U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

2.º CICLO

FCUP
ANO

U.PORTO

Development of cancer nanovaccines based on glyco-neoantigens

Beatriz Gusmão Frango Teixeira

FC

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

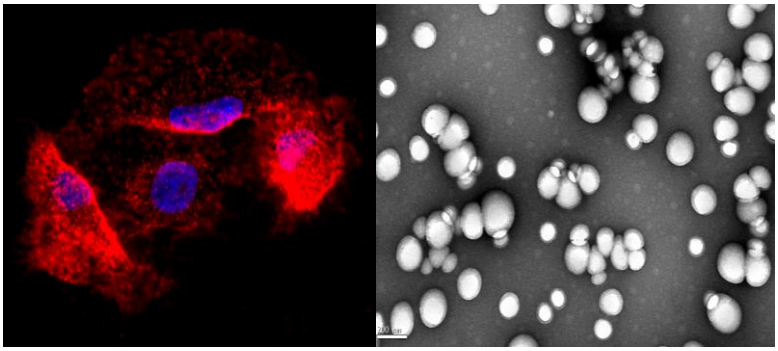
Development of cancer nanovaccines based on glyco-neoantigens

Beatriz Gusmão Frango Teixeira

Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto em
Área Científica

2020

MSc



Development of cancer nanovaccines based on glyco-neoantigens

Beatriz Gusmão Frango Teixeira
Mestrado em Bipquímica
Departamento de Química e Bioquímica
2020

Orientador
José Alexandre Ribeiro de Castro Ferreira
Coorientador
Lúcio José de Lara Santos

IPOPORTO
INSTITUTO PORTUGUÊS DE ONCOLOGIA DO PORTO FG, EPE

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

NOTA PRELIMINAR

Declaro que a presente tese é de minha autoria, tendo sido parcialmente utilizada previamente noutro curso desta instituição. As referências a outros autores (afirmações, ideias, pensamentos) respeitam escrupulosamente as regras da atribuição, e encontram-se devidamente indicadas no texto e nas referências bibliográficas, de acordo com as normas de referenciação. Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

Os trabalhos de investigação desta tese foram realizados no Grupo de Patologia e Terapêutica Experimental do Centro de Investigação do Instituto Português de Oncologia do Porto, EPE.

Declaro ainda que o presente trabalho teve o suporte financeiro do Instituto de Ciências Biomédicas Abel Salazar. Reconheço também o financiamento, pela Fundação para a Ciência e a Tecnologia, do Centro de Investigação do Instituto Português de Oncologia do Porto (PEst-OE/SAU/UI0776/201; CI-IPOP-29-2014; CI-IPOP-58-2015).

Contribution in the concept and accomplishment of the following research work during the year of development of Master degree's Thesis:

Scientific Papers:

I. Rui Freitas, Marta Relvas-Santos, **Beatriz Teixeira**, Rita Azevedo, Janine Soares, Elisabete Fernandes, Lúcio Lara Santos, André M. N. Silva, José Alexandre Ferreira: Single-pot enzymatic synthesis of cancer-associated MUC16 *O*-glycopeptides libraries and multivalent protein glycoconjugates: a step towards cancer glycovaccines – submitted to New Journal of Chemistry.

II. Andreia Peixoto, Dylan Ferreira, Rita Azevedo, Rui Freitas, Elisabete Fernandes, Marta Relvas-Santos, Cristiana Gaiteiro, Janine Soares, Sofia Cotton, **Beatriz Teixeira**, Sara Oliveira, Paula Paulo, Luís Lima, Carlos Palmeira, Maria José Oliveira, André M. N. Silva, Lúcio Lara Santos, José Alexandre Ferreira: Glycoproteomics identifies HOMER3 as a targetable biomarker triggered by hypoxia and glucose deprivation in bladder cancer – to be submitted to *Theranostics.*

Agradecimentos

No final desta etapa tenho de deixar os meus profundos agradecimentos a todos os que me ensinaram, apoiaram e que das mais variadas formas me permitiram alcançar os meus objectivos.

Neste ano particular, com todas as dificuldades que se impuseram, tive a imensa sorte de passar por este processo junto de uma equipa incrivelmente dedicada e empenhada, que se entrega à ciência e se lança aos objectivos de uma forma inspiradora, e sem a qual não posso dizer que estaria onde estou.

Ao meu orientador, o Doutor Alexandre Ferreira, que me facultou a oportunidade de conhecer a ciência pela sua perspectiva, e de aprender a olhar para as adversidades de forma optimista, procurando sempre novas respostas sob um olhar crítico. Foi um prazer enorme poder fazer este percurso ao seu lado.

Ao Professor Doutor Lúcio Lara Santos que me permitiu integrar esta equipa e realizar este trabalho sobre a alçada de um grupo que reflecte a sua própria versatilidade e empenho.

A toda a equipa do grupo de patologia e terapêutica experimental: acredito que um simples “obrigada” não seja suficiente. Tiveram paciência para me guiar e ajudar quando falhei, ao mesmo tempo que me apoiaram aquando das minhas vitórias. Foram professores e colegas de luta excepcionais com os quais pude sempre contar. Andreia, Rui, Cristiana, Dylan, Elisabete, Sofia, Marta, Janine, Sara, Rafaela e Doutor Luís Lima, estou-vos imensamente grata por tudo.

Um enorme obrigada também aos meus amigos pelo apoio incondicional, e um especial obrigada aqueles que, ainda que à distância, me acompanharam e me motivaram durante todo o meu percurso, e que mais que tudo foram motores do meu desenvolvimento pessoal e académico.

`A minha família deixo o maior dos agradecimentos por me terem permitido chegar aqui e por estarem sempre presentes. Aos meus pais, irmão e avós, espero um dia poder retribuir todo o apoio e carinho que me deram.

Sumário

O cancro da bexiga permanece um dos cancros com pior gestão clínica à escala mundial, sendo a quimioterapia e o uso de inibidores de “checkpoints” imunitários algumas das opções terapêuticas correntemente utilizadas. No entanto, muitas vezes os tratamentos recorrentes falham em prevenir recidivas e a progressão do tumor, sendo por isso necessárias várias intervenções que tornam o cancro da bexiga um dos cancros menos sustentáveis a nível económico. A glicosilação aberrante, presente nos estádios mais avançados do cancro da bexiga, é uma das modificações pós-traducionais mais comuns em proteínas de superfície celular. Uma característica recorrente das células tumorais com elevada capacidade de metastização, invasão e evasão do sistema imune é a sua expressão aberrante de espécies truncadas de O-glicanos. Caracterizados pelos antigénios T (antigénio de Thomsen-Friedenreich), Tn e Sialil-Tn, estes glicanos curtos são extensivamente encontrados em tecidos de tumores malignos mas geralmente ausentes em tecidos saudáveis. Nos últimos 3 anos, o nosso grupo desenvolveu caracterizações compreensivas dos perfis de glicoproteómica de células e tecidos de cancro da bexiga vs tecidos saudáveis, a partir dos quais conseguiu identificar glicoproteínas associadas a contextos tumorais específicos. Entre elas estão as moléculas de MUC16 e CD44 decoradas com glicanos Tn e STn que poderão ser utilizadas enquanto neoantigénios para o desenvolvimento de novas terapias. O uso de nanovacinas terapêuticas, associada ao uso destes neoantigénios específicos, tem ganho terreno enquanto alternativa aos tratamentos correntemente aplicados, inclusivamente promovendo resultados terapêuticos mais positivos e em alguns casos memória imunológica. Desta forma, este trabalho foca-se no desenvolvimento preliminar de formulações de nanovacinas pré-clínicas, utilizando glico-neoantigénios de cancro da bexiga.

O capítulo 1 procura aplicar alternativas para a falta de carbohidratos e glicoconjugados puros e bem definidos para futuras aplicações terapêuticas. Para isso recorre ao uso de um método “single-pot” com recurso a várias enzimas, associado a uma caracterização por espectrometria de massa, para gerar MUC16 e CD44 decoradas com Tn e STn e identificar as livrarias de glicopéptidos formadas. O capítulo 2 descreve a síntese de conjugados imunogénicos que resultam do acoplamento de KLH a glicopéptidos de MUC16-Tn e STn, prevendo a construção de uma vacina multivalente. Avaliou-se também o potencial de nanoencapsulação dos conjugados em partículas de

PLGA, um material biocompatível aprovado pela FDA, com o objectivo de aumentar a eficácia e estabilidade de futuras nanovacinas. Este capítulo apresenta ainda a caracterização físico-química das nanopartículas encapsuladas, através de técnicas como TEM, DLS e LDA, as quais demonstram a formação de partículas esféricas com índices de polidispersão razoáveis e carregadas negativamente. Também foi explorada a estabilidade das nanopartículas a diferentes pHs, permitindo avaliar, de forma preliminar, os perfis de libertação controlada das formulações utilizadas. Finalmente, as partículas de PLGA demonstraram selectividade de internalização para macrófagos comparativamente a células com fenótipo semelhantes a células epiteliais, e ainda perfis de citotoxicicdade aceitáveis. Esta segunda parte do trabalho permitiu estabelecer novas formulações para futuros testes *invitro* e *in vivo.*

A terceira parte deste trabalho procurou desenvolver um modelo celular capaz de produzir glicoepítopos específicos, surgindo assim enquanto alternativa para o uso de métodos químico-enzimáticos para a síntese de péptidos específicos de interesse. Para tal criou-se uma “cell factory” através de glico-bioengenharia para a produção estável e homogénea de glicoepítopos de CD44, decorados com glicanos presentes em ambientes tumorais. Recorrendo a tecnologias de edição genética por CRISPR/Cas9, células T24 de cancro da bexiga, que já expressam CD44, foram submetidas a um KO da enzima C1GALT1, impedindo a extensão de antigénios Tn em antigénios T, seguido de um KI para ST6GALNAC 1 para a expressão estável de STn. A validação do modelo foi feita com recurso a FACs e imunocitoquímica. Consequentemente, o modelo desenvolvido resultou no aumento significativo da expressão de CD44-STn quando comparado com células não editadas, tal como determinado por ensaios de PLA. Adicionalmente foi avaliada a manutenção dos níveis traducionais de CD44 e das suas isoformas através de análises de qRT-PCR, os quais demonstraram uma manutenção da expressão quando comparados com células “Wild Type”.

Com este trabalho estabelecemos as ferramentas moleculares para a produção estável de glicopéptidos conjugados com agentes imunogénicos, com relevância para o estabelecimento de futuras terapias. Além disso, fomos capazes de desenhar formulações de nanovacinas utilizando glicopéptidos caracterizados, e assim estabelecer as bases para o desenvolvimento de gliconanovacinas.

Palavras Chave: Cancro da Bexiga, Glicosilação, Tn, STn, CD44, MUC16, Nanovacinas de glicanos, neoantigénios.

Abstract

Bladder Cancer (BC) is one of the most poorly managed cancers in the world, with chemotherapy and immune-checkpoint inhibitors immunotherapy being current therapeutic options. However, mainstay treatments often fail to prevent tumour recurrence and progression, requiring multiple interventions, which makes BC one of the costliest diseases to treat. Aberrant glycosylation of advanced stage bladder cancer is one of the most common posttranslational modifications of cell surface proteins. A recurrent feature of cancer cells with higher capacity for metastasis, invasion and immune evasion is their aberrant expression of truncated O-glycans. These short glycoepitopes are mainly, T antigens, Tn antigen and Sialyl-Tn antigen, which can be extensively found on malignant tumour tissues, while being scarce in healthy tissues. Additionally, comprehensive glycoproteomics characterization of bladder cancer cell lines, tumours and healthy tissues conducted by our group over the past three years, identified several key cancer-specific glycoproteoforms carrying these alterations, namely MUC16- and CD44-Tn/STn. Accordingly, cancer specific glycoproteoforms of relevant cell membrane proteins, as CD44 and MUC16, are a key source of cancer neoantigens, which provide the necessary cancer specificity to support novel therapeutic strategies. Therapeutic nanovaccines based on cancer specific neoantigens start to emerge has promising strategies to address recurrent disease, frequently providing positive therapeutic outcomes and, in some cases, protective immunological memory. As such, the present work focuses on the development of preliminary pre-clinical nanovaccine formulations based on BC glyconeoantigens, envisaging future multivalent glyconanovaccines.

Chapter 1 addresses the lack of pure and structurally well-defined carbohydrates and glycoconjugates to develop therapeutic applications by employing a multi-enzymatic single-pot method coupled with mass spectrometry (MS) for generating STn and Tn MUC16 and CD44 glycopeptide libraries. Chapter 2 describes the synthesis of immunogenic constructs resulting from the engraftment of KLH immunogens to MUC16-Tn/STn glycopeptides, envisaging the assembly of a multivalent vaccine. We also address the potential of nanoencapsulation of glycoimmunogenic constructs by FDA-approved biocompatible PLGA polymers to improve nanovaccines stability and efficacy. This chapter provides a physico-chemical characterization of these nanoencapsulated constructs by Transmission electron microscopy (TEM), Dynamic Light Scattering (DLS) and laser Doppler anemometry (LDA). Briefly, the generated nanoformulations were

composed of polydisperse, spherical and negatively charged nanoparticles. Moreover, constructs stability at different physiological pHs was also explored, providing preliminary results on controlled construct release. Finally, PLGA nanoparticles displayed selective uptake by macrophages as opposed to epithelial-like cells, with acceptable internalization cytotoxicity profiles. Overall, the second part of the work establishes novel formulations for future tests *in vitro* and *in vivo.*

The third part of this work sets the molecular rational for the establishment of cell factories as alternative tools to chemical and enzymatic biosynthesis of relevant glycoepitopes. Specifically, a glycoengineered cell factory for stable production of glycoantigens, namely CD44 glycoepitopes expressing homogenous cancer-associated glycans, was designed through gene editing by CRISPR/Cas9 editing. First, T24 BC cells expressing CD44 were submitted to C1GALT1 knock-out, hampering Tn antigen extension to T antigen, following ST6GALNAC1 KI for stable STn overexpression. Cell model's phenotype validation was obtained by glycoproteome analysis through FACs and immunocytochemistry. Consequently, the developed cell factory significantly overexpressed CD44-STn glycoforms compared to non-edited cells as determined by *in situ* proximity ligation assays (PLA). Moreover, the resulting cell factory (C1GALT1 KO ST6GALNAC1 KI) has maintained the transcriptional levels of CD44 and its experimentally determined isoforms compared to the wild-type system, as determined by qRT-PCR.

Overall, we were able to establish the molecular tools for relevant glycopeptide stable production and coupling with immunogenic adjuvants. Moreover, novel nanovaccine formulations based on well characterized glycopeptide libraries were stablished, setting the necessary grounds for future glyconanovaccine development.

Keywords: Bladder Cancer, Glycosylation, Tn, STn, CD44, MUC16, neo-antigens, Glycan-based nanovaccines.

Index of Contents

Index of Figures

Introduction

Chapter I

Chapter II

Chapter III

Apendix

Index of tables

Abbreviations

APC Antigen presenting cell

ATP Adenosine triphosphate

B3GNT6 β1-3N-acetylglucosaminyltransferase 6

BC Bladder Cancer

BCG Bacillus Calmette-Guérin

BCL-xL B-cell lymphoma-extra large

C1GalT Glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1

C2GnT Core 2 beta-1,6-N-acetylglucosaminyl Transferase

CAM Cell Adhesion Molecules

CDKN2A cyclin-dependent kinase inhibitor 2A

CIS Carcinomas in Situ

COSMC1 C1GALT1 specific chaperone 1

CS chondroitin sulfate

CSC Cancer Stem Cells

DC Dendritic cells

ECM Extracellular matrix

EGFR Epidermal growth factor receptor

EMT epithelial to mesenchymal transition

EPR Enhanced Permeability and Retention

ER Endoplasmic Reticulum

FGFR3 Fibroblast growth factor receptor 3

G1 Grade 1

G2 Grade 2

G3 Grade 3

GAG GIlycosaminoglycan

Gal Galactose

GalNac N-acetylgalactosamine

GalNac-T N-acetylgalactosamine Transferase

GC Gemcitabine and Cisplatin

GCNT1 Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase

GlcNac N-acetylglucosamine

GPI Glycosylphosphatidylinositol

HA Hyaluronic Acid

HAS; HAS1;HAS2:HAS3 HA synthases, HA synthase 1/2/3

HER_2 Human epidermal growth factor receptor 2 or Receptor tyrosine-protein kinase erbB-2

HG High Grade

HGFR or c-Met hepatocyte growth factor receptor

ITAMs Immunoreceptor tyrosine-based activation motifs

ITIMs Immunoreceptor tyrosine-based inhibitory motifs

LG Low Grade

LSC Leukemic Stem Cells

MHC I or II Major Histocompability complex I or II

MIBC Muscle-Invasive Bladder Cancer

MMPs matrix metalloproteinases

MP Microparticle

MUC Mucin

Neu5AC N-acetylneuraminic acid

NK Natural Killer cells

NMIBC Non-Muscle-Invasive-Bladder Cancer

NP Nanoparticle

OPN Osteopontin

OST oligosaccharyl transferase

PD-1 L Programmed cell death protein 1 Ligand

PD-1 Programmed cell death protein 1

PE Pseudomonas exotoxin A

PEG Polyethylene glycol

P-glycoprotein permeability glycoprotein

PIK_3CA Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha

PLGA Poly (lactic-co-glycolic acid)

PTM post-translational modification

PTS Proline Threonine Serine

RB1 Retinoblastoma Transcriptional Corepressor 1

SAMP Self-associated-molecular patterns

SCCs squamous cell carcinomas

ST Sialyl-T

ST6GALNACI N-Acetylgalactosaminide Alpha-2,6-Sialyltransferase 1

STn Sialyl-Tn

T regs Regulatory T cells

TAAs Tumour Associated Antigens

TAMs Tumour Associated Macrophages

TCC Transitional Cell Carcinoma

TERT Telomerase Reverse Transcriptase

T-IC Tumour-Initiating Cells

TLR Toll like receptor

TP53 tumor protein p53 gene

TSAs Tumour Specific Antigens

TURB transurethral resection of the bladder

UDP-GlcNAc Uridine diphosphate *N*-acetylglucosamine

UDP-GlcUA Uridine diphosphate glucuronic acid dehydrogenase

Introduction

1.1 Bladder Cancer Epidemiology and pathophysiology

1.1.1. Epidemiology Stage and Grading

Bladder cancer (BC) is among the ten most common cancer types found in both men and women, with a worldwide incidence of 9.6 and 2.4 per 100,000 in men and women respectively. In 2018 it was responsible for up to 199,922 deaths displaying higher incidence and mortality rates in European Countries [1,2].

Among the several types of BC described: squamous cell carcinomas, adenocarcinomas and urothelial carcinomas, the aforementioned, also known as transitional cell carcinoma (TCC), account for the vast majority of diagnosed bladder cancers. Patients with BC usually exhibit non-muscle-invasive-bladder cancers (NMIBC) whereas only a smaller percentage is diagnosed with muscle-invasive bladder cancer (MIBC)[3]. Previously, Bladder cancer stratification was carried out according to a simple Grading system, where Grade 1 (G1) represented low-grade tumours, with less aggressive phenotypes, Grade 2 (G2) tumours were considered intermediate tumour states, while Grade 3 (G3) embodied high-grade tumours considered more aggressive.

According to the World Health Organization (WHO) 2016 grading system, non-invasive bladder cancers include Low-Grade (LG) tumours such as pTa and pT1 and High-Grade (HG) tumours like Carcinomas in Situ (CIS) and $T1^{HG}$. MIBC are considered HG tumours and classified from T2 to T4 depending on the depth of tissue invasion[4]. Low-Grade NMIBC tumours are characterized by hyperplasic epithelium growth while High-Grade NMIBC and MIBC exhibit dysplastic cell growth (**Figure 1**). Approximately 75% to 85% of diagnoses for BC are NMIBC, with a high incidence of pTa stage cancers. Out of all these patients around 80% will experience at least one recurrence and around 30% of cases will progress into MIBC[5]. This renders BC as a high cost disease, with a necessity for frequent follow-ups and long term treatments[6].

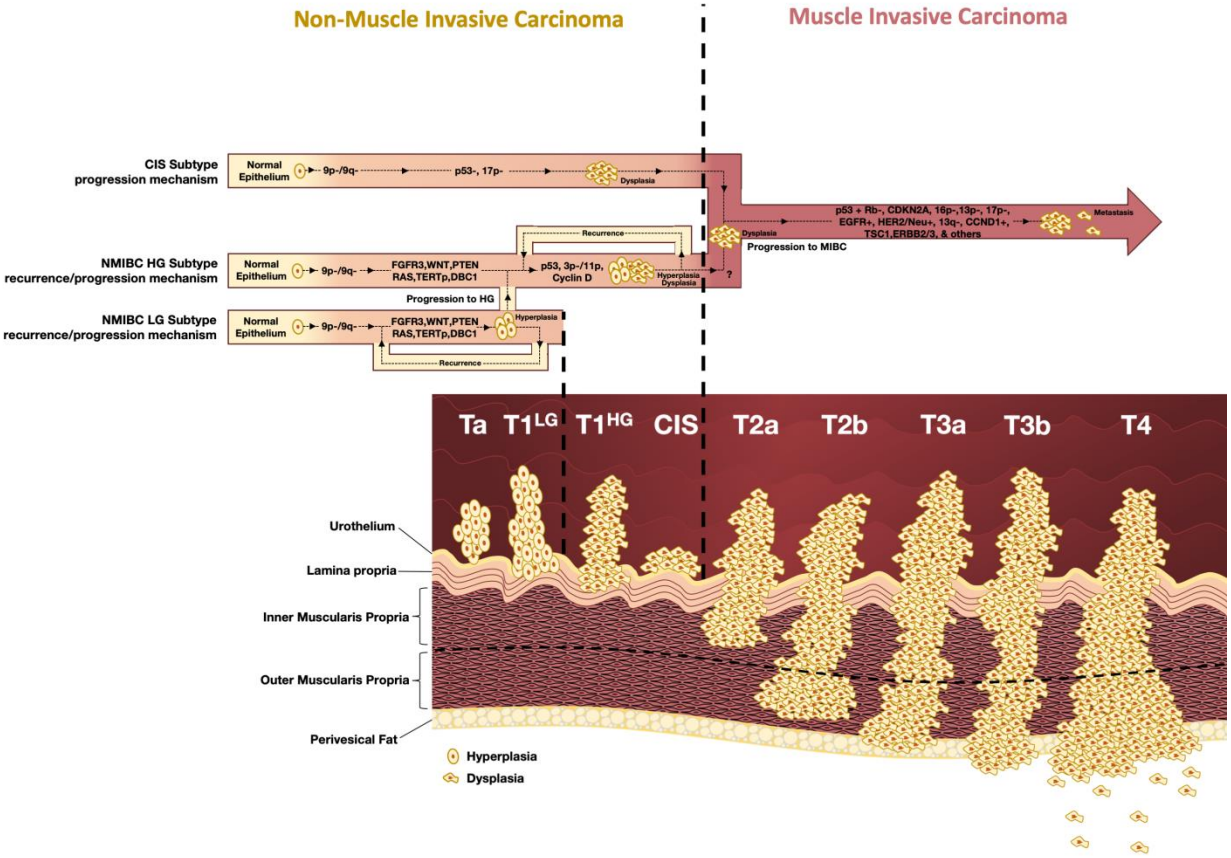


Figure 1. Visual representation of Bladder Cancer grading system according to the Wold Health Organization (WHO) grading system proposed in 2016. This figure depicts molecular and histological features of different Bladder Cancer types. Non-muscle invasive bladder cancers (NMIBCs) include Low Grade (LG) and High Grade (HG) tumours, namely pTa or Ta, pT1 or $T1^{LG}$ and $T1^{HG}$ and CIS, respectively. The LG NMIBCs display hyperplasic cell growth, and HG NMIBCs possess cells with dysplasic growth. HMuscle Invasive Bladder Cancers (MIBCs) display dysplasic cell growth and include different stages according to muscle invasion rates. Figure taken from *Batista R et all, 2020* review: *Biomarkers for Bladder Cancer Diagnosis and Surveillance: A Comprehensive Review*

1.1.2 Genetic Characterization and Risk Factors

NMIBC and MIBC are associated with different molecular and genetic backgrounds. Both cancer types can occur as a result of chromosome 9 loss, linked to early events of BC progression, nervertheless, NMIBCs are characterized by having few genomic rearrangements and a near-diploid karyotype, whereas, MIBCs are associated with aneuploidy events and several chromosomic rearrangements. In NMIBC it is frequent to find loss of 9p/9q chromosomal arms or *TERT* promotor mutations as key events for BC development[7,8]. However, the most recurrent alteration found in patients with these types of tumours are point mutations in *FGFR3* gene, which can be identified in up to 80% of pTa bladder cancer subtypes and are usually associated with more favourable outcomes[9].

These types of tumours can also be correlated with further activating mutations in *PIK$_3$CA* and upregulation of Cyclins, such as Cyclin D1 and Cyclin D3[9]. Among the non-muscle invasive bladder cancers, CIS are considered the most aggressive subtypes, characterized by cell dysplasia and increased potential for MIBC development. This progression has been associated with *TP53* mutations, also found in MIBC[10]. Regarding the more heterogeneous tumours found in MIBC, a vast majority show defects in cell cycle regulation proteins, with common inactivation of *TP53*, *RB1* and *CDKN2A*[9]. There have been reports showing that constitutively activate forms of EGFR are characteristic of more aggressive tumour development[11]. Usually, the identification of these mutations or modifications correlates with worse prognosis for diagnosed MIBC patients.

The most well-established carcinogenic agent causing BC is tobacco, also associated with several different cancer types[12]. However, other environmental and genetic factors contribute for the development of this disease. There are occupational risks, associated with exposure to aromatic amines and other components found in industrial areas, where processing of paints, petroleum products and dyes is carried out[13]. The exposure to ionizing radiation, as means to treat prostate cancer, and the use of cyclophosphamide, a chemotherapeutic agent, have also been associated with urothelial bladder cancer incidence[14–16]. It is clear that inflammation processes and the development of tumorigenic environments have a close relationship and can be a cause effect[17]. Recurrent infection with *Schistosoma haematobium* has been reported to result in chronic cystitis which in turn is considered a direct cause of bladder tumorigenesis[13]. Likewise there are particular important correlations between tobacco smoking, the release of inflammation molecules and cancer associated environments[18]. Genetic predisposition also plays an important role in the development of BC, with close relatives of diagnosed patients displaying higher risks of developing BC clinical pictures. It has been shown that genetically inherited slower N-acetyltransferase 2 (NAT2) and glutathione S-transferase mu 1 (GSTM1)–null variants increase the risk of developing urothelial bladder cancer[13]. These genetic variations are thought to contribute to increased susceptibility to external carcinogenic agents like tobacco[19]. Although genetic variations can be used as complementary diagnosis tools and as important determinants for tumour aggressiveness, they are inconclusive by themselves, specially, given the heterogeneity of this disease and of the mentioned genetic markers. Slow acetylation by NAT2 and GSTM-1 null genotypes is associated with both, invasive and non-invasive BC.

1.1.3. Bladder Cancer Management and Drawbacks

Treatment for BC depends on the degree of muscle invasion and relies on risk stratification to better address therapeutic approaches. For NMIBC, risk tables access grade, stage, multifocality, histology, among other factors to classify patients and their risk of tumour progression or recurrence[20,21]. Usually, for Low-Grade NMIBC, physicians use transurethral resection of the bladder (TURB) for diagnosis and treatment. This treatment can be complemented with intravesical chemotherapy, which has shown positive results for phase III clinical trials, and reduced recurrence in patients[22]. Induction intravesical immunotherapy (adjuvant), using *Bacillus Calmette-Guérin (BCG)* followed by maintenance therapy has also been applied after TURB, in patients with higher risks of tumour progression, and has shown higher chances of preventing recurrences[23]. However there are several reports of patients that endure these treatments, with un-responding reactions to the BCG approaches[24]. Higher grade tumours, such as CIS, should be approached in a more aggressive way with intravesicle immunotherapy and in some cases cystectomies[20,25]. MIBC are often treated with neoadjuvant chemotherapy (cisplatin) followed by radical cystectomy[26], which increases the survival rate for patients in approximately 15% when compared to cystectomy alone[27]. However, taking into account the toxicity of some adjuvant treatments there are important risk assessments to be made before applying these approaches[28]. For the last 20 years patients with tumours presenting metastatic behaviour have been treated with platinum based chemotherapy, with combinations of different neoadjuvants, these are usually composed of cisplatin in association with either gemcitabine or methotrexate/vinblastine/doxorubicin (MVAC - vinblastine, adriamycin and cisplatin or GC- gemcitabine and cisplatin)[29,30]. Even though these treatments represent central therapies for BC, median survival falls short of ideal, with patients displaying 12 to 15 months in both, clinical trials and the common practice[31,32]. Additionally, not all patients diagnosed with MIBC showing metastatic settings are prone for these types of therapies or show actual improvements under these regimes. As an alternative, immunotherapy, using checkpoint inhibitors associated with PD-1 or PD-1 Ligand, has been approved by the US Food and Drug Administration (FDA) in 2016[33–35]. Although these therapies seem to improve response rates, there is a reported unresponsiveness and development of resistance over the curse of time [36,37]. As presented, there have been several developments in alternative or substitute approaches, however, these are still far from optimal and there is a need to establish new therapies based on specific biomarkers, which could also improve prognosis.

Tumour heterogeneity, extensively described in human bladder cancer[38,39], is the major drawback when it comes to developing therapies and outline prognosis. There are several factors that influence this heterogeneity such as, genetic variations, microenvironment, immune settings, among others, which promote different types of alterations and tumorous profiles. Cancer Stem Cells (CSCs) have been hypothesized to be an important contributor for cancer associated heterogeneity, given their differentiation and self-renewal potential. Furthermore, considering that treatment potential of chemotherapy and radiotherapy resides in their capacity to damage rapid DNA replication, the quiescence CSCs and their slower growth rates are believed to endorse resistance, and promote prosperity against anticancer therapies. Additionally, CSCs have been found to display increased expression of ATP-binding cassette (ABC) transporter proteins, responsible for drug efflux, and so, promoting avoidance of anticancer drug effects[40]. Lately, metastatic behaviour, associated with more aggressive and complicated tumours, has also been attributed to the presence of cancer stem cells (CSCs)[41,42]. In light of these evidences, several studies have been design to further address the impact and presence of CSCs in cancer development[43]. In 2009, a group from Stanford Cancer Center, isolated and characterized a subpopulation of CSCs, from primary human bladder cancer, known as tumour-initiating cells (T-IC). T-IC $CD44^+$ cells displayed potential to induce xenograft tumours in vivo, with similar heterogeneity to the original tumour, whereas, $CD44^-$ cells, lacked the ability to implement xenografts in immune compromised mice[44]. CD44 is a molecular receptor, reported to characterize CSCs, which have been shown to display tumour inducing capabilities, in different malignancies like gastric and colorectal cancers[45,46]. On a different line of work, Mucin involvement in bladder cancer resistance to chemotherapy has also been shown, with interesting correlations with EMT (Epithelial to Mesenchymal Transition)[47,48]. Inclusively Mucins have been shown to upregulate CSC like phonotypes, typically associated with resistance to chemotherapy and relapses as previously mentioned[48,49]. Particularly MUC16, with preciously described roles for cisplatin resistance in cancer scenarios[50], has emerged as a novel marker in bladder cancer cells[51].

1.2. The Impact of CD44 and MUC16 on Cancer

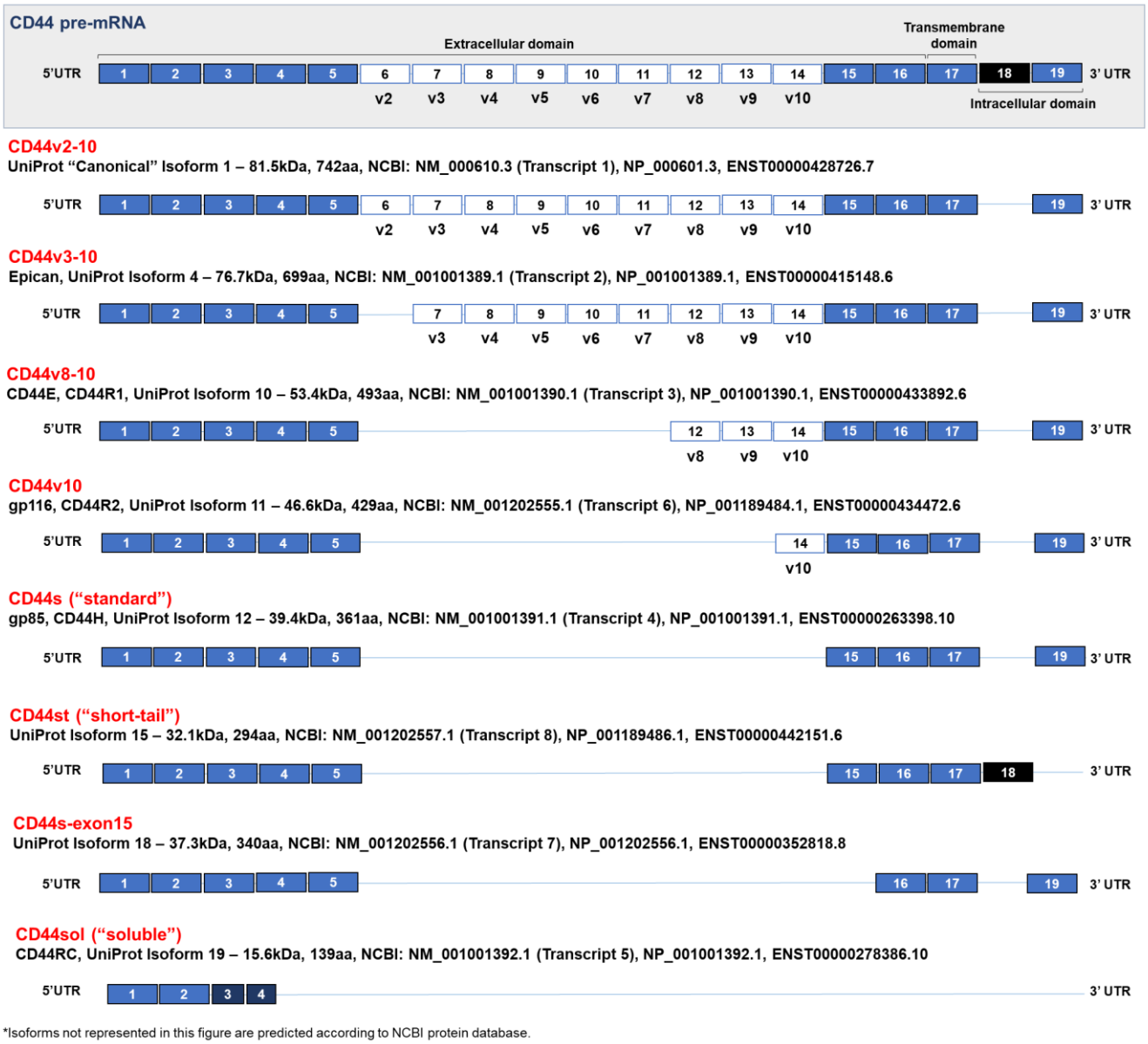
1.2.1. CD44 structure and biological role

CD44 is a cell surface glycoprotein, first described and identified as an antigen present on granulocytes and T lymphocytes of brain tissue[52]. Later on, this protein was also found on the surface of leukocytes, fibroblasts, epithelial and mesodermal cells, and was shown to impact inflammatory responses on different levels [53]. CD44 is a multifunctional receptor and a member of the CAM family (cell adhesion molecules) responsible for cell-cell and cell-matrix interactions, with important roles on migration and signalling pathways[54].

Encoded by one highly conserved gene on chromosome locus 11p13[55,56], CD44 transcripts can experience alternative splicing, resulting in a large variety of isoforms. The gene encoding CD44 possesses 20 known exons, the first 5 accounting for the non-variable region of the extracellular domain, and exons 6 to 15, encoding its' variable regions (V1-V10), which can suffer alternative splicing and thus generate different exon combinations. From here, exons 16 and 17, considered the constant region of the 3', encode the membrane proximal region of the extracellular domain, followed by exon 18 which accounts for the transmembrane domain, an important pathway for protein and co-factor interactions and lymphocyte homing[56,57]. The cytoplasmic domain of CD44 is also implicated in alternative splicing events, with expression of exon 19 resulting in a short cytoplasmic tail, with only 3 aa, and exon 20 expression being associated with a long cytoplasmic tail, with 70 aa found on all isoforms[55]. This 70 aa long intracellular tail of CD44 interacts with cytoskeletal protein adaptors, ERM (ezrin, radixin and moesin), with major structural and downstream signalling functions[58–60]. Furthermore, upon proteolytic cleavage, CD44-ICD (intracellular domain) has been found to mediate gene transcription via nuclear translocation, promoting expression of more CD44[57,61,62].

Out of all the isoforms, CD44v2-v10 is the largest one including all variable exons except for exon 1, which is not expressed in humans due to presence of a stop codon. CD44v3-v10 possesses one less exon and expresses the heparin-sulphate preotoglycan domain of CD44 similar to Epican[63,64]. Epican is a heparin/chondroitin/keratan sulfate proteoglycan present in the cell surface of keratinocytes and represents an important marker for epidermal differentiation and cell-cell adhesion [65,66]. Other isoforms, such as

the smaller CD44v8-v10 also referred to as CD44 E, are mainly found on epithelial cells and have been associated with treatment evasion, characteristic of CSCs[67,68].



*Isoforms not represented in this figure are predicted according to NCBI protein database.

Figure 2. Schematic representation of experimentally confirmed human CD44 pre-mRNA and respective isoforms. Blue filled boxes represent constant region exons, while white filled boxes represent exons of the variable region present in the designated CD44 isoform. Dark blue filled boxes with reduced box size represent truncated exons from the constant region. The blue line represents missing exon(s). Exon 18, filled black, contains an early 3'UTR and only makes part of CD44st isoform.

1.2.2. Relevance of CD44 ligands in cancer contexts

CD44s or CD44H (hematopoietic) is the most widely expressed and one of the shortest CD44 molecules, with approximately 85-90 kd. This conserved standard form is characterized by the lack of a variable region, containing transcripts with exons 1-5 and 16-20 spliced together. This form of CD44 is known to be the main receptor for hyaluronic acid (HA), a major component of the Extracellular matrix (ECM)[69]. HA is a glycosaminoglycan (GAG) involved in several pathways that determine matrix remodelling, tumour development, inflammation processes, morphogenesis and many others[70,71]. Unlike other smaller GAGs, synthesized in the Golgi apparatus and covalently associated with core proteins, HA is synthesized by one transmembrane HA synthase (HAS1, HAS2 or HAS3) that adds UDP-GlcUA and UDP-GlcNAc substrates to a extracellularly extruded HA[72–74]. Depending on the HA synthase different sized HA molecules can be formulated. HAS1 and HAS2 synthesize a long molecular HA with a high molecular weight (HMW) whereas HAS3 forges low weight molecules (LMW) with higher catalytic activities[75]. *In vitro* studies have shown that depending on its' weight, HA can promote or inhibit tumour settings[76,77]. It also entails different outcomes depending on its' interaction with receptor proteins, namely, CD44.This interaction between CD44 and HA depends on the receptors' activation sate, which can be modulated through intracellular phosphorylation, and/or glycosylation of the extracellular domain[78]. Activated CD44 is more prone to HA binding, and can be found overexpressed in solid tumours, with consequent expression of drug resistance proteins, like P-glycoprotein, and of Bcl-xL, anti-apoptotic gene[79]. In pancreatic cancer cells, small HA oligosaccharides have been associated with CD44 intracellular cleavage and tumour cell motility. Accordingly inhibition of CD44-HA binding in this context, overrides tumour cells' motility[80]. Expression of HA synthase 1 is associated with bladder cancer phenotypes. Bladder cancer cells, HT1376, transfected with anti-sense HAS1 displayed slower growth and invasion rates compared to sense HAS1 transfected cells. Moreover the absence of HAS1 resulted in downregulation of CD44 isoforms, (CD44v3, CD44v6 and CD44E)[81]. Given the reported importance of CD44-HA binding in tumour development, this interaction has been thoroughly studied as a potential therapeutic target[71].

CD44 binds HA with the help of a “link module” motif (BX_7B) in the amino terminal extracellular domain[82]. This motif can be found on exons 2 and 3 of CD44 transcripts and is present in all isoforms of this molecule. It would be expected that as a result of this

binding motif's prevalent transcription, all CD44 forms would bind HA, however this is not the case as molecules expressing this region have been shown to not bind HA, or bind it with lower affinity and without promoting the same functional outcomes[69,83]. This is a result of CD44's heterogeneity, deeply associated with its' wide array of post-translational modifications, such as N-linked and O-linked glycosylation and glycosaminoglycan's addition (GAGs side chains), which are dependent on cell type, cell environment and cell activation profiles[83,84]. The NH_2 terminal extracellular region of CD44 contains five potential N-glycosylation sites, in the BX_7B HA binding motif, and its' membrane proximal region also displays a large number of Serine and Threonine residues which can function as sites for O-Glycosylation[85]. The presence of glycosylation structures in these regions impacts the binding of CD44 to HA and has been shown to act as either, an important promoter or an inhibitor for this interaction, depending on the cell line evaluated. It was reported that CD44 N-glycosylation inhibition by tunicamycin, resulted in decreased HA adhesion, independently of CD44 density[86]. Accordingly, in melanoma cells, mutated N-glycosylation sites resulted in decreased HA recognition by CD44[85]. A study evaluating the mechanisms by which CD44 glycosylation influences HA binding in Chinese hamster ovary cells, found that for nonbinding clones, treatment with tunicamycin, neuraminidase or endoglycosidases mixed with glycosidases, increased HA recognition through CD44[87]. Overall N-glycosylation seems to regulate HA binding to some extent and it does so differentially for each cell line[88]. O-glycosylation also impacts the binding between CD44 and HA, as shown by *Bennett et al.* studies on melanoma cell transfectants. Cells expressing CD44H bound HA more efficiently when compared to CD44E, however, the absence of O-glycosylation structures granted CD44E an HA binding capacity similar to CD44H. Unsurprisingly, a CD44 fusion protein with increased O-glycosylation potential, displayed lower binding potentials for HA[89]. Another study reported that the presence of truncated O-glycans on CD44, increased HA binding capacity[90]. As such, it is important to properly understand O-glycosylation mechanisms, and how these different O-glycans can modulate the activity of CD44 and other potential cancer antigens

Despite HA being CD44s' major ligand it is not the only molecule this receptor binds and interacts with. Osteopontin (OPN) is an extracellular matrix protein able to bind CD44 and mediate tumour progression and metastasis[91,92]. There have been correlations between expression of both molecules in tissues of gastric cancer patients[93], and reports showing that OPN-CD44 expression and signalling promotes stem cell-like properties and resistance to radiotherapy[94]. Inclusively, CD44 knock-down or inhibition resulted in attenuated OPN secretion and signalling, with consequent decrease in clonogenicity of colorectal cancer cells[95]. There are other know ligands for CD44, such as, chondroitin

sulfate (CS), whose interaction with CD44 has been associated with apoptosis protection in chronic lymphocytic leukemia cells[96]; or fibrin, a ligand for CD44 in colon carcinoma cells. There are also important associations between matrix metalloproteinases (MMPs) activity and localization regarding CD44 signalling in cancer environments. MMPs are ECM proteins involved in its' degradation and reformulation. The interplay between MMPs and CD44 has been previously shown in both human melanoma cells and mouse mammary carcinomas, where CD44 seems to anchor MMP-9 on carcinoma cells, and cluster disruption results in reduction of tumour invasiveness[97].

CD44 and its ligands seem to be tightly related to cancer and/or inflammation processes; however pin pointing one specific molecular target becomes a difficult task when we consider CD44s' heterogeneity, accounting for several different isoforms associated with different cell types and environments, different post-translational modifications and different ligand associations.

1.2.3. CD44 differential forms and Cancer

1.2.3.1. CD44 isoforms and Cancer Specificities

Many studies have been conducted in order to properly address the relevance of specific CD44 forms in different cancer subsets. For instance, CD44v6 isoform is represented in the large majority of squamous cell carcinomas (SCCs), comprising epithelial tumours of the airway, digestive and genital tract as well as skin, with reported CD44v6 expression as high as 90% on head and neck, lung and skin tumours[98]. Accordingly, it has been found on HSC-3 cells, from head and neck tumours, and in pancreatic cancer cells. In HSC-3 cells, antibody inhibition of CD44v6, resulted in increased susceptibility to cisplatin treatments and decreased proliferation, whereas for pancreatic tumours its expression converted non-metastatic carcinoma rat cells into metastatic ones upon transfection[99]. By quantitative-RT-PCR *Olsson et al.* evaluated the expression of different CD44 isoforms in different breast cancer subtypes (distinguished by the different clinical markers displayed). They found that for each breast cancer cell subtype there were different CD44 isoforms associated, proving that expression of CD44 forms is specific for a given tumorous cell context[100]. All these highlight isoform specificity for a given cancer scenario and denote alternative interactions for CD44 isoforms depending on their environment.

In colorectal cancer, isoforms v6 and v3 have been both associated with metastatic phenotypes, believed to correlate with c-Met receptor/HGFR[101–103]. Regardless of evidences indicating an important role for CD44v6 in colorectal tumour progression, other studies have shown this molecule is not a determinant player for these events[104]. CD44v10 is also believed to mediate important events on pancreatic cancer, like CD44v6 does but to a lesser extent[99], and is also implicated in leukaemia and lymphomas[105]. CD44v10 is associated with modulations in HA interactions, potentiating alterations of CD44 and cytoskeletal interactions as well as binding to Ankyrin, which in return induces cell migration and invasive behaviours[106]. However, a study has presented CD44v10 as an anti-metastatic glycoprotein in pancreatic cancer, leaving some doubt as to its real potential. The lack of consensus regarding the roles each CD44 variant interprets in cancer can be attributed to overall scenario specificities: the same molecule in different environments will interact with different receptors and exert alternative functions. Additionally, one cannot forget that isoforms share multiple variable regions and therefore, expression rates and functional results may uphold cumulative effects of more than one isoform. Likewise all isoforms can express the constant region characteristic of CD44s, and so depending on specificity of the identification method results can be biased.

.

1.2.3.2. Importance of short CD44 standard forms in oncogenic development

In 2011 *Brown et al.* showed that, in breast cancer, both *in vivo* and *in vitro* systems required transition from CD44v expression to CD44s for cells to undergo epithelial to mesenchymal transition (EMT), which is an important mechanism underlying cancer cell metastasis and recurrence[107]. Similar associations have been reporter for pancreatic adenocarcinoma cells whose expression of CD44s and display of EMT phenotypes was linked to higher invasion rates and gemcitabine resistance *in vivo*[108]. CD44s/h influence on cancer progression was also previously evaluated in Namalwa cells, from Burkitts' lymphoma, with studies demonstrating this molecules' impact on metastasis and tumorigenicity [109,110]. Higher expression levels of CD44s have also been correlated with lower medians of survival for patients with pancreatic cancer, compared to patients with low CD44s levels. For this case, blocking of CD44s with specific antibodies inhibited tumour initiation and recurrence[111], something already described for Namalwa cells from Burkitts' lymphoma[109].

1.2.3.3. The relevance of CD44 standard form in Bladder Cancer

As previously mentioned CD44 is considered a marker for CSCs, extensively related to tumour initiating events and therapy resistance. There are studies confirming the tumour initiating capacities of CD44, which show that subpopulations of $CD44^+$ bladder cancer cells display T-ICs characteristics[112]. Particularly, cancer stem like populations have been identified in high-grade T24 bladder cancer cell lines, improving the relevance this marker plays in bladder cancer scenarios[113,114]. CD44s is also associated with EMT transitions as previously described. In Bladder Cancer scenarios, $CD44^+$ cells exhibited higher invasion ability and increased EMT compared to CD44- cells, favoured by IL-6[115]. CD44 has also been proposed as a predictive marker for bladder cancer progression, There are several correlations between tumour grade and CD44s immunostaining that could corroborate these predictive capabilities, however its efficiency in these regards is still controversial[116–118]. Even though CD44s is found on G1 and G2 transitional cell carcinomas (TCCs), its manifestation is more heterogeneous in HT1197 cells and high grade, G3, tumours. Nonetheless, there were areas of squamous differentiation in high-grade tumours which displayed high reactivity for CD44s[119]. Another study showed that CD44s could be found on HT1197 and 5637 cells through a RT-PCR analysis, and that cell positive staining for this molecule was associated to cell phenotype and differentiation status[120].

1.2.4. Mucin family in Cancer disease development

Mucins are a family of proteins that comprise up to 21 different MUCs which can be classified as membrane bound or secreted Mucins. Among this family of proteins, there are forms expressing Proline/Threonine/Serine rich domains (PTS domains), which are highly associated with PTMs, particularly with O-glycosylation extension and modification. They can be found on epithelial surfaces of the gastrointestinal, respiratory and genitourinary tract and serve as protection against infectious agents, physical and chemical damage, comprehending the major components of the glycocalyx in mucosal tissues. The secreted mucus gel that involves and protects epithelial cells in mucosas is mainly made up of secreted mucins along with other defensive structures, such as antibodies, defensins and lysozymes. These molecules form a tri dimensional mesh with antimicrobial activity and opsonisation characteristics. The expression of mucins in

healthy tissues is restricted to the apical membranes of exposed epithelial cells. However, following tumour events there is a deregulated expression of these molecules in cell surfaces that mediate profitable tumour interactions. Among this large family of molecules, many MUCs, secreted and membrane bound, have been shown to potentiate chemoresistance in cancer cells[121], broadening the array of relevant cancer molecules available as biomarkers and therapeutic targets.

Mucins 1 and 4 are membrane bound MUCs that have been extensively studied in cancer contexts. Just like other members in this family these molecules have been shown to be involved in cancer treatment resistance to drugs such as, cisplatin, doxorubicin, paclitaxel and gemcitabine, in several different cancer types[122]. However, these molecules can be associated with several other key tumorous features such as, metastasis, evasion of immune surveillance, signalling, tumour-cell interactions and growth[123–126].

MUC1 has been reported to have anti-apoptotic properties that counteract the apoptotic action of gemcitabine in fibroblasts[127]. For breast cancer cells, overexpression of MUC1 is associated with less cell susceptibility for antibody HER_2 treatments. Overexpression of HER_2 in breast cancer patients often entails the use of antibodies against this receptor (trastuzumab) as a treatment option. However resistance to treatments is recurrent but has been shown to be improved with simultaneous administration of MUC1 inhibition molecules. This reveals a possible role for MUC1 in breast cancer resistance mechanism[126]. Accordingly, breast cancer cells with developing resistences to lapatinib and trastuzumab treatments upregulate MUC4, which when silenced can reverse resistance[128,129]. Mucin evaluation can be a predictive marker for cancer resistance to neoadjuvant chemotherapies and radiotherapies. The overexpression of these mucins seems to masks cancer cell surfaces and decreases their accessibility. Ovarian tumours express MUC16 on their cell surfaces which can be shed from tumour cells and proceed to inhibit the cytotoxic mechanisms of Natural Killer (NK) cells. Additionally, the presence of membrane bound MUC16 can inhibit immune synapses between tumour and NK cells, necessary for effector immunologic responses[130]. Inclusively, MUC1, 4 and 16 have been shown to trigger endothelial to mesenchymal cell transitions with further disruption between cell-cell interactions and invasion[48,131–133].

All these features associated with mucins in tumour scenarios, from chemoresistance, apoptotic inhibition, immune evasion to shifts to EMT and CSC phonotypes, render these molecules as a reliable target for biomarker development and applications for treatment designs[134].

1.2.4.1. MUC16 in cancer settings

MUC16 also known as CA125 (when shed from cell surface) is the largest MUC and one of the longest human proteins. Despite its major size this molecule can be further extended by O-glycosylaton at these Ser/Thr rich domains. MUC16 is encoded at chromosome 19 and is a transmembrane protein with MUC family characteristic PTS domain in the N-terminal. This tandem region is composed of approximately 60 repeats with 156 amino acids each. Its cytosolic domain (CT) can be phosphorylated and this leads to MUC16 sheding[135]. This domain contains a specific sequence of a.a, (RRRKK) predicted to bind ERM proteins which could potential serve as intermediators for MUC16 interaction with the cytoskeleton [136]. This sequence could also serve as an nuclear localization signal NLS), however it has been shown that MUC16 transfected MUC16 domain don't mediate nuclear internalization[137]. Inclusively, in pancreatic cancer cells, this domain has been shown to interact with Janus Kinase 2 (JAK2) and increase proliferation, this domain enhances nuclear localization of JAK 2 and promotes metastatic and stem like properties[138]

CA 125 is a biomarker for ovarian cancer, but its aberrant expression has been identified in breast, lung, pancreatic and other cancers[139]. For breast cancer cell lines MUC16 knockdown (KD) was shown to inhibit cell growth by promoting apoptosis. Furthermore, MUC16 KD suppressed adhesion and invasive abilities of cancer cells and reduced their colony-forming abilities.

In ovarian cancers its expression has been associated to metastasis, unregulated cell growth and also inhibition of immune cells[140,141]. Particularly, the presence of natural human antibodies against CA125 in pathological conditions, suggest a possible target for immunotherapy. More recently, MUC16 has been implicated in Bladder cancer, with reported high levels of serum MUC16[142] and identification in several bladder cancer cell lines[51].

The O-glycosylation pathway is altered during cancer and its deregulation influences the functions of many glycosylated proteins. As mentioned, MUC16 is rich in high repeats of regions contain many Ser/Thr O-glycosites. Given that, around 77 % of the total MUC16 weight has been attributed to the presence of carbohydrates, disturbances in glycosylation are expected to cause major alterations in the interactions and properties of this molecule. Accordingly Glycosylated forms of MUC16 have been shown to bind E- and L-seletins resulting in altered adhesion pathways and tumour associated features[143]. Additionally different levels of glycosylation found on MUC16 have been proposed as

specific biomarkers[144]. The recent evidence points to MUC16 as a molecule of interest for cancer treatment. Despite it being mainly studied in ovarian cancers several new implications for this molecule have appeared.

1.3. General Aspects of Glycosylation

Addressing the importance and function of cancer associated molecules compels fundamental knowledge about protein modifications, which can majorly influence the behaviour of a given cancer specific agent. Both CD44 and MUC16 can be extensively submitted to glycosylation, a phenomenon that can alter cell interactions, recognition, among many other processes, as further described.

Glycosylation is pos-translational modification (PTM) found on most membrane associated proteins. It is an enzymatic process characterized by the addition of carbohydrates to lipids, proteins or other saccharides. It's responsible for originating highly diverse constructs, given that different monosaccharides can be added in different positions, associated with different aglycones (protein, lipid, etc) and glycosydic linkages (N-, O- and C-linked Glycosylation, as well as GPI anchor attachment and phosphoglycosylation). One protein can be covalently bound to more than one glycan, either through a nitrogen linkage (Asparagine), where the process is called N-glycosylation, or through an oxygen linkage (Serine or Threonine), in which case it's known as O-glycosylation. All these diverse glycosylation structures differentially mediate events of cell to cell adhesion, differentiation, migration, folding, immune recognition and signal transduction.

N-glycosylation in eukaryotes usually starts co-translationally with translocation of the un-folded peptide to the lumen of the endoplasmic reticulum. Here, the first 14 sugar residues of the N-glycan are synthesized (Glc3Man9GlcNac2) with the help of a lipid anchor and are transferred "en bloc" to the recently nascent protein by an oligosaccharyl transferase (OST). These structures are then "trimmed" in the ER (glucosidase dependent pathway) or Golgi apparatus (glucosidase independent pathway) thus attaining a proper folded conformation, controlled by ER chaperones in the glucosidase dependent pathway. Followed by elongation and substitution of additional sugars in the medial and trans-Golgi,

resulting in different types of glycans. N-glycosylation affects not only the protein folding but its integrity and function. For example, IgG are important immune effectors affected by glycosylation, which can modulate their affinity for Fc receptors and their function upon interaction[145–147]. As such, it is important to preserve glycosylation status when developing immune therapeutic strategies, for instance regarding antitumor mAbs development[148]. Mature N-glycans comprise Lewis (Le) blood related antigens and ABH (O) blood group determinant antigens, important in pathogenesis of some diseases and essential for transfusion compatibility[149,150].

In contrast to N-glycosylation, where a complex structure is transferred to a specific protein motif, O-glycosylation involves a step by step addition of sugar residues to characteristic amino acids. The typical O-glycosylation cascade starts with N-acetylgalactosamine (GalNac) transfer from an UDP-GalNac sugar donor to a serine or threonine (Ser/Thr) residue of an acceptor protein. This reaction is catalysed by enzymes of the GalNac transferase (GalNac-T) family, across the golgi apparatus [151,152]. There are several known GalNac-Transferase isoforms with different affinities for a given substrate and tissue specific contexts. Fundamentally, expression of specific GalNac T isoforms on a given cell, are a key regulator of the O-glycoproteome found on glycoproteins of that cell. This glycosylation system is also known as mucin-type O-glycosylation since it is extensively, but not exclusively, found on mucins.

O-GalNac glycans are further elongated to one of four core structures, which can subsequently be extended into a branched or linear mature O-GalNac glycan. The single addition of GalNac to a proteins' Ser/Thr residue establishes the Tn antigen, which can be modified through incorporation of a galactose residue (Gal) by a β1-3 Galactosyltransferase (C1GalT), originating the most common O-GalNac-glycan, T antigen (core 1 or Thomsen-Friedenreich antigen). For C1GalT1 to properly employ its' transferase activity it requires the presence of COSMC1 (C1GALT1C1), an ER chaperone that ensures C1GALT1 folding. Jurkat cells and colon cancer leukemic stem cells (LSCs) lack COSMC1, essential for C1GALT1 activity and therefore are characterized by having higher expression of Tn. Extension into core 2 structures implies addition of β1-6N-acetylglucosamine (GlcNac) to the GalNac residue of core 1 by Core 2 β1,6-*N*-acetylglucosaminyltransferases 1, 2 and 3 (C2GnT or GCNT-1-3), associated with different cell and tissue contexts. Synthesis of Core 3 structures results from incorporation

of GlcNac to Tn antigens in a β1-3 specific manner, promoted by β1-3N-acetylglucosaminyltransferase 6 (C3GnT or B3GNT6). Core 4 structures emerge when C2GnT enzymes promote new addition of β1-6 GlcNac residues to core 3 glycans. All these O-GalNac cores can be further extended to form more complex O-glycans by addition of sugars such as fucose or galactose and addition of sialic acids or sulfate groups, with some these elongation processes originating O-GalNac glycans with ABO or Lewis group determinants (**Figure 3**). Sialylation of O-GalNac glycans occurs by enzymatic activity of sialyltransferases. α 2-6 sialyltransferase (ST6GALNACI) is mainly responsible for sialylation of Tn and T antigens blocking any further elongation into more complex structures. These truncated O-glycans, T, Tn, ST and STn, can be found in increased levels on cancer cells[153–155].

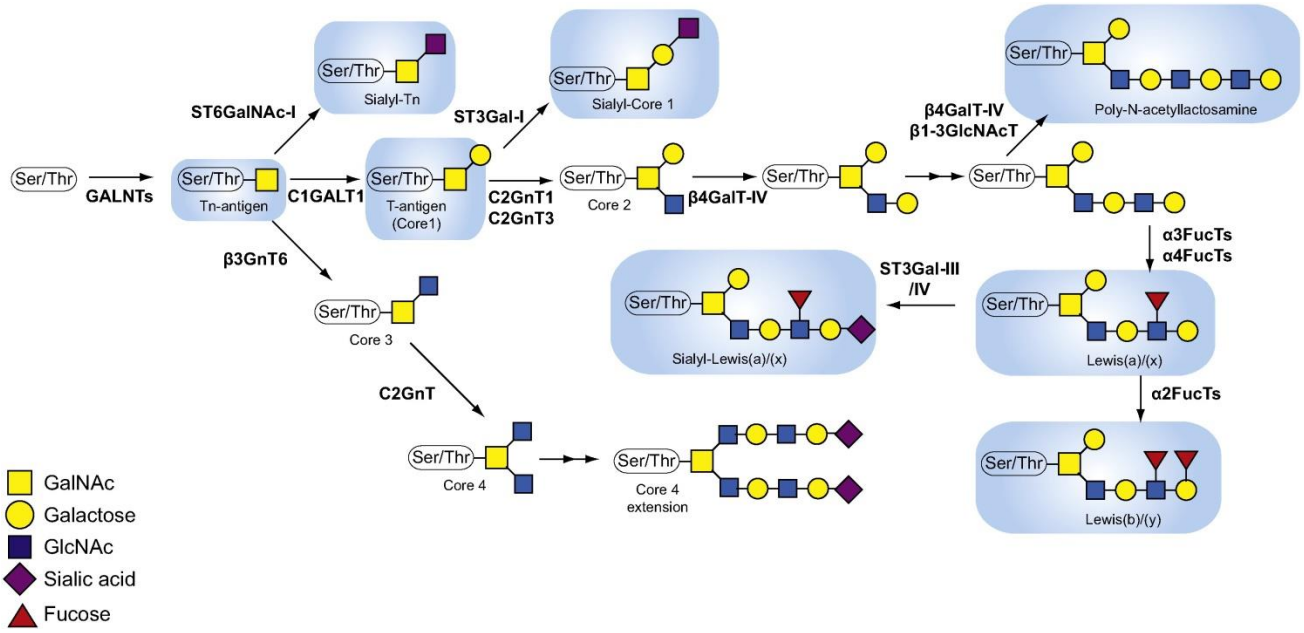


Figure 3. O-GalNAc glycosylation biosynthetic pathway. Cancer-associated structures are highlighted with blue boxes. Figure taken from *Chia J. et al*, *2016, Short O-GalNAc glycans: regulation and role in tumor development and clinical perspectives*

1.3.1. Truncated O-Glycans in Cancer

Alterations in cancer cells are associated with premature stops in O-glycosylation biosynthesis and so display aberrant glycosylation of cell surface proteins, with prevalent expression of T, Tn and STn glycans.

1.3.1.1. Impact of T and Tn antigen Expression

T and Tn antigens are considered hallmarks of epithelial cancers and are associated with alterations in cellular adhesion and metastasis development[156,157]. T antigen has been proposed to influence adhesion and therefore metastasis potential, through interactions with galactins and through MUC1[158,159]. On the other hand, Tn antigen has been reported to promote EMT on colorectal cancer cells, associated with metastasis and invasion[160]. There are several mechanism responsible for these short O-glycan forms: upregulation of sialyltransferases, *COSMC* mutations or modifications, shifts on glycosyltransferase localization or modulations of their activities, pH fluctuations, among others. Radhakrishanan et all, showed that *COMSC* disruption in pancreatic cells results in loss of *C1GALT1* enzymatic activity and increased presence of Tn truncated o-glycans. These events cause enhanced tumour growth and increased invasion in culture and xenografts, with other oncogenic features being established upon COSMC loss[161]. Recently the expression of Tn antigen has also been shown to be associated with immune modulation, specifically with its suppression[162], which could possibly be indicative of a new role for Tn in cancer environment.

In healthy tissues T and Tn antigens are usually not found due to carbohydrate masking, however in disease contexts there are events that promote exposure of these antigens. Virus or bacterial infections show specific enzymes, known as sialidases, that expose these antigens, and in tumour scenarios, there seem to be events that prevent elongation and masking of these antigens. In fact, apart from tumorous tissues, staining for Tn has been found on embryonic nervous tissues[163]. Given the prevalence and role these antigens play in cancer environments and their tissue specificity, there have been several therapy developments addressing the use of these molecules as targets. Additionally, the fact that there are normal human antibodies against T and Tn O-glycans, that can be found in higher concentrations on patients with more benign disease

development[164] , increases the value inhibitory therapeutics against these antigens might have.

1.3.1.2. Sialylated Glycan Interactions as Cancer Promoter Agents

Sialic acids are negatively charged molecules known to modulate recognition and adhesion events. Such is the case for selectin ligands, Calcium dependent lectins, found on leukocytes, endothelium and platelets, which can be functionally altered depending on the level of sialylation. Sialic acids were shown to be essential for lymphocyte homing to the lymph nodes due to their recognition by L-selectins expressed on several leukocytes[165,166]. Additionally, proper selectin ligand sialylation is required for leukocyte adhesion through E-selectin, a phenomenon that occurs upon endothelium activation and mediates leukocyte rolling. The same can be seen for P-selectins on platelets, which mediate leukocyte adhesion in a sialic acid dependent manner. Cancer metastases involve extravasation of tumorous cells from blood to healthy tissues, in which case, selectins are believed to play important roles as extravasation mediators, given their contribution in tethering and rolling interactions. Cancer cells accomplish this feature through altered expression of sialo-carbohydrates, which resemble those on migrating leukocytes. Some evidences have even shown that not only do tumorous cells change in order to bind E-selectins more efficiently; there is an upregulation of E-selectin expression near vascular endothelium, particularly close to metastases sites. Accordingly, transfection of antisense sequences, blocking expression of important selectin ligand sialylated structures, decreases tumorigenicity of colon carcinoma cells and prevents colonization to the liver[167,168]. There is growing awareness regarding the influence microenvironment and its components, like lymphocytes, platelets as well as other immune cells, have on cancer development. These cells usually promote cytokine release which may upregulate selectin expression on endothelium, plus, tumour interaction with these leukocytes and platelets through selectins may endorse tumour resilience and spread[169,170].

Sialic acids are dual molecules that can work as masks, protecting several important receptors from harmful recognition or as specific sites for recognition, as

previously described. These molecules mediate self-recognition events as SAMPs (Self-associated-molecular patterns) through interactions with inhibitory receptors that dampen immune responses[171]. The aforementioned modulations were shown to be mediated by sialic acid-binding to immunoglobulin-like lectins (Siglecs). Siglecs are a family of lectins, found primarily on innate immune cells and as such are associated with modulation of these responses. Siglecs containing ITIM motifs (Immunoreceptor tyrosine-based inhibitory motifs) inhibit activation signals from ITAMs (immunoreceptor tyrosine-based activation motifs) by recruiting either inositol phosphatases or tyrosine phosphatases, known to hinder kinase dependent pathways. Among these ITIM IgG like lectins we can find, siglecs 9, siglecs 1, siglecs 2 (CD22), 3 (CD33), 7 and siglec 10. When activated these receptors mediate inhibitory cascades that may be useful to prevent autoreactive reactions[172], however, in a cancer context they can tilt the odds in favour of tumour prevalence.

Cancer cells display increased levels of sialylated glycans which can interact with these Ig-like receptors to promote receptor inhibition through sialo masking, and immune modulation. There are reports showing inhibition of Natural Killer cells (NKs) either through sialo negative charge repulsions or through sialo-glycan interactions with siglecs 7 and 9, expressed on NKs[173–175]. In other studies, hyper-sialylated tumours were compared with low sialo expressing tumours regarding their influence on T cell populations. T-regs were present in higher amounts in hyper-sialylated tumours, and the absence of NK cells in this context resulted in lack of CD4+ and CD8+ T effector cells[176]. It has been reported that immune system cells, especially myeloid cells, can be modulated by sialic acids to promote cancer development. For example, monocytes are plastic cells that polarize depending on stimuli; sialic acids can interact with siglec 9 on monocyrtes and induce tumour associated macrophages phenotypes (TAMs), shown to increase IL-10 production and decrease the pro-inflammatory cytokine, TNF-α. Accordingly, CD33 was found to enhance expansion of Myeloid-derived suppresser cells (MDSCs), which goes to show how these interaction between siglecs and sialic acids modulate the tumour environments to support cancer thriving.

1.3.2. Short O-glycans in Bladder Cancer

In Bladder Cancer the presence of Sialylated glycans is characteristically associated with recurrence and tumour development[177]. Manifestation of these short STn glycans on over 70 % of higher grade bladder cancers tissues (NMIBCs and MIBCs) has been described and it has been associated with expression of *ST6GalNacI*, believed to be responsible for increased levels of STn glycans in various cancer types[178,179]. The expression patterns show increased STn in non-proliferative areas of the tumour, known to be resistant to treatment agents which increase patient survival. Our group has also shown that STn can be found in metastized ganglia and distant metastasis reflecting a role for STn in cancer motility and mestastization. In MIBC STn expression is a marker of poor prognosis and conjugation with mTOR pathway intervenients might help increase this diagnosis potential, additionally mTOR pathway inhibition was proposed to potentiate therapeutic results for MIBCs[180]. Considering the role Sialylated structures play in immune evasion, STn expression was studied to address its influence on Dendritic Cells (DCs) activation. Dendritic cells are essential for translating innate immunity signals into adaptive immunity actions. DC maturation is associated with loss of phagocytic abilities and upregulation of MHC class I and II molecules, which intermediate antigen presentation to T cells, capable of mounting anti-cancer responses. This maturation state is also characterized by release of specific cytokines and upregulation of co-stimulatory molecules. In cancer contexts DC are usually found in non-mature states, which fail to override tumour cell development and contribute for the creation of tolerogenic and permissive cancer environments. This study showed that cancer cells expressing STn and specific STn antigens, namely, CD44 and MUC1, impact the mature state of DCs and their ability to prime T cells. STn seems to induce immature phenotypes on DCs and prevent normal release of Th1, pro-inflammatory associated cytokines, like TNF-α[181].

Mechanism underlying the role of Tn and T antigen in the context of urothelial cancer have been less evident. However, these molecules have been described as markers for tumour recurrence and phenotypically more aggressive phenotypes[182,183], suggesting a role in distinctive and more hostile BC scenarios.

The absence of truncated short O-glycans in healthy tissues together with their extensively reported implication in cancer settings, particularly in BC, renders these molecules as interesting tumour specific antigens (TSAs) for therapeutical developments. Tackling these molecules with specific antibodies could represent important therapies for BC, however, antibodies addressing Tn and STn in therapeutical contexts are still lacking

and are insufficiently developed to successfully design clinical applications[184]. One recent studied has shown positive indications regarding this topic, identifying and purifying human anti Tn antibodies that recognized several carcinoma tissues[185]. Another alternative already approached, was the formulation of vaccines using truncated O-glycans, which displayed poor results in phase III Clinical trials, possibly due to the lack of immunogenic potential of STn-epitopes and their later described immunogenic suppression[162,181].

1.4. Therapeutical Approaches

1.4.1. Therapeutic alternatives for Cancer Treatment

Regarding cancer therapies several innovative approaches have been explored; from antibody therapies, to immunomodulation, gene therapy and thermal ablation all contribute to higher diversity and widen the array of possible therapeutics, paving the way for complementary personalized treatments.

Conventional Cancer treatment strategies display low specificities for cancer cells and end up damaging healthy cells. Therefore, targeted therapies have been extensively pursued so that cancer treatments can be more effective and less toxic to patients.

Immunotherapies have emerged as possible alternatives to conventional therapies, given their increased specificity and overall reduced toxicity. Among the various therapies available we can find, anti-cancer antibody administration (mAbs), adoptive Cell Transfers (ACT), chimeric antigen receptor T cells (CAR-T cells), Dendritic Cell Therapies, interleukin administrations, among others. Additionally, some of these therapies can be conjugated promoting more efficient results.

Antibodies, whether by themselves or as conjugates, can be used to mark cancer cells leaving them vulnerable to immune system components, or even to inhibit specific cell surface proteins involved in cancer progression. There are several FDA approved anti-cancer antibodies, such as, trastuzumab emtansine, nivolumab or the most widely used biotherapeutic antibody, rituximab. Rituximab is an IgG monoclonal anti-CD20 antibody used as therapy for B-cell malignancies. Its single administration has been shown to induce remission in patients with B-cell lymphomas. However, patients are known to relapse and so a regime of combinatory cytotoxic therapies and rituximab seem

to be the best approach, improving response rates and overall survival. Nonetheless, overtime, a large fraction of patients seem to develop resistance to rituximab based therapies, resulting in lymphoma relapses[186].

T-cells are considered the effectors of the immune system. These cells interact with APCs, namely DCs, and though major histocompatibility complex (MHC-I and MHC-II) interactions, are activated to produce effective responses against the probed antigen. In cancer setting there are mechanisms that hinder T-cell responses, by inhibiting decisive checkpoints like: DC maturation, antigen presentation, T-cell activation/proliferation, among others. This way, CAR T-cell therapies emerge as a means to circumvent these phenomena. CAR T-cells therapies consists on inducing autologous T-cells *ex vivo* with molecular features capable of intersecting tumour cells. These types of therapies seem to be gaining ground and have already been used in clinical practice with enthusiastic results. Patients with end-stage acute Lymphocytic Leukaemia were treated with CAR T-cells which led to recovery of up to 92% of patients[187]. Despite these promising results, treatments with CAR T-cells are highly effective in hematological tumours but are harder to wire for treatment of solid tumours. These tumours display increased genetic instability and are characterized by immunomodulatory environments that compromise CAR T-cell trafficking and effectiveness. Of note, the long term effects of CAR T-cells are still not fully disclosed and as such, further analysis needs to be conducted in order to properly address the fate and role these cells might have.

1.4.2 Nanoparticles in Clinical Tumour Settings

As previously mentioned conventional cancer therapies fail to focus their effects solely on cancer cells, and so, nanomedicine as emerged as a means to explore new vehicles for cancer localized antigen immunization or cytotoxic delivery. Nanoparticles are small systems with physicochemical properties that allow for targeted and effective therapeutics. There are several formulations one can chose depending on the goal in mind. These targeted therapies can employ organic nanoparticles, displaying a wide array of construct options, from liposomal formulations with confirmed clinical relevance[188], to polymer based strategies or a combinations of both. Inorganic nanoparticles, such as gold nanoparticles[189], or quantum dots have also displayed high applicability, particularly for imaging purposes.

Nanoparticles can accumulate in tumour tissues in a passive or active way. According to their small size and surface charge, nanoparticles utilize the concept of enhanced permeability and retention (EPR) effect, and passively accumulate in tumour tissues. The leaky and loosely vasculature of tumour sites facilitates extravasation of nanoparticles into, which cannot be seen in normal tissues with tight and less permissive blood vessels. On top of that, tumour settings lack efficient lymphatic drainage which allows for longer retention periods of these particles.

Active targeting can enhance the uptake of nanoparticles by specific cells through association with surface ligands, which will promote uptake by binding membrane receptors on targeted cells. These ligands can be antibodies, peptides or whole proteins that will bias the interactions and outcomes of nanoparticles based therapies.

Therapies using nanoparticles as delivery systems benefit from controlled drug/antigen release, and protect these compounds from immediate enzymatic or chemical degradation and renal excretion, at the same time protecting the host from off target effects, due to their targeted specificity.

Characterization of nanoparticles is crucial for understanding their physical and chemical properties, which affect load-bearing amounts, interactions with cellular components and overall particle applicability in a given setting. As such, many parameters need to be evaluated when formulating nanoparticles, such as, size and structural properties, shape, surface charge (by zeta potentials), growth kinetics, agglomeration states, among others. Also, depending on the type of formulation, different techniques may be more appropriate for the evaluation of these singatures[190].

1.4.2.1 PLGA Nanoparticles

Poly (lactic-co-glycolic acid (PLGA) nanoparticles are one of the most widely used materials for nanoconstruct formulation. Given their biocompatibility and biodegradability, they have been recognized by the FDA, and are extensively sought as carriers for several clinical applications. PLGA nanoparticles can be modulated to act on different therapy settings by conjugation of its functional groups with diverse molecules, such as, antibodies, drugs, targeting ligands for immune enhancement, or fluorophores for imaging purposes. Additionally, one can alter the composition of PLGA based nanoparticles and further modulate drug loading and encapsulation, release rates, cell

interactions and others. One example are PEG functionalized PLGA nanoparticles, which acquire hydrophilic characteristics that warrant evasion from clearance systems, thus increasing particle circulation periods. Furthermore, the methods employed for nanoparticle formulation, can also extensively influence PLGA NPs characteristics[191].

The wide variety of possible polymer formulations demonstrates the plasticity of PLGA based nanoparticles for clinical purposes; however, it also entails several optimization steps and rigorous evaluation of particle characteristics.

PLGA nanoparticles conjugated to Bombesin peptide, loaded with docetaxel were found to successfully target gastrin-releasing peptide receptor, found overexpressed in several cancer cells (such as gastric, ovarian and breast cancers)[192,193]. PLGA has also been demonstrated to encapsulate drugs like, cisplatin, doxorubicin, or paclitaxel and induce antitumor effects *in vitro* and *in vivo*[194]. PLGA constructs can include antibodies for precision targeting. Particles loaded with Pseudomonas exotoxin A (PE)-based immunotoxins and conjugated with antibody fragments for HER 2 (Epidermal growth factor receptor - extensively described in breast tumours), were successfully administered to breast cancer cells and displayed potent cytotoxicity responses. However, *in vivo*, these particles displayed lower immunogenicity for PE when compared to single PE-HER antibody induction and were well tolerated in mice[195].

Lectin and Carbohydrate interactions are very specific and so can be used to modulate nanoparticle interactions, especially given their described role in tumour environments. *Mo et al,* formulated PLGA particles conjugated with wheat germ agglutinin (WGA) loaded with paclitaxel, design to target lung cancer cells specifically. These nanoparticles displayed higher cytotoxic responses when compared to controls without WGA, associated with higher particle uptake[196]. Inversely PLGA formulation can carry carbohydrates, with emerging interest in the development of nanoparticles with sialic acids at their surface. *Bondioli et al*, was able to create PLGA nanoparticles coated with N-acetylneuraminic acid (Neu5AC), however these particles were phagocytosed by monocytes, clarifying the need for particle optimization, with hypothesized improvement with use of hydrophilic polymers[197]. In this drug delivery context, phagocytosis of nanoparticles inhibits the therapeutic potential of such approaches; however, it can prove to be an advantage in different settings, such as nanovaccine formulations.

1.4.2.2. Overview of cancer therapeutic vaccines

Cancer vaccines envisage the mounting of robust immune responses that may induce tumour remission. There are several types of approaches for this end considering the vast diversity of immune cells, and their immunomodulation potential.

As previously mentioned, T-cell effectors are finely controlled by DCs. In cancer environments it has been reported that these APCs are induced to display immature and tolerogenic features, hindering their capacity to mount effective immune responses against cancer cells. This way, DCs have been evaluated for their potential in cancer therapies. Vaccines targeting these cells can be based on *ex vivo* DC stimulation and re-introduction into the host, potentiating individual and specific patient therapies. On the other hand, DCs can be modified by administration of specific ligands with stimulatory effects. DC stimulation can be achieved by introducing TLR ligands into vaccines formulations, by using co-stimulatory molecules (CD40L, CD80L and CD86L) or through peptide/protein administration of tumour antigens (TAAs or TSAs)[198]. The use of tumour antigens usually requires inclusion of immunostimulants to produce effective responses; however, it can enable immune memory and potentiate long lasting anti-cancer responses, reducing the need for long costly treatments[199]. Neoantigens are TSAs (Tumour specific antigens), absent from healthy tissues but characteristic of tumour tissues, usually associated with high mutational loads. These neoantigens have been associated with immune escape and immune checkpoint inhibition and show promising results in clinical settings, as reviewed by *Hollingsworth et al, 2019*. Neoantigens are usually considered patient specific, given that, for each individual, tumorous environments promote different sets of mutational profiles. However, there are common oncogenic mutations that can produce neoantigens largely found on cancer specific tissues, and so these neoantigens hold enormous potential for future therapies[199].

1.4.2.3. PLGA based Nanovaccines

PLGA has also been used to formulate nano vaccine constructs. The rational for these therapies involves targeting immune cells with specific cancer antigens, consequently potentiating robust and efficient immune responses. Inclusively, these constructs containing PLGA and cancer specific antigens can be coupled to immunogenic molecules that enhance immunological responses, by activating DCs, macorphages or NKs. Employment of nanovaccines reduces adjuvant dosage and subsequently the prices for vaccine construction, while maintaining long lasting immunity[200,201], and minimizing adjuvant associated toxicity. Furthermore, the gradual release that encapsulated antigens (proteins/peptides) acquire when associated to PLGA polymers, facilitates the induction of immune responses.

Initial interactions between NPs with targeted cells involve particle uptake, usually through endocytosis. This internalization mechanism is generally associated with lysosomal fusion, which promotes particle degradation by lowering the compartments pH and by enzymatic activity of lysosomal proteases, leading to antigen presentation via MHC II. On the other hand, there are other pathways for NP internalization that can promote antigen presentation via MHC I. Caveolin-mediated endocytosis (CVME) and macropinocytosis are alternative internalization pathways described by authors to be more suitable for antigen presentation via MHC I, given their propensity to release cargo into the cytoplasm without lysosomal degradation[202]. The uptake of PLGA particles can be favoured by positively charged MPs/NPs that increase interactions between negative cell surface of immune cells and cationic particles. Furthermore, NP surface charge can also influence the uptake pathway [202]. The immune responses promoted upon NP internalization, can additionally be influenced by particle size. *Kanchan and Panda* demonstrated that nano and microparticles (NPs and MPs) elicit different immune responses: while MPs seem to activate Th2 responses, promoting release of IL-4 and upregulation of MHC II; NPs induce higher release of IFN-γ and upregulation of MHC I, with less antibody titters reported [203]. Also, NPs seem to display preferable uptake by DCs whereas MPs are more commonly uptaken by macrophages[204,205]. There is a general consensus that nanoparticles smaller than 200 nm are better for DC uptake and preferably induce CD8+ T-cell responses[204]. PLGA nanoparticles can be further modulated to induce specific T cell responses, by creating pH sensitive particles, which facilitate or hinder cargo release in APC lysosomal environments, upon particle

internalization (**Figure 4**). Coupling NPs with specific immunostimulants (TLR ligands, or bacteria toxins like CTB) can also bias the type of immune response elicited.

It is not surprising that PLGA nanoparticles are so sought out for medical applications, specifically in heterogeneous and dynamic systems like cancer. The high versatility and easy manipulation can rapidly produce adaptable nanovehicles for cancer vaccines.

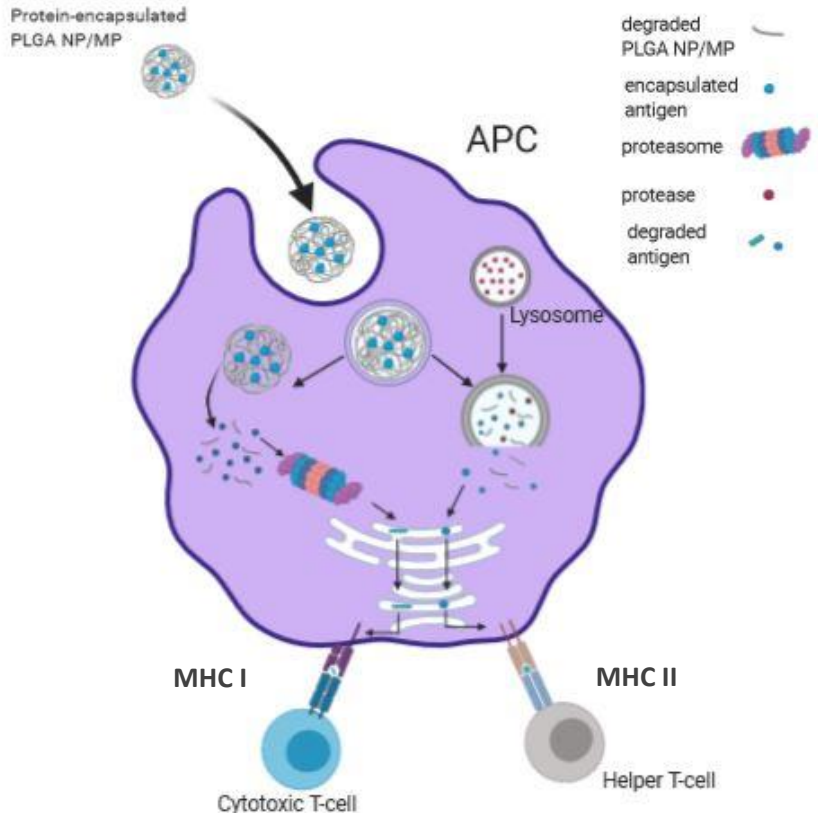


Figure 4.Schematic representation of intracellular trafficking of protein/peptide - loaded PLGA particles in antigen presenting cells (mostly dendritic cell). Desired proteins are encapsulated into PLGA particles and internalized by APCs. If early endosomes fuse with lysosomes than the peptides are degraded by lysosomal proteases and are later presented to T-cells via MHC II. If NPs/MPs escape early endosomes before lysosomal fusion PLGA particles release the encapsulated antigen into the cytoplasm, and via proteasome degradation, antigens are presented to T cytotoxic cells via MHC I. (Figure created with BioRender.com)

II. AIMS AND SCOPES

Bladder Cancer remains one of the deadliest genitourinary cancers due to significant intra and inter-tumoral molecular heterogeneity, which constitutes a significant challenge for the introduction of effective therapies. Immunotherapy has long been introduced in bladder cancer clinical practice with promising results, delaying and even avoiding disease progression and dissemination. Namely, intravesical instillations with BCG were shown to effectively mount local immune responses that avoid the relapse and progression of superficial bladder tumours, whereas the recent introduction of immune check-point inhibitors has provided novel tools to address more aggressive metastatic forms of the disease. Nevertheless, improvements in survival are still modest and substantial acquired resistance as well as side effects support the additional investment in novel, safer and more effective immunotherapies.

Cancer vaccines exploiting cancer neoantigens have emerged as a powerful and safer cancer therapy in pre-clinical studies and early phase clinical trials. Moreover, cancer vaccines were able to induce strong protective immunological memory against cancer cells in relevant animal models, suggesting that it may hold potential to avoid relapse and control disease dissemination in clinical settings. Recently, nano-encapsulation of neoantigens coupled with immunostimulatory molecules in biocompatible polymers, such as PLGA, has contributed to enhance therapeutic efficacy. Nanoencapsulation has been particularly effective in protecting antigen payloads from degradation in circulation and in increasing its uptake by antigen presenting cells, especially after grafting the surface of the nanoparticle with immune cells receptor agonists. However, the most critical bottleneck for effective cancer vaccines remains the precise identification of neoantigens (i.e. foreigner molecular signatures present only in cancer cells and absent from normal tissues).

Envisaging the identification of cancer neoantigens, our group has been exploiting the altered glycosylation of proteins at the surface of more aggressive cancer cells. We have devoted particular emphasis to glycoproteins carrying short-chain O-glycans such as the Tn and STn antigens, which are rarely observed in healthy tissues, with few exceptions in specialized secretory cells facing the lumen of the gastrointestinal and respiratory tracts. On the other hand, these glycans are abundantly expressed in cancer tissues, circulating tumour cells and metastasis, and are generally associated with worst prognosis. Moreover, short-chain O-glycans play a functional role in the establishment of main cancer hallmarks, with emphasis on enhanced invasive capacity and immune

escape. Using targeted glycoproteomics, our group has been attempting to characterize glycoproteins involved in cancer progression as well as specific glycosites carrying these post-translational modifications for cancer bispecificity. This led to the identification of several MUC16 and CD44 Tn/STn-glycoproteoforms expressed solely in cancer cells and not the healthy urothelium and, to our knowledge, other healthy organs.

Therefore, this thesis main objective is to create the molecular foundations for bladder cancer nanovaccines exploiting glycoproteomics data generated by our research group. The main goal is to create nanovaccine constructs for future pre-clinical validation. Cancer-specific Tn and STn glycopeptides derived from MUC16 and CD44, which are relevant glycoproteins associated with bladder cancer aggressiveness, will be used as antigens towards this end. This encompasses the following specific objectives:

1. Establishment of chemoenzymatic methods for the synthesis of MUC16 and CD44 cancer glycoepitopes;
2. Development of biocompatible PLGA nanovaccines carrying immunogenic glycoantigens payloads;
3. Establishment of glycoengineered cell factories for stable production of glycoantigens expressing homogenous cancer-associated glycans.

The thesis includes three research chapters that respond to these objectives. Chapter 1 describes a simple chemoenzymatic method for the generation of CD44 and MUC16 glycoepitope libraries, carrying different types of short-chain O-glycans. It also uses liquid chromatography coupled to tandem mass spectrometry to provide a thorough characterization of the synthesis products, including precise annotation of occupied glycosites. Chapter 2 describes the preparation of immunogenic constructs resulting from the engraftment of KLH to MUC16-Tn/STn glycopeptides, envisaging the assembly of a multivalent vaccine. It also addresses the potential of nanoencapsulation to improve the efficacy of nanovaccines, providing a physico-chemical characterization of these constructs, and their stability at different pHs, thus establishing novel formulations for future tests in vitro and in vivo. The last chapter provides an alternative approach to potentially scale up glycoepitope production based on bladder cancer cell factories expressing relevant CD44-STn glycosignatures associated with aggressiveness but not reflected in healthy tissues. It describes the process of glycoengineering T24 bladder cancer cells to express simple cell STn glycophenotypes with minor implications on CD44 expression patterns compared to wild type cells. Our strategy involved using CRISPR-Cas9 to knock-out C1GALT1, a key enzyme involved in O-glycan extension. The cells were also led to overexpress ST6GALNAC1, guarantying high yields of Tn antigen

sialylation. The maintenance of the CD44 expression pattern was confirmed by RT-PCR for most relevant CD44 variants in this cell line and the glycosylation of CD44 with STn was determined by in situ proximity ligation assays.

Bibliography

1. International Agency for Research on Cancer. Bladder Globocan 2018. *Globocan* 1–2 (2019).
2. Globocan Observatory, W. Cancer Today - World. *Int. Agency Res. Cancer* **876**, 2018–2019 (2019).
3. Witjes, J. A. *et al.* EAU guidelines on muscle-invasive and metastatic bladder cancer: Summary of the 2013 guidelines. *Eur. Urol.* **65**, 778–792 (2014).
4. Compérat, E. M. *et al.* Grading of Urothelial Carcinoma and The New "World Health Organisation Classification of Tumours of the Urinary System and Male Genital Organs 2016". *Eur. Urol. Focus* **5**, 457–466 (2019).
5. Xylinas, E. *et al.* Urine markers for detection and surveillance of bladder cancer. *Urol. Oncol. Semin. Orig. Investig.* **32**, 222–229 (2014).
6. van Rhijn, B. W. G. *et al.* Recurrence and Progression of Disease in Non-Muscle-Invasive Bladder Cancer: From Epidemiology to Treatment Strategy. *Eur. Urol.* **56**, 430–442 (2009).
7. Batista, R. *et al.* TERT promoter mutation as a potential predictive biomarker in BCG-treated bladder cancer patients. *Int. J. Mol. Sci.* **21**, (2020).
8. Kimura, F. *et al.* Destabilization of chromosome 9 in transitional cell carcinoma of the urinary bladder. *Br. J. Cancer* **85**, 1887–1893 (2001).
9. Knowles, M. A. & Hurst, C. D. Molecular biology of bladder cancer: New insights into pathogenesis and clinical diversity. *Nat. Rev. Cancer* **15**, 25–41 (2015).
10. Claude Billerey,* Dominique Chopin,†Marie-He ́le` ne Aubriot-Lorton,‡David Ricol,‡Sixtina Gil Diez de Medina,†‡Bas Van Rhijn,§Marie-Pierre Bralet, Marie-Aude Lefrere-Belda, Jean-Baptiste Lahaye,†‡Claude C. Abbou,†Jacky Bonaventure, iElie Serge Zafrani, Th, andFrancois R. Frequent FGFR3 Mutations in Papillary Non-Invasive Bladder (pTa) Tumors. *Am. J. ofPathology, Vol. 158, No. 6, June 2001 Copyr. © Am. Soc. Investig. Pathol. Short* **Vol. 158**, 95–97 (2001).
11. Mason, R. A. *et al.* EGFR pathway polymorphisms and bladder cancer susceptibility and prognosis. *Carcinogenesis* **30**, 1155–1160 (2009).
12. Christensen, C. H. *et al.* Association of cigarette, cigar, and pipe use with mortality risk in the US population. *JAMA Intern. Med.* **178**, 469–476 (2018).
13. Burger, M. *et al.* Epidemiology and risk factors of urothelial bladder cancer. *Eur. Urol.* **63**, 234–241 (2013).
14. Nilsson, S. & Ullén, A. Chemotherapy-induced bladder cancer. *Scand. J. Urol. Nephrol.* **42**, 89–92 (2008).
15. Monach, P. A., Arnold, L. M. & Merkel, P. A. Incidence and prevention of bladder toxicity from cyclophosphamide in the treatment of rheumatic diseases: A data-driven review. *Arthritis Rheum.* **62**, 9–21 (2010).
16. Suriano, F., Altobelli, E., Sergi, F. & Buscarini, M. Bladder cancer after radiotherapy for prostate cancer. *Rev. Urol.* **15**, 108–12 (2013).
17. Sui, X., Lei, L., Chen, L., Xie, T. & Li, X. Inflammatory microenvironment in the initiation and progression of bladder cancer. *Oncotarget* **8**, 93279–93294 (2017).
18. Pezzuto, A., Citarella, F., Croghan, I. & Tonini, G. The effects of cigarette smoking extracts on cell cycle and tumor spread: Novel evidence. *Futur. Sci. OA* **5**, (2019).
19. Gu, J., Liang, D., Wang, Y., Lu, C. & Wu, X. Effects of N-acetyl transferase 1 and 2 polymorphisms on bladder cancer risk in Caucasians. *Mutat. Res. - Genet. Toxicol. Environ. Mutagen.* **581**, 97–104 (2005).
20. Chang, S. S. *et al.* Diagnosis and Treatment of Non-Muscle Invasive Bladder Cancer: AUA/SUO Guideline. *J. Urol.* **196**, 1021–1029 (2016).
21. Sylvester, R. J. *et al.* Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: A combined analysis of

2596 patients from seven EORTC trials. *Eur. Urol.* **49**, 466–477 (2006).
22. Bosschieter, J. *et al.* Value of an Immediate Intravesical Instillation of Mitomycin C in Patients with Non–muscle-invasive Bladder Cancer: A Prospective Multicentre Randomised Study in 2243 patients [Figure presented]. *Eur. Urol.* **73**, 226–232 (2018).
23. Böhle, A., Jocham, D. & Bock, P. R. Intravesical bacillus Calmette-Guerin versus mitomycin C for superficial bladder cancer: A formal meta-analysis of comparative studies on recurrence and toxicity. *J. Urol.* **169**, 90–95 (2003).
24. Jarow, J. P. *et al.* Clinical trial design for the development of new therapies for nonmuscle-invasive bladder cancer: Report of a food and drug administration and american urological association public workshop. *Urology* **83**, 262–265 (2014).
25. Babjuk, M. *et al.* EAU Guidelines on Non–Muscle-invasive Urothelial Carcinoma of the Bladder: Update 2016. *Eur. Urol.* **71**, 447–461 (2017).
26. Chang, S. S. *et al.* Treatment of Non-Metastatic Muscle-Invasive Bladder Cancer: AUA/ASCO/ASTRO/SUO Guideline. *J. Urol.* **198**, 552–559 (2017).
27. Grossman, H. B. *et al.* Neoadjuvant chemotherapy plus cystectomy compared with cystectomy alone for locally advanced bladder cancer. *N. Engl. J. Med.* **349**, 859–866 (2003).
28. Karam, J. A. & Kamat, A. M. Optimal timing of chemotherapy and cystectomy. *F1000 Med. Rep.* **2**, 23–26 (2010).
29. Calabrò, F. & Sternberg, C. N. State-of-the-art management of metastatic disease at initial presentation or recurrence. *World J. Urol.* **24**, 543–556 (2006).
30. Study, P. I. I. I. *et al.* Metastatic Bladder Cancer : Results of a Large ,. *Society* **17**, 3068–3077 (2000).
31. Von Der Maase, H. *et al.* Long-term survival results of a randomized trial comparing gemcitabine plus cisplatin, with methotrexate, vinblastine, doxorubicin, plus cisplatin in patients with bladder cancer. *J. Clin. Oncol.* **23**, 4602–4608 (2005).
32. Robinson, A. G., Wei, X., Vera-Badillo, F. E., Mackillop, W. J. & Booth, C. M. Palliative Chemotherapy for Bladder Cancer: Treatment Delivery and Outcomes in the General Population. *Clin. Genitourin. Cancer* **15**, e535–e541 (2017).
33. Resch, I., Shariat, S. F. & Gust, K. M. PD-1 and PD-L1 inhibitors after platinum-based chemotherapy or in first-line therapy in cisplatin-ineligible patients: Dramatic improvement of prognosis and overall survival after decades of hopelessness in patients with metastatic urothelial cancer. *Memo - Mag. Eur. Med. Oncol.* **11**, 43–46 (2018).
34. Rosenberg, J. E. *et al.* Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: A single-arm, multicentre, phase 2 trial. *Lancet* **387**, 1909–1920 (2016).
35. TECENTRIQ (atezolizumab) PI. **21**, 1–9 (2020).
36. O'Donnell, J. S., Long, G. V., Scolyer, R. A., Teng, M. W. L. & Smyth, M. J. Resistance to PD1/PDL1 checkpoint inhibition. *Cancer Treat. Rev.* **52**, 71–81 (2017).
37. Einstein, D. J. & Sonpavde, G. Treatment Approaches for Cisplatin-Ineligible Patients with Invasive Bladder Cancer. *Curr. Treat. Options Oncol.* **20**, (2019).
38. Ma, G. *et al.* Precision medicine and bladder cancer heterogeneity. *Bull. Cancer* **105**, 925–931 (2018).
39. Meeks, J. J. *et al.* Genomic heterogeneity in bladder cancer: challenges and possible solutions to improve outcomes. *Nat. Rev. Urol.* **17**, 259–270 (2020).
40. Dean, M. ABC transporters, drug resistance, and cancer stem cells. *J. Mammary Gland Biol. Neoplasia* **14**, 3–9 (2009).
41. Ayob, A. Z. & Ramasamy, T. S. Cancer stem cells as key drivers of tumour progression. *J. Biomed. Sci.* **25**, 1–18 (2018).
42. Abugomaa, A., Elbadawy, M., Yamawaki, H., Usui, T. & Sasaki, K. Emerging Roles

of Cancer Stem Cells in Bladder Cancer Progression, Tumorigenesis, and Resistance to Chemotherapy: A Potential Therapeutic Target for Bladder Cancer. *Cells* **9**, 235 (2020).
43. Gaisa, N. T. *et al.* The human urothelium consists of multiple clonal units, each maintained by a stem cell. *J. Pathol.* **225**, 163–171 (2011).
44. Chan, K. S. *et al.* Identification, molecular characterization, clinical prognosis, and therapeutic targeting of human bladder tumor-initiating cells. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 14016–14021 (2009).
45. Takaishi, S. *et al.* Identification of gastric cancer stem cells using the cell surface marker CD44. *Stem Cells* **27**, 1006–1020 (2009).
46. Dalerba, P. *et al.* Phenotypic characterization of human colorectal cancer stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 10158–10163 (2007).
47. Fujii, T., Shimada, K., Anai, S., Fujimoto, K. & Konishi, N. MUC1 expression. **104**, 321–327 (2013).
48. Matte, I., Albert, A., Boivin, M., Beaudin, J. & Piche, A. Downregulation of cell surface CA125 / MUC16 induces epithelial-to-mesenchymal transition and restores EGFR signalling in NIH : OVCAR3 ovarian carcinoma cells. 989–999 (2011) doi:10.1038/bjc.2011.34.
49. Stroopinsky, D. *et al.* MUC1 Is a Potential Target for the Treatment of Acute Myeloid Leukemia Stem Cells. 5569–5580 (2013) doi:10.1158/0008-5472.CAN-13-0677.
50. Lakshmanan, I. *et al.* MUC16 regulates TSPYL5 for lung cancer cell growth and chemoresistance by suppressing p53. *Clin. Cancer Res.* **23**, 3906–3917 (2017).
51. Cotton, S. *et al.* Targeted O-glycoproteomics explored increased sialylation and identified MUC16 as a poor prognosis biomarker in advanced-stage bladder tumours. *Mol. Oncol.* **11**, 895–912 (2017).
52. Dalchau, R., Kirkley, J. & Fabre, J. W. Monoclonal antibody to a human leukocyte-specific membrane glycoprotein probably homologous to the leukocyte-common (L-C) antigen of the rat. *Eur. J. Immunol.* **10**, 737–744 (1980).
53. Johnson, P. & Ruffell, B. CD44 and its role in inflammation and inflammatory diseases. *Inflamm. Allergy - Drug Targets* **8**, 208–220 (2009).
54. Naor, D., Nedvetzki, S., Golan, I., Melnik, L. & Faitelson, Y. CD44 in cancer. *Crit. Rev. Clin. Lab. Sci.* **39**, 527–579 (2002).
55. Screaton, G. R. *et al.* Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 12160–12164 (1992).
56. UNDERHILL, C. CD44: The hyaluronan receptor. *J. Cell Sci. 103, 293-298* **298**, 83–123 (1992).
57. Williams, K., Motiani, K., Giridhar, P. V. & Kasper, S. CD44 integrates signaling in normal stem cell, cancer stem cell and (pre)metastatic niches. *Exp. Biol. Med.* **238**, 324–338 (2013).
58. Turley, E. A., Noble, P. W. & Bourguignon, L. Y. W. Signaling properties of hyaluronan receptors. *J. Biol. Chem.* **277**, 4589–4592 (2002).
59. Mrass, P. *et al.* CD44 Mediates Successful Interstitial Navigation by Killer T Cells and Enables Efficient Antitumor Immunity. *Immunity* **29**, 971–985 (2008).
60. Thorne, R. F., Legg, J. W. & Isacke, C. M. The role of the CD44 transmembrane and cytoplasmic domains in co-ordinating adhesive and signalling events. *J. Cell Sci.* **117**, 373–380 (2004).
61. Okamoto, I. *et al.* Proteolytic release of CD44 intracellular domain and its role in the CD44 signaling pathway. *J. Cell Biol.* **155**, 755–762 (2001).
62. Cichy, J. & Puré, E. The liberation of CD44. *J. Cell Biol.* **161**, 839–843 (2003).
63. Harrison, G. M., Davies, G., Martin, T. A., Mason, M. D. & Jiang, W. G. The influence of CD44v3-v10 on adhesion, invasion and MMP-14 expression in prostate cancer cells. *Oncol. Rep.* **15**, 199–206 (2006).

64. Kugelman, L. C., Ganguly, S., Haggerty, J. G., Weissman, S. M. & Milstone, L. M. The core protein of epican, a heparan sulfate proteoglycan on keratinocytes, is an alternative form of CD44. *J. Invest. Dermatol.* **99**, 381–385 (1992).
65. Zhou, J., Haggerty, J. G. & Milstone, L. M. Growth and differentiation regulate CD44 expression on human keratinocytes. *Vitr. Cell. Dev. Biol. - Anim.* **35**, 228–235 (1999).
66. Milstone, L. M., Hough-Monroe, L., Kugelman, L. C., Bender, J. R. & Haggerty, J. G. Epican, a heparan/chondroitin sulfate proteoglycan form of CD44, mediates cell-cell adhesion. *J. Cell Sci.* **107**, 3183–3190 (1994).
67. Hagiwara, M. *et al.* Variant isoforms of CD44 involves acquisition of chemoresistance to cisplatin and has potential as a novel indicator for identifying a cisplatin-resistant population in urothelial cancer. *BMC Cancer* **18**, 1–11 (2018).
68. Zeng, Y. *et al.* Stress-response protein RBM3 attenuates the stem-like properties of prostate cancer cells by interfering with CD44 variant splicing. *Cancer Res.* **73**, 4123–4133 (2013).
69. Bartolazzi, A. *et al.* Regulation of growth and dissemination of a human lymphoma by CD44 splice variants. *J. Cell Sci.* **108**, 1723–1733 (1995).
70. Camenisch, T. D. *et al.* Camenisch Has2 JClinInvest2001. **106**, 1–12 (2000).
71. Misra, S. *et al.* Hyaluronan-CD44 interactions as potential targets for cancer therapy. *FEBS J.* **278**, 1429–1443 (2011).
72. Weigel, P. H., Hascall, V. C. & Tammi, M. Hyaluronan synthases. *J. Biol. Chem.* **272**, 13997–14000 (1997).
73. Toole, B. P. Hyaluronan: From extracellular glue to pericellular cue. *Nat. Rev. Cancer* **4**, 528–539 (2004).
74. Lee, J. Y. & Spicer, A. P. Functions Biophysical and structural ECM organizer. *Curr. Opin. Cell Biol.* **12**, 581–586 (2000).
75. Itano, N. *et al.* Three isoforms of mammalian hyaluronan synthases have distinct enzymatic properties. *J. Biol. Chem.* **274**, 25085–25092 (1999).
76. Toole, B., Ghatak, S. & Misra, S. Hyaluronan Oligosaccharides as a Potential Anticancer Therapeutic. *Curr. Pharm. Biotechnol.* **9**, 249–252 (2008).
77. Ghatak, S., Misra, S. & Toole, B. P. Hyaluronan oligosaccharides inhibit anchorage-independent growth of tumor cells by suppressing the phosphoinositide 3-kinase/Akt cell survival pathway. *J. Biol. Chem.* **277**, 38013–38020 (2002).
78. Skelton, T. P., Zeng, C., Nocks, A. & Stamenkovic, I. Glycosylation provides both stimulatory and inhibitory effects on cell surface and soluble CD44 binding to hyaluronan. *J. Cell Biol.* **140**, 431–446 (1998).
79. Bourguignon, L. Y. W., Xia, W. & Wong, G. Hyaluronan-mediated CD44 interaction with p300 and SIRT1 regulates β-catenin signaling and NFκB-specific transcription activity leading to MDR1 and Bcl-xL gene expression and chemoresistance in breast tumor cells. *J. Biol. Chem.* **284**, 2657–2671 (2009).
80. Sugahara, K. N. *et al.* Tumor cells enhance their own CD44 cleavage and motility by generating hyaluronan fragments. *J. Biol. Chem.* **281**, 5861–5868 (2006).
81. Golshani, R. *et al.* Hyaluronic acid synthase-1 expression regulates bladder cancer growth, invasion, and angiogenesis through CD44. *Cancer Res.* **68**, 483–491 (2008).
82. Toole, B. P. Hyaluronan in morphogenesis. *Semin. Cell Dev. Biol.* **12**, 79–87 (2001).
83. Lesley, J., Hyman, R., English, N., Catterall, J. B. & Turner, G. A. CD44 in inflammation and metastasis. *Glycoconj. J.* **14**, 611–622 (1997).
84. English, N. M., Lesley, J. F. & Hyman, R. Site-specific de-N-glycosylation of CD44 can activate hyaluronan binding, and CD44 activation states show distinct threshold densities for hyaluronan binding. *Cancer Res.* **58**, 3736–3742 (1998).
85. Bartolazzi, A., Nocks, A., Aruffo, A., Spring, F. & Stamenkovic, I. Glycosylation of CD44 is implicated in CD44-mediated cell adhesion to hyaluronan. *J. Cell Biol.* **132**,

1199–1208 (1996).
86. Catterall, J. B., Jones, L. M. H. & Turner, G. A. Membrane protein glycosylation and CD44 content in the adhesion of human ovarian cancer cells to hyaluronan. *Clin. Exp. Metastasis* **17**, 583–591 (1999).
87. Katoh, S., Zheng, Z., Oritani, K., Shimozato, T. & Kincade, P. W. Glycosylation of CD44 negatively regulates its recognition of hyaluronan. *J. Exp. Med.* **182**, 419–429 (1995).
88. Lesley, J., English, N., Perschl, A., Gregoroff, J. & Hyman, R. Variant cell lines selected for alterations in the function of the hyaluronan receptor CD44 show differences in glycosylation. *J. Exp. Med.* **182**, 431–437 (1995).
89. Bennett, K. L. *et al.* Regulation of CD44 binding to hyaluronan by glycosylation of variably spliced exons. *J. Cell Biol.* **131**, 1623–1633 (1995).
90. Mereiter, S. *et al.* O-glycan truncation enhances cancer-related functions of CD44 in gastric cancer. *FEBS Lett.* **593**, 1675–1689 (2019).
91. Weber, G. F., Ashkar, S., Glimcher, M. J. & Cantor, H. Receptor-ligand interaction between CD44 and osteopontin (Eta-1). *Science (80-.).* **271**, 509–512 (1996).
92. Tuck, A. B., Chambers, A. F. & Allan, A. L. Osteopontin overexpression in breast cancer: Knowledge gained and possible implications for clinical management. *J. Cell. Biochem.* **102**, 859–868 (2007).
93. Qiu, Y. *et al.* Genetic association of osteopontin (OPN) and its receptor CD44 genes with susceptibility to Chinese gastric cancer patients. *J. Cancer Res. Clin. Oncol.* **140**, 2143–2156 (2014).
94. Alexander Pietras1, 2, 3, 4, Amanda M. Katz3, 4, Elin J. Ekström5, Boyoung Wee3, 4, J. J., Halliday3, 4, Kenneth L. Pitter3, 4, Jillian L. Werbeck3, 4, Nduka M. Amankulor6, J. T. & Huse4, 7, 8, and Eric C. Holland1, 2. Osteopontin-CD44 signaling in the glioma perivascular niche enhances cancer stem cell phenotypes and promotes aggressive tumor growth. *Cell Stem Cell. 2014 March 6; 14(3) 357–369.* (2014) doi:doi:10.1016/j.stem.2014.01.005.
95. Rao, G. *et al.* Reciprocal interactions between tumor-associated macrophages and CD44-positive cancer cells via osteopontin/CD44 promote tumorigenicity in colorectal cancer. *Clin. Cancer Res.* **19**, 785–797 (2013).
96. Fedorchenko, O. *et al.* CD44 regulates the apoptotic response and promotes disease development in chronic lymphocytic leukemia. *Blood* **121**, 4126–4136 (2013).
97. Yu, Q. & Stamenkovic, I. Localization of matrix metalloproteinase 9 to the cell surface provides a mechanism for CD44-mediated tumor invasion. *Genes Dev.* **13**, 35–48 (1999).
98. Heider, K. H., Kuthan, H., Stehle, G. & Munzert, G. CD44v6: A target for antibody-based cancer therapy. *Cancer Immunol. Immunother.* **53**, 567–579 (2004).
99. Gnthert, U. *et al.* A New Variant of Glycoprotein CD44 Confers Metastatic Potential to Rat Carcinoma Cells. **65**, 1–12 (2011).
100. Olsson, E. *et al.* CD44 isoforms are heterogeneously expressed in breast cancer and correlate with tumor subtypes and cancer stem cell markers. *BMC Cancer* **11**, (2011).
101. Wielenga, V. J. M. *et al.* Expression of c-Met and heparan-sulfate proteoglycan forms of CD44 in colorectal cancer. *Am. J. Pathol.* **157**, 1563–1573 (2000).
102. Orian-Rousseau, V., Chen, L., Sleeman, J. P., Herrlich, P. & Ponta, H. CD44 is required for two consecutive steps in HGF/c-Met signaling. *Genes Dev.* **16**, 3074–3086 (2002).
103. Odenbreit, S. *et al.* Translocation of Helicobacter pylori CagA into gastric epithelial cells by type IV secretion. *Science (80-.).* **287**, 1497–1500 (2000).
104. Morrin, M. & Delaney, P. V. CD44v6 is not relevant in colorectal tumour progression. *Int. J. Colorectal Dis.* **17**, 30–36 (2002).
105. Megaptche, A. P., Erb, U., Büchler, M. W. & Zöller, M. CD44v10, osteopontin and

lymphoma growth retardation by a CD44v10-specific antibody. *Immunol. Cell Biol.* **92**, 709–720 (2014).
106. Bourguignon, L. Y. W., Zhu, D. & Zhu, H. CD44 isoform-cytoskeleton interaction in oncogenic signaling and tumor progression Lilly Y.W. Bourguignon, Dan Zhu and Hongbo Zhu. *Front. Biosci.* (1998).
107. Brown, R. L., Reinke, L. M., Damerow, M. S. & Perez, D. CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. **121**, (2011).
108. Zhao, S. *et al.* CD44 expression level and isoform contributes to pancreatic cancer cell plasticity, invasiveness and response to therapy. *Clin Cancer Res* **22**, 5592–5604 (2016).
109. Sy, M. S., Guo, Y. J. & Stamenkovic, I. Inhibition of tumor growth in vivo with a soluble cd44-immunoglobulin fusion protein. *J. Exp. Med.* **176**, 623–627 (1992).
110. Man, B., Sy, S. & Stamenkovic, I. Distinct Effects o f Two C D 4 4 Isoforms on T u m o r Growth In Vivo By Man Sun Sy, yaoJun Guo, and Ivan Stamenkovic. **174**, (1991).
111. Ling Lia,b,*, Xinbao Haoa,c,*, Jun Qind,*, Wenhua Tanga,e, Fengtian Hea, Amber Smithe, MinZhanga, Diane M. Simeonef, Xiaotan T. Qiaog, Zhi-Nan Chenb,§, Theodore S. Lawrencea, andLiang Xua,e. Antibody against CD44s Inhibits Pancreatic Tumor Initiation andPost-Radiation Recurrence in Mice. *Gastroenterology* **146**, 1–7 (2014).
112. Tatokoro, M. *et al.* Potential role of Hsp90 inhibitors in overcoming cisplatin resistance of bladder cancer-initiating cells. *Int. J. Cancer* **131**, 987–996 (2012).
113. Ojha, R., Jha, V., Singh, S. K. & Bhattacharyya, S. Autophagy inhibition suppresses the tumorigenic potential of cancer stem cell enriched side population in bladder cancer. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1842**, 2073–2086 (2014).
114. Ning, Z. F. *et al.* Subpopulations of stem-like cells in side population cells from the human bladder transitional cell cancer cell line T24. *J. Int. Med. Res.* **37**, 621–630 (2009).
115. Wu, C. Te, Lin, W. Y., Chen, W. C. & Chen, M. F. Predictive Value of CD44 in Muscle-Invasive Bladder Cancer and Its Relationship with IL-6 Signaling. *Ann. Surg. Oncol.* **25**, 3518–3526 (2018).
116. Sugino, T. *et al.* Progressive loss of CD44 gene expression in invasive bladder cancer. *Am. J. Pathol.* **149**, 873–882 (1996).
117. Sugiyama, M. *et al.* Non-invasive detection of bladder cancer by identification of abnormal CD44 proteins in exfoliated cancer cells in urine. *J. Clin. Pathol. - Clin. Mol. Pathol.* **48**, 142–147 (1995).
118. Stavropoulos, N. E. *et al.* CD44 standard form expression as a predictor of progression in high risk superficial bladder tumors. *Int. Urol. Nephrol.* **33**, 479–483 (2001).
119. Kuncová, J., Urban, M. & Mandys, V. Expression of CD44s and CD44v6 in transitional cell carcinomas of the urinary bladder: Comparison with tumour grade, proliferative activity and p53 immunoreactivity of tumour cells. *Apmis* **115**, 1194–1205 (2007).
120. Kuncová, J., Kostrouch, Z., Viale, M., Revoltella, R. & Mandys, V. Expression of CD44v6 correlates with cell proliferation and cellular atypia in urothelial carcinoma cell lines 5637 and HT1197. *Folia Biol. (Praha).* **51**, 3–11 (2005).
121. Jonckheere, N., Skrypek, N. & Seuningen, I. Van. Mucins and tumor resistance to chemotherapeutic drug. *BBA - Rev. Cancer* **1846**, 142–151 (2014).
122. Jonckheere, N., Skrypek, N. & Van Seuningen, I. Mucins and tumor resistance to chemotherapeutic drugs. *Biochim. Biophys. Acta - Rev. Cancer* **1846**, 142–151 (2014).
123. Singh, A. P. *et al.* Aberrant expression of transmembrane mucins, MUCI and MUC4, in human prostate carcinomas. *Prostate* **66**, 421–429 (2006).

124. Chaturvedi, P. *et al.* MUC4 mucin potentiates pancreatic tumor cell proliferation, survival, and invasive properties and interferes with its interaction to extracellular matrix proteins. *Mol. Cancer Res.* **5**, 309–320 (2007).
125. Jonckheere, N. & Seuningen, I. Van. The membrane-bound mucins : From cell signalling to transcriptional regulation and expression in epithelial cancers. *Biochimie* **92**, 1–11 (2010).
126. Schroeder, J. A. *et al.* MUC1 overexpression results in mammary gland tumorigenesis and prolonged alveolar differentiation. *Oncogene* **23**, 5739–5747 (2004).
127. Raina, D., Kharbanda, S. & Kufe, D. The MUC1 oncoprotein activates the anti-apoptotic phosphoinositide 3-kinase/Akt and Bcl-xL pathways in rat 3Y1 fibroblasts. *J. Biol. Chem.* **279**, 20607–20612 (2004).
128. Shi, M. *et al.* Catecholamine-Induced β2-Adrenergic Receptor Activation Mediates Desensitization of Gastric Cancer Cells to Trastuzumab by Upregulating MUC4 Expression. *J. Immunol.* **190**, 5600–5608 (2013).
129. Chen, A. C. *et al.* Upregulation of mucin4 in ER-positive/HER2-overexpressing breast cancer xenografts with acquired resistance to endocrine and HER2-targeted therapies. *Breast Cancer Res. Treat.* **134**, 583–593 (2012).
130. Gubbels, J. A. A. *et al.* MUC16 provides immune protection by inhibiting synapse formation between NK and ovarian tumor cells. *Mol. Cancer* **9**, 1–14 (2010).
131. Ponnusamy, M. P. *et al.* MUC4 mucin-induced epithelial to mesenchymal transition: A novel mechanism for metastasis of human ovarian cancer cells. *Oncogene* **29**, 5741–5754 (2010).
132. Horn, G., Gaziel, A., Wreschner, D. H., Smorodinsky, N. I. & Ehrlich, M. ERK and PI3K regulate different aspects of the epithelial to mesenchymal transition of mammary tumor cells induced by truncated MUC1. *Exp. Cell Res.* **315**, 1490–1504 (2009).
133. Roy, L. D. *et al.* MUC1 enhances invasiveness of pancreatic cancer cells by inducing epithelial to mesenchymal transition. *Oncogene* **30**, 1449–1459 (2011).
134. Beatson, R. E., Taylor-Papadimitriou, J. & Burchell, J. M. MUC1 immunotherapy. *Immunotherapy* **2**, 305–327 (2010).
135. Fendrick, J. L. *et al.* Characterization of CA 125 synthesized by the human epithelial amnion WISH cell line. *Tumor Biol.* **14**, 310–318 (1993).
136. Blalock, T. D. *et al.* Functions of MUC16 in corneal epithelial cells. *Investig. Ophthalmol. Vis. Sci.* **48**, 4509–4518 (2007).
137. Das, S. *et al.* Membrane proximal ectodomain cleavage of MUC16 occurs in the acidifying Golgi/post-Golgi compartments. *Sci. Rep.* **5**, 1–11 (2015).
138. Das, S. *et al.* Carboxyl-terminal domain of MUC16 imparts tumorigenic and metastatic functions through nuclear translocation of JAK2 to pancreatic cancer cells. *Oncotarget* **6**, 5772–5787 (2015).
139. Kim, N., Hong, Y., Kwon, D. & Yoon, S. Somatic Mutaome Profile in Human Cancer Tissues. *Genomics Inform.* **11**, 239 (2013).
140. Thériault, C. *et al.* MUC16 (CA125) regulates epithelial ovarian cancer cell growth, tumorigenesis and metastasis. *Gynecol. Oncol.* **121**, 434–43 (2011).
141. Belisle, J. A. *et al.* Identification of Siglec-9 as the receptor for MUC16 on human NK cells, B cells, and monocytes. *Mol. Cancer* **9**, 1–14 (2010).
142. Manvar, A. M., Wallen, E. M., Pruthi, R. S. & Nielsen, M. E. Prognostic value of CA 125 in transitional cell carcinoma of the bladder. *Expert Rev. Anticancer Ther.* **10**, 1877–1881 (2010).
143. Chen, S., Dallas, M. R. & Balzer, E. M. Mucin 16 is a functional selectin ligand on pancreatic cancer cells. 1349–1359 doi:10.1096/fj.11-195669.
144. Akita, K. *et al.* Different levels of sialyl-Tn antigen expressed on MUC16 in patients with endometriosis and ovarian cancer. *Int. J. Gynecol. Cancer* **22**, 531–538 (2012).

145. Radaev, S. & Sun, P. D. Recognition of IgG by Fcγ Receptor: The role of Fc glycosylation and the binding of peptide inhibitors. *J. Biol. Chem.* **276**, 16478–16483 (2001).
146. Shields, R. L. *et al.* Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human FcγRIII and antibody-dependent cellular toxicity. *J. Biol. Chem.* **277**, 26733–26740 (2002).
147. Schuster, M. *et al.* Improved effector functions of a therapeutic monoclonal Lewis Y-specific antibody by glycoform engineering. *Cancer Res.* **65**, 7934–7941 (2005).
148. Strome, S. E., Sausville, E. A. & Mann, D. A Mechanistic Perspective of Monoclonal Antibodies in Cancer Therapy Beyond Target-Related Effects. *Oncologist* **12**, 1084–1095 (2007).
149. Chakrani, Z., Robinson, K. & Taye, B. Association Between ABO Blood Groups and Helicobacter pylori Infection: A Meta-Analysis. *Sci. Rep.* **8**, 1–11 (2018).
150. Anstee, D. J. The relationship between blood groups and disease. *Blood* **115**, 4635–4643 (2010).
151. Röttger, S. *et al.* Localization of three human polypeptide GalNAc-transferases in HeLa cells suggests initiation of O-linked glycosylation throughout the Golgi apparatus. *J. Cell Sci.* **111**, 45–60 (1998).
152. Gill, D. J., Clausen, H. & Bard, F. Location, location, location: New insights into O-GalNAc protein glycosylation. *Trends Cell Biol.* **21**, 149–158 (2011).
153. Grosso, M., Vitarelli, E., Giuffrè, G., Tuccari, G. & Barresi, G. Expression of Tn, sialosyl-Tn and T antigens in human foetal large intestine. *Eur. J. Histochem.* **44**, 359–363 (2000).
154. Itzkowitz, S. *et al.* Expression of Tn, sialosyl Tn, and T antigens in human pancreas. *Gastroenterology* **100**, 1691–1700 (1991).
155. Yokoyama, M. *et al.* Thomsen-Friedenreich antigen in bladder cancer tissues detected by monoclonal antibody. *Hinyokika Kiyo.* **34**, 255–258 (1988).
156. Kishikawa, T., Ghazizadeh, M., Sasaki, Y. & Springer, G. F. Specific role of T and Tn tumor-associated antigens in adhesion between a human breast carcinoma cell line and a normal human breast epithelial cell line. *Japanese J. Cancer Res.* **90**, 326–332 (1999).
157. Danussi, C. *et al.* A newly generated functional antibody identifies Tn antigen as a novel determinant in the cancer cell-lymphatic endothelium interaction. *Glycobiology* **19**, 1056–1067 (2009).
158. Yu, L. G. *et al.* Galectin-3 interaction with Thomsen-Friedenreich disaccharide on cancer-associated MUC1 causes increased cancer cell endothelial adhesion. *J. Biol. Chem.* **282**, 773–781 (2007).
159. Santos-Silva, F. *et al.* Thomsen-Friedenreich antigen expression in gastric carcinomas is associated with MUC1 mucin VNTR polymorphism. *Glycobiology* **15**, 511–517 (2005).
160. Liu, Z. *et al.* Tn antigen promotes human colorectal cancer metastasis via H-Ras mediated epithelial-mesenchymal transition activation. *J. Cell. Mol. Med.* **23**, 2083–2092 (2019).
161. Radhakrishnan, P. *et al.* Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4066–E4075 (2014).
162. Cornelissen, L. A. M. *et al.* Tn Antigen Expression Contributes to an Immune Suppressive Microenvironment and Drives Tumor Growth in Colorectal Cancer. *Front. Oncol.* **10**, 1–15 (2020).
163. Akita, K. *et al.* Developmental expression of a unique carbohydrate antigen, Tn antigen, in mouse central nervous tissues. *J. Neurosci. Res.* **65**, 595–603 (2001).
164. Smorodin, E., Sergeyev, B., Klaamas, K., Chuzmarov, V. & Kurtenkov, O. The relation of the level of serum anti-TF, -Tn and -alpha-gal IgG to survival in gastrointestinal cancer patients. *Int. J. Med. Sci.* **10**, 1674–1682 (2013).

165. SD Rosen, MS Singer, TA Yednock, L. S. Involvement of Sialic Acid on Endothelial Cells in Organ-Specific Lymphocyte Recirculation. *Science (80-.).* **228**, 98–100 (1985).
166. Cummings, R. D. & Smith, D. F. The selectin family of carbohydrate-binding proteins: Structure and importance of carbohydrate ligands for cell adhesion. *BioEssays* **14**, 849–856 (1992).
167. Hiller, K. M. *et al.* Transfection of α(1,3)fucosyltransferase antisense sequences impairs the proliferative and tumorigenic ability of human colon carcinoma cells. *Mol. Carcinog.* **27**, 280–288 (2000).
168. Weston, B. W. *et al.* Expression of human α(1,3)fucosyltransferase antisense sequences inhibits selectin-mediated adhesion and liver metastasis of colon carcinoma cells. *Cancer Res.* **59**, 2127–2135 (1999).
169. Nieswandt, B., Hafner, M., Echtenacher, B. & Männel, D. N. Lysis of tumor cells by natural killer cells in mice is impeded by platelets. *Cancer Res.* **59**, 1295–1300 (1999).
170. Kim, Y. J., Borsig, L., Han, H. L., Varki, N. M. & Varki, A. Distinct selectin ligands on colon carcinoma mucins can mediate pathological interactions among platelets, leukocytes, and endothelium. *Am. J. Pathol.* **155**, 461–472 (1999).
171. Varki, A. Letter to the Glyco-Forum: Since there are PAMPs and DAMPs, there must be SAMPs? Glycan 'self-associated molecular patterns' dampen innate immunity, but pathogens can mimic them. *Glycobiology* **21**, 1121–1124 (2011).
172. Jellusova, J., Wellmann, U., Amann, K., Winkler, T. H. & Nitschke, L. CD22 × Siglec-G Double-Deficient Mice Have Massively Increased B1 Cell Numbers and Develop Systemic Autoimmunity. *J. Immunol.* **184**, 3618–3627 (2010).
173. Büll, C., den Brok, M. H. & Adema, G. J. Sweet escape: Sialic acids in tumor immune evasion. *Biochim. Biophys. Acta - Rev. Cancer* **1846**, 238–246 (2014).
174. Camilla Jandus, 1 Kayluz Frias Boligan, 1 Obinna Chijioke, 2 He Liu, 1 Meike Dahlhaus, 3, 4Thomas Démoulins, 5 Christoph Schneider, 1 Marc Wehrli, 1 Robert E. Hunger, 6Gabriela M. Baerlocher, 3, 4 Hans-Uwe Simon, 1 Pedro Romero, 7Christian Münz, 2 and Stephan von Gunten1. Interactions between Siglec-7/9 receptors and ligands influence NK cell–dependent tumor immunosurveillance. *J. Clin. Invest.* **124**, 316 (2014).
175. Jandus, C. *et al.* Siglec-7/-9 ligands shield tumor cells from NK cell attack. *J. Immunother. Cancer* **1**, P157 (2013).
176. Perdicchio, M. *et al.* Tumor sialylation impedes T cell mediated anti-tumor responses while promoting tumor associated-regulatory T cells. *Oncotarget* **7**, 8771–8782 (2016).
177. JA, F. *et al.* Overexpression of tumour-associated carbohydrate antigen sialyl-Tn in advanced bladder tumours. *Mol. Oncol.* **7**, 719–731 (2013).
178. Sewell, R. *et al.* The ST6GalNAc-I sialyltransferase localizes throughout the golgi and is responsible for the synthesis of the tumor-associated sialyl-Tn O-glycan in human breast cancer. *J. Biol. Chem.* **281**, 3586–3594 (2006).
179. Vázquez-Martín, C., Cuevas, E., Gil-Martín, E. & Fernández-Briera, A. Correlation analysis between tumorous associated antigen sialyl-Tn expression and ST6GalNAc I activity in human colon adenocarcinoma. *Oncology* **67**, 159–165 (2004).
180. Costa, C. *et al.* Abnormal protein glycosylation and activated PI3K/Akt/mTOR pathway: Role in bladder cancer prognosis and targeted therapeutics. *PLoS One* **10**, 1–19 (2015).
181. Carrascal, M. A. *et al.* Sialyl Tn-expressing bladder cancer cells induce a tolerogenic phenotype in innate and adaptive immune cells. *Mol. Oncol.* **8**, 753–765 (2014).
182. Langkilde, N. C., Wolf, H., Clausen, H., Kjeldsen, T. & Ørntoft, T. F. Nuclear volume and expression of T-antigen, sialosyl-Tn-antigen, and Tn-antigen in carcinoma of

the human bladder. Relation to tumor recurrence and progression. *Cancer* **69**, 219–227 (1992).
183. Summers, J. L. *et al.* Prognosis in Carcinoma of the Urinary Bladder Based upon Tissue Blood Group ABH and Thomsen-Friedenreich Antigen Status and Karyotype of the Initial Tumor. *Cancer Res.* **43**, 934 (1983).
184. Loureiro, L. R. *et al.* Challenges in antibody development against Tn and sialyl-Tn antigens. *Biomolecules* **5**, 1783–1809 (2015).
185. Zlocowski, N. *et al.* Purified human anti-Tn and anti-T antibodies specifically recognize carcinoma tissues. *Sci. Rep.* **9**, 1–11 (2019).
186. Stolz, C. & Schuler, M. Molecular mechanisms of resistance to Rituximab and pharmacologic strategies for its circumvention. *Leuk. Lymphoma* **50**, 873–885 (2009).
187. Miliotou, A. N. & Papadopoulou, L. C. CAR T-cell Therapy: A New Era in Cancer Immunotherapy. *Curr. Pharm. Biotechnol.* **19**, 5–18 (2018).
188. Hsin-I Chang and Ming-Kung Yeh. Clinical Dev Liposome. *Int. J. Nanomed.* **7**, 49–60 (2012).
189. Martinez-Finkelshtein, A. & Van Assche, W. WHAT IS...A Multiple Orthogonal Polynomial? *Not. Am. Math. Soc.* **63**, 1029–1031 (2016).
190. Modena, M. M., Rühle, B., Burg, T. P. & Wuttke, S. Nanoparticle Characterization: What to Measure? *Adv. Mater.* **31**, 1–26 (2019).
191. Swider, E. *et al.* Customizing poly(lactic-co-glycolic acid) particles for biomedical applications. *Acta Biomater.* **73**, 38–51 (2018).
192. Kulhari, H., Pooja, D., Shrivastava, S., V.G.M, N. & Sistla, R. Peptide conjugated polymeric nanoparticles as a carrier for targeted delivery of docetaxel. *Colloids Surfaces B Biointerfaces* **117**, 166–173 (2014).
193. Kulhari, H. *et al.* Bombesin-conjugated nanoparticles improve the cytotoxic efficacy of docetaxel against gastrin-releasing but androgen-independent prostate cancer. *Nanomedicine* **10**, 2847–2859 (2015).
194. Dinarvand, R., Sepehri, N., Manoochehri, S., Rouhani, H. & Atyabi, F. Polylactide-co-glycolide nanoparticles for controlled delivery of anticancer agents. *Int. J. Nanomedicine* **6**, 877–895 (2011).
195. Gao, J. *et al.* PE38KDEL-loaded anti-HER2 nanoparticles inhibit breast tumor progression with reduced toxicity and immunogenicity. *Breast Cancer Res. Treat.* **115**, 29–41 (2009).
196. Mo, Y. & Lim, L. Y. Preparation and in vitro anticancer activity of wheat germ agglutinin (WGA)-conjugated PLGA nanoparticles loaded with paclitaxel and isopropyl myristate. *J. Control. Release* **107**, 30–42 (2005).
197. Bondioli, L. *et al.* PLGA nanoparticles surface decorated with the sialic acid, N-acetylneuraminic acid. *Biomaterials* **31**, 3395–3403 (2010).
198. Guo, C. *et al. Therapeutic cancer vaccines. Past, present, and future*. *Advances in Cancer Research* vol. 119 (2013).
199. Robert E. Hollingsworth and & Kathrin Jansen. Turning the corner on therapeutic cancer vaccines. *npj Vaccines* **4**, 1–10 (2019).
200. Feng, L. *et al.* Pharmaceutical and immunological evaluation of a single-dose hepatitis B vaccine using PLGA microspheres. *J. Control. Release* **112**, 35–42 (2006).
201. Saini, V. *et al.* Comparison of humoral and cell-mediated immune responses to cationic PLGA microspheres containing recombinant hepatitis B antigen. *Int. J. Pharm.* **408**, 50–57 (2011).
202. Manzanares, D. & Ceña, V. Endocytosis: The nanoparticle and submicron nanocompounds gateway into the cell. *Pharmaceutics* **12**, 1–22 (2020).
203. Kanchan, V. & Panda, A. K. Interactions of antigen-loaded polylactide particles with macrophages and their correlation with the immune response. *Biomaterials* **28**, 5344–5357 (2007).

204. Koerner, J., Horvath, D. & Groettrup, M. Harnessing Dendritic Cells for Poly (D,L-lactide-co-glycolide) Microspheres (PLGA MS)-Mediated Anti-tumor Therapy. *Front. Immunol.* **10**, 1–16 (2019).
205. Pacheco, P., White, D. & Sulchek, T. Effects of Microparticle Size and Fc Density on Macrophage Phagocytosis. *PLoS One* **8**, 1–9 (2013).

CHAPTER I | Single-pot enzymatic synthesis and characterization of cancer-associated MUC16 and CD44 O-glycopeptide libraries

Abstract

Most aggressive human solid tumours overexpress abnormally glycosylated proteins such as MUC16 and CD44 carrying short-chain O-glycans, namely the Tn and STn antigens. These glycoproteoforms are absent from most healthy human cells and highly expressed in cancer, offering enormous potential for vaccine development. Therefore, the present chapter is devoted to the enzymatic synthesis and characterization of cancer-specific MUC16- and CD44-Tn and/or STn glycoforms resulting from extensive glycoproteomics characterization of bladder tumours by our group. The adopted strategy involved the glycosylation of a synthetic 20-mer tandem repeat sequence from MUC16 and a 30-mer peptide from the standard and short forms of CD44s/CD44st using glycosyltransferases and sugars nucleotides, added in stepwise manner in a single pot reaction. Briefly, a recombinant GalNAc-T1, one of the first glycosyltransferases in O-glycans biosynthesis was first used, together with UDP-GalNAc, to produce Tn-glycopeptides with different degrees of glycosylation. STn-glycopeptides were then produced from Tn glycoepitopes by sialylation by ST6GalNAc-1 in the presence of CMP-Neu5Ac. Tn-glycopeptides were purified by Vicia Villosa agglutinin an STn-glycopeptides by TiO2 affinity chromatography and subsequently characterized by nanoLC-MS/MS regarding the number of glycosites and their location in the peptide chain. Collectively, it was possible to develop multiple and well characterized MUC16-Tn/STn as well as CD44-Tn glycoepitopes. However, we were not successful synthesizing CD44-STn, suggesting difficulties in the sialylation of the CD44 chain by this method, supporting the introduction of alternative strategies such as cell factories addressed in detail in Chapter 3. Nevertheless, several glycoepitopes libraries were produced to support the construction of multivalent cancer vaccine constructs in chapter 2.

1. Introduction

The use of glycopeptides as therapeutic targets has represented a new set of potential approaches in clinical settings. Their involvement in cancer specific environments and functional roles in tumour progression endorses their possible use in the development of novel therapeutics with particular interest in glyco-based-vaccines. Nonetheless, there have been some major drawbacks regarding the production of carbohydrates for these ends. In general, the production of pure and properly characterized glycoconjugates is hardly ever obtained, partially due to the small amounts of complexed glycans available which are difficult to isolate and characterize[1,2]. Chemical synthesis can produce well defined glycopeptides and has been explored inclusively for vaccine construction[3]. However, its' associated expensivness and time consuming protocols allied to the fact that precise glycosylation requires protection and deprotection steps that are usually difficult to perform, hinders sustainable development of lab based therapeutic designs[3].As such, chemo-enzymatic approaches emerge as an alternative source of glycan production. The use of glycosyltransferases for the progressive build up of more complex structures, using peptides and/or oligosaccharides as scaffolds, enables the production of relevant and defined glycans which can be used to further establish glycoconjugate libraries[4].The use of enzymatic methods in place of chemical approaches is also advantageous as it produces highly regio or stereospecific glycopeptides disregarding the use of laborious protection and deprotection reactions. It also represents a more environmentally sustainable approach given that it doesn't rely on the use of organic solvents[5,6]. Inclusively, these enzymatic approaches have been successfully used for the biosynthesis of O- and N-glycosites[7].

Unsurprisingly the use of chemo-enzymatic methods has been increasing when it comes to production of gloycopeptides, in part due to its accessible prices, but also due to the wide array of recombinant transferases available for peptide customization[8]. However, despite the fact that substrate specific enzymes with well-defined kinetic profiles are accessible, obtaining specific O-glycosylation patterns is not always easy. For this purpose, the initial glycosylation step has to be tightly tuned and further elongation into the desired structures has to be assured. O-glycosylation starts with the covalent linkage between a GalNac residue and specific amino acids present in a given peptide, namely serine and threonine. This reaction is catalysed by polypeptide N-acetylgalactosaminyltransferases (GalNAc-Ts) using a nucleotide sugar donor UDP-GalNAc[9]. In a chemo-enzymatic approach several GalNac-Ts can be conjugated to produce the desired O-GalNac glycosites due to their complementary activities, however,

it is important to note that most GalNac-Ts can target multiple sites of glycosylation in a given peptide sequence, and as such, will produce a wide arrays of glycosylation patterns. Respectfully, the use of bioinformatics tools like, ISOGlyP[10] and NetOGlyc[11], can help guide the process of GalNac-T selection and the choice of possible transferase combinations.

Further peptide elongation can be achieved by the use of successive glycosyltransferases, resulting in progressively more complex and diverse structures settled by sialic acid addition. Using this approach we demonstrate that these methods can successfully synthesize Tn/STn-MUC16 and CD44-Tn glycopeptides commonly overexpressed by cancer cells[12,13]. Moreover, we were able to obtain glycopeptide specific libraries and further delineate the favourable glycosites associated to our peptides. However, CD44 peptides could not be sialylated using this strategy, entailing experimental optimization for sialylation of specific peptides.

2. Methods and Materials

2.1 CD44s and MUC16 epitope formulation and Bioinformatics analysis

CD44 peptide was designed from sequences of CD44 short forms, CD44s (NP_001001391.1) and CD44st (NP_001189486.1). The 30-mer peptide sequence was analysed by bioinformatics tools, NetOGlyc, (http://www.cbs.dtu.dk/services/NetOGlyc/) to predict possible O-glycosylation sites in vitro. Subsequently, GalNAc-T1 expression in bladder cancer was evaluated using transcriptomic data available in Oncomine (www.oncomine.org)[14]. This analysis was performed addressing expression in healthy tissue and a p-value ≤ 0.05, a 2 fold change and all gene rank were considered for this study[15].

The selected CD44 sequence was evaluated using ISOGlyP I (Isoform Specific O-Glycosylation Prediction) (https://isoglyp.utep.edu/)[10], a bioinformatics tool to determine the fitness of human GalNAc-T1 (EC 2.4.1.41) to O-glycosylate this particular sequence of amino acids. The results obtained by this predictive tool are displayed as enhancement value products (EVP), which indicate the propensity of a given site for glycosylation. These values were obtained by using random glycopeptide libraries, containing forms similar or equal to GAGA(X)nT(X)nAGAGK (n=4 or 5), in order to quantitively determine the preferences of ppGalNac-Ts[16]. Values greater than 1 are indicative of GalNac-T preference for that specific substrate while values smaller than 1, indicate a less

preferable glycosylation site for a given transferase. EVPs equal to 1 indicate neither a preference nor an unfavourable site for Glycosylation[10,11]. It is important to note that residues with values lower than 1 could still be glycosylated given more incubation time and/or transferase amount.

From the MUC16 protein sequence found on NCBI (NP_078966.2) a 20 mer peptide chain was selected supported by NetOGlyc analysis. This sequence was inserted in the tandem repeat region of the MUC16 sequence. Seemingly to CD44 we used ISOGlyP to address probable sites for glycosylation with GalNacT1 within our MUC16 peptide sequence.

2.2. Enzymatic Synthesis of MUC16 and CD44-glycopeptide epitopes

A MUC16 20-mer peptide (VDVGTSGTPSSSPSPTTAGP) from the TR region and its cysteine tagged form were purchased from GenScript Biotech. A CD44 30-mer peptide (DSPWITDSTDRIPATRDQDTFHPSGGSHTT), as well as a C-terminal Cysteine-tagged version (DSPWITDSTDRIPATRDQDTFHPSGGSHTTC) were purchased from GenScript Biotech. These two peptides were glycosylated to display Tn and STn epitopes. Tn epitopes were obtained by preparing 100 µl glycosylation batches containing: glycosylation buffer (125 mM sodium cacodylate, 50 mM MnCl2, pH 7.4; Sigma-Aldrich), 0.5 mM UDP-GalNAc (Sigma-Aldrich), 50 µg of CD44 or CD44-CysTag peptides (from 10µg/µl) and 0.025 µg GalNAc-T1 (R&D Systems); and incubated them at 37ºC with 350 rpm O/N. STn epitopes were obtained by vaccum dry of Tn samples, followed by an O/N incubation (37ºC ; 350 rpm) with a 100 µl mix containing: glycosylation buffer (125 mM sodium cacodylate, 50 mM MnCl2, pH 7.4; Sigma-Aldrich), 0.5 mM CMP-Neu5Ac (1mg/ml; Merck Millipore) and 40 µg ST6GalNAc I (20µg/ µl; R&D Systems), a human sialyltransferase (EC 2.4.99.3). Tn peptides were “purified” using VVA agarose (Agarose Bound Vicia Villosa Lectin; Vector Laboratories) enriched columns (pierce spin columns from Thermo Fisher Scientific) whereas, STn peptides were isolated using the High-Select™ TiO2 Phosphopeptide Enrichment Kit (Thermo Fisher Scientific), according to the manufacturer instructions. Samples were quantified using a Fluorometric peptide assay Kit (PierceTM Quantitative Fluorometric Peptide Assay; Thermo Fisher Scientific).

2.3. NanoLC-ESI-LTQ-Orbitrap Mass Spectrometry

Glycopeptide synthesis was characterized by NanoLC-ESI-MS/MS on a nanoLC system (Dionex 3000 Ultimate nano-LC) coupled online to an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific) equipped with a nano-electrospray ion source (EASY-Spray source, Thermo Fisher Scientific). Eluent A was aqueous formic acid (0.2%) and eluent B was formic acid (0.2%) in acetonitrile. Samples (10 µl) were injected directly into a trapping column (C18 PepMap 100, 5 µm particle size) and washed with an isocratic flux of 95% eluent A and 5% eluent B at a flow rate of 30 µl/min. After 3 minutes, the flux was redirected to the analytical column (EASY-Spray C18 PepMap, 100 Å, 150 mm × 75µm ID and 3 µm particle size) at a flow rate of 0.3 µl/min. Column temperature was set at 35°C. Glycopeptide separation occurred using a linear gradient of 5–40% eluent B over 30 min. Column wash and re-equilibration proceeded with 40-90% eluent B over 5 min, 5 min with 90% eluent B and 15 min at 5% eluent B before the following injection. The mass spectrometer was operated in the positive ion mode, with a spray voltage of 1.9 kV and a transfer capillary temperature of 250°C. Tube lens voltage was set to 120 V. MS full scans were acquired at an Orbitrap resolution of 60,000 for an m/z range from 300 to 2000. Data were recorded with Xcalibur software version 2.1 (Thermo Fisher Scientific). Reported ions are presented in terms of relative abundance in relation to the total glycopeptides abundance. Tandem MS data were acquired in the linear ion trap using a data dependent method with dynamic exclusion (90 s) and the 3 most intense ions were selected for collision induced dissociation (CID). CID settings were 35% normalized collision energy, 2 Da isolation window, 30 ms activation time and an activation Q of 0.250. Automatic Gain Control was enabled and target values were 1.00 e+6 for the Orbitrap and 1.00 e+4 for LTQ MS analysis. Data were recorded with Xcalibur software version 2.1 (Thermo Fisher Scientific). For annotation of glycosylated sites, the data were analysed using the Sequest HT search engine (Proteome Discoverer 1.4, Thermo Fisher Scientific), supported by Byonic search engine (Byonic 3.8.13, Protein Metrics). The sequence FASTA of MUC16 peptide used in this study was defined as protein database. Variable modification of serine and threonine residues with HexNAc (+203.0794, GalNAc, Tn antigen) was considered for glycosites annotation of MUC16 peptide (monoisotopic mass 1800.8378) modified with the Tn antigen.

3. Results Discussion

Both MUC16- and CD44-Tn/STn glycoforms are found in bladder tumours displaying more aggressive phenotypes but, to our knowledge, not in healthy human tissues. Moreover, these glycoforms are associated with worse prognosis in BC and their associated glycans play a functional role in chemoresistance, immune evasion and overall disease progression. Therefore, these molecules could potentially represent important targets for therapeutics and immunotherapy, especially due to their cell-surface nature. As such, this work focused on revisiting the enzymatic synthesis of Tn- and STn-glycopeptides under development by the group, providing a more comprehensive molecular characterization of the synthesis products, with emphasis on precise glycosites identification. For the first time, attempts were also made to produce CD44-Tn/STn glycoepitopes. The overall goal was to generate libraries of Tn/Stn-glycoepitopes for glycovaccines construction in chapter 2. As general strategy, commercial synthetic peptides corresponding to a synthetic 20-mer tandem repeat sequence from MUC16 and a 30-mer peptide from the standard and short forms of CD44s/CD44st were elected for this study. The sequences were custom designed to mimic peptides identified in bladder tumours in previous glycoproteomics studies. A recombinant GalNAc-T1, one of the first glycosyltransferases in *O*-glycans biosynthesis was first used, together with UDP-GalNAc, to produce Tn-glycopeptides with different degrees of glycosylation. GalNAc-T1 was elected for this study because it showed increased expression in more aggressive urothelial tumours when compared to healthy tissues, according to Oncomine[17]. STn-glycopeptides were then produced from Tn glycoepitopes by *O*-6 GalNAc sialylation with ST6GalNAc-1 in the presence of CMP-Neu5Ac. The enzymatic reactions were conducted in a stepwise manner in a single pot reaction. Intermediate CD44 Tn-glycopeptide products were isolated by VVA lectin affinity chromatography and characterized by nanoLC-MS/MS. STn-glycopeptides were pre-purified by TiO_2 chromatography, which has been widely used to isolate sialylated peptides/proteins and also characterized by nanoLC-MS/MS.

3.1. Bioinformatics for MUC16 glycosites estimation

To better dimension the analytical work and adopt the best enrichment strategies for downstream nanoLC-MS/MS, we started by estimating the number of potential glycosites generated by GalNAc-T1 using bioinformatics tools. According to NetOGlyc 4.0,

the number of potential O-glycosylation sites in the elected MUC16 peptide was 9; however only residues S6, S10, S11, S12, S14, T16 and T17 displayed high probability for O-glycosylation. We then used ISOGlyP to gain insights on the specific glycosites that may be generated with GalNAcT1. According to this predicting tool, C1GALT1 preferentially targets the T17 residue and, to a lesser extent, T8, T16, T5 and S12, while other Ser/Thr residues displayed lower GalNac-T1 specificity values. This suggests that, further on, peptides displaying distinct glycosylation profiles would be expected, which could represent an advantage when seeking to develop multivalent vaccines that mirror the heterologous truncated O-glycosylation found in cancer environments.

3.2. Characterization of MUC16-Tn glycosites

The first part of the study was focused on characterizing the MUC16-Tn glycopetides generated by C1Gal-T1. According to nanoLC-MS analysis, after GalNac-T1 glycosylation, around 70% of MUC16 peptides in our sample displayed O-glycosylation structures, with mono-glycosylated species being predominant over di, tri, tetra and penta-glycosylated peptides (**Table 1**). Moreover, in this analysis, only five glycosites were determined, contrary to the seven glycosites predicted by bioinformatics tools. Furthermore, glycosylated and unglycosylated species seem to be adequately isolated by the chromatographic conditions applied, given the differences in their retention times (**Table 1**).

Table 1. C18 reverse phase nanoLC-ESI-MS profiles for reaction products resulting from the glycosylation of the 20 mer MUC16 peptide by GalNAc-T1. For MUC16 the main products correspond to the unglycosylated form and mono-glycosylated peptides, followed by di- and tri-glycosylated species

***O*-glycans**	***Glycopeptides Monoisotopic mass [M+2H]2+/ [M+3H]3+***	**Apex Retention time (min)**	**Relative abundance (%)**
unglycosylated	901.43	16.31	29.26
1 Tn	1002.97	15.61	38.08
2 Tn	1104.51	15.09	19.61
3 Tn	804.37	14.75	11.71
4 Tn	872.06	14.39	1.30
5 Tn	939.75	13.66	0.04

These glycosylated MUC16-Tn peptides will serve as substrates for further elongation steps and formation of more complex glycopeptides, and as such, we elaborated a comprehensive characterization of their main glycosites, based on characteristic b and y glycopeptide fragments from MS/MS spectra. Glycosite signatures were established using the SEQUEST HT algorithm supported by Byonic. For MUC16-Tn, T16 was the preferential glycosite, with much less evident amounts of peptide glycosylation at S6, 11, 12, and 14 and T5 and 17 (**Figure 5**). These experimental observations contrasted with the initial bioinformatic predictions using ISOGlyP, which pointed to T17 as the preferential glycosite, and displayed EVPs lower than 1 for residues S6, 11, 14. This portraits the importance of experimental validation in the context of *in silico* approaches. Additionally, we observed a predominant second glycosylation site at T17, which represented the main glycosylation site for Transferase 1 according to ISOGlyPs analysis. Although less abundant, there are glycospecies showing simultaneous glycosylations at S6 and S10, S12 and S14, T5 and S14, and T8 and S14/T16. Due to low abundance, we could not precisely identify the third preferential glycosites in tri-glycosylated peptides (**Table 1**, **Figure 5**). Collectively these findings reveal the predominance of glycopeptides with a single glycosite a T16 and, to less extent, a second glycosylation at T17.

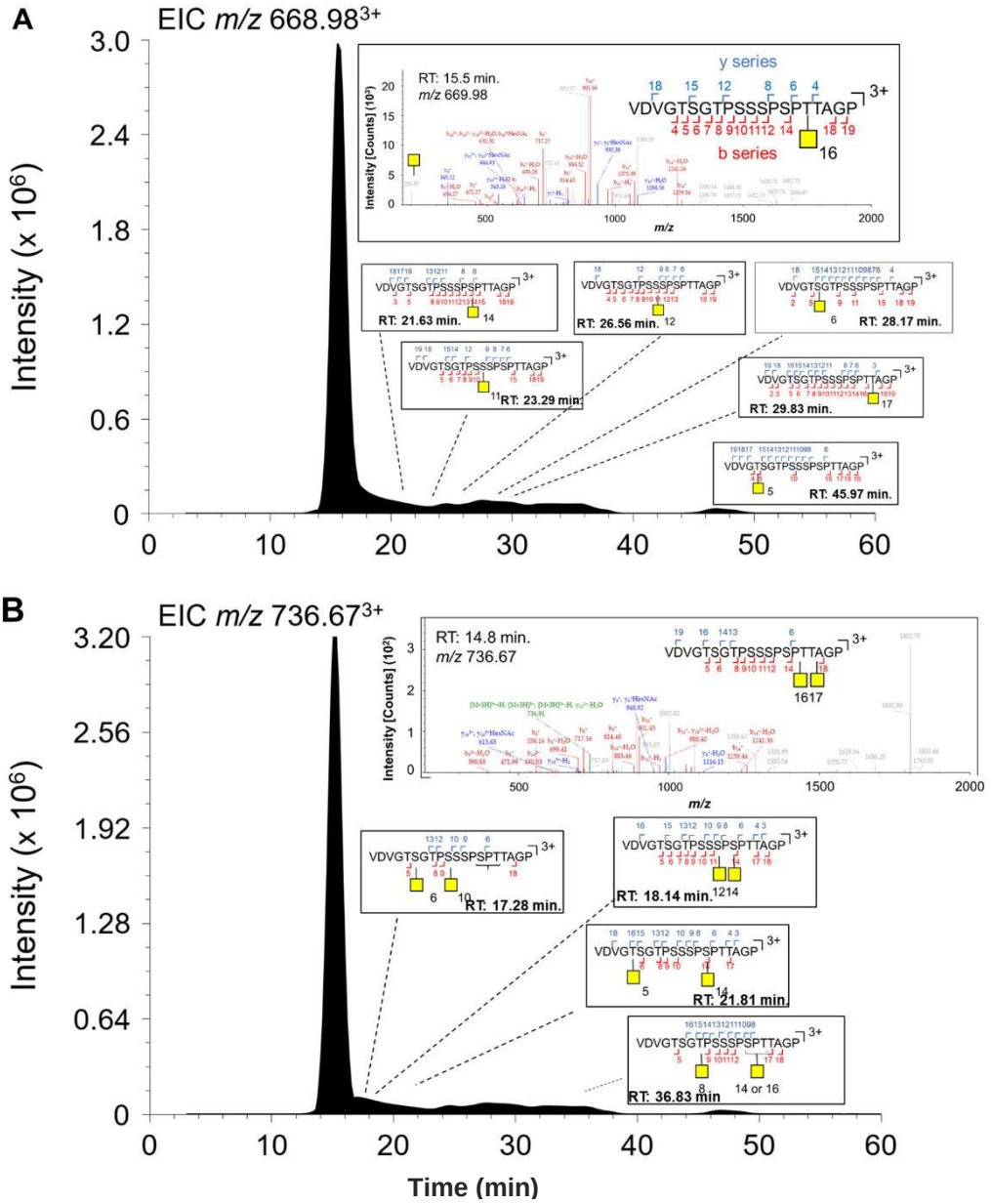


Figure 5. Glycosites identification for 20 mer MUC16 peptide after glycosylation by GalNAc-T1. A) Identification of T16 as the main glycosite in MUC16-Tn glycopeptides. T16 appears as the preferential glycosite, with low abundant glycopeptides substituted at T5, S6, S11, S12, S14, T17. B) Identification of T16 and T17 as the main glycosites in MUC16-2Tn glycopeptides. The syntheses of S6/10, S12/14, T5/S14 and T8/S14 or T16 were also observed, although in exceptionally low amounts.

3.3. Synthesis and enrichment of sialo-MUC16 glycopeptides

MUC16-Tn peptides were further processed to carry STn epitopes by sialylation of MUC16-Tn with ST6GalNAc-1, and subsequently enriched for sialylated species, using TiO_2 coated columns. TiO_2 is often used for enrichment of phosphorylated peptides; nevertheless, sialoglycopeptides have also shown high affinity for TiO_2. This has been

attributed to multiple interactions between the carboxyl groups on TiO_2 resins and the hydroxyl groups of sialic acids at the nonreducing end of the glycoepitopes[21]. According to MS analysis, no unglycosylated species were observed, demonstrating the potential of this approach to obtain fully glycosylated species. Moreover, the reaction generated neutral and hybrid species, showing potential for development of multivalent and highly heterogenous glycolibraries. These products were then submitted to TiO_2 enrichment, leading to an increase in sialo and Tn glycosylated hybrids at the expenses neutral species (**Table 2**). Ultimately TiO_2 enriched samples showed more STn decorated glycopeptides, which may potentially mimic relevant cancer associated glycosylation profiles.

Table 2. Products from the single-pot enzymatic synthesis of MUC16-STn glycopeptides, without purification and after TiO2 enrichment, analysed by C18 reverse phase nanoLC-ESI-MS. Neutral glycopeptides represent Tn-peptides; Hybrid glycopeptides are characterized by structures containing both Tn and STn; Sialylated glycopeptides only show sialylated structures.

***O*-glycans**	**Glycopeptides Monoisotopic mass $[M+2H]^{2+}$ / $[M+3H]^{3+}$**	**Apex Retention time (min)**	**Without enrichment**	**Enrichment with TiO_2**
			Relative abundance (%)	**Relative abundance (%)**
Neutral glycopeptides				
1 Tn	1002.97 / 668.98	16.29	31.65	6.38
2 Tn	1104.51 / 736.67	15.73	23.73	3.36
3 Tn	1206.05 / 804.37	14.83	3.28	1.01
4 Tn	1307.59 / 872.06	14.45	3.18	0.49
5 Tn	1409.13 / 939.75	14.07	1.05	0.34
Hybrid glycopeptides				
1 Tn + 1 STn	1250.05 / 833.71	16.10	14.08	9.71
1 Tn + 2 STn	1497.14 / 998.43	15.56	0.45	3.97
1 Tn + 3 STn	1744.23 / 1163.16	15.02	0.06	3.86
2 Tn + 1 STn	1351.59 / 901.40	15.21	3.03	5.15
2 Tn + 2 STn	1598.68 / 1066.12	14.83	0.85	9.50
2 Tn + 3 STn	1845.77 / 1230.85	14.64	0.03	2.92
2 Tn + 4 STn	2092.86 / 1395.57	14.64	0.00	1.07
3 Tn + 1 STn	1453.13 / 969.09	14.64	2.24	7.32
3 Tn + 2 STn	1700.22 / 1133.82	14.45	0.26	5.04
3 Tn + 3 STn	1947.31 / 1298.54	14.25	0.04	4.51
3 Tn + 4 STn	2194.40 / 1463.27	14.45	0.00	0.37
4 Tn + 1 STn	1554.67 / 1036.78	14.25	1.01	3.47
4 Tn + 2 STn	1801.76 / 1201.51	14.07	0.26	8.09
4 Tn + 3 STn	2048.85 / 1366.23	14.07	0.01	1.84
5 Tn + 1 STn	1656.21 / 1104.48	13.89	0.00	5.52
5 Tn + 2 STn	1903.30 / 1269.20	13.89	0.10	4.41
Sialylated glycopeptides				
1 STn	1148.51 / 766.01	16.67	12.02	4.90
2 STn	1395.60 / 930.74	16.29	0.61	1.30
3 STn	1642.69 / 1095.46	15.73	0.03	0.80
4 STn	1889.78 / 1260.19	15.40	0.00	0.63

3.4. CD44s Bioinformatic Characterization

Solid tumours often overexpress a wide number of CD44 proteoforms resulting from alternative splicing. Moreover, CD44-Tn/STn glycoforms play key roles in cancer, contributing for its aggressive phenotype and progression and more recently have been associated with inhibition of immune responses. Building on these observations, our group has been conducted glycoproteomics studies to identify CD44-Tn/STn proteoglycoforms more associated with aggressive disease. This showed that invasive tumours express mostly shorter CD44 forms, particularly CD44st (short tail) and CD44s (standard). Accordingly, we designed a 30-mer peptide sequence common to these CD44 proteoforms for constructing CD44 glycopeptides using the same strategy described for MUC16-Tn/STn synthesis.

Based on the same workflow used for MUC16 glycopeptides, we started by using bioinformatics tools to predict the potential number and localization of the glycosites generated by GalNAc-T1. According to NetOGlyc our CD44 peptide may display up to 9 *O*-glycosites (S2, T6, S8, T9, T15, T20, S24, S27, T29 and T30); however, NetOGlyc analysis also indicated that only 7 had high probability of exhibiting O-glycosylation (S2, T6, S8, T9, T15, T20 and S24). Using the same rational applied for MUC16 synthesis, we used ISOGlyP to evaluate *in silico* glycosites distribution governed by GalNAc-T1. This approach identified residues T6 and T9 as those showing higher probability of glycosylation, followed by residues T29 and T30. As previously observed for MUC16, this approach supported the synthesis of multiple glycosylated peptides, supporting the rationale for a multivalent library reflecting tumour heterogeneity.

3.5. Characterization of CD44-Tn glycosites

To confirm CD44-Tn glycopeptide synthesis we analysed samples by nanoLC-ESI-MS. The analysis of these samples after C18 reverse phase separation showed formation of CD44-Tn species with a yield of approximately 70% (**Figure 6**). Overall, CD44 monoglycosylated species (54%) were predominant in relation to di and tri-glycosylated peptides (11 and 5%, respectively). Trace amounts of tretra-glycosylated MUC16 could also be observed (0.4%). Following the same protocol used for MUC16, we have also characterized the main glycosites for CD44-Tn from MS/MS spectra. The main glycosylated residues were T6 and T9. Trace amounts glycosylation could also be

observed in S8 and T15 (**Figure 7A**). These observations are partially in agreement with the predictive bioinformatics evaluation obtained using ISOGlyP. However, this prediction did not contemplate the glycosylation of residues S8 and T15, once again denoting the importance of experimental validation. For di-glycosylated forms, the most predominant glycosites were T6 and T9 and, to less extent, T9 and T15 (**Figure 7B**). Finally, for tri-glycosylated species we were able to identify T6, T9 and T15 as main glycosites (**Figure 7C**). Despite indications of presence of tetra-glycosylated peptides (**Figure 6**), we were not able to identify CD44-4Tn forms due to their very low abundances (**Figure 7**). Overall, it seems that mono glycosylated species with Tn on residues T6 and T9 are predominant in our CD44 sample. In summary, we have identified monoglycosylated CD44 as the dominant glycoform resulting from this synthesis, followed by di- and tri-glycosylated peptides. Facing these observations, we attempted to enrich the sample for more glycosylated species by VVA lectin chromatography. This increased the abundance of glycosylated species and decreased by 10% the presence of unglycosylated species; however, it also significantly enriched the sample for monoglycosylated peptides at the expenses of more glycosylated species (**Table 3**). Notably, the CD44 peptide elected for this study has up to 9 potential glycosites, implying the need to improve this synthesis to obtain a broader number of glycospecies. Future approaches may include longer incubation periods with GalNAc-T1 or considering the inclusion of other GalNAc-Ts to increase the number of glycosites, broadening the representation of our CD44 glycolibrary.

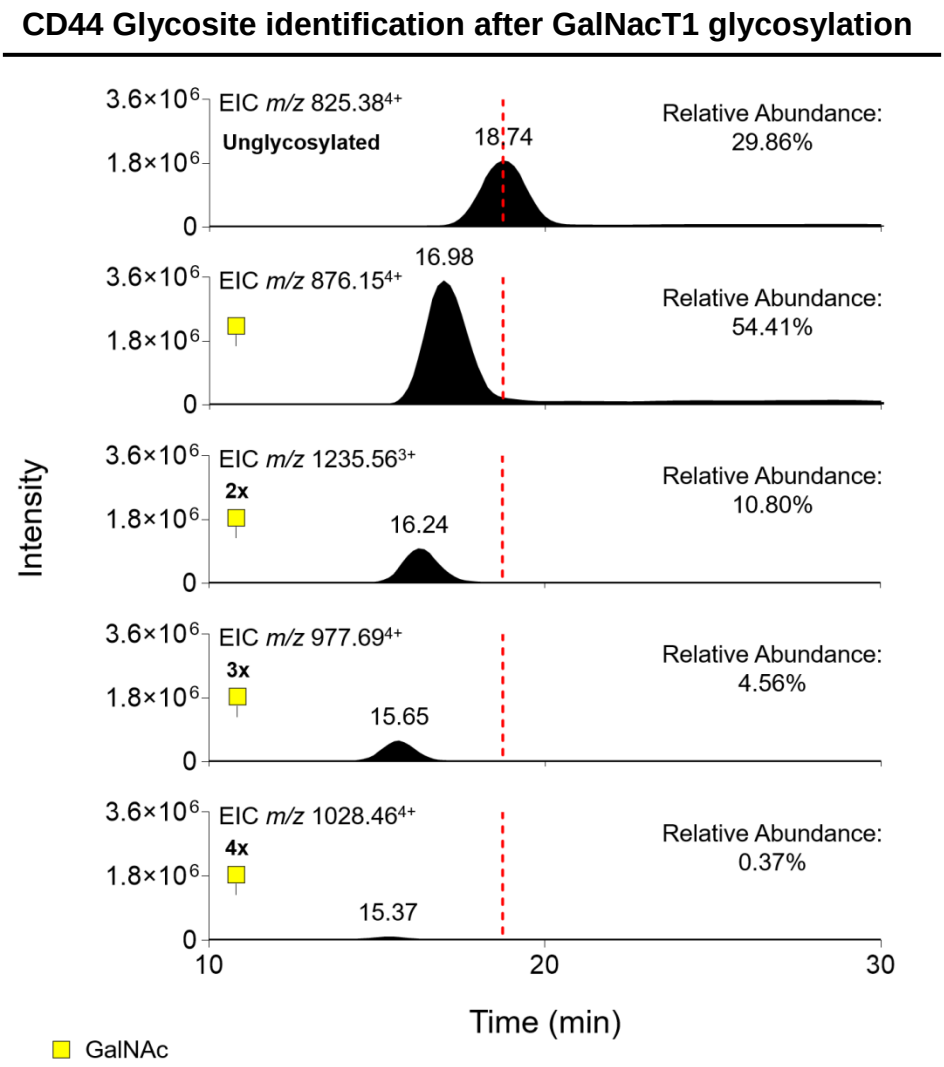


Figure 6. C18 reverse phase nanoLC-ESI-MS profiles for reaction products resulting from the glycosylation of the 30 mer CD44 peptide by GalNAc-T1. CD44 main glycosylated forms are mono-glycosylated peptides, followed by unglicosylated, di-glycosylated and tri-glycosylated forms, respectively. Peptides with 4 glycosylated sites can be detected in low abundances.

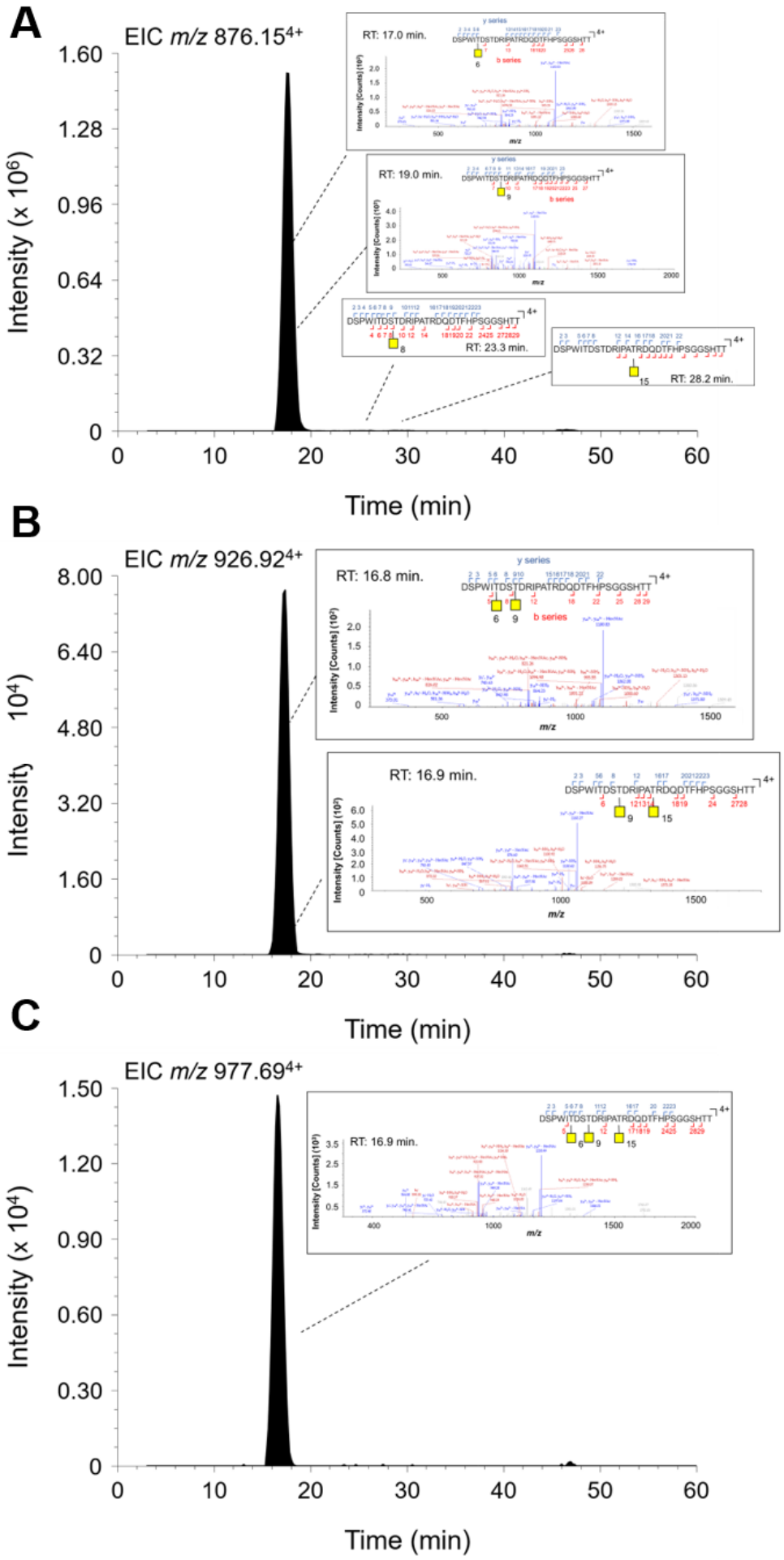

A
EIC m/z 876.15⁴⁺
RT: 17.0 min.
RT: 19.0 min.
RT: 23.3 min.
RT: 28.2 min.
Intensity (x 10⁶)
Time (min)
B
EIC m/z 926.92⁴⁺
RT: 16.8 min.
RT: 16.9 min.
Intensity (x 10⁴)
Time (min)
C
EIC m/z 977.69⁴⁺
RT: 16.9 min.
Intensity (x 10⁴)
Time (min)

Figure 7. Glycosite identification for 30 mer CD44 peptide after glycosylation by GalNAc-T1. A) Identification of T6 and T9 as the main glycosites in CD44-Tn glycopeptides. T6 appears as the preferential glycosite, with lower abundance of glycopeptides substituted at S8 and T15. **B) Identification of T6 and T9 as the main glycosites in CD44-2Tn glycopeptides.** The presence of T9/T15 were also observed but in lower abundance. **C) Identification of T6, T9 and T15 as the mains glycosites in CD44-3Tn.**

Table 3. Products from the single-pot enzymatic synthesis of CD44-Tn glycopeptides, without purification and after VVA enrichment, analysed by C18 reverse phase nanoLC-ESI-MS.

***O*-glycans**	**Glycopeptides $[M+2H]^{2+}$ / $[M+3H]^{3+}$ / $[M+4H]^{4+}$ / $[M+5H]^{5+}$**	**Retention time (min)**	**Without enrichment**	**Enrichment with VVA**
			Relative abundance (%)	**Relative abundance (%)**
Neutral glycopeptides				
Unglycosylated	1649.75 / 1100.17 / 825.38 / 660.51	18.55	29.86	19.55
1 Tn	1751.29 / 1167.87 / 876.15 / 701.12	17.50	54.41	76.49
2 Tn	1852.83 / 1235.56 / 926.92 / 741.74	17.35	10.80	3.04
3 Tn	1954.37 / 1303.25 / 977.69 / 782.35	16.57	4.56	0.70
4 Tn	2055.91 / 1370.94 / 1028.46 / 822.97	16.42	0.37	0.22

3.6. Characterization of synthesized CD44-STn glycopeptides

To obtain sialylated CD44 structures we used the same strategy applied for MUC16 based on the addition of ST6GalNAc-1 and CMP-Neu5Ac to the CD44-Tn reaction medium. CD44 peptides were sequentially submitted to TiO_2 enrichment. However, unlike MUC16, we were not able to identify sialylated peptides in our sample. In order to further address this issue we would need to analyse possible experimental flaws that might have compromised our system and results. Nonetheless, we can speculate that sialylation of this peptide would require the use of a different set of enzymes and protocol adaptations. As such, subsequent optimization would be required to address this issue.

4. Concluding Remarks

The synthesis of well characterized glycopeptides libraries is of key importance to support the development of cancer vaccines. Herein, we explore an easier, environmentally friendly, less time consuming and less expensive strategy for *O*-glycopetides synthesis, foreseeing the development of cancer vaccines in Chapter 2. Emphasis was set on the development of MUC16- and CD44-Tn/STn libraries reflecting a myriad of glycopatterns observed in bladder and, potentially, other solid tumours. The overall strategy builds on technology implemented in our group, which consists of bringing together in the same reaction medium synthetic peptides of interest, recombinant human glycosyltransferases, and sugar nucleotides, added in a stepwise manner, in accordance with glycans biosynthesis pathways *in vivo*. Similar approaches have also been used in the past to generate MUC1-Tn, -STn, antigens[18], allowing to overcome extensive, costly and non-environmental friendly chemical reactions as well as considerable low synthesis yields. For the initial reaction step that defines the number and distribution of glycosites in the peptide chains, we have elected GalNAc-T1 out of family of 20 possible glycosyltransferases encoded by the human genome. GalNAc-T1 was chosen because it was significantly overexpressed in more aggressive bladder tumours, according to a bioinformatics prediction using Oncomine. This option is expected to generate glycopeptide libraries that better reflect the nature of the glycosylation patterns found in tumours. The Tn-glycopeptides were subsequently sialylated by ST6GalNAc-1, also reported as overexpressed in invasive bladder tumours. We then engaged in a detailed characterization of the reaction products by nanoLC-MS/MS, which showed complex mixtures of both MUC16-Tn and STn and CD44-Tn glycopeptides with different degrees of glycosylation and heterogeneous glycosites occupation. High glycosylation yields were also observed for all reactions (approximately 70%). Notably, for MUC16 sialylation we were able to produce both fully sialylated glycopeptides as well as glycopeptides presenting both Tn and STn antigens. The introduction of TiO_2 affinity chromatography was also very effective to enrich these mixtures for sialylated species. Moreover, we concluded that further separation of both neutral and sialylated peptides into individual species could be achieved by C18 reverse phase chromatography; however, we believe that the mixtures reflect better the dynamic and heterogenous nature of glycosylation pathways *in vivo*. However, we were unable to produce CD44-STn using this chemo-enzymatic strategy. We hypothesize that the glycosites in the peptide chain might be inaccessible to ST6GalNAc-1. Modulating the length of the initial peptide, electing other GalNAc-Ts for the initial glycosylation step, changing the distribution of glycosites, or

optimizing the different enzymatic reactions in terms of enzyme and substrate concentrations may partially circumvented this limitation, which should be explored in future studies. The introduction of alternative methods for production of CD44-STn glycoepitopes such as cell factories that may be used also to scale-up their production maybe a valuable alternative, which has been explored in more detail in chapter 3.

In summary, we have created multivalent MUC16 and CD44 glycopeptide libraries reflecting, at least in part, the glycosylation observed by us in bladder tumours and possibly many other solid tumours exhibiting the same molecular features. These libraries will be crucial to support the development of vaccine constructs that constitute the main objective of chapter 2.

Bibliography

1. Jason E. Hudak1, 4 and Carolyn R. Bertozzi. Glycotherapy: New Advances Inspire a Reemergence of Glycansin Medicine. **21**, 16–37 (2014).

2. Zhang, P. *et al.* Challenges of glycosylation analysis and control: An integrated approach to producing optimal and consistent therapeutic drugs. *Drug Discov. Today* **21**, 740–765 (2016).

3. Ajit Varki, Richard D Cummings, Jeffrey D Esko, Hudson H Freeze, Pamela Stanley, Carolyn R Bertozzi, Gerald W Hart, M. E. E. *Essentials of Glycobiology*. (2009).

4. Schmaltz, R. M., Hanson, S. R. & Wong, C. Enzymes in the Synthesis of Glycoconjugates. 4259–4307 (2011) doi:10.1021/cr200113w.

5. Jihen Ati, 1 Pierre Lafite, 1 and Richard Danielloucorresponding author1. Enzymatic synthesis of glycosides: from natural O- and N-glycosides to rare C- and S-glycosides. *Beilstein J Org Chem* **13**, (2017).

6. Shoda, S., Uyama, H., Kadokawa, J., Kimura, S. & Kobayashi, S. Enzymes as Green Catalysts for Precision Macromolecular Synthesis. (2016) doi:10.1021/acs.chemrev.5b00472.

7. Huang, Z. *et al.* Strategy for Designing a Synthetic Tumor Vaccine: Multi-Component, Multivalency and Antigen Modification. 549–562 (2014) doi:10.3390/vaccines2030549.

8. Polypeptide, N. *et al.* CARBOHYDRATES , LIPIDS , AND OTHER NATURAL PRODUCTS : Substrate Specificities of Three Members of the Human UDP- N -Acetyl- α Substrate Specificities of Three Members of the Human UDP- N -Acetyl- ␣ - D -galactosamine : Polypeptide N -Acetylgalactosaminylt. (1997) doi:10.1074/jbc.272.38.23503.

9. Pratt, M. R. *et al.* Deconvoluting the Functions of Polypeptide N - ␣ - Acetylgalactosaminyltransferase Family Members by Glycopeptide Substrate Profiling. **11**, 1009–1016 (2004).

10. Jonathon E Mohl1,2 Thomas A Gerken1,3 Ming-Ying Leung1, 2. ISOGlyP: de novopredictionofisoform specific mucin-type O-glycosylation. (2020).

11. Kong, Y. *et al.* Probing polypeptide GalNAc-transferase isoform substrate specificities by in vitro analysis. *Glycobiology* **25**, 55–65 (2015).

12. Yoshimura, Y., Denda-nagai, K., Takahashi, Y. & Nagashima, I. Products of Chemoenzymatic Synthesis Representing MUC1 Tandem Repeat Unit with T- , ST- or STn-antigen Revealed Distinct Specificities of Anti-MUC1 Antibodies. 1–12 (2019) doi:10.1038/s41598-019-53052-1.

13. Naor, D., Sionov, R. V. & Ish-Shalom, D. CD44: Structure, function, and association with the malignant process. *Adv. Cancer Res.* **71**, 241–319 (1997).

14. Rhodes, D. R. *et al.* Oncomine 3 . 0 : Genes , Pathways , and Networks in a Collection of 18 , 000 Cancer Gene Expression Profiles 1. **9**, 166–180 (2007).

15. Azevedo, R., Silva, A. M. N., Reis, C. A., Lara, L. & Alexandre, J. In silico approaches for

unveiling novel glycobiomarkers in cancer. **171**, 95–106 (2018).

16. Gerken, T. A., Raman, J., Fritz, T. A. & Jamison, O. Identification of common and unique peptide substrate preferences for the UDP-GalNAc:polypeptide α-N-acetylgalactosaminyltransferases T1 and T2 derived from oriented random peptide substrates. *J. Biol. Chem.* **281**, 32403–32416 (2006).

17. Azevedo, R., Silva, A. M. N., Reis, C. A., Santos, L. L. & Ferreira, J. A. In silico approaches for unveiling novel glycobiomarkers in cancer. *J. Proteomics* **171**, 95–106 (2018).

18. Malekan, H. *et al.* One-pot multi-enzyme (OPME) chemoenzymatic synthesis of sialyl-Tn-MUC1 and sialyl-T-MUC1 glycopeptides containing natural or non-natural sialic acid. *Bioorganic Med. Chem.* **21**, 4778–4785 (2013).

19. Ricardo, S. *et al.* Detection of glyco-mucin profiles improves specificity of MUC16 and MUC1 biomarkers in ovarian serous tumours. *Mol. Oncol.* **9**, 503–512 (2015).

20. Singh, R. *et al.* Cell surface-expressed Thomsen-Friedenreich antigen in colon cancer is predominantly carried on high molecular weight splice variants of CD44. *Glycobiology* **11**, 587–592 (2001).

21. Palmisano, G. *et al.* Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat. Protoc.* **5**, 1974–1982 (2010).

CHAPTER II | Glyco-nanovaccines: a carrier of immunogenic glycoantigens for immunotherapy

Abstract

Bladder cancer (BC) remains a poorly managed cancer, with mainstay chemotherapy and immunotherapy treatments often failing to prevent tumour progression and recurrence. To address this hurdles novel more effective and safer therapeutic approaches are needed. Advanced stage bladder cancer cells frequently overexpress membrane proteins carrying aberrant short-chain *O*-glycans, given raise to cancer specific molecular fingerprints with theragnostic potential. Particularly, MUC16-STn glycoforms, typically found in ovarian cancers, have been previously reported by our group in a subset of advanced-stage bladder tumours facing worst prognosis, vowing for the clinical relevance of MUC16 glycopeptides. Accordingly, this work was set to design MUC16-Tn/STn glycopeptide based nanovaccines envisaging novel immunotherapeutic tools for BC.

Well characterized MUC16-Tn/STn glycopeptides were coupled to keyhole limpet hemocyanin (KLH) carrier proteins to render them more immunogenic, following PLGA nanoparticle encapsulation. Produced monodispersed nanoformulations showed spherical morphologies, with sizes bellow 200nm and a negative net charge. Additionally, all nanoformulations provided controlled release of glycoconjugates at physiological pH, supporting potential for parenteral administration. Moreover, PLGA nanoparticles displayed preferred uptake by macrophages compared to epithelial-like cell lines with no impact on cellular viability, vowing for the potential of glycan-based nanovaccines to target antigen presenting cells.

We anticipate that these preliminary findings may constitute an important milestone to establish novel glycan-based nanovaccines for bladder cancer.

1. Introduction

Bladder cancer (BC) is one of the most common genitourinary malignancies, with 549,393 new cases and 199,922 cancer-related deaths in 2018 [1]. Its high prevalence and recurrence rates, allied to frequent progression despite local therapy [2], makes BC the most expensive cancer to treat [3,4]. Locally advanced tumours are commonly subjected to repeated cycles of chemotherapy to reduce tumor size, risk of metastization and control disease dissemination [5]. However, half of patients develop metastasis within 5 years after curative treatment and many cannot tolerate or complete systemic treatment due to severe treatment side-effects[6]. The treatment of advanced metastatic urothelial carcinoma has recently evolved with the approval of five checkpoint inhibitors [7]. However, improvements in survival are still modest and substantial acquired resistance as well as side effects support the additional investment in novel, safer and more effective immunotherapies for improved treatment outcomes and reduce toxicity. In this context, focus has been set on the development of cancer vaccines which are expected to induce tumor specific immune responses able to either eliminate malignant cells or keep it under constant restraint, delaying tumor recurrence and prolonging survival.

Nanoscale delivery vehicles have improved the efficacy of bioactive agents, while being relevant intermediates in the crosstalk with the immune system, with several formulations being available for different tumour types [8,9]. Namely, nanoparticles (NPs) increase the circulation time of the conjugated or entrapped therapeutic compounds by exploiting the tortuous and poorly differentiated vasculature of solid tumors that, in contrast to healthy tissues, allow the extravasation of formulations with sizes up to several hundred nanometers [10]. Solid tumors also lack functional lymphatic systems, making them unable to eliminate nanomaterials. Consequently, long-circulating nanomedicines tend to accumulate in tumors over time, a mechanism known as enhanced permeability and retention effect [11,12]. Moreover, nanomedicines have improved the pharmacokinetics and efficacy of bioactive agents, while reducing systemic toxicity [13,14]; nevertheless, few studies explored their potential to treat bladder cancer in comparison to other models [15,16]. In particular, poly (lactic-co-glycolic acid) (PLGA) is one of the most biocompatible polymers for NPs formulation envisaging controlled release. Its attractive properties include: (i) biodegradability and biocompatibility; (i) well established formulations and production methods adapted to various types of small or macromolecules of either hydrophilic or hydrophobic nature; (ii) approval by EMA and FDA, (iv) protection of drugs

from degradation, (v) controlled release, (vi) possibility to graft the surface with different functional groups and/or biomolecules capable of providing stealthiness and/or improving interaction with biological materials/milieus and (vii) possibility to target NPs to specific cells and organs, significantly improving its efficiency [17].

Malignant transformation is accompanied by significant alterations in cell surface glycosylation, holding tremendous potential for cancer detection and targeted therapeutics [18,19]. It has been long described that advanced stage bladder tumours, its metastasis and even circulating tumour cells overexpress the STn antigen, a simple mucin type *O*-glycan responsible for arresting *O*-glycosylation extension [20,21]. On the same note, glycoproteomic analysis of advanced bladder tumours based on enzymatic treatments, lectin-affinity chromatography enrichment and nanoLC-ESI-MS/MS analysis resulted in the identification of several key cancer-associated glycoproteins (MUC16, CD44, integrins) carrying altered glycosylation. Of particular interest were MUC16 STn+-glycoforms, characteristic of ovarian cancers, which were found in a subset of advanced-stage bladder tumours facing the worst prognosis [22]. Given these findings, herein we describe the synthesis of immunogenic constructs resulting from the engraftment of KLH imunogens to MUC16-Tn/STn glycopeptides, envisaging the assembly of a multivalent bladder cancer vaccine. We also address the potential of nanoencapsulation to improve the efficacy of nanovaccines, providing a physico-chemical characterization of these constructs, and their stability at different pHs, thus establishing novel formulations for future tests *in vitro* and *in vivo.*

2. Material and methods

2.1. MUC16 glycopeptides conjugation to KLH

Cysteine-tagged MUC16-STn and MUC16-Tn glycopeptides were conjugated to keyhole limpet hemocyanin (KLH, Sigma-Aldrich), to render immunogenic glycoepitopes. Briefly, 0.01mg/µL KLH in 0.01 M phosphate buffer (pH 7, Sigma-Aldrich) were incubated for 30 min at room temperature under agitation with 0.015mg/µL 3-maleimidobenzoic acid N-hydroxysuccinimide ester (MBS, Sigma-Aldrich) in dimethylformamide (DMF, Sigma-Aldrich). Posteriorly, KLH/MBS solution was passed through a PD-10 desalting column (Sigma-Aldrich) conditioned with 0.05 M phosphate buffer (pH 6). Desalted KLH/MBS solution was incubated with 10 µg of glycopeptides in 25 µL DMF, pH 7.0, and incubated at 4°C overnight under agitation, following evaporation in a SpeedVac.

2.2. Nuclear Magnetic Resonance (NMR)

KLH, KLH-MUC16-Tn and KLH-MUC16-STn peptides were dissolved in 60% Dimethyl Sulfoxide-d6 in D2O and analyzed for 1H by NMR. NMR spectra were acquired at room temperature, on a Brüker AMX 300 spectrometer operating at 400.13 MHz. The chemical shifts are expressed in δ (ppm) values relative to tetramethylsilane (TMS) as internal reference and coupling constants (J) are given in Hz.

2.3. Dot Blot

The products of conjugation with KLH were analyzed by dot blot in a 0.45 µm polyvinylidene difluoride (PVDF; ThermoFisher Scientific) membrane. Membranes were blocked using Carbo-free Blocking Solution (1x; Vector Laboratories) and blotted for Tn antigen using biotinylated VVA lectin (1:5000) (Vector Laboratories) for 1 hour at room temperature. Subsequently, membranes were incubated with vectastain elite ABC reagent, Peroxidase, R.T.U. (Vector Laboratories) followed by incubation with ECL prime (GE Healthcare AmershamTM) and signal detection in a ChemiDoc Imaging System (BioRad)

2.4. Nanoparticle production

PLGA nanoparticles containing encapsulated glycopeptides were generated by a modified water-in-oil-in-water (w/o/w) double emulsion-solvent evaporation technique [23,24]. Briefly, 20 mg of 50:50 (Lactide:Glycolide) - Poly(Lactide-co-Glycolide) (PLGA, Corbion-Purac Biomaterials) was dissolved in 2mL ethyl acetate (Sigma-Aldrich), following the addition of 50 µg KLH conjugated to 20 µg of glycosylated or non-glycosylated peptide. Emulsification occurred at 70% amplitude for 30s using a Vibra-Cell™ ultrasonic processor in an ice bath. The second emulsion was achieved by adding 2% Pluronic F-127 (4mL) (Sigma-Aldrich) to the first emulsion, following Vibra-Cell™ ultrasonic homogenization for 60s and addition of the resulting mixture to 15mL of the same surfactant. Ethyl acetate evaporation from the final solution occurred for 3h under magnetic stirring at 250 rpm. Nanoparticles were isolated by centrifugation at 10 000 g for 45 min. Fluorescent PLGA-FITC NPs were obtained using a modified nanoprecipitation method [25]. Accordingly, 16mg PLGA-FITC plus 4mg PLGA 50:50 was dissolved in 1 mL acetone (Sigma-Aldrich). This organic solution was transferred through a 25 G needle to an aqueous solution of 1% Pluronic F-127 (15 mL) and solvent evaporations occurred under magnetic stirring at 300 rpm for 5h. A washing step with Milli-Q water was performed.

2.3. Nanoparticles physicochemical and morphological characterization

The size and polydispersity index (PdI) were determined by Dynamic Light Scattering (DLS), while surface charge of NPs was determined by laser Doppler anemometry (LDA). After dispersion, NPs were diluted in 10mM sodium chloride solution (pH 7.0) (Sigma-Aldrich) and DLS and LDA were performed with a Nano ZS Zetasizer (Malvern, Worcestershire, UK) at 25 °C. Surface morphology and NPs size confirmation were performed by transmission electron microscopy (TEM) using a JEM-1400 microscope (JEOL, Tokyo, Japan) at an acceleration voltage of 80 kV.

2.4. Association efficacy

Supernatants from the encapsulation of KLH-MUC16 conjugates were quantified using the Pierce Micro BCA Protein Assay Kit (ThermoFisher Scientific) according to manufacturer's instructions to determine association efficacy. Association Efficacy (AE) was determined using the following equation:

$$\text{AE (\%)} = \frac{\text{Initial mass of loaded protein} - \text{mass of protein in the supernatant}}{\text{Total mass of loaded protein}} \times 100$$

2.5. Glycoconjugate controlled release

Envisaging analysis of possible oral and parenteral administration of nanoformulations, the release profiles of glycopeptide-loaded PLGA-NPs were determined in PBS at physiological pH 7.4 and gastric pH 3.0 under agitation at 37°C. Briefly, 50µl of NPs suspensions supernatant were collected at 0, 0,5, 1, 2, 4, 8, 12, 24, 48, 72 and 96h and replaced with the same amount of buffer. The supernatants were centrifuged at 10,000g for 20min and quantified using the Pierce Micro BCA Protein Assay Kit.

2.6. Cell culture conditions

The human bladder cancer cell line T24, gastric cancer cell line AGS and the monocyte cell line THP-1 were purchased from ATCC and cultured with RPMI 1640+GlutaMAX™-I medium (Gibco, Life Technologies), supplemented with 10% heat-

inactivated FBS (Gibco, Life Technologies) and 1% penicillin–streptomycin (10,000 Units/mL penicillin; 10,000 mg/mL streptomycin; Gibco, Life Technologies). Cell lines were cultured at 37 °C in a 5% CO2 humidified atmosphere. THP-1 cells were differentiated into M0 macrophages using 100ng/mL phorbol myristate acetate (PMA) for 48h, following 24h in standard culture medium prior to any experiment.

2.7. Nanoparticles cytotoxicity profiles

The red-fluorescent nuclear and chromosome counterstain Propidium iodide (PI) was used to detect dead cells after nanoformulation exposure by Flow cytometry using a FC500 (Beckman Coulter) cytometer. Data analysis was performed through Infinicyt^{TM} (Cytognos) Software.

2.8. Nanoparticle internalization assays

The interactions between NPs, cancer cell lines (T24 and AGS) and macrophages were evaluated by fluorescence microscopy and flow cytometry, using and PLGA-FITC NPs. For fluorescent microscopy assays, cells were seeded in 12-well plates and incubated for 4h at 37°C with 0.004mg/mL, 0,04 mg/mL or 0,4 mg/mL PLGA-FITC NPs in OptiMEM. Cells were fixed with 4% PFA and stained with CellMaskTM Orange 0.5x (Invitrogen) for 6min at room temperature. The uptake of fluorescent NPs was evaluated under inverted fluorescence microscopy (Leica DMI 6000 from Leica), detecting PLGAc NPs through the GFP channel and CellMaskTM stained cells through the TXR channel. Tilescan analysis was performed using the LAS X software. For flow cytometry analysis, tumour cells were detached through a non-enzymatic method (Versene 1x, Gibco) and macrophages were detached with Acutase cell detachment solution (Gibco). An FC500 (Beckman Coulter) cytometer was used and data analysis was performed through InfinicytTM (Cytognos) Software.

2.9. Statistical analysis

All experiments were performed in triplicate and represented as mean ± standard deviations (SD). Statistical differences were appointed using Student's T-test unpaired. The level of significance was set at probabilities of "*" $p < 0.05$; "**" $p < 0.01$; "***" $p < 0.001$; and "****" $p < 0.0001$.

3. Results and Discussion

The present work focusses on the development of novel biocompatible PLGA-based NPs loaded with glycoantigen payloads, envisaging future multivalent vaccine design. The overall objective of this preliminary work was to render relevant and well-characterized Tn and STn-MUC16 glycopeptides more immunogenic through coupling with immunostimulant carriers for improved immunological response. Furthermore, nanoencapsulation of such immunoconjugates was attempted in order to improve their stability and possibly provide controlled release under specific physiological conditions. Moreover, such encapsulation strategies usually allow the delivery of glycoantigens while providing means to surpass systemic administration issues, such as indiscriminate cytotoxicity and potential off-target effects.

3.1. Glycopeptide Conjugation with immunogenic carriers

Using the MUC16 glycopeptide library successfully developed and characterized in chapter 1, we started by coupling MUC16-Tn/Stn glycopeptides to the carrier protein KLH (keyhole limpet hemocyanin). Because peptides alone are generally too small to elicit effective immune responses, carrier proteins like KLH, which contain several epitopes, are used to facilitate the induction of cellular and humoral immune responses[27]. Specifically, glycopeptides were conjugated to KLH via a maleimide linker, using a well-established and validated click-chemistry reaction routinely used in our lab [25]. For rapid validation of conjugation, ^{1}H NMR spectroscopy was used. The NMR spectra for KLH, despite diluted, showed typical signals attributable to amino acids alpha, aliphatic and methyl protons (**Figure 8A**). The conjugation of KLH to MUC16-Tn and STn glycopeptides significantly changed the ^{1}H NMR profile, supporting glycopeptide-KLH conjugation. Namely, both spectra presented novel signals at chemical shifts typical of sugars anomeric protons at δ 4.5-6.5 ppm, ring protons between δ 3.5-4.5 ppm and *N*-acetyl groups of GalNAc δ 2.0-3.0 ppm (**Figure 8B** and **C**). Several other less classic signals may be attributed to an increase in the number of amino acids derived from MUC16 peptide. Moreover, **Figure 8C** also exhibits new signals that may correspond to the introduction of the Neu5Ac residue between δ 2.0-3.0 ppm, typical of H3 protons at different chemical environments. A Dot Blot analysis was used to further address peptide conjugation (**Figure 8D**). Blotting for VVA demonstrated a major dot in the slot loaded with KLH-MBS-MUC16-Tn, whereas

KLH and KLH-MBS slots displayed no reactivity for the lectin, as expected. These observations further suggest glycopeptide coupling with the immunogenic carrier.

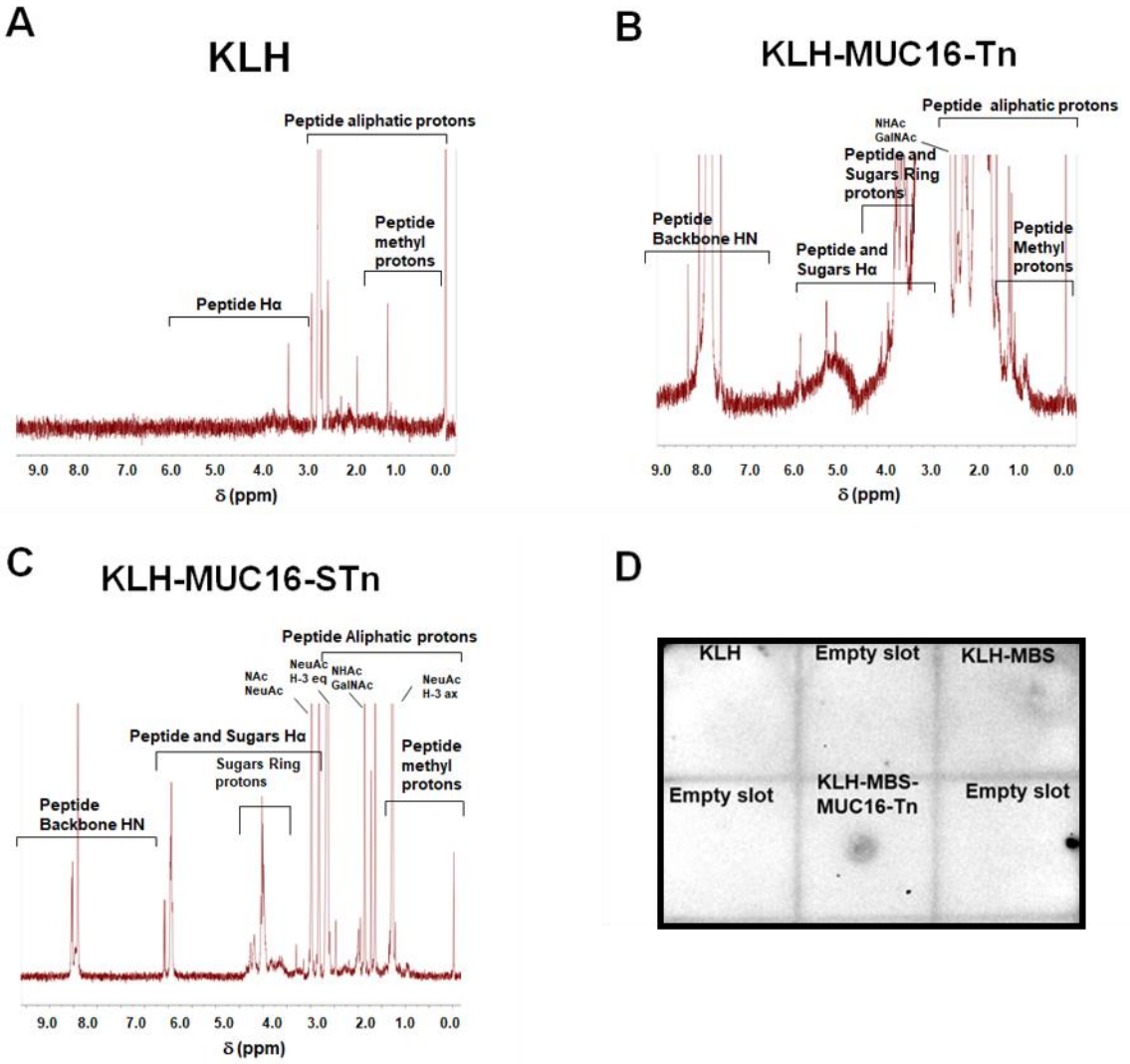


Figure 8. ^{1}H NMR spectra and Dot Blot analysis of KLH, KLH-MUC16-TN and STn conjugation profiles. **A)** Spectra profile for KLH molecules; **B)** KLH-MUC16-Tn spectra profiles display different characteristics compared to the KLH RMN analysis, associated with the presence of glycopepetide structures; **C)** Spectra displaying the profile of KLH-MUC16-STn samples; **D)** Dot Blot analysis using VVA blotting for Tn structures; KLH -MUC16-Tn shows a distinct reactivity for the lectin, whereas KLH and KLH-MBS display no reactivity for Tn blotting.

3.1. Nanoparticle Characterization

Biocompatible Poly (lactic-co-glycolic acid) (PLGA) nanoparticle formulations have been shown to shield encapsulated materials from degradation in circulation, while

promoting controlled release of NPs content [24]. Moreover, its versatile chemical nature opens an avenue for NPs functionalization to optimize specific cellular targeting. Accordingly, envisaging future glyconanovaccines based on MUC16 glycopeptides, we proceeded with the encapsulation of glycopeptide-carrier constructs into PLGA nanoparticles.

PLGA NPs were produced through the double emulsion-solvent evaporation technique and modified nanoprecipitation method. Different PLGA formulations, including empty controls and FITC fluorescent particles, were analysed according to their physiochemical and morphological properties, such as mean size, size distribution, surface charge and association efficiency (AE) (**Table 4**). Nanoparticle size, as determined by dynamic light scattering (DLS), ranged between 143 and 153 nm. Transmission electron microscopy (TEM) was used to confirm the aspheric shape of NPs as well as the size range **(Figure 9)**, which was invariably smaller than 200 nm. The size of nanoparticles can influence specific uptake and accumulation in relevant tissues. Specifically, particles smaller than 200 nm have been described to rapidly drain to lymph nodes (LN), where they actively prime antigen presenting cells as dendritic cells (DCs) and consequently promote T-cell effector responses [26,27]. These observations support that any of the developed NPs formulations carrying immunogenic glycopeptides could potentially mount a successful DC response.

Furthermore, the polydispersity index (PdI) of empty and glycoconjugate loaded NPs revealed some formulation heterogeneity. Empty and KLH-MUC16-STn loaded NPs provided highly monodisperse formulations (PdI <0.1) [28], while KLH-MUC16-Tn and KLH-MUC16 loaded NPs showed increased polydispersity indexes suggesting mild particle aggregation. KLH-MUC16-STn constructs exhibited the lowest PdI values and the smallest size particles, which could relate to STn associated negative charges enhancing particles electrostatic repulsions[29].

All formulations were negatively charged and displayed zeta potentials within a desirable interval of ±20 mV for nanosuspensions stability [30] (**Table 4**).

The association efficacy (AE) of glycopeptides to PLGA-NPs was indirectly determined by protein quantification of encapsulation supernatants. Interestingly, the association efficacy of glycopeptides to NPs increases with the complexity of the encapsulated glycopeptide. This might be due to increased hydrophilicity of sialylated peptides compared to non-sialylated or non-glycosylated counterparts, since the water-in-oil-in-water (w/o/w) double emulsion-solvent evaporation technique favours hydrophilic compound encapsulation.

Overall, we were able to develop a novel PLGA nanoparticle formulation comprising the successful encapsulation of imunoconjugated glycopeptides. Moreover, the sizes, polydispersity indexes, surface charges and association efficiencies of the generated NPs offer promising upfront characteristics for possible dendritic cell priming in future pre-clinical contexts.

Table 4. ***Properties of empty, encapsulated, and fluorescent PLGA-FITC NPs,*** *including particles mean size in nm, polydispersity index (PdI), surface charge (mV) and encapsulation efficiency (AE%). The values are represented as mean values ± SD (n = 3).*

Nano-construct	Encapsulated peptides	Size (nm)	PDI	Surface Charge (mV)	AE (%)
PLGA	Empty	148.5±0.379	0.088±0.010	-3.50±0.309	NA
	KLH-MUC16	149.3±1.620	0.160±0.018	-6.0±1.28	53,3
	KLH-MUn16 Tn	153.5±1.002	0.126±0.017	-4.61±0.194	75,6
	KLH-MUC16 sTn	143.6±0.265	0.042±0.006	-4.41±0.509	90,4
PLGA-FITC	Empty	303.3±8.5	0.304±0.006	-0.32±0.0236	NA

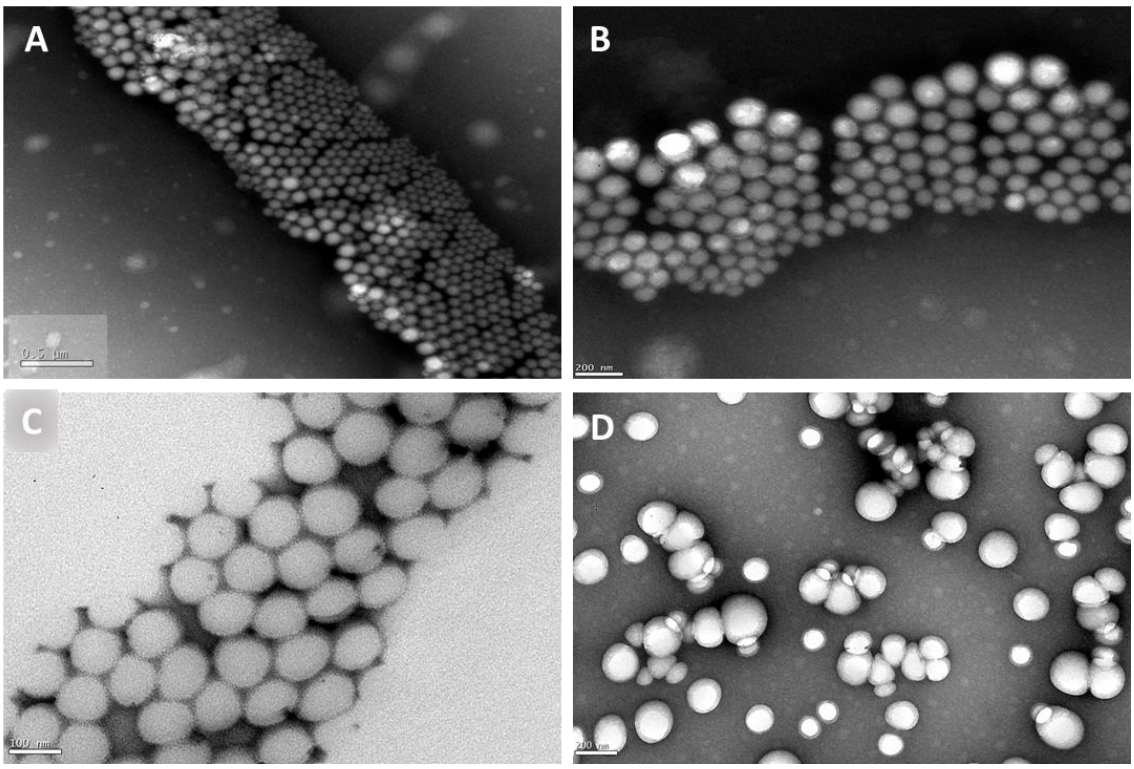


Figure 9*.* ***Transmission Electron Microscopy (TEM) analysis of.*** *A) Empty PLGA nanoparticles; B) KLH-MUC16 PLGA particles; C) KLH-MUC16-Tn PLGA particles;D) KLH-MUC16-STn PLGA particles. Scale bar marking 200 nm.*

3.2. Glycoconjugate controlled release

Envisaging the parenteral or the oral administration of glyconanovaccines, the release profiles of glycopeptide-loaded PLGA-NPs were determined at the gastric pH 3 and at the physiological pH 7.4 **(Figure 10)**. At pH 3, both glycopeptides and non-glycosylated peptide experience a burst release of close to 100 % of NPs content in the first 20 hours. This data suggests that these nanoformulations would not be ideal for oral administration, since upon relative short-term interactions with gastric environments they would release most of the loaded peptide, failing to reach the circulatory system with high peptide loads. At physiological pH 7.4, a controlled released of peptides occurs for all formulations; however, at considerably different rates. Specifically, compared to the non-sialylated glycopeptide (**Figure 10 B**), KLH-MUC16 (**Figure 10 A**) and KLH-MUC16-STn (**Figure 10 C**) constructs seem to be released throughout a longer period of time, only reaching a 50% loading release after 100 hours assay. From a clinical application standpoint, rapid release of antigens can lead to tolerance, whereas prolonged release frequently induces potent immune responses [31]. Accordingly, the described tendency for increased release rates of KLH-MUC16-Tn *in vitro* could possibly indicate less probability of APC priming, whereas, the controlled over time release of KLH-MUC16 and KLH-MUC16-STn constructs could render these immunogenic peptides more fit for immune stimulation in the context of vaccine development [31].

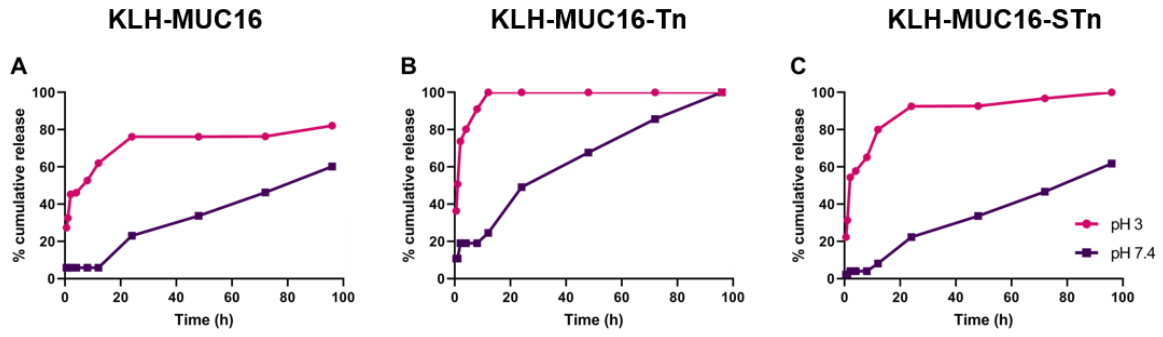


Figure 10. Glycoconjugate cumulative release at pH 3 and pH 7.4. Conditioned media was collected at 0h, 0,5h, 1h, 2h, 4h, 8h, 12h, 24h, 48h, 72h and 96h **A) Release profile of encapsulated KLH-MUC16 peptides from PLGA nanoparticles.** At pH 3, a burst release occurs at pH 3 within the first 24h, following a very slow and incomplete release during the next 76h. At pH 7.4 a very slow peptide release was recorded up until 12h assay, followed by an increasing but incomplete release during the next 76h**; B) Release profile of encapsulated KLH-MUC16-Tn peptides from PLGA nanoparticles**: at pH 3; a complete release is achieved withing the first 24h, while at pH 7.4 a controlled and increasing peptide release occurs within 96h **C) Release profile of encapsulated KLH-MUC16-STn peptides from PLGA nanoparticles**: At pH 3. Close to 100% of NPs content is released within the first 24h, followed small release increments until full release at 96h; At pH 7.4, sialylated glycopeptides are slowly released throughout 100h assay, achieving only approximately 60% of total nanoparticle content release.

3.3. Cellular Uptake Assays

The interactions between NPs, epithelial-like (T24 and AGS) and human monocytic (THP-1) cell lines were evaluated by fluorescence microscopy and flow cytometry, using empty PLGA-FITC NPs **(Figure 11)**. Increasing concentrations of fluorescent NPs were tested to determine the optimal concentration for NP internalization associated with the lowest cytotoxicity rates. Nanoparticle formulations up to 0.4 mg/mL do not comprise significant cytotoxicity, as cell viability for all nanoformulations was significantly above 70%, which is regarded as toxicity threshold according to ISO 10993-5 [32] (**Figure 11 A right graph**). Moreover, increasing nanoparticle concentrations have provided increased nanoparticle uptake as determined by FACS (**Figure 11 A left graph**) and fluorescent microscopy (**Figure 11B**). Moreover, for 0.4 mg/mL NPs nanoformulations, NPs seem to be easily internalized by THP-1 derived macrophages **(Figure 11A left graph)** as opposed to epithelial-like cells, which seem to mainly adsorb NPs at the cell surface without proper NPs engulfing (**Figure 11B**). These findings suggest PLGA NPs selectivity for phagocytic cells. Future studies shall include loaded NPs formulations for *in vitro* testing of dendritic cell activation, to broaden our knowledge on glyconanoparticles effect on professional antigen presenting cells.

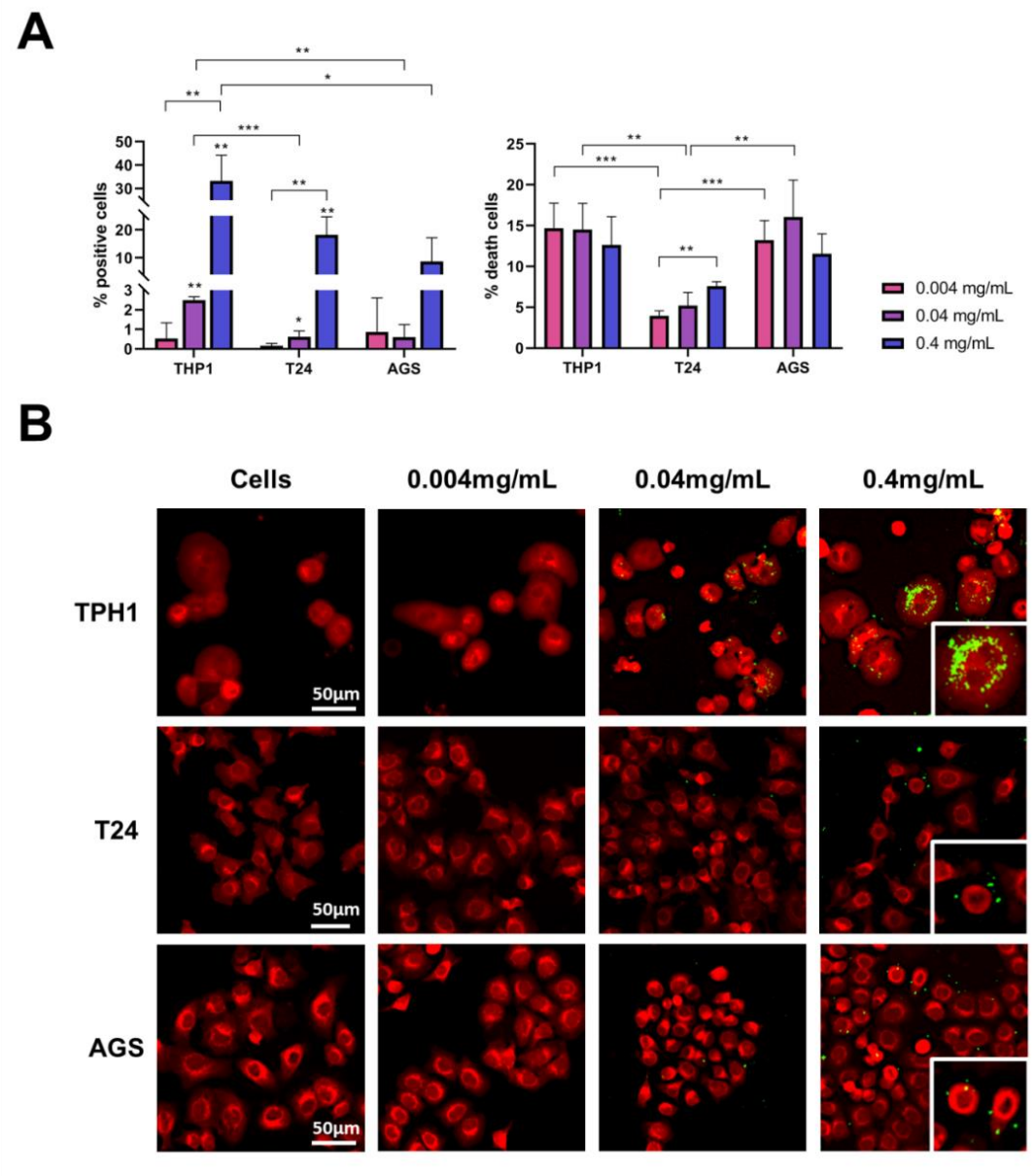


Figure 11. **A) Flow cytometry analysis of interactions between PLGA-FITC NPs with epithelial-like cell lines (AGS and T24, respectively) and differentiated THP-1 macrophages (M0)**. Crescent PLGA-FITC NPs concentrations were tested to determine formulations threshold cytotoxicity's and cellular uptake. THP-1 derived macrophages displayed increased PLGA-FITC NPs internalization compared to epithelial-like cell lines **(left graph)**. Propidium iodide exclusion flow cytometry has allowed the assessment of cytotoxicity profiles of the different NPs formulations, all of which were regarded non-significantly cytotoxic according with ISO 10993-5 **(right graph)**. **B) Fluorescence microscopy analysis of PLGA-FITC NPs interactions with epithelial and phagocytic cells.** Cells were stained with CellMask™ deep red plasma membrane stain and NPs are marked green. THP-1 derived macrophages actively internalize PLGA-FITC nanoparticle at working concentrations of 0,4 mg/mL, while epithelial-like cells seem to mainly adsorb nanoparticles at the cell surface.

4. Concluding Remarks

Bladder cancer remains one of the most common and deathly cancers of the genitourinary tract. Moreover, conventional therapy modest outcomes and severe toxicity of chemotherapy agents prompted the search for novel, less toxic and more effective target therapeutics, including immunotherapy tools. As such, the present work devoted to the development of novel biocompatible PLGA nanovaccines based on relevant immunogenic glycopeptide encapsulation. Cancer vaccines are expected to induce a tumor specific immune response able to either eliminate malignant cells or keep them under constant restraint, delaying tumor recurrence and prolonging survival. Both prophylactic and therapeutic vaccine-based cancer therapies have been proposed to enhance a specific immune response to tumor cells, mostly through dendritic cells (DCs) priming and consequent activation of cellular and human responses against the presented cancer antigen. Accordingly, DCs become a striking target for cancer vaccination. Of note, the herein developed glyconanovaccine formulations ranged from 140 to 150 nm particle size, making them suitable for DC priming. Particularly, while larger particles must rely on migratory DCs for the uptake at the administration site, delivery systems smaller particles can effectively drain to lymph nodes where DC priming may promptly occur [26]. Moreover, all nanoformulations provided controlled release of glycoconjugates at physiological pH, supporting potential for parenteral administration. In addition, PLGA nanoparticles displayed preferred uptake by macrophages compared to epithelial-like cell lines with no impact on cellular viability, vowing for the potential of glycan-based nanovaccines to target antigen presenting cells. Future studies should now focus on further *in vitro* testing of nanoformulations efficiency, namely, to induce DC uptake and activation. Furthermore, translating these findings into pre-clinical tests in relevant animal models, possibly by exploiting patient derived xenografts, would be of key importance to validate the clinical applicability of the developed tools.

Bibliography

1. International Agency for Research on Cancer. Bladder Globocan 2018. *Globocan* 1–2 (2019).

2. van Rhijn, B. W. G. *et al.* Recurrence and Progression of Disease in Non-Muscle-Invasive Bladder Cancer: From Epidemiology to Treatment Strategy. *Eur. Urol.* **56**, 430–442 (2009).

3. Leal, J., Luengo-Fernandez, R., Sullivan, R. & Witjes, J. A. Economic Burden of Bladder Cancer Across the European Union. *Eur. Urol.* **69**, 438–447 (2016).

4. Svatek, R. S. *et al.* The economics of bladder cancer: Costs and considerations of caring for this disease. *Eur. Urol.* **66**, 253–262 (2014).

5. Raghavan, D. Chemotherapy for invasive bladder cancer: Five simple rules learned over 30 years. *Bl. Cancer* **1**, 3–13 (2015).

6. Einstein, D. J. & Sonpavde, G. Treatment Approaches for Cisplatin-Ineligible Patients with Invasive Bladder Cancer. *Curr. Treat. Options Oncol.* **20**, (2019).

7. Pooja Ghatalia, Matthew Zibelman, D. M. G. and E. P. Approved checkpoint inhibitors in bladder cancer: which drug should be used when? *Ther. Adv. Med. Oncol. Rev.* **10**, 259–261 (2018).

8. Senapati, S., Mahanta, A. K., Kumar, S. & Maiti, P. Controlled drug delivery vehicles for cancer treatment and their performance. *Signal Transduct. Target. Ther.* **3**, 1–19 (2018).

9. Qiu, H. *et al.* Nanomedicine approaches to improve cancer immunotherapy. 1–23 (2018) doi:10.1002/wnan.1456.Nanomedicine.

10. Greish, K. Enhanced permeability and retention of macromolecular drugs in solid tumors: A royal gate for targeted anticancer nanomedicines. *J. Drug Target.* **15**, 457–464 (2007).

11. Golombek, S. K. *et al.* Tumor targeting via EPR: Strategies to enhance patient responses. *Adv. Drug Deliv. Rev.* **130**, 17–38 (2018).

12. Matsumura, Y. & Maeda, H. A New Concept for Macromolecular Therapeutics in Cancer Chemotherapy: Mechanism of Tumoritropic Accumulation of Proteins and the Antitumor Agent Smancs. *Cancer Res.* **46**, 6387–6392 (1986).

13. Patra, J. K. *et al.* Nano based drug delivery systems: Recent developments and future prospects 10 Technology 1007 Nanotechnology 03 Chemical Sciences 0306 Physical Chemistry (incl. Structural) 03 Chemical Sciences 0303 Macromolecular and Materials Chemistry 11 Medical and He. *J. Nanobiotechnology* **16**, 1–33 (2018).

14. Adjei, I. M., Temples, M. N., Brown, S. B. & Sharma, B. Targeted nanomedicine to treat bone metastasis. *Pharmaceutics* **10**, (2018).

15. Fernandes, E. *et al.* New trends in guided nanotherapies for digestive cancers: A systematic review. *J. Control. Release* **209**, 288–307 (2015).

16. Tomlinson, B., Lin, T. Y., Dall'Era, M. & Pan, C. X. Nanotechnology in bladder cancer: Current state of development and clinical practice. *Nanomedicine* **10**, 1189–1201 (2015).

17. Martins, C., Sousa, F., Araújo, F. & Sarmento, B. Functionalizing PLGA and PLGA Derivatives for Drug Delivery and Tissue Regeneration Applications. *Adv. Healthc. Mater.* **7**, 1–24 (2018).

18. Peixoto, A., Relvas-Santos, M., Azevedo, R., Lara Santos, L. & Ferreira, J. A. Protein glycosylation and tumor microenvironment alterations driving cancer hallmarks. *Front. Oncol.* **9**, 1–24 (2019).

19. Pinho, S. S. & Reis, C. A. Glycosylation in cancer: Mechanisms and clinical implications. *Nat. Rev. Cancer* **15**, 540–555 (2015).

20. JA, F. *et al.* Overexpression of tumour-associated carbohydrate antigen sialyl-Tn in advanced bladder tumours. *Mol. Oncol.* **7**, 719–731 (2013).

21. Neves, M. *et al.* Exploring sialyl-Tn expression in microfluidic-isolated circulating tumour cells: A novel biomarker and an analytical tool for precision oncology applications. *N. Biotechnol.* **49**, 77–87 (2019).

22. Cotton, S. *et al.* Targeted O-glycoproteomics explored increased sialylation and identified MUC16 as a poor prognosis biomarker in advanced-stage bladder tumours. *Mol. Oncol.* **11**, 895–912 (2017).

23. Gomes, M. J., Kennedy, P. J., Martins, S. & Sarmento, B. Delivery of siRNA silencing P-gp in peptide-functionalized nanoparticles causes efflux modulation at the blood-brain barrier. *Nanomedicine* **12**, 1385–1399 (2017).

24. Sousa, F. *et al.* A new paradigm for antiangiogenic therapy through controlled release of bevacizumab from PLGA nanoparticles. *Sci. Rep.* **7**, 1–13 (2017).

25. Fessi, H., Puisieux, F., Devissaguet, J. P., Ammoury, N. & Benita, S. Nanocapsule formation by interfacial polymer deposition following solvent displacement. *Int. J. Pharm.* **55**, 1–4 (1989).

26. Manolova, V. *et al.* Nanoparticles target distinct dendritic cell populations according to their size. *Eur. J. Immunol.* **38**, 1404–1413 (2008).

27. Koerner, J., Horvath, D. & Groettrup, M. Harnessing Dendritic Cells for Poly (D,L-lactide-co-glycolide) Microspheres (PLGA MS)-Mediated Anti-tumor Therapy. *Front. Immunol.* **10**, 1–16 (2019).

28. Clayton, K. N., Salameh, J. W., Wereley, S. T. & Kinzer-Ursem, T. L. Physical characterization of nanoparticle size and surface modification using particle scattering diffusometry. *Biomicrofluidics* **10**, 1–14 (2016).

29. Munkley, J. The role of sialyl-Tn in cancer. *Int. J. Mol. Sci.* **17**, (2016).

30. Gupta, V. & Trivedi, P. *In vitro and in vivo characterization of pharmaceutical topical nanocarriers containing anticancer drugs for skin cancer treatment*. *Lipid Nanocarriers for Drug Targeting* (Elsevier Inc., 2018). doi:10.1016/B978-0-12-813687-4.00015-3.

31. Li, W., Joshi, M. D., Singhania, S., Ramsey, K. H. & Murthy, A. K. Peptide vaccine: Progress and challenges. *Vaccines* **2**, 515–536 (2014).

32. (CSA), C. S. A. ISO 10993-5 in vitro cytotoxicity. *Int. Organ.* **2007**, 1–11 (2009).

CHAPTER III | Glycoengineered cell factory for production of CD44-STn glycoepitopes for multivalent glycan-based bladder cancer vaccines

Abstract

Bladder cancer (BC) is amongst the most common and deadliest genitourinary cancers, known to overexpress heavily sialylated glycoproteins, driving disease progression, and stemming from a premature stop in protein glycosylation. Moreover, cancer associated sialylated glycoproteoforms not observed in healthy urothelium offer a valuable means for selective targeting of malignant cells or the development of highly specific cancer vaccines. However, the clinical application of glycan-based therapeutic solutions has been significantly hampered by the poor bioavailability of pure and structurally characterized carbohydrates and glycoconjugates for pre-clinical studies and development of clinical grade tools. Herein, we propose the stabloshment of glycoengineered cell factories for stable production of glycoantigens, namely CD44 glycopeptides expressing homogenous cancer-associated glycans, envisaging future glycoconjugate-based vaccines.

T24 BC cells constitutively expressing CD44 proteoforms were genetically edited using CRISPR Cas9 technology to provide overexpression of Tn and STn glycoforms, while abrogating extension of *O*-glycan chains beyond T antigens. The membrane expression of STn antigen was further potentiated by *ST6GALNAC1* gene knock-in. Cell model's phenotype validation was obtained by FACs and immunocytochemistry. The resulting cell factory (C1GALT1 KO ST6GALNAC1 KI) has maintained the transcriptional levels of CD44 and its experimentally determined isoforms compared to the wild-type system, as determined by qRT-PCR. Moreover, *in situ* Proximity Ligation assays (PLA) have demonstrated that the developed cell model significantly overexpressed CD44-STn glycoforms compared to non-edited cells.

In summary, this work sets the molecular rational for the establishment of cell factories as alternative tools to chemical and enzymatic biosynthesis of relevant glycopeptides. Future studies should focus on further cell factory refinement and the context-oriented glycomapping of relevant CD44 glycopeptides, envisaging upcoming multivalent glycan-based vaccines.

1. Introduction

Bladder cancer is one of the deadliest genitourinary cancers, with chemotherapy and immune checkpoint inhibitors immunotherapy being current therapeutic options[1–3]. However, chemotherapy often fails to prevent tumour recurrence, and checkpoint inhibitors immunotherapy resistance is becoming an increasingly manifesting reality. Given this hurdles, safer and more effective immunotherapies are warranted.

Glycosylation constitutes a key posttranslational modification of proteins, significantly increasing protein complexity and directly impacting protein function [4]. In addition, state-of-the-art knowledge supports the relevance of glycans in pathogen recognition, inflammation, innate immune responses, and cancer [5]. Carbohydrate-based vaccines have been at the forefront of efforts to bring carbohydrate chemistry into clinically relevant platforms. However, a foremost hurdle in glycobiology and glycan-based therapeutics is the lack of pure and structurally well-defined carbohydrates and glycoconjugates [6,7]. These complexes are often found in low concentrations and in highly heterogeneous forms, negatively impacting isolation protocols and characterization. Currently, well-defined oligosaccharides and glycopeptides are often obtained by chemical- or enzymatic synthesis, which remains costly, time-consuming, and low yield [8–10]. Given these challenges, cellular engineering for the enhanced production of well-defined glycopeptides has emerged as a cost-effective and easily scalable alternative[11,12]. Overall, the successful high-yield process for production of a given protein requires high transcription and translation of the gene, correct folding, the desired post-translational modifications, and limited or no degradation of the product [13]. Accordingly, the user must choose the cell factory based on the properties and applications of the desired protein. Envisaging glycopeptide production, engineering the post-translational modification machinery of cells constitutively expressing the preferred protein would be the most straightforward methodology for optimized glycoproteoforms purification titers. Moreover, choosing a cell factory constitutively expressing the required protein could circumvent the need for genetic editing to overexpress it, while minimizing stress responses associated to induced protein overexpression, which may result in product degradation[14–16]. Given these challenges, the dynamic design of cell models and genome editing tools hold great promise for the development of relevant cell factories for glycopeptide production.

CD44 is a heavily glycosylated membrane receptor playing a key role in cell adhesion, signal transduction and cytoskeleton remodelling. Moreover, it is frequently explored for stem cell identification, while being associated with chemoresistance and metastasis[17]. In bladder cancer, both CD44 and its splicing variants have been implicated

in cancer progression and resistance to standard therapy[18]. Preliminary data from our group has uncovered specific CD44-STn glycoforms expressed in bladder cancer and absent from the normal urothelium. The cancer specific nature of CD44-STn glycoforms and its cell surface nature makes them ideal targets for guided therapeutics and immunotherapy, including cancer glycovaccines. In summary, increased levels of CD44-STn glycoforms have been associated with the severity of disease and as part of the molecular signature of more malignant bladder cancer sub-populations. These events not only amplify structural alterations that stem from deregulations in glycosylation pathways but also synergically contribute to a net effect favouring disease progression. Nevertheless, a comprehensive and context-oriented glycomapping of relevant glycopeptides has not been provided yet, which would be crucial for achieving highly specific cancer targets holding true therapeutic potential.

This chapter establishes the rational for glycoengineered cell factories for stable production of glycoantigens, namely CD44 glycopeptides expressing homogenous cancer-associated glycans, envisaging future glycoconjugate-based vaccines.

2. Material and Methods

2.1. Cell lines and culture conditions

T24 Human Bladder Cancer Cell Line, purchased from American Type Culture Collection (ATCC), was maintained using RPMI 1640+GlutaMAX™- I medium supplemented with 1% penicillin-streptomycin (10 000 Units/ml penicillin; 10 000 mg/ml streptomycin; Gibco, Life Technologies) and 10% heat-inactivated FBS (Gibco, Life Technologies). These cells were cultured at 37º C in a 5% CO_2 humidified atmosphere.

2.2. CRISPR Cas-9 gene editing

At approximately 70% confluence, T24 cells, previously plated onto a 24 well plate, were transfected using Lipofectamine™ CRISPRMAX™ Transfection Reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. A recombinant *Streptococcus pyogenes* Cas9 (GeneArt™ Platinum Cas9 Nuclease, Thermo Fisher Scientific) together with a single-guided RNA (GTAAAGCAGGGCTACATGAG sgRNA) were used to generate site-specific DSBs in the *C1GALT1* gene *in vitro*. For STn

overexpression, human ST6GalNAc1 (hST6GALNAC1 [NM_018414.5]) knock-in (KI) was performed using jetPRIME® transfection reagent (PolyPlus Transfection), according to manufacturer's instructions. Simultaneously, a mock control containing a 300 bp stuffer ORF was stablished. The cellular pool of positively transfected cells was enriched with through puromycin selection.

2.3. C1GALT1 mutation analysis

Using a GRS Genomic DNA Kit (GRISP Research Solutions) genomic DNA from T24 C1GALT1 KO and respective transfection controls was purified. Mutation screening of the *C1GALT1* gene was performed by Sanger sequencing using the primers: 5'-3' Forward: CCGGCCCTCAAAACCTAGAG and Reverse: TGCATCTCCCCAGTGCTAAG, both designed using Primer-BLAST[19].

PCR amplification of the *C1GALT1* amplicon was prepared using the following reagent setup: 2 µl of PCR Buffer; 1 µl of 10 mM dNTP Mix; 2µl of 25 mM. $MgCl_2$; 1 µl of Fw and Rev primers; 0.5 µl of template DNA; 0.5 µl of the AmpliTaq Gold DNA Polymerase (5U/µl) and ddH_2O to make up a 20 µl reaction. In a thermocycler, the following program was applied: initial denaturation step at 95ºC for 15 min, followed by 35 cycles of denaturation at 95ºC for 30 seconds; annealing reaction for 30 seconds at the appropriate temperature; extension step at 72ºC for 45 seconds; and a final extension step at 72°C for 9 min. Using an electrophoresis 2% (w/v) agarose gel, stained with SYBR safe DNA gel stain (Thermo Fisher Scientific) we confirmed the amplification process. The PCR products were further purified using an exonuclease I (20µg/µl) (Thermo Fisher Scientific) and Fast Thermosensitive Alkaline Phosphatase (1µg/µl) (Thermo Fisher Scientific) mix, in a ratio of 1:2. Using 2 µl ExoSap mix (permits degradation of unincorporated primers and nucleotides) per 5 µl of PCR product, samples were incubated for 50 min at 37º in a thermocycler for proper enzyme activation. Enzyme inactivation was accomplished by applying an 85C° cycle for 15 min.

The obtained DNA samples were sequenced using BigDye® Terminator v3.1 Cycle Sequencing Kit from Applied Biosystems® (Life Technologies) according to the following reaction: 0.5 µl BigDye® Mix; 3.4 µl of Buffer Sequencing 5x; 0.5 µl of Primer; 4.78 µl of ddH_2O; 1 µl (75ng) of PCR Product. The products were sequenced following a denaturation step at 96 ºC for 10 min; 35 cycles of denaturation at 96ºC for 10 sec; annealing for 5 sec at 52ºC and an extension step at 60ºC for 6 min. The sequenced

products were purified using Illustra Sephadex® G-50 fine columns (GE Healthcare, Life Sciences, Cleveland, USA), according to the manufacturer´s instructions. To stabilize the DNA, the purified DNA samples were eluted in 15 µl of deionized formamide (Applied Biosystems). Sequencing was achieved using a 3500 Genetic Analyzer (Thermo Fisher Scientific), and our CRISPR editing results were analysed using ICE from Synthego[20]. We determined the size and the frequency of Cas9 induced indels using Indel Detection by Amplicon Analysis (IDAA). For this purpose, three primers were used to analyse Cas9 targeted sites: a forward primer, a reverse primer (above mentioned) and a Fam Forward primer, with the same sequence as de Forward primer but with a labelled 6-FAM, making amplicons fluorescent. Selected amplified products were diluted 1:20; 1 ul of diluted amplified product was mixed with Hi-Formamide (Applied Biosystems) and appropriate dye size standard (GeneScan 600 LIZ or 500 ROX, Applied Biosystems). The mixture was analysed in a fragment analyser (3500 Genetic Analyzer, Applied Biosystems), revealing the size of the amplicons (bp), and thereby the inflicted indels, as well as their frequency in relative fluorescence units (RFU). IDAA was performed with an ABI 3130 instrument.

2.4. Flow cytometry for O-linked glycans

Semiconfluent cells were detached using the non-enzymatic Versene 1x reagent (ThermoFisher, Waltham, MA, USA), following a 2% paraformaldehyde (Sigma-Aldrich) fixation step for 15 min at room temperature. Tn antigen detection was achieved through fluorescein labelled *Vicia villosa lectin* (VVL, VVA-FITC) staining using a 0.01mg/mL dilution in PBS 2% FBS buffer for 1h at room temperature under mild agitation and obscurity. STn detection was performed using the same method after a 37ºC overnight enzymatic digestion of cells with Neuraminidase (NeuAse) from *Clostridium perfringens* (Sigma-Aldrich) (70mU NeuAse/ 10^6 cells in sodium acetate buffer). Samples were analysed using a FC500 Beckman Coulter Flow Cytometer coupled with CXP analysis software.

2.5. Immunofluorescence for short-chain *O*-glycans detection

In order to evaluate short O-glycans expression, T24 glycoengineered cell models were cultured at low density and fixed with 2% paraformaldehyde (PFA; Sigma-Aldrich), following immunofluorescent staining similar to the above-mentioned protocol defined for

flow cytometry. *Vicia Vilosa* (VVA, 0.01 mg/ml), *Peanut Agglutinin* (PNA, 0.01mg/ml) and *Griffonia Simplicifolia* II (GSL II, 0.02mg/mL) lectins were used for Tn, T and core 3 antigens screening, respectively. VVA and PNA lectins were incubated for 1h at room temparature in PBS 2% FBS under mild agitation and obscurity, while GSL II was incubated in a 10mM HEPES, 0.15M NaCl, 0.1mM CaCl2, pH=7.5 buffer. The sialylated forms of Tn and T antigen were detected using the same method after a 4h enzymatic digestion of samples at 37ºC using 50mU/mL α-neuraminidase from *Clostridium perfringens*. After antigen staining, cells were incubated with 4',6-Diamidino-2-Phenylindole Dihydrochloride (DAPI, Thermo Fisher Scientific) for 10 minutes at room temperature in obscurity. All images were acquired using a Leica DMI6000 FFW microscope following Las X software analysis.

2.6. *O*-glycomics

O-glycome characterization of T24 cells was carried out using a Cellular O-glycome Reporter/Amplification assay, as previously described[20,21]. Briefly, semiconfluent cells were incubated with 150 µM of peracetylated benzyl 2-acetamido-2-deoxy-α-D-galactopyranoside (Sigma-Aldrich) in complete cell culture medium for 24h. Following the incubation step, conditioned media was collected and processed using 10 kDa centrifugal filters (Amicon, Ultra-4; Merk Millipore) for sample deproteination. Benzylated O-glycans were further purified by solid-phase extraction using Sep-Pak 3 cc C18 cartridges (Waters), following glycan permethylation and analysis by nano LC-ESI-MS/MS[21]. Structures were assigned based on previous knowledge on O-glycans chromatography retention times and m/z ratios.

2.7. *In situ* proximity ligation assay

The detection of CD44 STn+-glycoforms was made by *in situ* proximity ligation assay (PLA) using the Duolink *in situ* detection reagent Red (Sigma-Aldrich), according to the manufacturer's instructions. Briefly, cells were fixed with 4% paraformaldehyde (PFA; Sigma Aldrich), blocked for 1h at 37°C with Blocking Reagent (Sigma Aldrich) and simultaneously incubated with 0.5µg/mL of anti-STn monoclonal antibody (B72.3 + CC49; Abcam) as well as monoclonal anti-CD44 (1:8000; Abcam) overnight at 4°C in a humidified chamber. In parallel, an enzymatic control with 70mU Neuraminidase from

Clostridium perfringens (Sigma-Aldrich) was performed. After incubation with primary antibodies, PLA probes were incubated for 1h at 37°C and ligation step was performed for 30 minutes at 37°C, followed by a 2h incubation at 37°C with an amplification solution. Finally, cells were stained with 4,6-diamidino-2-phenylindole (DAPI, Thermo Fisher Scientific) for 10 min at room temperature. All images were acquired using a Leica DMI6000 FFW microscope and Las X software (Leica).

2.8. RT-PCR analysis

TriPure isolation Reagent (Roche Diagnostics GmbH, Mannheim, Germany) was used to extract total RNA from cultured cells according with manufacturer's instructions. The purity and quantity of RNA extracts were determined based on the A260/A280 ratio using a Nanodrop™ lite spectrophotometer (Thermo Scientific, USA). Up to 2 µg total RNA was reverse transcribed into complementary DNA (cDNA) with random primers, using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA) in a Mycycler Thermal cycler (BioRad, Hercules, CA). The amplification conditions were the following: 25°C for 10 minutes, 37°C for 120 minutes, and reverse transcriptase inactivation at 85°C for 5 minutes. The relative expression levels of CD44 and its splicing variants (CD44v2-v10, CD44v3-v10, CD44v8-v10, CD44s and CD44soluble) was determined using TaqMan assays (Applied Biosystems, Perkin Elmer, CA, USA; assays IDs indicated in **Table 5**) and a 7500 Sequence Detector (Applied Biosystems, Perkin Elmer, CA, USA). The proposed nomenclature for experimentally observed CD44 isoforms has been extensively described in recent[22] by our group and will be herein employed accordingly. B2M and HPRT (Applied Biosystems, Perkin Elmer, CA, USA) housekeeping genes were used for normalization. PCR amplification was performed using Xpert Fast Probe Master Mix (Grisp, Portugal) and thermal cycling conditions were 10 minutes at 95°C followed by 45 cycles of 15 seconds at 95°C and 1 minute at 60°C. All samples were run in duplicate. Relative mRNA gene expression was calculated with the $2^{-\Delta Ct}$ formula.

Table 5. TaqMan assays IDs.

Gene	TaqMan Gene Expression Assay ID
CD44 Total	Hs01075864_m1

CD44v2-v10	Hs01075866_m1
CD44v3-v10	Hs 01081480_m1
CD44v8-v10	Hs01081475_m1
CD44s	Hs01081473_m1
CD44soluble	Hs01081469_m1
B2M	Hs00984230_m1
HPRT	Hs99999909_m1

2.9. Statistical analysis

The *Mann-Whitney* nonparametric test was used for transcript expression analysis. For glycomics analysis, unpaired T-tests were employed. Differences were considered significant when $p < 0.05$. The coefficient of variation (CV) was used to measure the dispersion of data points around the mean for glycans under nanoLC-ESI-MS/MS analysis.

3. Results and Discussion

CD44 is known to have high impact on cancer development[17,18]. In BC, CD44 is reported to be associated with resistance to chemotherapy and tumour initiating events[23,24], possibly constituting a relevant target for therapeutic approaches. Furthermore, the high molecular heterogeneity of CD44 molecules and its extensive glycosylation remodelling during malignant transformation may offer the necessary molecular grounds for intervention[22]. Previous work from our group has identified CD44-STn glycoproteoforms in bladder tumours, which are absent in healthy urothelium, conferring the necessary cancer bispecificity for further therapeutics development. T24 bladder cancer cells were also identified as potential sources of CD44 glycoproteoforms due to their constitutive expression of CD44 and a myriad of splicing variants, including relevant CD44 short forms as CD44s and CD44st. As such, these cells constitute a valuable starting point for the design of a CD44 glycopeptide cell factory. Envisaging homogeneous glycosylation of CD44 glycopeptides, with emphasis on Tn and STn

glycoforms, CRISPR/Cas9- mediated gene editing of the short-chain O-glycan biosynthesis machinery was performed.

3.1. Glycoengineered bladder cancer cell models

Bladder cancer T24 cells were first submitted to *O*-glycomics nanoLC-MS/MS characterization (**Figure 12**). This cell line mainly expressed fucosylated (m/z 746.40) and, particularly, mono- (m/z 933.48) and di-sialylated T antigens (m/z 1294.66). Evidence of core 3 (m/z 613.33) and more extended core 2 O-glycans (m/z 991.53; 1178.61; 1195.62; 1382.71; 1369.70; 1556.79; 1743.89) could also be observed. On the other hand, T24 cells expressed only vestigial amounts of STn (m/z 729.38), in accordance with previous observations for most human cancer cell lines[25,26] .

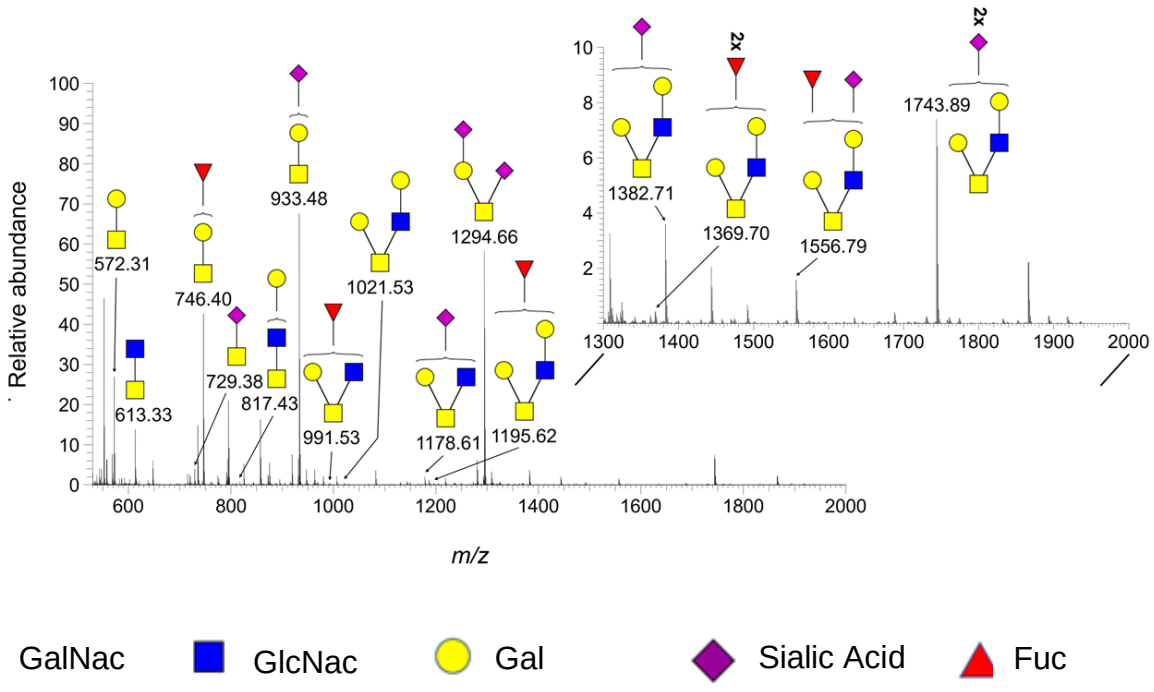


Figure 12. O-Glycomics analysis of T24 cells using Cellular O-Glycome Reporter/Amplification MALDI-TOF-MS; T24 cells mainly express fucosylated (m/z 746.40), mono- (m/z 933.48) and di-sialylated T antigens (m/z 1294.66). Core 3 can also be identified (m/z 613.33) as well as extended core 2 O-glycans (m/z 991.53; 1178.61; 1195.62; 1382.71; 1369.70; 1556.79; 1743.89); STn is expressed only on vestigial amounts (m/z 729.38).

Collectively, these observations support the need to reduce the complexity of cell-surface O-glycosylation while maximizing STn expression. Envisaging *O*-glycome

homogenization, a C1GALT1 Knock-Out was produced using CRISPR-Cas9 technology and a validated gRNA[27], preventing O-glycan elongation and inducing expression of truncated *O*-glycans like Tn and STn (**Figure 13 A**). To minimize of target effects provided by the gene editing strategy, three C1GALT1 knock-out clones were selected according to their distinct indel profile, as determined by Indel Detection by Amplicon Analysis (IDAA) **(Figure S1A)**. Sanger sequencing further allowed detection of induced mutation sites in different coding alleles **(Figure S1B)**. C1GALT KO cells were then screened for O-glycan expression by immunocytochemistry (**Figure 13 C**) and flow cytometry (**Figure 13 B**) using glycan specific lectins. Neuraminidase (NeuAse) treatment prior to VVA and PNA staining was performed to assess the expression levels of Tn and T antigen sialylated forms; namely STn and ST antigens. C1GALT1 KO cells invariably overexpressed Tn antigen compared to wild type cells (**Figure 13 C and D**), while losing T antigens expression as a direct consequence of T-synthase inactivating mutation (**Figure 13 C**). Mutated cells mildly expressed core 3 antigens, as observed for wild-type cells. PNGaseF controls to GSL II staining allowed us to exclude complex N-glycans contribution to GlcNAc signal, leading to the conclusion that most terminal GlcNAc epitopes derive from *O*-glycans. Interestingly, the increase in bioavailability of Tn epitopes provided by C1GALT1 KO did not translate in considerable STn expression increase (**Figure 13 C**). Accordingly, to obtain STn antigen stabilization at the cell surface, which has been shown to be highly dependent on the microenvironmental context of BC cells [28], human ST6GALNAC1 has been Knocked-in (KI) in C1GALT1 KO cells (**Figure 13 C**). Both immunofluorescence and flow cytometry analysis have confirmed increased STn levels as a consequence of ST6GalNAc1 overexpression compared to both wild type and mock controls (**Figure 13 C and D**). Low levels of core 3 antigens were maintained in the edited cells, as well as abundant Tn antigen.

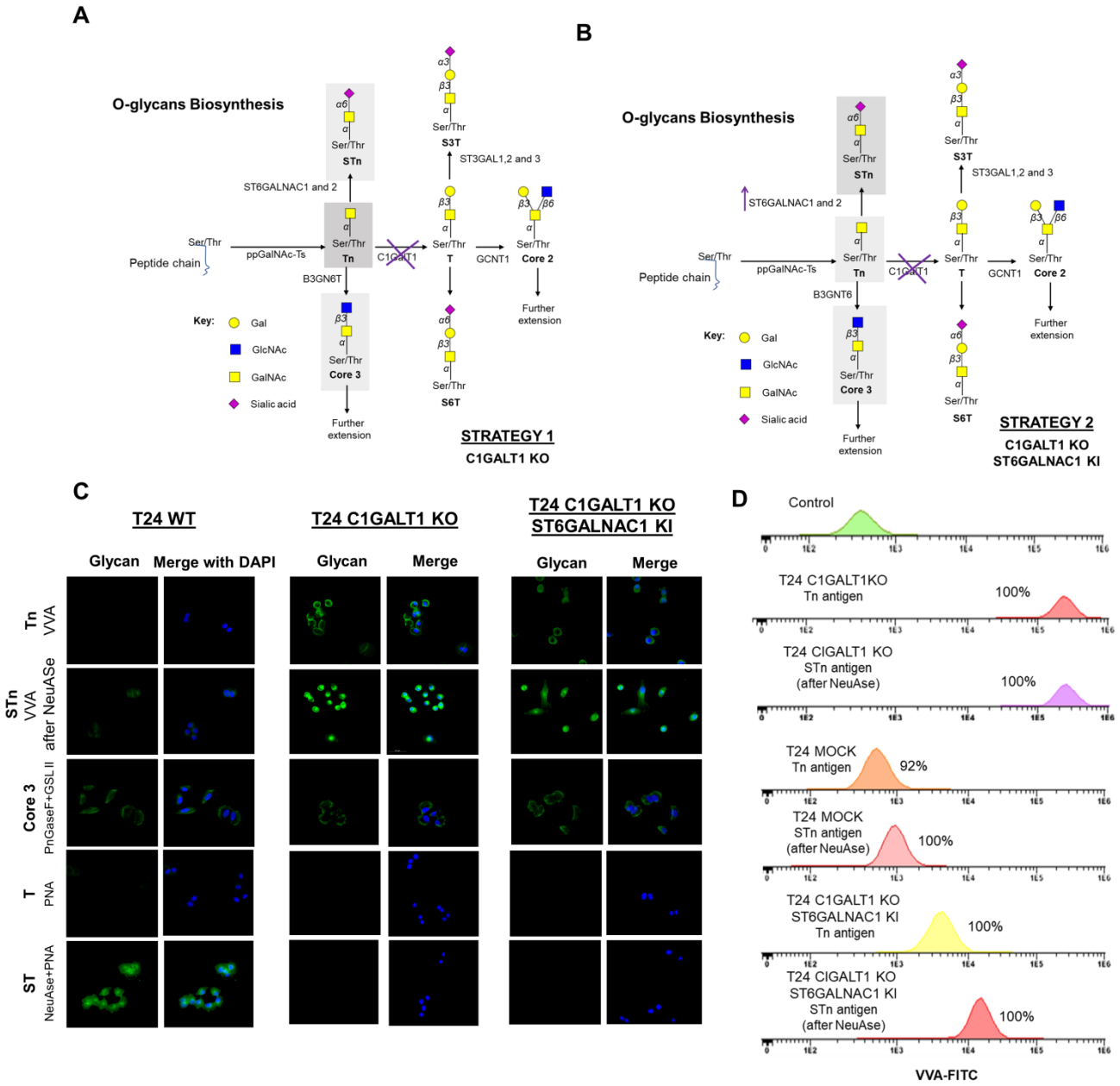


Figure 13. Strategy planning for increased STn expression and immunocytochemistry and flow cytometry analysis of cell models. A) Representation of the O-glycan biosynthetic pathway with the main products and targeted strategy for the induction of truncated O-glycan expression by C1GALT1 KO. **B)** Strategy representation of truncated O-glycan and Sialyl-Tn induction by C1GALT1 KO and ST6 KI, respectively. **C)** Immunocytochemistry analysis of T24 WT, T24 C1GALT1 KO and T24 C1GALT1 KO ST6 KI cells, with staining for Tn by VVA, STn by VVA after NeuAse treatment, Core 3 by GSL II after PNGaseF treatment, T using PNA and ST by PNA after NeuAse. **D)** Flow cytometry analysis for STn and Tn antigens using VVA.

Overall, the knock-out of simple mucin-type *O*-glycosylation glycosyltransferases as well as insertion of the *ST6GALNAC1* gene has allowed the development of a

customized cell factory for production of homogeneous STn expressing glycoproteins. Additionally, an alternative suppression of O-glycosylation pathways that lead to core 3 biosynthesis could be envisaged in future cell factory optimization steps, possibly further potentiating STn overexpression by eliminating competing pathways using Tn antigen as substrate.

3.2. CD44 STn$^+$ glycoforms cell factories

In situ proximity ligation (PLA) assays have been used to simultaneously detect CD44 and STn whenever both epitopes occur in close proximity. Accordingly, in the native wild-type system, the close co-localization of CD44 and STn was modest, possibly due to scarce membrane expression of STn in wild-type cells [25,26]. These findings have reinforced the need for genomic modulation of T24 cells towards increased expression of relevant CD44 glycoforms. Hence, the induced expression of STn biosynthesis enzyme ST6GALNAC1 and concomitant knock-out of core 1 (C1GALT1) have increased cell surface sialylation. This translated in a very strong prevalence of co-localization of CD44 and STn, suggesting the existence of CD44 STn$^+$-glycoforms in all established cell factory cells (**Figure 17.A**). Negative controls in the absence of monoclonal antibodies were set, confirming the specificity of PLA signals. Moreover, the presence of cellular heterogeneity comprising moderate to high CD44 and STn co-localization **(Figure 14 B)** opens an avenue for further population purification envisaging higher CD44-STn extraction titers.

The impact of glycogene editing on CD44 and its experimentally observed isoforms expression was also explored by Real-Time qRT-PCR (quantitative Reverse Transcription PCR). Of note, the overall amount of *CD44total* transcripts do not vary between T24 wild-type cells and glycoengineered cell factories **(Figure 14 C)**, reinforcing that genome editing has not translated in stress response and product degradation at the transcriptomic level [29]. Similarly, CD44v2-v10, CD44v3-v10, CD44s and CD44soluble transcript levels did not differ between the established cell factory and the wild-type system, with the exception CD44v8-v10 isoform transcripts which become upregulated in T24 C1GALT1 KO ST6GALNAC1 KI cells (**Figure 14 D**).

Overall, the developed cell factory ensures stable levels of total CD44 and splicing variants with emphasis on CD44-STn glycoforms, offering a promising tool for high-yield purification of relevant glycoproteoforms of CD44 envisaging future glycoconjugate-based vaccines.

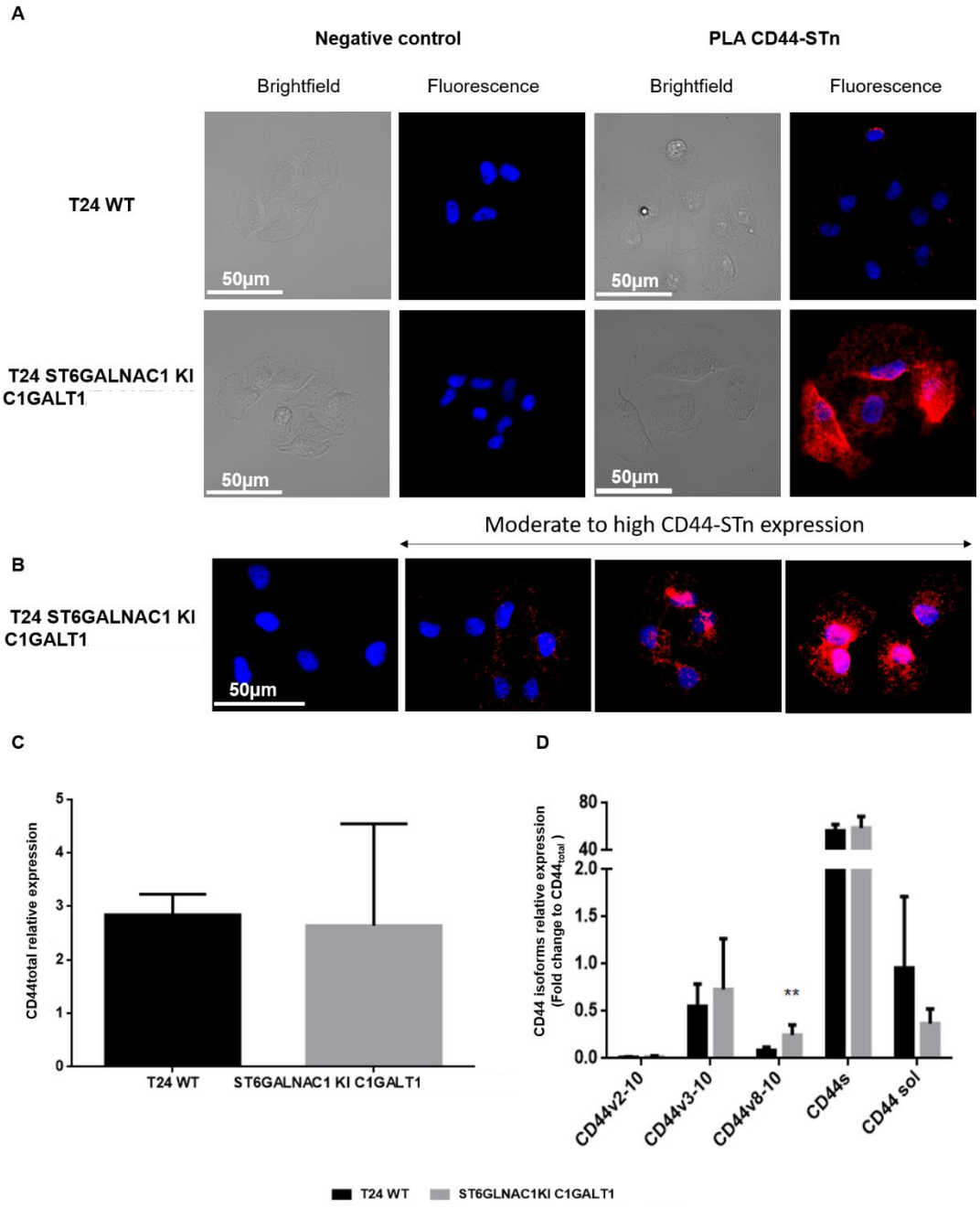


Figure 14. A) Proximity ligation assay (PLA) allowing the simultaneous detection of the CD44 protein and STn whenever there is close proximity (red dots) **(A and B)**. Negative controls were set without monoclonal antibodies, showing the specificity of PLA signals. ST6GALNAc KI C1GALT1 cells showing high prevalence of CD44 STn+ glycoforms compared to the wild-type system **(A).** ST6GALNAc KI C1GALT1 cells showing heterogeneous expression of CD44 STn isoforms spanning from moderate to high expression **B)**. qRT-PCR data of CD44total and CD44 experimentally determined isoforms transcript levels in Wild type and gene edited cells **(C and D)**. Total CD44 **C)** as well as its splicing variants **D)** transcript levels do not vary between the native system and ST6GALNAc KI C1GALT1 KO cells, suggesting that gene editing has not translated in the activation of stress responses and CD44 degradation. **p < 0.01 (Mann-Whitney test).

4. Concluding remarks

Advanced stage bladder cancer (BC) is amongst the most common and deadliest genitourinary cancers, with current therapeutic options failing to prevent disease progression and providing modest survival rates [30]. Over-forty years of glycobiology have provided the necessary insight on bladder cancer associated deregulation of glycosylation pathways [31–33]. Accordingly, BC is known to overexpress heavily sialylated glycoproteins driving disease progression and stemming from a premature stop in protein glycosylation. Moreover, cancer specific sialylated glycoproteoforms not observed in healthy urothelium offer a valuable means for selective targeting of malignant cells or the development of highly specific cancer vaccines. Particularly, preliminary data from our group has uncovered cancer specific cell surface CD44-STn glycoforms associated with the severity of disease and as part of the molecular signature of more malignant bladder cancer sub-populations.

The clinical application of glycan-based therapeutic solutions has been significantly hampered by the poor bioavailability of pure and structurally well-defined carbohydrates and glycoconjugates for pre-clinical studies and development of clinical grade tools. Herein, we developed glycoengineered cell factories for stable production of CD44 glycopeptides expressing homogenous cancer-associated glycans, mostly STn glycoforms, as opposed to glycopeptide enzymatic synthesis. We have selected an eucaryotic bladder cancer cell model constitutively expressing the target protein CD44 and several other experimentally determined isoforms derived from alternative splicing processes. This has allowed limiting genomic manipulation of the wild-type system to the glycosylation machinery. Manipulation of simple mucin-type *O*-glycan biosynthesis pathways was achieved by CRISPR Cas9 technology, emphasizing Tn and STn antigens synthesis while abrogating extension of T antigens. This has culminated in significant production of homogenous CD44-STn glycoforms with no detrimental effects on CD44 transcriptomic regulation nor product degradation, as demonstrated by qRT-PCR and PLA. Furthermore, the key advantage of these cell factories is its ability to use cheap and abundant raw materials, all provided by cell culture conditions, to catalyse multistep reactions, which may be hard, expensive and time consuming to control in a step wise manner in *in vitro* synthesis. Moreover, the establishment of a stable production system simplifies the glycopeptide production process, which allied to the fine-tuning of downstream recovery decreases environmental and economic costs. Ultimately, cheap raw materials, efficient bioproduction, and the ability to reuse the immortalized cell factories indefinitely reduces the glycopeptide cost and makes whole-cell biocatalysis very

cost-competitive. Accordingly, this work provided an advantageous cell factory tool to be further explored for glycopeptide recovery. Future studies should focus on further cell factory refinement, optimization of CD44 extraction methods and glycolmapping of extracted glycopeptides, envisaging extensive product characterization for forthcoming glycan-based vaccines assembly.

Bibliography

1. Calabrò, F. & Sternberg, C. N. State-of-the-art management of metastatic disease at initial presentation or recurrence. *World J. Urol.* **24**, 543–556 (2006).

2. Resch, I., Shariat, S. F. & Gust, K. M. PD-1 and PD-L1 inhibitors after platinum-based chemotherapy or in first-line therapy in cisplatin-ineligible patients: Dramatic improvement of prognosis and overall survival after decades of hopelessness in patients with metastatic urothelial cancer. *Memo - Mag. Eur. Med. Oncol.* **11**, 43–46 (2018).

3. Rosenberg, J. E. *et al.* Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: A single-arm, multicentre, phase 2 trial. *Lancet* **387**, 1909–1920 (2016).

4. Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: Diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **13**, 448–462 (2012).

5. Shade, K. T., Conroy, M. E. & Anthony, R. M. IgE Glycosylation in Health and Disease. *Curr. Top. Microbiol. Immunol.* **423**, 77–93 (2019).

6. Jason E. Hudak1, 4 and Carolyn R. Bertozzi. Glycotherapy: New Advances Inspire a Reemergence of Glycansin Medicine. **21**, 16–37 (2014).

7. Zhang, P. *et al.* Challenges of glycosylation analysis and control: An integrated approach to producing optimal and consistent therapeutic drugs. *Drug Discov. Today* **21**, 740–765 (2016).

8. Lepenies, B., Yin, J. & Seeberger, P. H. Applications of synthetic carbohydrates to chemical biology. *Curr. Opin. Chem. Biol.* **14**, 404–411 (2010).

9. Palcic, M. M. Glycosyltransferases as biocatalysts. *Curr. Opin. Chem. Biol.* **15**, 226–233 (2011).

10. Das, R. & Mukhopadhyay, B. Chemical O-Glycosylations: An Overview. *ChemistryOpen* **5**, 401–433 (2016).

11. Natarajan, A. *et al.* Engineering orthogonal human O-linked glycoprotein biosynthesis in bacteria. *Nat. Chem. Biol.* **16**, 1062–1070 (2020).

12. Mueller, P. *et al.* High level in vivo mucin-type glycosylation in Escherichia coli. *Microb. Cell Fact.* **17**, 1–15 (2018).

13. Westers, L., Westers, H. & Quax, W. J. Bacillus subtilis as cell factory for pharmaceutical proteins: A biotechnological approach to optimize the host organism. *Biochim. Biophys. Acta - Mol. Cell Res.* **1694**, 299–310 (2004).

14. Gill, R. T., Valdes, J. J. & Bentley, W. E. A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in Escherichia coli. *Metab. Eng.* **2**, 178–189 (2000).

15. Dong, H., Nilsson, L. & Kurland, C. G. Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction. *J. Bacteriol.* **177**, 1497–1504 (1995).

16. Schweder, T. *et al.* Role of the general stress response during strong overexpression of a heterologous gene in Escherichia coli. *Appl. Microbiol. Biotechnol.* **58**, 330–337 (2002).

17. Williams, K., Motiani, K., Giridhar, P. V. & Kasper, S. CD44 integrates signaling in normal stem cell, cancer stem cell and (pre)metastatic niches. *Exp. Biol. Med.* **238**, 324–338 (2013).

18. Golshani, R. *et al.* Hyaluronic acid synthase-1 expression regulates bladder cancer growth, invasion, and angiogenesis through CD44. *Cancer Res.* **68**, 483–491 (2008).

19. Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134 (2012).

20. Kudelka, M. R. *et al.* Cellular O-Glycome Reporter/Amplification to explore O-glycans of living cells. *Nat. Methods* **13**, 81–86 (2015).

21. Fernandes, E. *et al.* Nucleolin-sle a glycoforms as E-selectin ligands and potentially targetable biomarkers at the cell surface of gastric cancer cells. *Cancers (Basel).* **12**, (2020).

22. Azevedo, R. *et al.* CD44 glycoprotein in cancer: A molecular conundrum hampering clinical applications. *Clin. Proteomics* **15**, 1–5 (2018).

23. Wu, C. Te, Lin, W. Y., Chen, W. C. & Chen, M. F. Predictive Value of CD44 in Muscle-Invasive Bladder Cancer and Its Relationship with IL-6 Signaling. *Ann. Surg. Oncol.* **25**, 3518–3526 (2018).

24. Tatokoro, M. *et al.* Potential role of Hsp90 inhibitors in overcoming cisplatin resistance of bladder cancer-initiating cells. *Int. J. Cancer* **131**, 987–996 (2012).

25. JA, F. *et al.* Overexpression of tumour-associated carbohydrate antigen sialyl-Tn in advanced bladder tumours. *Mol. Oncol.* **7**, 719–731 (2013).

26. Ferreira, J. A. *et al.* Mechanisms of cisplatin resistance and targeting of cancer stem cells: Adding glycosylation to the equation. *Drug Resist. Updat.* **24**, 34–54 (2016).

27. Narimatsu, Y. *et al.* A validated gRNA library for CRISPR/Cas9 targeting of the human glycosyltransferase genome. *Glycobiology* **28**, 295–305 (2018).

28. Peixoto, A., Relvas-Santos, M., Azevedo, R., Lara Santos, L. & Ferreira, J. A. Protein glycosylation and tumor microenvironment alterations driving cancer hallmarks. *Front. Oncol.* **9**, 1–24 (2019).

29. Flick, K. & Kaiser, P. Protein degradation and the stress response. *Semin. Cell Dev. Biol.* **23**, 515–522 (2012).

30. Compérat, E. M. *et al.* Grading of Urothelial Carcinoma and The New "World Health Organisation Classification of Tumours of the Urinary System and Male Genital Organs 2016". *Eur. Urol. Focus* **5**, 457–466 (2019).

31. Carrascal, M. A. *et al.* Sialyl Tn-expressing bladder cancer cells induce a tolerogenic phenotype in innate and adaptive immune cells. *Mol. Oncol.* **8**, 753–765 (2014).

32. Carrascal, M. Role of sialylated glycans in cancer progression : sialyl-Tn and sialyl-LewisX - Tese de Doutoramento. (2017).

33. Neves, M. *et al.* Exploring sialyl-Tn expression in microfluidic-isolated circulating tumour cells: A novel biomarker and an analytical tool for precision oncology applications. *N. Biotechnol.* **49**, 77–87 (2019).

IV Concluding Remarks and Future perspectives

Bladder Cancer treatments presently employed for disease control seem to be inefficient for many patients, who often develop recurrences with progression into more complicated cancer phenotypes. These treatments usually use a combination of drugs (MVAC - vinblastine, adriamycin and cisplatin or GC- gemcitabine and cisplatin) to induce disease regression [29]. However, many patients show resistance to these drug based therapies with overall, no improvement rates, which has urged the employment of novel therapies. Immunotherapies, utilizing check point inhibitors, have been used as alternatives for the current BC treatments. Nonetheless these approaches have shown to induce resistance in a given set of patients and cannot be generally utilized. As such, therapeutic anticancer nanovaccines seem to mediate an increasingly interesting new path for bladder cancer therapies.

Tumour cells display glycoproteins at the cell surface carrying aberrant short-chain O-glycans such as the Tn and STn antigens, which play a key role driving invasion, resistance to chemotherapy and immune escape. Most aggressive bladder tumours, circulating tumour cells and corresponding lymph node as well as distant metastasis overexpress either one or both these glycoepitopes. On the other hand, they are absent from the healthy urothelium, suggesting potential for cancer detection but also targeted therapeutics, including immunotherapy by vaccination. However, the presence of these antigens, even though in low abundance in some specific healthy cells of the gastrointestinal and respiratory tracts has challenged this concept. As such, our group has been devoted to a comprehensive characterization of the bladder cancer and human glycoproteome and has identified cancer-specific MUC16 and CD44-Tn/STn glycoproteoforms that pave the way towards the development of cancer vaccines. Foreseeing this goal, this work was devoted to creating molecular and cellular approaches for glycoepitopes production as well as prototyping different vaccine constructs for pre-clinical assays in the near future.

Following a single pot enzymatic approach we produced libraries of glycoepitopes with different surface O-glycans. These libraries were extensively characterized and highlighted the synthesis of CD44 neutral peptides (Tn), and MUC16 neutral, hybrid and sialylated peptides (Peptide decoration with: Tn antigens only, Tn and STn antigens, STn glycans only; respectively). This strategy entails the use of transferases for the production of synthetic glycopeptides and as such can be modulated to fit different epitopes that require specific glycosylation profiles. These optimization characteristics could rectify the

lack of identifiable CD44 sialylated peptides. Overall, this approached described the sequential enzymatic synthesis of glyconeoantigens that can be applied for the construction of further relevant epitopes in vaccine formulations. Envisaging the application of these synthethic glyconeoantigens in future vaccine constructs we successfully coupled MUC16 glycoantigens with the immunogenic protein Keyhole limpet hemocyanin (KLH)[256], commonly used to attribute immunogenic potential to otherwise poor immunogenic antigens. The evidences of coupling between these two molecules enable the use of low immunogenic peptides for further applicability in clinical vaccine preparations. MUC16 molecules can be found on several tumour tissues and so the employment of a MUC16-based glycovaccine could benefit several cancer diseases [135,136]. Particularly the presence of Sialyl-Tn and Tn epitopes can further increase accuracy of glycosylated antigens found on tumour tissues, with more aggressive and immunomodulatory phenotypes[140,279]. Therefore, we persuited encapsulation of these conjugates in PLGA formulations, assessing possible uses of these glyconeoantigens for nanovaccine systems, and their applicability in the context of cancer remission. PLGA nanovaccines are highly versatile vehicles characterized by having great compatibility properties for nanovaccine administrations, which include promoting controlled releases of cargo, for appropriate immune stimulation, and favourable interactions with APCs, as well as biodegradability and biocompatibility. PLGA particles entrapped with enriched KLH-MUC16-Tn and KLH-MUC16-STn glycoconjugates were characterized for their physicochemical properties. NPs displayed mild negative charges, homogenous suspensions and sizes believed to be appropriate for APC presentation[202]. Further confirmed by macrophages selective uptake of PLGA-FITC particles that mimicked the overall properties found on our encapsulated NPs. The concentrations of NPs necessary for appropriate internalization displayed negligible cytotoxicity. Our particles seem to be more appropriate for parenteral administrations than for oral routes of administration, given their controlled and over time release rates at physiological pH vs release bursts seen for particles in gastric like pH environments. Furthermore it would be interesting to evaluate the behaviour of our NPs at lysosomal like pH (pH 5.0), which could provide some indications as to what antigen presentation pathway our antigens would follow. Particles encapsulated with KLH-MUC16-STn seem to display the most appropriate characteristics for nanovaccination, although no major changes were described.

Although synthetic epitopes can be obtained with high clinical qualities using the methodologies previously described, the implementation of systems that could potentiate glycopeptide yields in a cheaper and more *in vivo* like way could further facilitate the generalized usage of glycoepitopes for nanovaccine applications. As such, we developed

a simple cell model that potentiates the expression of Sialyl Tn glycans, envisaging the particular production of CD44 glycoantigens, which are characteristic of T24 high grade bladder cancer cell lines. Despite their wide spread expression in bladder cancer tissues, T24 do not report particular high expression of STn glycans and so genomic silencing of O-glycosylation pathways and further overexpression of *ST6GALNAC I* were exploited for the establishment of a bladder cancer model primed for STn expression. The final sequential cell product was evaluated regarding maintenance of CD44 expression profiles for which no major alterations were detected. Additionally a PLA assay confirmed the presence of CD44-STn glycopeptides in our cell model. As such, future approaches should seek to enrich cell population with cells expressing increased levels of CD44, optimizing the yields of glycoepitopes obtained for further characterization and research applications. This single cell model could be further improved to better fit the strategies of encapsulation and conjugation here described. It would be interesting to induce cellular expression of CD44 molecules with Cystein Tags that could rapidly be used for further conjugation with immunogenic proteins, such as KLH. This would result in less time consuming preparations of possible immunogenic CD44 constructs. Furthermore our particles could be further addressed regarding their capacity to internalize DC and induce their activation *in vitro*.

Bibliography

1. Calabrò, F. & Sternberg, C. N. State-of-the-art management of metastatic disease at initial presentation or recurrence. *World J. Urol.* **24**, 543–556 (2006).

2. Hossain, M. K. & Wall, K. A. Immunological evaluation of recent MUC1 glycopeptide cancer vaccines. *Vaccines* **4**, 1–13 (2016).

3. Kim, N., Hong, Y., Kwon, D. & Yoon, S. Somatic Mutaome Profile in Human Cancer Tissues. *Genomics Inform.* **11**, 239 (2013).

4. Das, S. *et al.* Carboxyl-terminal domain of MUC16 imparts tumorigenic and metastatic functions through nuclear translocation of JAK2 to pancreatic cancer cells. *Oncotarget* **6**, 5772–5787 (2015).

5. Cotton, S. *et al.* Targeted O-glycoproteomics explored increased sialylation and identified MUC16 as a poor prognosis biomarker in advanced-stage bladder tumours. *Mol. Oncol.* **11**, 895–912 (2017).

6. Belisle, J. A. *et al.* Identification of Siglec-9 as the receptor for MUC16 on human NK cells, B cells, and monocytes. *Mol. Cancer* **9**, (2010).

7. Koerner, J., Horvath, D. & Groettrup, M. Harnessing Dendritic Cells for Poly (D,L-lactide-co-glycolide) Microspheres (PLGA MS)-Mediated Anti-tumor Therapy. *Front. Immunol.* **10**, 1–16 (2019).

Supplementary material

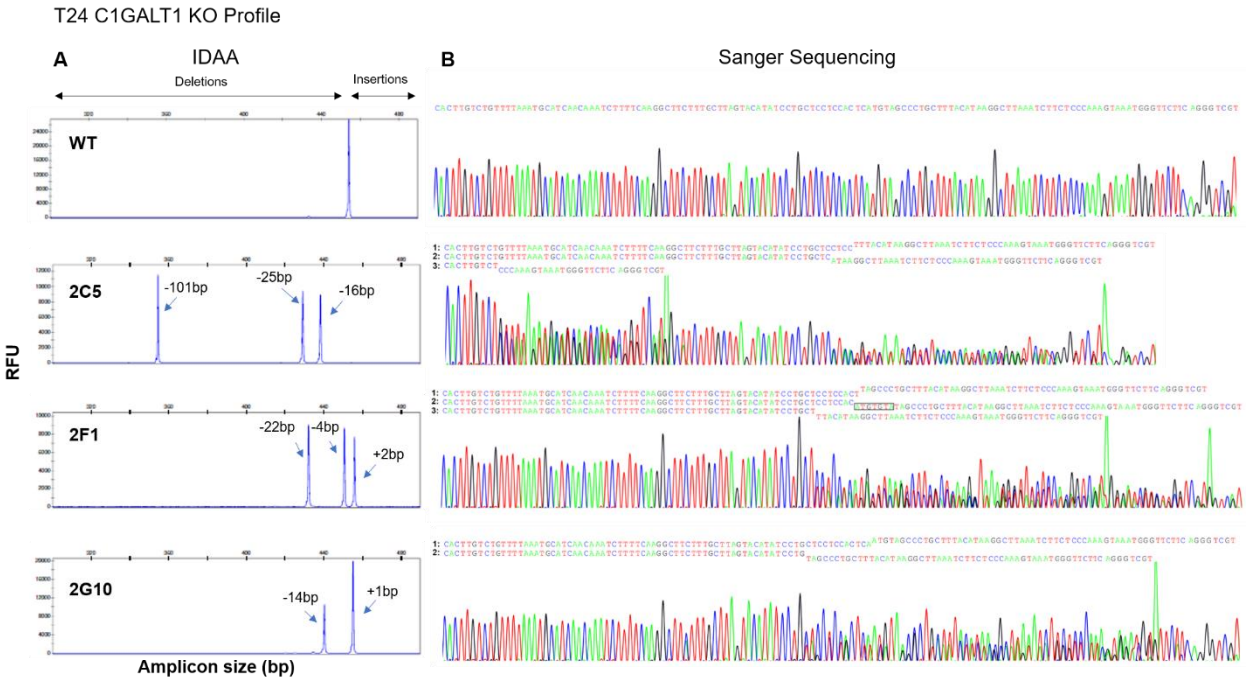


Figure S1. Genomic profiling of glycoengineered C1GALT1 Knock-out models. T24 cells were glycoengineered to knock-out human C1GALT1 and three clones with diverse indels were selected for proof-of-concept experiments. Induced indel mutations were characterized by Indel Detection by Amplicon Analysis (IDAA) **(A)** and Sanger Sequencing **(B)**. Electropherograms of Sanger Sequencing with the reverse primer are shown, along with the corresponding sequencing readings (1, 2 and 3). Accordingly, the 2C5 clone is characterized by three different DNA sequences, with the following observed variants and predicted consequences: **1:** a c.610_625del [p.(Gln204GlufsTer7)], leading to a 16bp deletion and the replacement of Gln 204 by Glu, resulting in a frame-shift introducing a stop codon 7 a.a. ahead; **2:** a c.605_629del [p.(Val202GlufsTer6)], resulting in a 25bp deletion and the replacement of Val 202 by Glu, leading to a frame-shift introducing a stop codon 6 a.a. ahead; and **3:** a c.589_689del [p.(Arg197GlnfsTer5)], in which a 101bp deletion results in the replacement of Arg 197 by Gln, leading to a frame-shift introducing a stop codon 5 a.a. ahead. Similarly, clone 2F1 is also characterized by three different DNA sequences, with the following observed variants and predicted consequences: **1:** a c.618_621del [p.(Tyr206Ter)], where a 4bp deletion results in the replacement of Tyr 206 by a stop codon; **2:** a c.618_622delinsTACACAT [p.(Met207ThrfsTer10)], where a 5bp deletion and a 7bp insertion results in the replacement of Met 207 by Thr, inducing a frame-shift and introducing a stop codon 10 a.a. ahead; and **3:** a c.614_635del [p.(Gly205AspfsTer4)], where a 22bp deletion results in the replacement of Gly 205 by Asp, leading to a frame-shift introducing a stop codon 4 a.a. ahead. Finally, clone 2G10 is characterized by two different DNA sequences, with the following observed variants and predicted consequences: **1:** a c.620dup [p.(Met207IlefsTer19)], where the duplication of nucleotide 620 results in the replacement of Met 207 by Ile, leading to a frame-shift introducing a stop codon 19 a.a. ahead; and **2:** a c.620_633del [p.(Met207ArgfsTer14)], resulting in a 14bp deletion and the replacement of Met 207 by Arg, resulting in a frame-shift introducing a stop codon 14 a.a ahead.