, 94, 32-36.

[74] Kato, K.; Takeuchi, H.; Kanoh, A.; Miyahara, N.; Nemoto-Sasaki, Y.; Morimoto.Tomita, M.; Matsubara, A.; Ohashi, Y.; Waki, M.; Usami, K.; Mandel, U.; Clausen, H.; Higashi, N.; Irimura, T. *Glycoconj. J.* **2010**, 27, 267-276.

[75] Clausen, H.; Bennett, E. P. *Glycobiology* **1996**, 6, 635-646.

[76] Yi, B.; Zhang, Z.; Zhang, M.; Schwartz-Albiez, R.; Cao, Y. *Oncology Reports* **2013**, 30, 1841-1847.

[77] Liu, J.; Yi, B.; Zhang, Z.; Cao, Y. *Front. Med.* **2016**, 10, 204-211.

[78] Fiedler, W.; DeDosso, S.; Cresta, S.; Weidmann, J.; Tessari, A.; Salzberg, M.; Dietrich, B.; Baumeister, H.; Goletz, S.; Gianni, L.; Sessa, C. *European Journal of Cancer* **2016**, 63, 55-63.

[79] Gulley, J. L.; Madan, R. A.; Tsang, K. Y.; Jochems, C.; Marté, J. L.; Farsaci, B.; Tucker, J. A.; Hodge, J. W.; Liewehr, D. J.; Steinberg, S. M.; Heery, C. R.; Schlom, J. *Cancer Immunol. Res.* **2014**, 2, 133-141.

[80] Topp, M. S.; Gökbuget, N.; Zugmaier, G.; Klappers, P.; Stelljes, M.; Neumann, S.; Viardot, A.; Marks, R.; Diedrich, H.; Faul, C.; Reichle, A.; Horst, H.-A.; Brüggemann, M.; Wessiepe, D.; Holland, C.; Alekar, S.; Mergen, N.; Einsele, H.; Hoelzer, D.; Bargou, R. C. *J. Clin. Oncol.* **2014**, 32, 4134-4140.

[81] Unione, L.; Galante, S.; Díaz, D.; Cañada, F. J.; Jiménez-Barbero, J. *Med. Chem. Commun.* **2014**, 5, 1280-1289.

[82] Carvalho, A. L.; Silva, T. S.; Romão, M. J.; Cabrita, E. J.; Marcelo, F. *Structural elucidation of macromolecules;* (In) Sahar Iftakhar (ed.): Essential Techniques for Medical and Life Scientists: A Guide to Comtemporary Methods and Current Applications with the Protocols, Bentham Science Publishers. *In Press*.

[83] Koharudin, L. M. I.; Gronenborn, A. M. In *Structural Glycobiology*; Yuriev, E.; Ramsland, P. A., Eds.; CRC Press, 2012; pp 29-47.

[84] Mayer, M.; Meyer, B. *Angew. Chem. Int. Ed.* **1999**, 38, 1784-1788.

[85] Meyer, B.; Peters, T. *Angew. Chem. Int. Ed.* **2003**, 42, 864-890.

[86] Ju, T.; Aryal, R. P.; Kudelka, M. R.; Wang, Y.; Cummings, R. D. *Cancer Biomarkers* **2014**, 14, 63-81.

[87] Diniz, A.; Dias, J. S.; Jiménez-Barbero, J.; Marcelo, F.; Cabrita, E. J. *Chem. Eur. J.* **2017**, 23, 1-9.

[88] Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. *Protein Identification and Analysis Tools on the ExPASy Server;* (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005). pp. 571-607.

[89] Keller, R. *The computer aided resonance assignment tutorial CARA*; Cantina Verlag, Goldau, Switz, 2004.

[90] Vraken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins* 2005, 59, 687-696.

[91] Umemoto, K.; Leffler, H. *J. Biomol.* **2001**, 20, 91-92.

[92] Karsten, U.; Serttas, N.; Paulsen, H.; Danielczyk, A.; Goletz, S. *Glycosylation* **2004**, 14, 681-692.

[93] Bailey-Kellogg, C.; Widge, A.; Kelley III, J. J.; Berardi, M. J.; Bushweller, J. H.; Donald, B. R. *Journal of Computational Biology* **2000**, 7, 537-558.

[94] The PyMOL Molecular Graphics System Version 1.1, Schrödinger, LLC.

[95] Brokx, R. D.; Revers, L.; Zhang, Q.; Yang, S.; Mal, T. K.; Ikura, M.; Gariépy, J. *Biochemistry* **2003**, 42, 13817-12825.

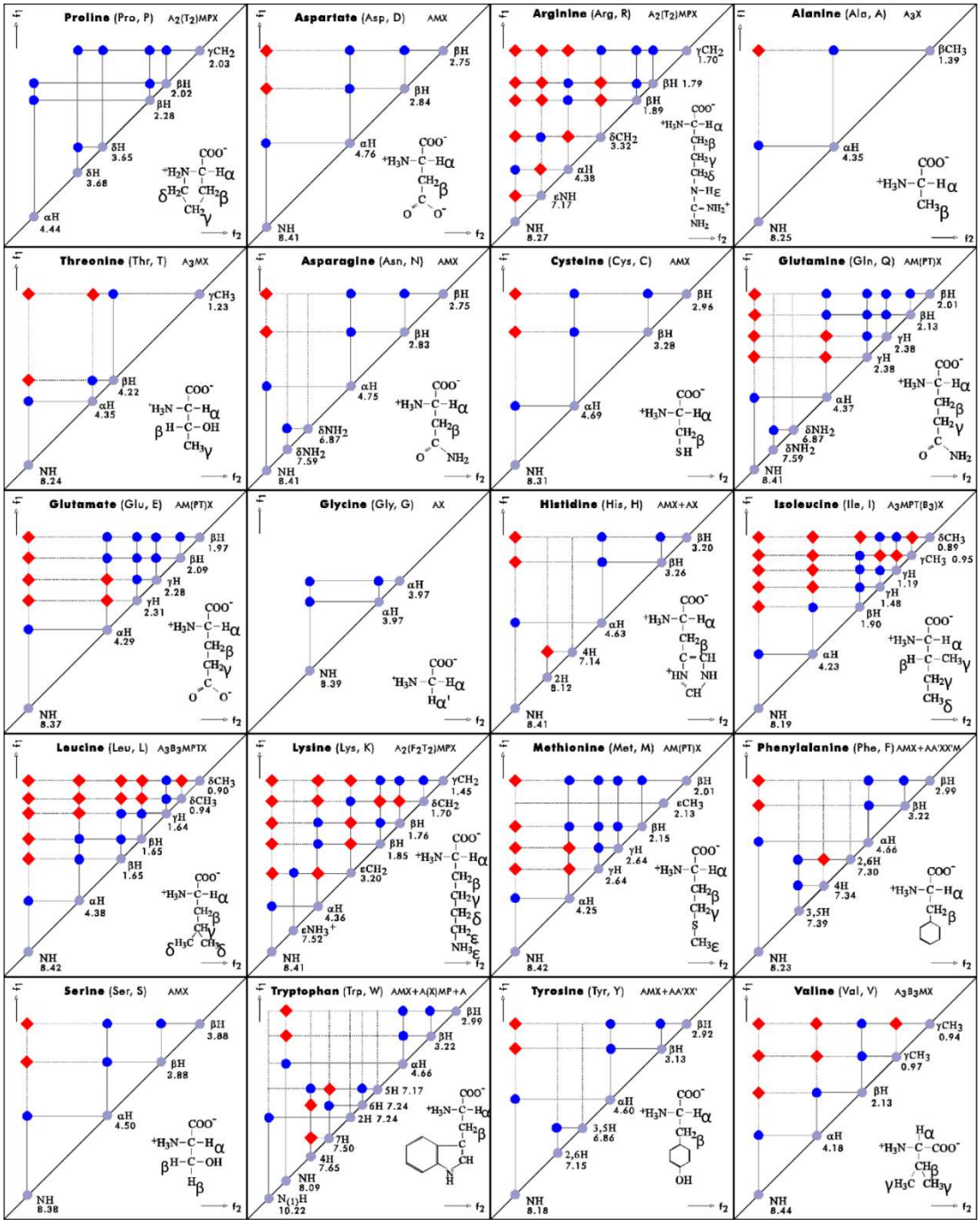
[96] Bennett, E. P.; Hassan, H.; Mandel, U.; Mirgorodskaya, E.; Roepstorff, P.; Burchell, Taylor-Papadimitriou, J.; Hollingsworth, M. A.; Merkx, G.; van Kessel, A. G.; Eiberg, H.; Steffensen, R.; Clausen, H. *J. Biol. Chem.* **1998**, 278, 30472-30481.

[97] Revoredo, L.; Wang, S.; Bennett, E. P.; Clausen, H.; Moremen, K. W.; Jarvis, D. L.; ten Hagen, K. G.; Tabak, L. A.; Gerken, T. A. *Glycobiology* **2016**, 26, 360-376.

6. Appendix

6.1. **Appendix 6.1-** Table with the characteristic chemical shift pattern for each amino acid.

Table 6.1: Characteristic chemical shift pattern for each amino acid in $^{1}H,^{1}H$-TOCSY.



6.2. **Appendix 6.2-** Assignment of the ^{1}H-NMR spectrum for the glycopeptide APD(Hnv)*RP (* corresponds to the site of Tn-glycosylation)

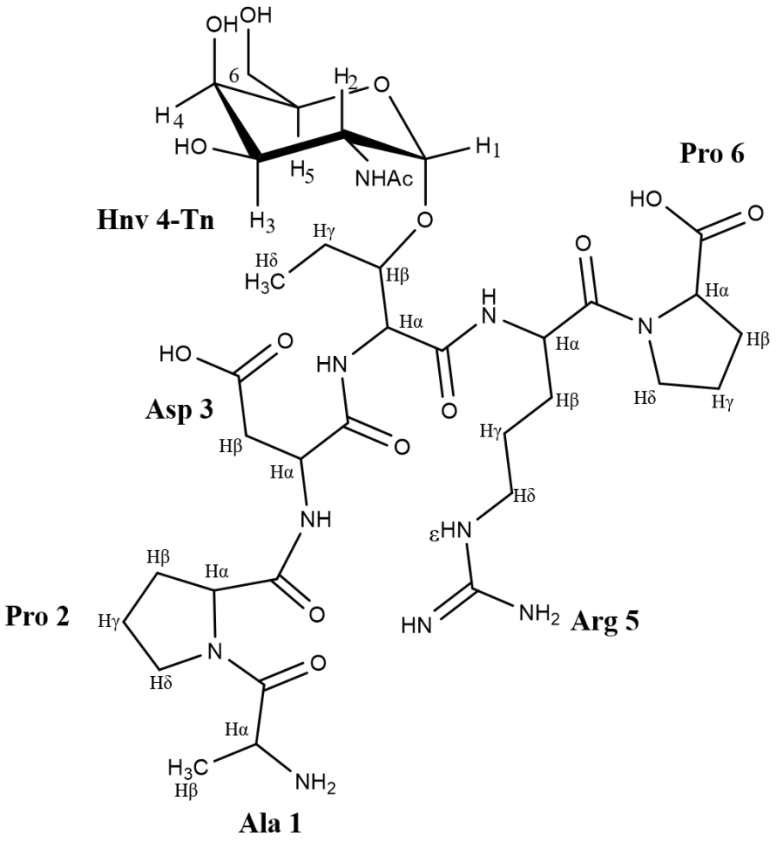


Figure 6.1: Structure of the Tn-glycopeptide mimetic with the sequence APD(Hnv)*RP (* indicates the site of Tn-glycosylation).

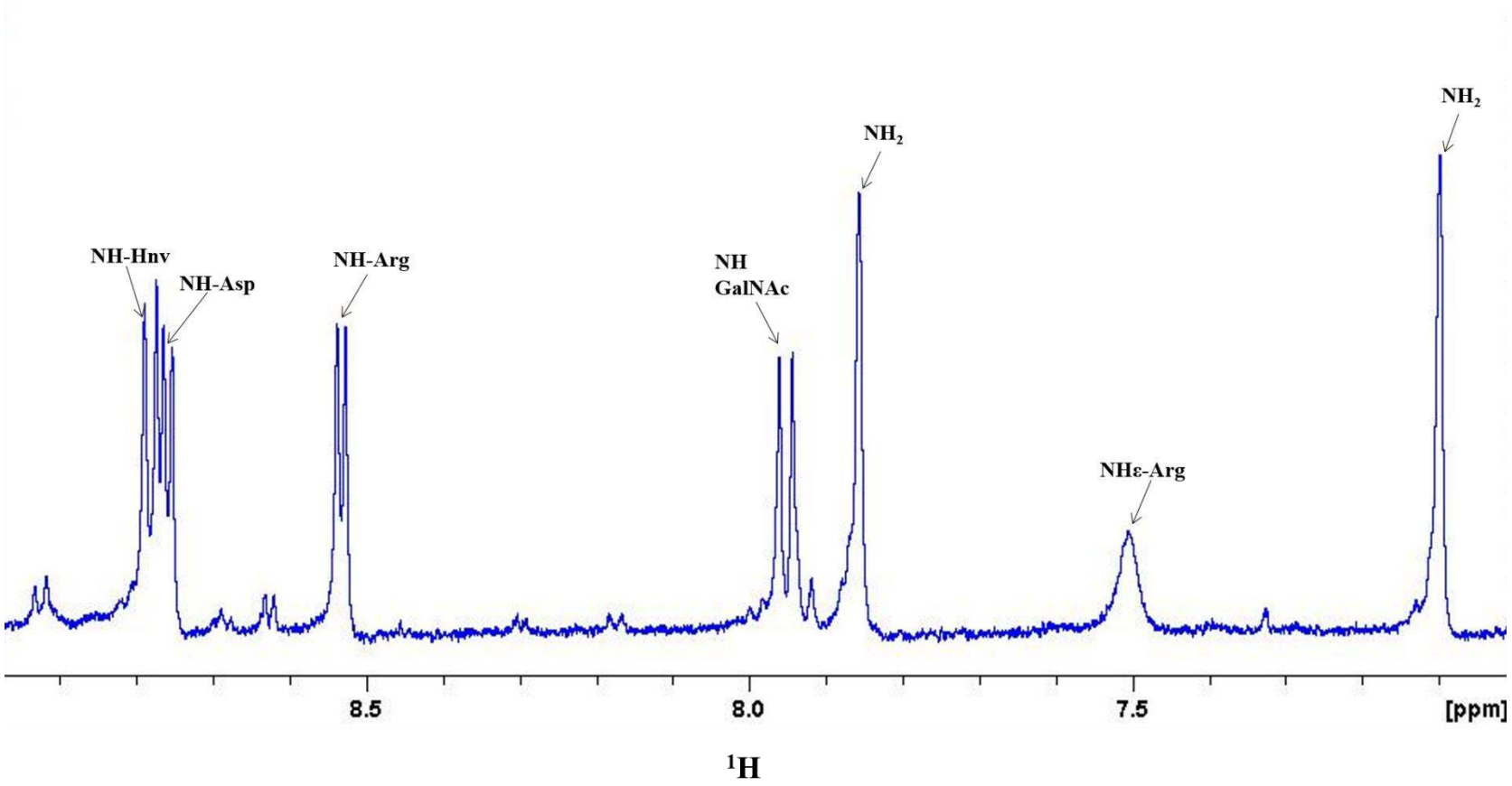


Figure 6.2: ^{1}H-NMR spectrum assignment of the region NH (9 ppm to 7 ppm) for the glycopeptide APD(Hnv)*RP (* indicates the site of Tn-glycosylation).

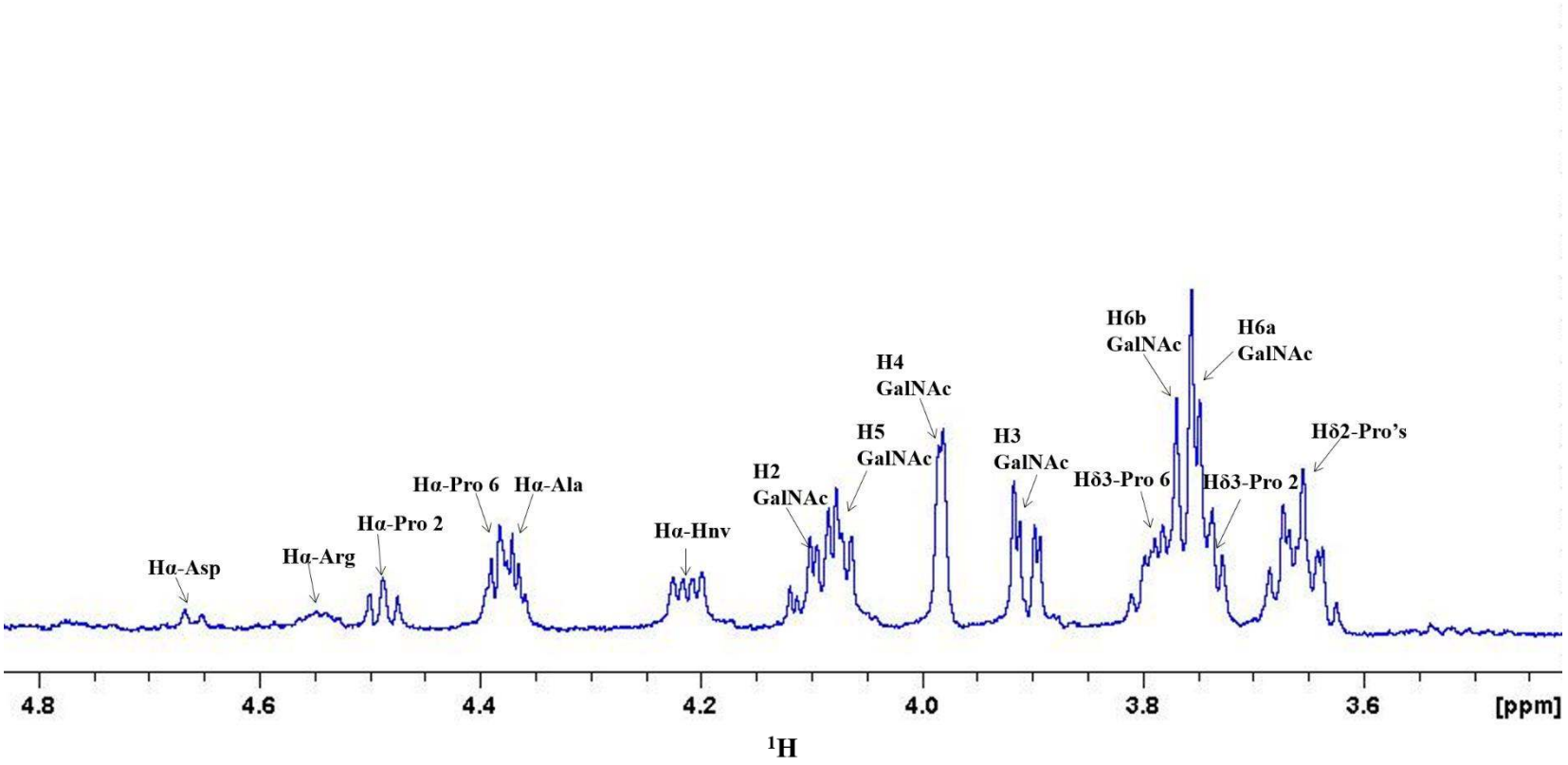


Figure 6.3: [1]H-NMR spectrum assignment for the region 4.8 to 3.4 ppm for the glycopeptide APD(Hnv)*RP (* indicates the site of Tn-gçycosylation).

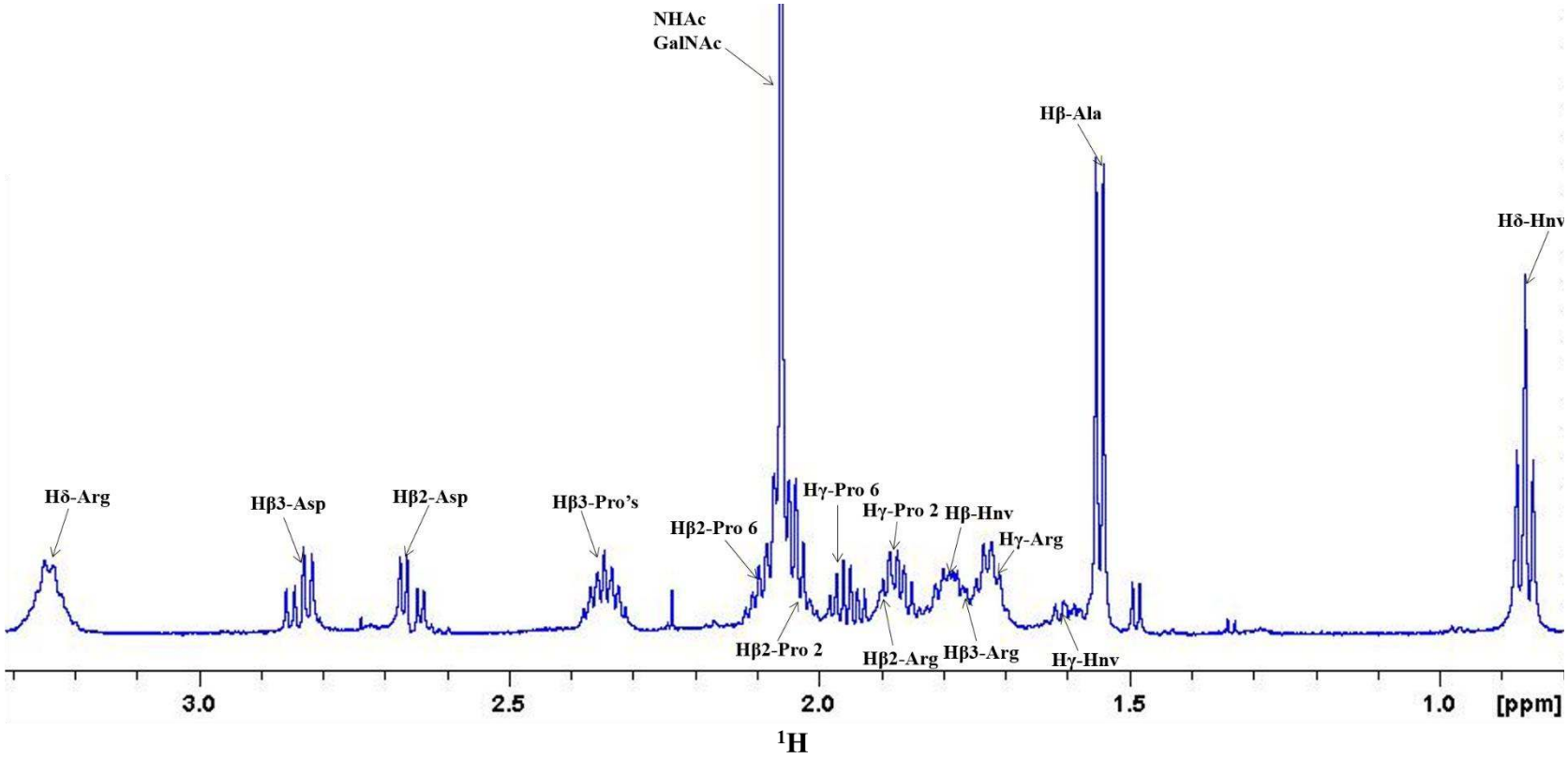


Figure 6.4: ^{1}H-NMR spectrum assignment for the region 3.3 ppm to 0.8 ppm for the glycopeptide APD(Hnv)* RP (*indicates the site of Tn-glycosylation).

6.3. **Appendix 6.3-** Expression vector used for the expression of ^{15}N labelled Gal-3 CRD

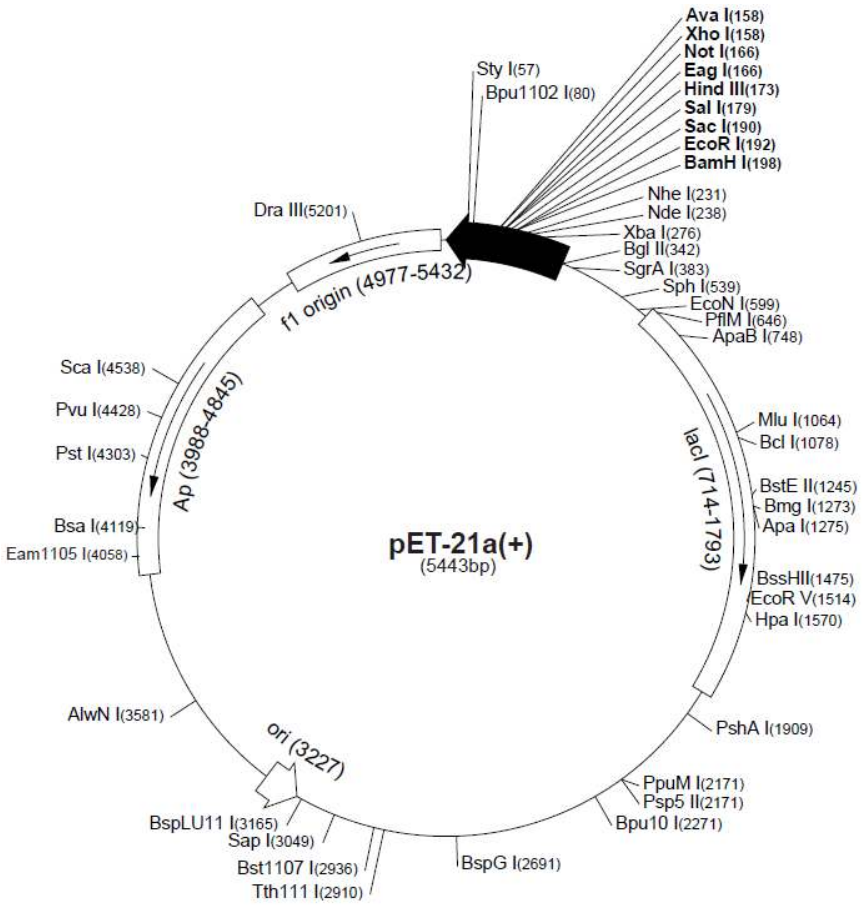


Figure 6.5: Scheme of the expression vector pET-21, obtained from NZYTech. This expression vector contains 5443 bp and Ampicillin resistance.

Recombinant protein sequence for Gal-3 CRD:

MLIVPYNLPLPGGVVPRMLITILGTVKPNANRIALDFQRGNDVAFHFNPRFNERNRRVIVCNTKLDNNWGR
EERQSVFPFESGKPFKIQVLVEPDHFKVAVNDAHLLQYNHRVKKLNEISKLGISGDIDLTSASYTMI

6.4. **Appendix 6.4-**Composition of the LB medium and M9 minimum medium used in the expression of Gal-3 CRD

Table 6.2: Composition of the LB Medium.

Solution	Reagents	Concentration
LB medium	Yeast Extract	5 g/L
	Tryptone	10 g/L
	NaCl	10 g/L

Table 6.3: Composition for the M9 Minimum Medium.

Solutions	Reagents	Concentration
10x M9 salts pH=7.5	$NaHPO_4.7H_2O$	60 g/L
	KH_2PO_4	30 g/L
	NaCl	5 g/L
M9 medium	$MgSO_4$	2 M
	$CaCl_2$	0.1 M
	Glucose	2 g/L
	Thiamine-HCl	10 g/L
	$FeSO_4$	0.1 M
	Ampicillin	100 mg/mL
	$^{15}NH^{+}_{4}Cl$ (^{15}N 99 % Cambridge Isotope Laboratories)	1 g/L

6.5. **Appendix 6.5-** Assignment of the ^{1}H-NMR spectrum for the TF-antigen

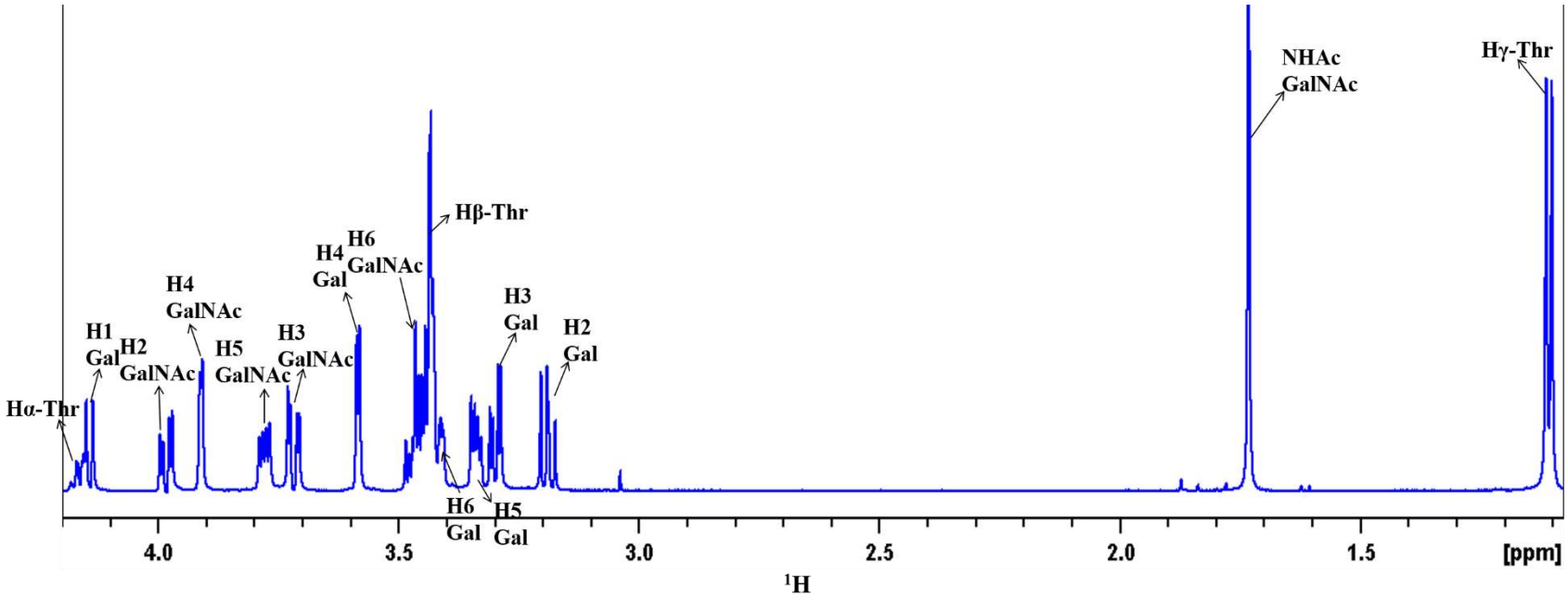


Figure 6.6: ^{1}H-NMR spectrum assignment of the TF-Thr.

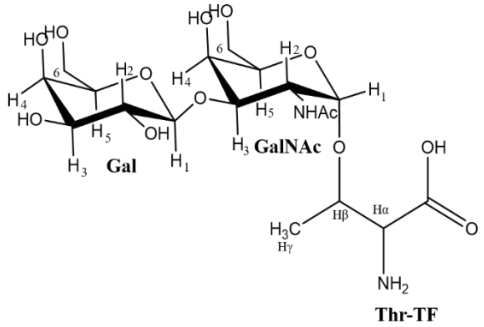


Figure 6.7: Structure for the TF-Thr.

6.6. **Appendix 6.6-** Assignment of the ^{1}H-NMR spectrum for the TF-glycopeptide PDT*RP (* corresponds to the site of TF-glycosylation)

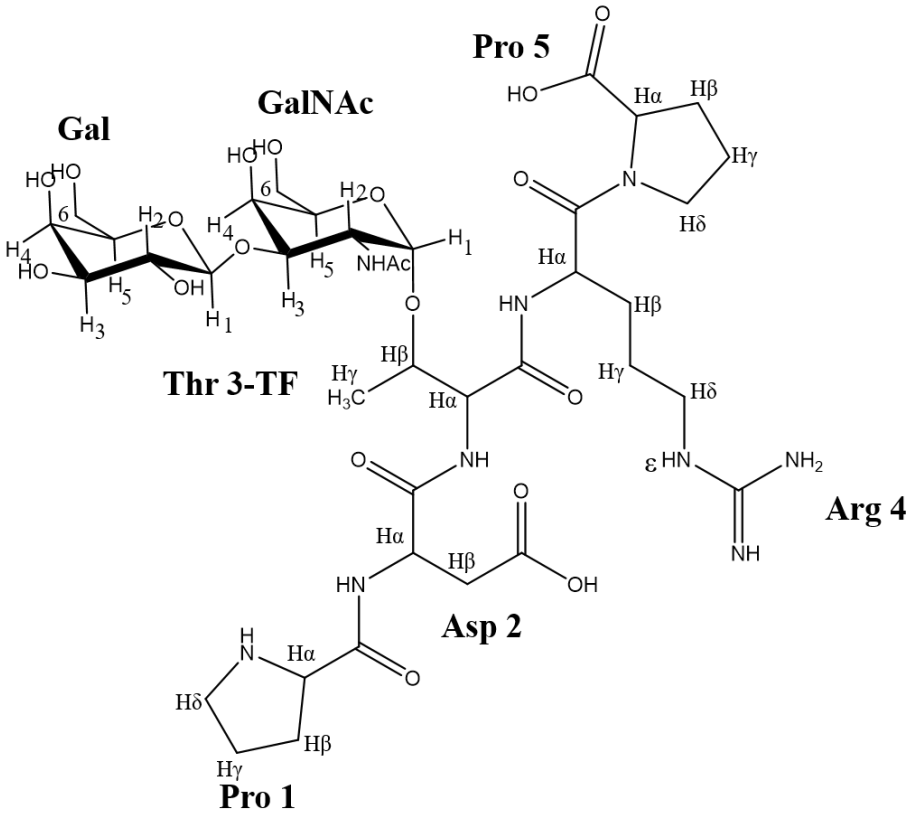


Figure 6.8: Structure of the TF-glycopeptide with the sequence PDT*RP (* indicates the site of TF-glycosylation).

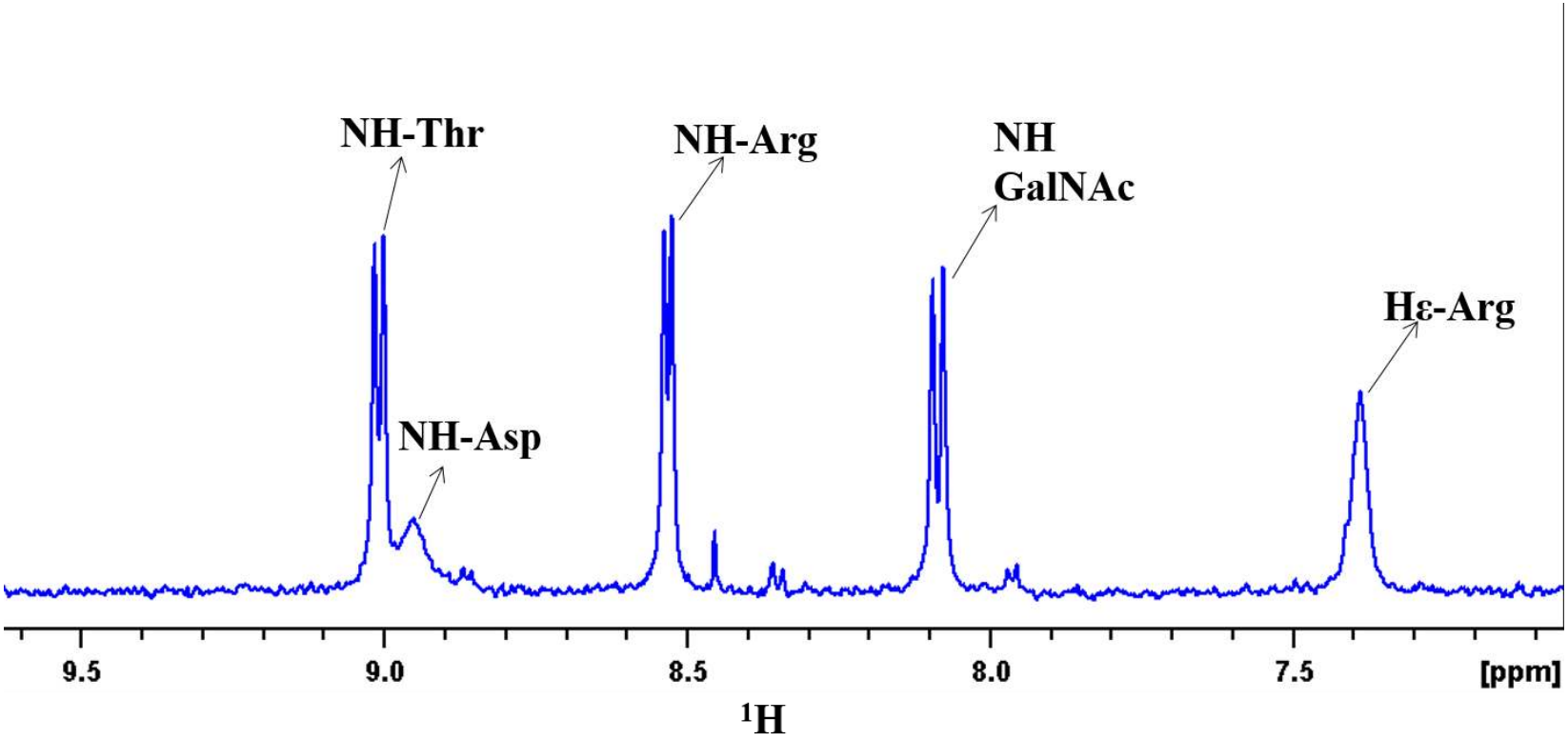


Figure 6.9: ^{1}H-NMR spectrum assignment for the NH region, corresponding to the region of 9.5 ppm to 7.0 ppm, for the TF-glycopeptide PDT*RP (* indicates the site of TF-glycosylation).

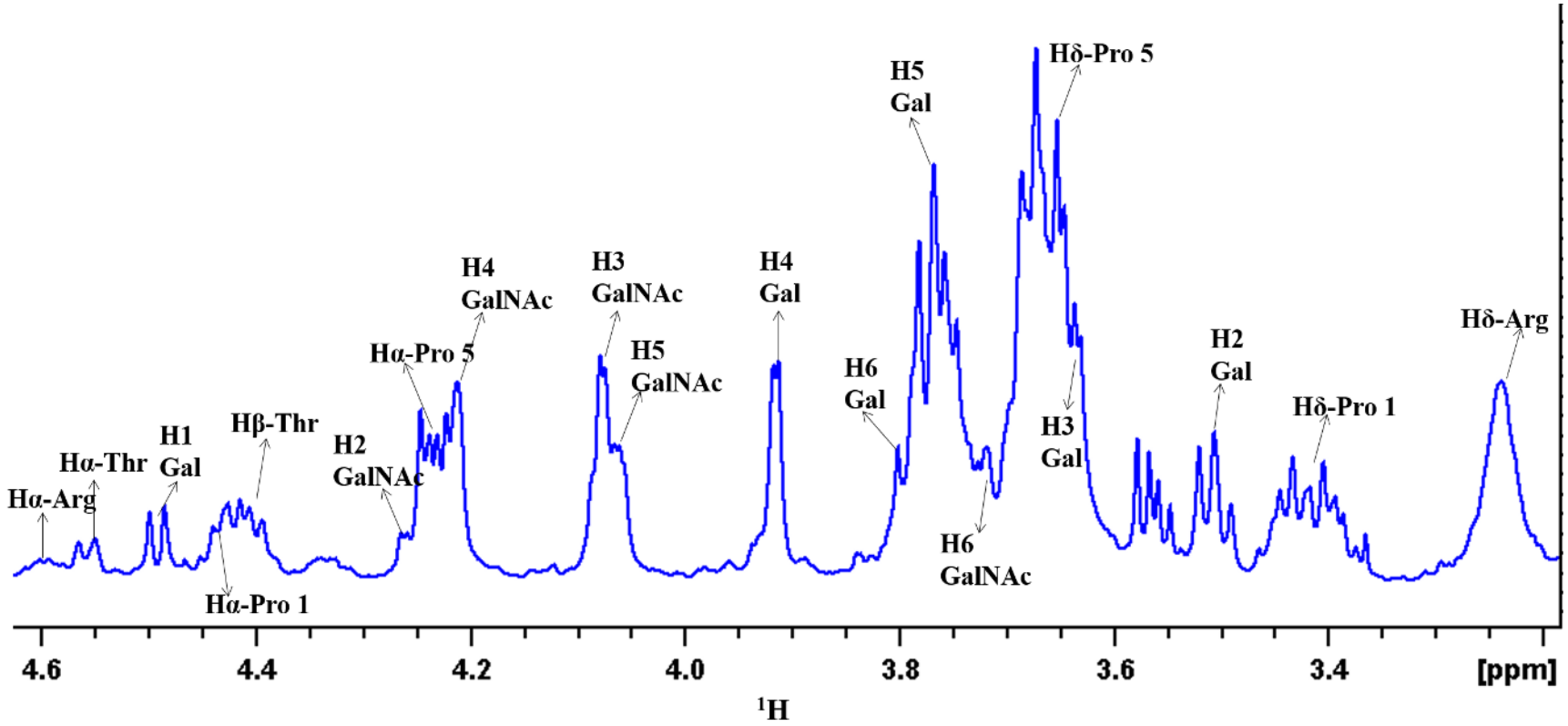


Figure 6.10: ^{1}H-NMR spectrum assignment for the region of 4.6 ppm to 3.2 ppm, for the TF-glycopeptide PDT*RP (* indicates the site of TF-glycosylation).

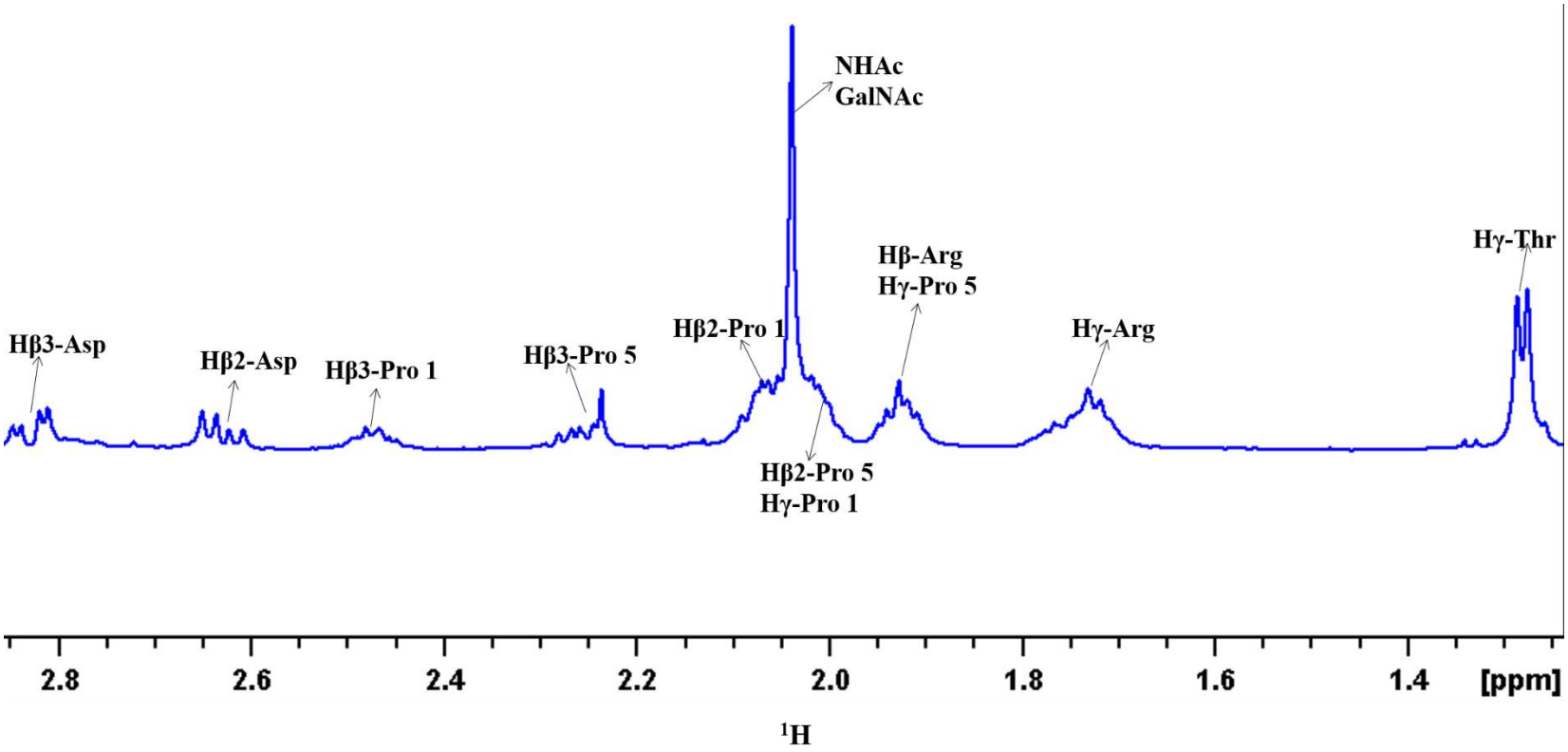


Figure 6.11: ^{1}H-NMR spectrum assignment for the region of 2.8 ppm to 1.2 ppm, for the TF-glycopeptide PDT*RP (* indicates the site of TF-glycosylation).

6.7. **Appendix 6.7-** Expression vector used for the expression of ^{15}N labelled MUC1-4TR

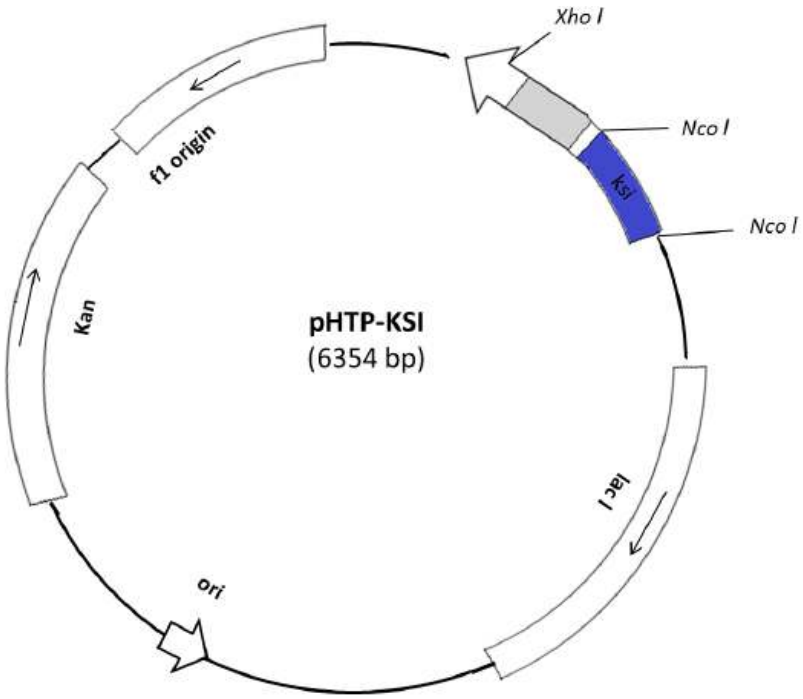


Figure 6.12: Scheme of the expression vector pHTP-KSI, obtained from NZYTech. This expression vector contains 6354 bp and Kanamycin resistance.

Recombinant Protein sequence for MUC1-4TR

MGHTPEHITAVVQRFVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAIREFYANSLKLPL
AVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFRFNGAGKVVSIRALFGEKNIHACQAM
GSSHHHHHHSSGPQQGLRENLYFQGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAH
GVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAH

The KSI tag is the blue sequence; the His-tag is the orange sequence; the TEV protease recognition sequence corresponds to the green sequence and the sequence underlined corresponds to MUC1-4TR.

6.8. **Appendix 6.8-** Composition of the LB medium and M9 minimum medium used in the expression of MUC1-4TR

Table 6.4*:* Composition of LB Medium.

Solution	Reagents	Concentration
LB Medium	Yeast Extract	5 g/L
	Tryptone	10 g/L
	Kanamycin	50 µg/mL
	NaCl	10 g/L

Table 6.5: Composition of M9 Minimum Medium.

Solutions	Reagents	Concentration
10x M9 salts pH=7.5	$NaHPO_4.7H_2O$	60 g/L
	KH_2PO_4	30 g/L
	NaCl	5 g/L
M9 Medium	$MgSO_4$	2 M
	$CaCl_2$	0.1 M
	Glucose	2 g/L
	Thiamine-HCl	10 g/L
	$FeSO_4$	0.1 M
	Ampicillin	100 mg/mL
	Kanamycin	50 µm/mL
	$^{15}NH_4Cl$ (^{15}N 99 % Cambridge Isotope Laboratories)	1 g/L

6.9. **Appendix 6.9-** Assignment of the ^{1}H-NMR spectrum for the T3-Tn peptide

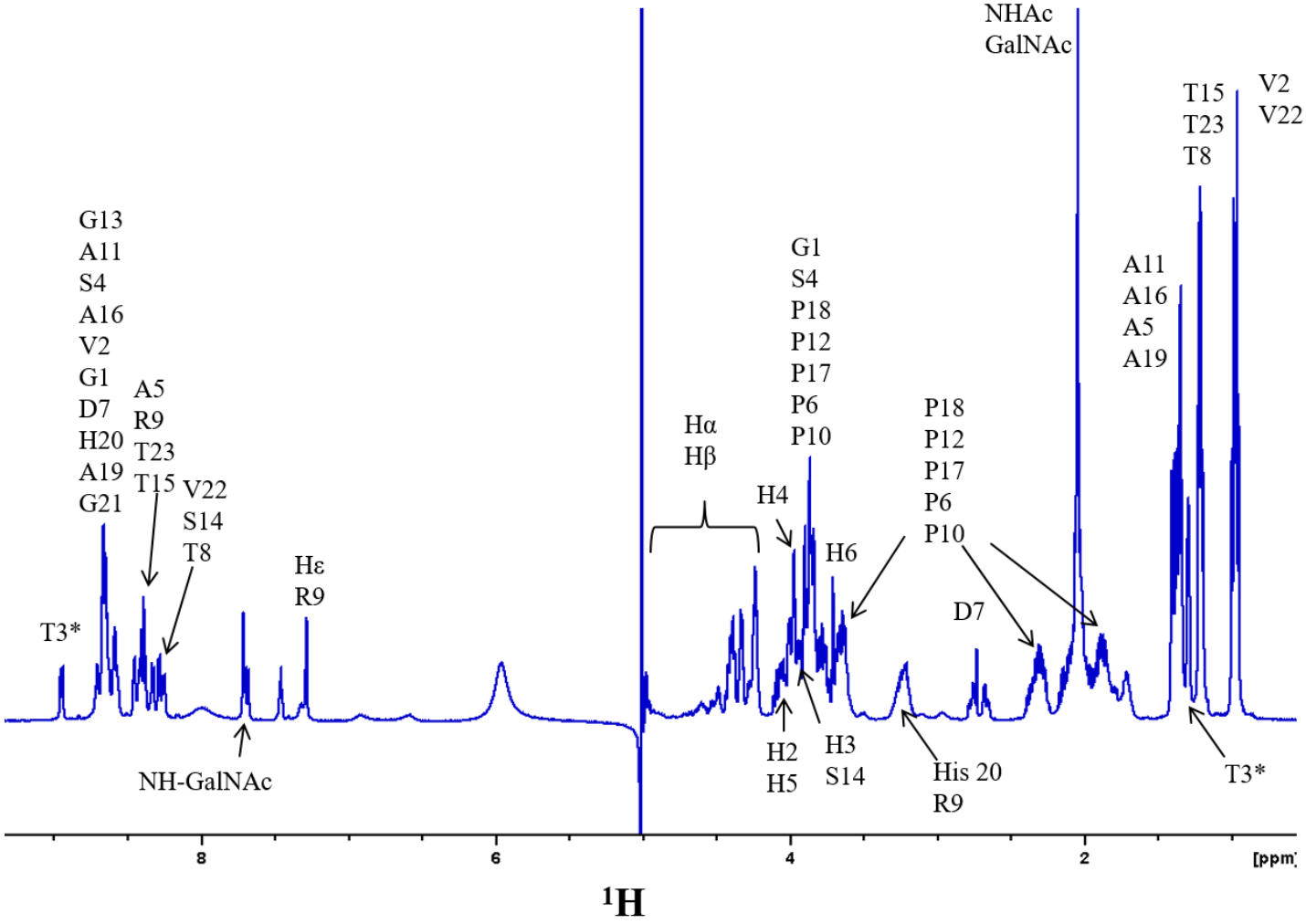


Figure 6.13: ^{1}H-NMR spectrum assignment for the glycopeptide T3-Tn.

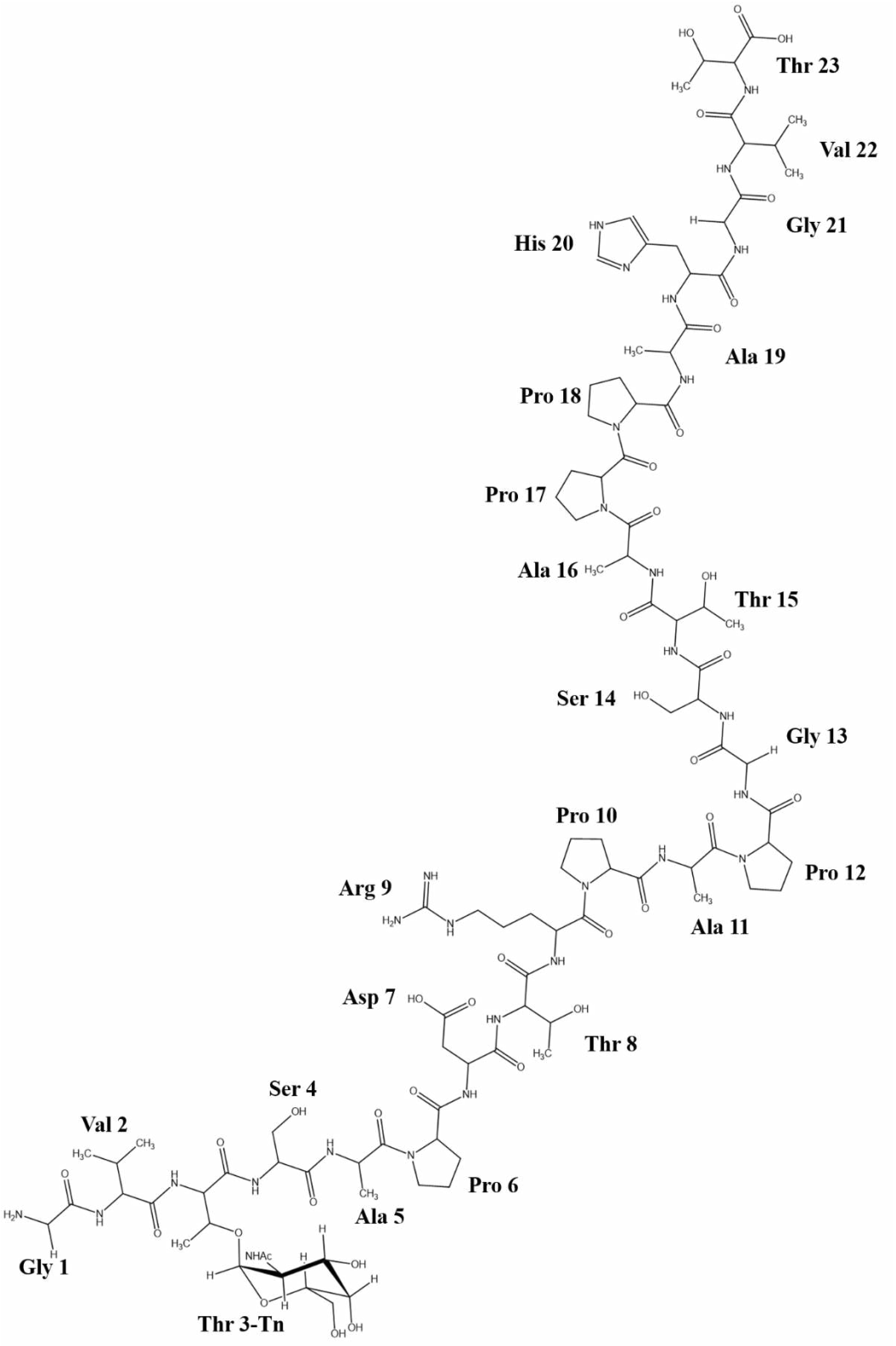


Figure 6.14: Structure of the Tn-glycopeptide T3-Tn.