INVESTIGATION OF SIGLEC8 AS A THERAPEUTIC TARGET TO TREAT ASTHMA

Angela Garcia-Perez

THIS THESIS IS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)

MARCH, 2018
UNIVERSITY OF WARWICK

Table of Contents

Acknowledgements

Firstly, I would like to thank my supervisor Ann Dixon for her guidance, advice, encouragement and support in the disappointing and exciting moments through this project for three years and a half. I also would like to thank our industrial collaborators from Mologic Ltd., specially Paul Davis, John Wilkins, Gita Parekh, Michael Johnson and James Schouten for their kind welcome during my industrial placement and collaboration in the production of soluble protein target, together with Richard Williamson from the University of Kent, who provided us with a very useful refolding protocol.

I also would like to express my gratitude to the old and new Dixon group members: Leo Bowsher, Muhammad Hasan, Dhadchi Jeyaharan, Christine Lockey, Christopher Thoroughgood, Rihannon Brooks, Eleanor Jayawant, Chandni Mistry and Shradda Vadodaria for their comradeship and help in the laboratory.

Thank you to other Chemical Biology and Warwick University colleagues and staff: Matt Gibson for letting me use his Octet instrument and Anne Sophie Gleinich for teaching me how to set up the Octet experiments, Marco Pinto Corujo for advice about protein biophysical characterization, Steven Huband for performing the SAXS measurements and Dave Scott from the University of Nottingham for the SAXS data 3D-modelling.

Finally, I would like to express my gratitude to my parents, brothers and my boyfriend Julian for their encouragement and unconditional support which was fundamental to reach this point, and to other friends as Javier Fernandez-Garcia, Antonio Exposito-Serrano and Marcelo Alves da Silva for providing me with good advice along this path.

Thank you to the Chemical Biology Research Facility and the Chemistry Department at Warwick University, the EPSRC and Mologic Ltd. for my PhD studentship and this great opportunity to work in such an interesting and relevant project.

‘Imagination is more important than knowledge. For knowledge is limited, whereas imagination embraces the entire world, stimulating progress, giving birth to evolution’. Albert Einstein.

Declaration

The Siglec8 Ig-like V-type domains cloning, expression and purification protocols were developed by the author of this thesis, Angela Garcia-Perez at the Chemical Biology research facility at the University of Warwick.

The refolding protocol was provided by Richard Williamson from the University of Kent and Mologic Ltd and subsequently adapted to the Siglec8 Ig-like V-type domains by Angela Garcia-Perez at Warwick who also prepared all protein samples.

Fluorescence, Absorbance and Circular Dichroism measurements were performed by Angela Garcia-Perez in the Biophysics laboratory on the 6th floor of the Chemistry department at the University of Warwick.

We paid a service charge for the LC-ESI-MS/MS measurements performed at the Mass Spectrometry facility in the School of Life Sciences at the University of Warwick. Angela Garcia-Perez prepared a protein sample according to the facility protocols and the staff carried out the measurements.

SAXS measurements were performed by Steven Huband in the Physics department at the University of Warwick, Eleanor Jayawant estimated the ratio of the ssNMR Siglec8 resolved structure using the software Gromacs and Dave Scott from the University of Nottingham did the SAXS data 3D modelling.

The small mimetic molecules targeting Siglec8 were designed by Angela Garcia-Perez and purchased from Sigma, Insightbio and Carbosynth.

Octet experiments were performed by Angela Garcia-Perez with the kind collaboration of Anne Sophie Gleinich using Matt Gibson's Octet instrument at the University of Warwick.

Equilibrium Dialysis experiments were carried out by Angela Garcia-Perez.

1D H^+ NMR samples were prepared by Angela Garcia-Perez and measured by Ann Dixon with the collaboration of Christine Lockey at the 700 MHz NMR instrument at the University of Warwick.

Abstract

Siglecs (Sialic Acid Immunoglobulin-like Lectins) are immune system protein receptors that play important roles in the down-regulation of cellular signals upon Sialic Acid ligand binding. At present, the human Siglec family is composed of 16 members expressed by specific immune system cellular types. Due to their highly restrictive expression patterns and key regulatory function, Siglecs have been pointed out as interesting pharmaceutical targets to treat certain immune diseases where those specific cellular types are involved.

This work is focused on Siglec8 which is expressed on the surface of mast cells and eosinophils and has demonstrated to induce their apoptosis upon ligand binding. This effect could be exploited to develop therapeutics that target and reduce the levels of eosinophils in the inflammatory diseases such as asthma where they are the main effectors. The aim of this thesis is to design some Siglec8-ligand mimetic small molecules, test their binding to the receptor and obtain affinity data about the interaction.

We managed to express the Siglec8 Ig-like V-type carbohydrate binding domains in *Escherichia coli*, refold and purify them in solution and perform a biophysical characterization by means of different techniques such as Fluorescence, Absorbance, Mass Spectrometry (MS), Circular Dichroism (CD) and Single-Angle Neutron Scattering (SAXS). We also designed nine small Siglec8-ligand mimetic molecules and tested their binding to the receptor domains using the Octet technology and equilibrium dialysis-1 D H^+ NMR measurements. We found the ligand Phospho-Tyrosine to be our best Siglec8 binder and estimated a dissociation constant around 100 μM for the interaction. Thus, the binding information obtained in this work could lead to the development of higher affinity and specific molecules that target Siglec8.

This thesis also encompassed an industrial collaboration with the company Mologic Ltd where anti-Siglec8 antibodies where developed with chronic lung disease diagnostic purposes.

Abbreviations

ACOS, Asthma-Chronic Obstructive pulmonary disease overlap Syndrome.

ATP, Adenosine Triphosphate.

B, fraction of ligand bound to receptor.

B', fraction of occupied receptor.

CADD, Computer-Aided Drug Discovery.

CD, Circular Dichroism.

COPD, Chronic Obstructive Pulmonary Disease.

DDS, (4,4-dimethyl-4-silapentane-1-sulfonic acid).

DNA, Deoxyribonucleic Acid.

DTT, Dithiothreitol.

ε, molar absorptivity coefficient.

EDTA, Ethylenediaminetetraacetic acid.

ERK, Extracellular receptor-stimulated kinase.

FDS, Food and Drugs administration.

Gal6S, D-Galactose 6-Sulphate.

GdnHCl, Guanidine hydrochloride.

GlcNAc, Acetyl-D-Glucosamine.

GPI, Glycosylphosphatidylinositol anchored proteins.

GSH, Reduced glutathione.

GSSG, Oxidised glutathione.

HPLC, High performance liquid chromatography.

HTS, High throughput screening.

ITAMs, Tyrosine-based activation motifs.

ITIMs, Tyrosine-based activation motifs.

I, Intensity.

IPTG, Isopropyl β-D-1-thiogalactopyranoside.

K_D, dissociation constant.

L, Ligand.

LB, Lysogeny Broth.

λ, Wavelength.

MAPK, Mitogen-activated protein kinase.

MBL, mannose-binding lectin

M9, Minimal media.

MW, Molecular weight.

MWCO, Molecular Weight Cut Off.

NeuAc, N-acetylneuraminic acid

NHS, National Health System.

NK, Natural Killer.

NMR, Nuclear Magnetic Resonance.

OD, Optical Density.

PCR, Polymerase Chain Reaction.

pI, Isoelectric point.

R, Receptor.

RGA, Receiver gain.

RNA, Ribonucleic Acid

ROS, Reactive Oxygen Species.

SAXS, Single Angle X-Ray Scattering.

SDS-PAGE, Sodium Dodecyl Sulfate Polyacrylamide

Siglec, Sialic acid Immunoglobulin like Lectin.

TEV, Tobacco Etch Virus protease)

UV, Ultraviolet.

2x YT, Yeast Extract Tryptone.

Figures

4. BIOPHYSICAL CHARACTERIZATION OF SIGLEC8 IG-LIKE V-TYPE DOMAINS

5. EXPLORING SIGLEC8 AS A THERAPEUTIC TARGET.

Tables

Chapter 1.
INTRODUCTION

1. Introduction

1.1 Importance of protein-glycoconjugate interactions in human health and disease.

The term glycan is defined by IUPAC as "compound consisting of several monosaccharides linked glycosidically". In the human body, glycan biomolecules can be found associated with other glycans, proteins or lipids forming structures such as oligosaccharides, polysaccharides, glycoproteins, glycopeptides, glycolipids and proteoglycans (Marchetti, *et al.*, 2016). Many of these structures participate in the molecular events that take place to support vital biological processes such as cell-cell interaction, signal transduction, inflammation, viral entry and host-bacterial recognition, which are relevant in disease and defence.

Glycans achieve this through specific interaction with proteins such as lectins, antibodies or enzymes (Smith and Helenius, 2004). Lectins are carbohydrate-binding proteins highly specific for sugar moieties, which serve many different biological functions in animals. Viruses infect specific cell types by binding to receptors in cells' surfaces, i.e. influenza virus displays in its surface the lectin hemagglutinin that recognizes sialic acid residues present on cell-surface glycoproteins, and this interaction leads to the virus internalization into the cell. Another example is the mannose-binding lectin (MBL), which recognizes carbohydrate patterns found in the surface of pathogenic microorganisms and initiates the activation of the immune response, leading to the production of specific anti-pathogen carbohydrate antibodies. Furthermore, in several inflammatory autoimmune diseases as systemic lupus erythematosus, effector T cell populations with altered glycosylation are highly upregulated leading to an excessive activation of other immune cellular effectors and response (van Kooyk and Rabinovich, 2008). Figure 1.1 displays the different kinds of physiological protein-glycan interactions that can occur in the body: cell-cell, cell-virus, cell-antibody, cell-bacteria or cell-toxin.

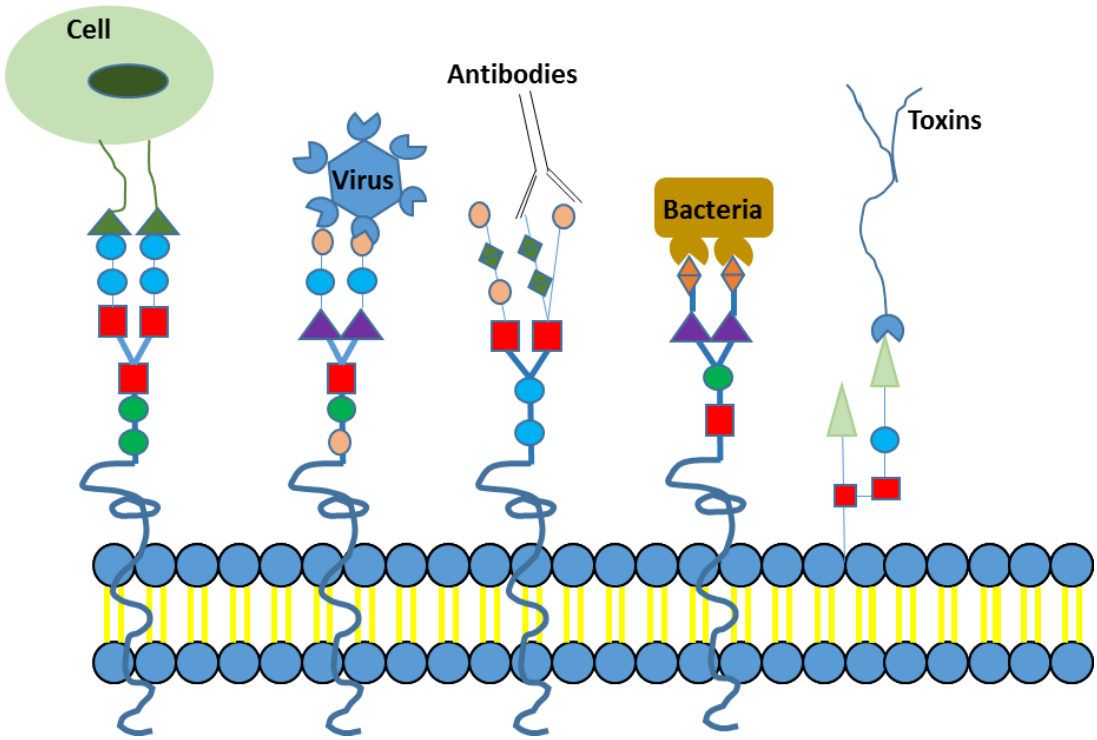


Figure 1.1. Different protein-glycan interactions (cell-cell, cell-virus, cell-antibody, cell-bacteria, cell-toxin) which are relevant for the human immune system (van Kooyk and Rabinovich, 2008).

Although in mammals there are only around ten monosaccharides which associate to build longer glycans, they can combine to form a huge variety of three-dimensional structures, linking through different positions of the sugar units and forming branched structures that adopt a wide variety of spatial orientations. They also may undergo different modifications with non-carbohydrate substituents like sulphate, phosphate or acetyl groups (Nieto, *et al.*, 2011). Figure 1.2 represents the different linking positions and some possible modifications for the sialic acid family, which includes 43 derivatives of the nine-carbon sugar neuraminic acid. Those glycans normally are found as components of oligosaccharide chains occupying terminal positions where they are very exposed and develop important functions.

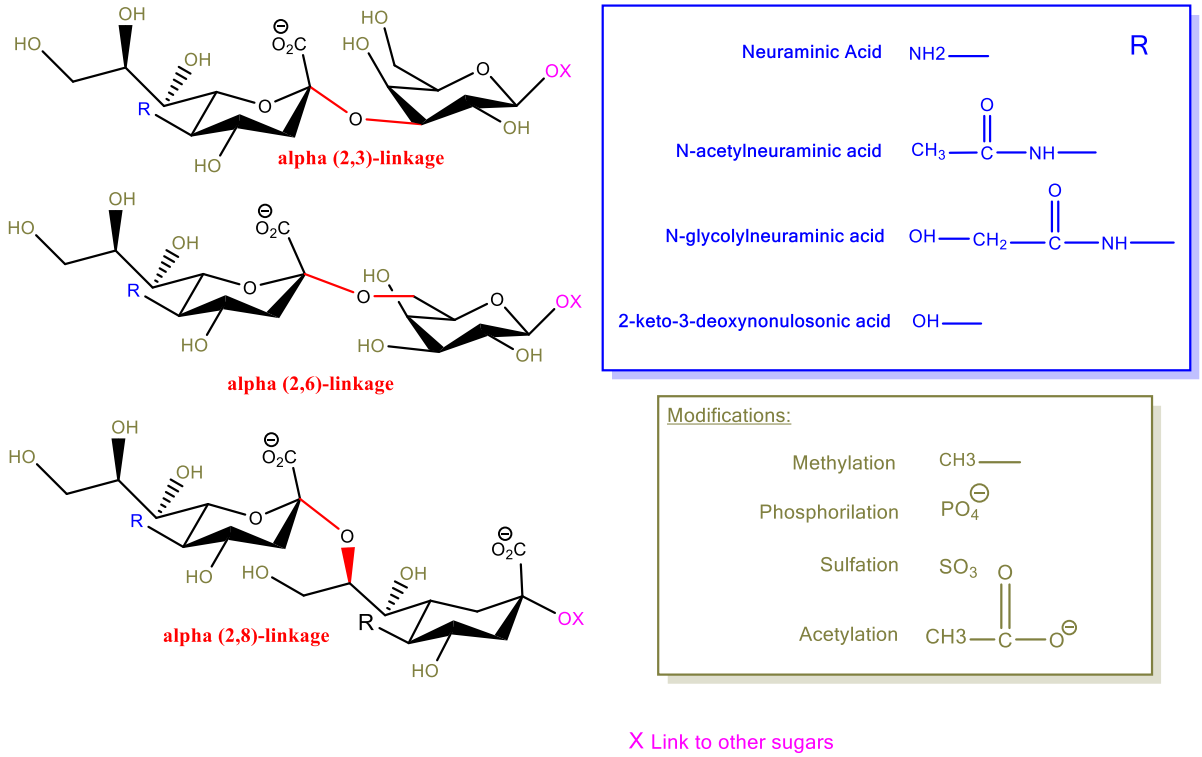


Figure 1.2. Different linkages and modifications that the carbohydrate family of sialic acids can undergo. Red: sialic acids can link to other sugars in different positions. Blue: different substitutions that give place to different sialic acid molecules. Green: different modifications that sialic acids may undergo and affect their reactivity.

This structural variability and complexity make glycans important carriers of biological information, which is transformed into biological activity when they interact in a very specific way with their protein receptor targets. An interesting example is the important biological effect of the human transmembrane lectin receptor expressed by endothelial cells DC-SIGNR on colon cancer illustrated in Figure 1.3 (Na, *et al.*, 2017). Patients with liver metastasis show higher levels of expression of DC-SIGNR compared with patients without metastasis. Both, *in vitro* and *in vivo* experiments suggest that the inhibition of DC-SIGNR activity or downstream regulated events could be a therapeutic intervention for liver metastasis formation. In liver metastasis formation, colon cancer cells detach from the primary sites, circulate in the surrounding tissues, and enter the blood vessels where they adhere to endothelial cells through the interaction between DC-SIGNR and DC-SIGNR glycan ligands expressed on their own surface. The interaction induces cellular expression changes leading to activation of biochemical pathways that make the cells able to extravasate, migrate and invade the liver, forming a metastatic tumour.

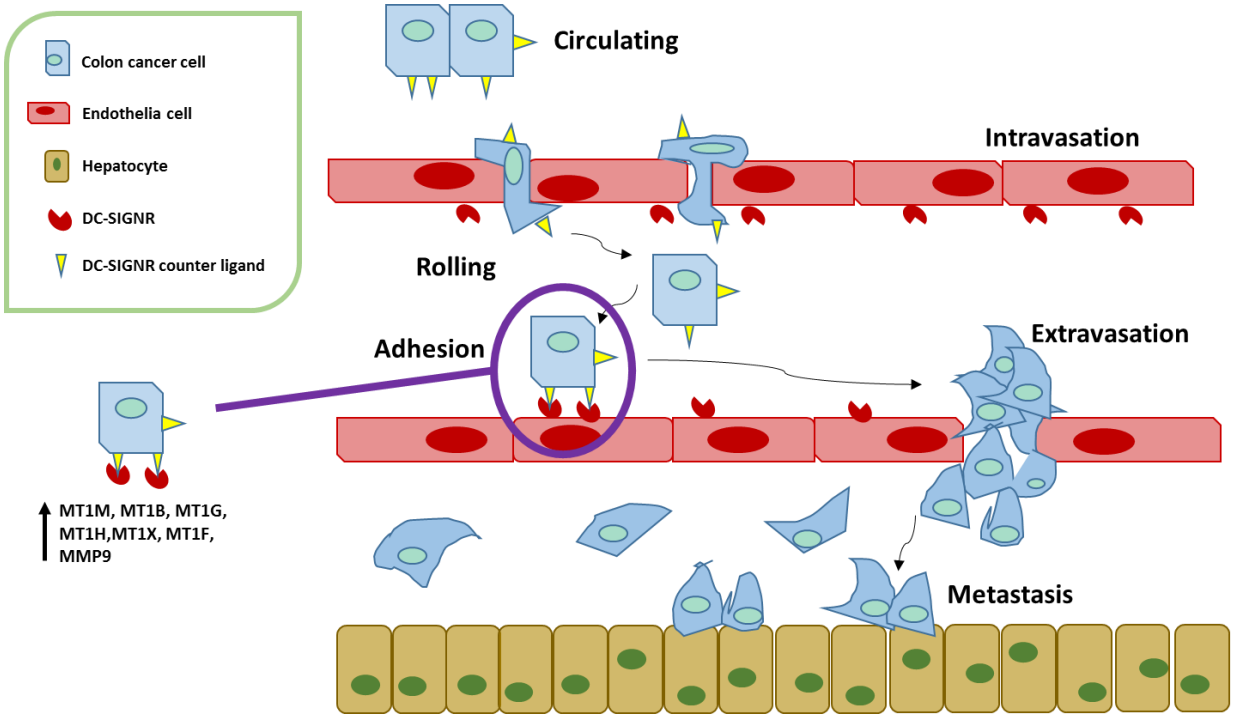


Figure 1.3. Representation of the biological effect of DC-SIGNR on colon cancer. Circulating colon cancer cells display DC-SIGNR ligands on their surface and interact with DC-SIGNR expressed by endothelial cells. The interaction triggers expression changes which enable the cancer cells to extravasate and adhere to other tissues inducing metastasis (Na, *et al.*, 2017).

Understanding the different aspects of carbohydrate-protein interactions, including the identification of ligand-receptor pairs, their three-dimensional structures, the dynamics and molecular basis of the interactions and the subsequent cellular signals, constitutes one of the main challenges of modern structural glycoscience. This is relevant for the discovery and development of new pharmaceuticals to treat those diseases in which glycan interactions are involved (Marchetti, *et al.*, 2016).

1.2 Structure and signalling of sialic acid-Siglec binding.

Sialic acid-binding immunoglobulin-like lectins (Siglecs) participate in biologically critical protein-carbohydrate interactions. Siglecs are cell-surface transmembrane receptors that belong to the lectin family. They express in specific immune system cell types and participate in cell-cell communication and cellular signalling downregulation by interacting with their endogenous ligands, some sialic acids (Macauley, *et al.*, 2014).

Sialoadhesin (also known as Siglec1) was initially described as a macrophage adhesion receptor recognizing sialic acids, and later shown to be a member of the immunoglobulin (Ig) superfamily. CD22 (Siglec2) was initially characterized as a B cell inhibitory receptor of the Ig superfamily and afterwards it was discovered to bind

sialic acids. The structural homology of these proteins with CD33 (Siglec3) and the myelin-associated glycoprotein (Siglec4) led to the establishment of the Siglec family (Crocker, *et al.*, 2007). To date, fourteen Siglecs have been identified in humans and nine in mice containing between 2 to 17 extracellular immunoglobulin domains (Macauley, *et al.*, 2014).

Figure 1.4 represents the general structure of Siglec membrane protein receptors (Macauley, *et al.*, 2014). The extracellular region of all Siglecs include an amino-terminal V-set domain that contains the sialic acid-binding site and a variable number of Ig-like C2-type domains. The intracytoplasmic domains of most Siglecs are tyrosine-based motifs, but a few Siglecs such as Siglec14, Siglec15 and Siglec16 lack of them. Tyrosine-based inhibition motifs (ITIMs) are conserved sequences of amino acids found in the cytoplasmic tails of many receptors of the immune system. The characteristic sequence of amino acids for ITIM domains is S/I/V/LxYxxI/V/L. After receptor-ligand interaction, these motifs become phosphorylated by enzymes of the Src kinase family, allowing them to recruit other enzymes such as phosphotyrosine-phosphatases or inositol-phosphatase which decrease the activation of molecules involved in cytoplasmic cell signalling. Table 1.1 summarizes the structure and disease relevance of each of the Siglec family members discovered until now (Macauley, *et al.*, 2014).

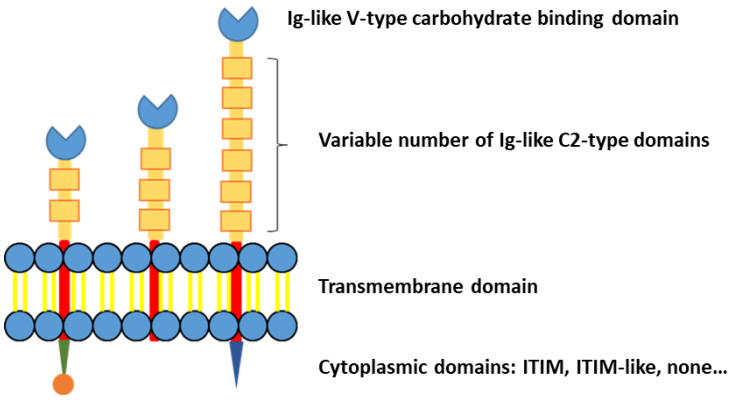


Figure 1.4. Schematic representation of the general structure of Siglec receptors. All Siglecs have an extracellular sialic acid-binding terminal Ig-like V-type domain, followed by a variable number of Ig-like C2-type domains and the transmembrane domain. Some Siglecs do not have any cytoplasmic domain and others have inhibition tyrosine based intra-cytoplasmic domains (ITIMs) (Macauley, et al., 2014).

Table 1.1. Summary of the Siglec family members discovered and some of their known features (Macauley, *et al.*, 2014).

Human Siglec	Expression	N Ig domains	Tyrosine motifs	Disease Relevance
1 (Sialoadhesin or CD169)	Macrophages and activated monocytes	17	None	Autoimmunity and infections (HIV-1, GBS…)
2 (CD22)	B cells	7	ITIM, ITIM-like and GRB2	Lymphoma, leukaemia, SLE and rheumatoid arthritis
3 (CD33)	Myeloid progenitors, macrophages, monocytes, microglia and granulocytes	2	ITIM and ITIM-like	Acute myeloid leukaemia and Alzheimer's disease
4 (MAG)	Oligodendrocytes and Schwann cells	5	FYN kinase site	Neurodegeneration
5	Neutrophils, monocytes and B cells	4	ITIM and ITIM-like	GBS infection
6	Trophoblasts and B cells	3	ITIM and ITIM-like	Pre-eclampsia
7	NK cells, monocytes and masts cells	3	ITIM and ITIM-like	Cancer and C. jejuni infection
8	**Eosinophils, mast cells and basophils**	**3**	**ITIM and ITIM-like**	**Eosinophilia and asthma**
9	Neutrophils, monocytes, DCs and NK cells	3	ITIM and ITIM-like	Chronic Obstructive Lung Inflammation
10	B cells, monocytes and eosinophils	5	ITIM, ITIM-like and GRB2	Lymphoma, leukaemia, eosinophilia and allergy
11	Macrophages and microglia	5	ITIM and ITIM-like	Neuroinflammation
14	Neutrophils and monocytes	3	None	GBS infection and COPD
15	Osteoclasts, macrophages and DCs	2	None	Osteopetrosis
16	Macrophages and microglia	5	None	Unknown

The native ligands for Siglecs are sialic acids, sugars with a nine-carbon backbone which can undergo different chemical modifications and establish different linkages with other sugars, as represented in Figure 1.2, leading to more than 50 naturally occurring structures (Macauley, *et al.*, 2014). Most Siglecs are endocytic receptors that either constitutively cycle between the cell surface and intracellular endosomes or are induced to undergo endocytosis upon ligand binding when plasma membrane vesicles containing receptor with the bound ligand are internalised into the cell (both types of endocytosis are represented in Figure 1.5).

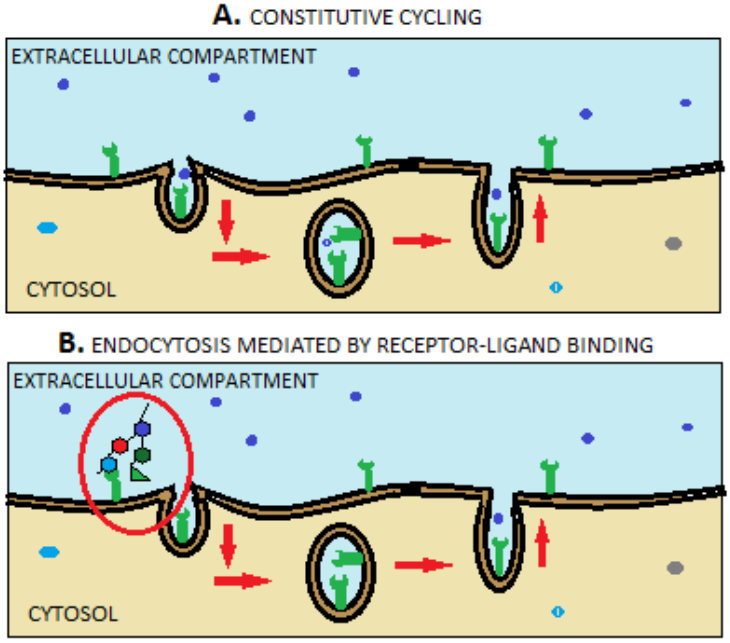


Figure 1.5. **A.** Constitutive or continuous cycling of transmembrane receptors between the cell surface and vesicles. **B.** Endocytosis mediated by receptor-ligand binding. Cycling between the cell surface and vesicles is induced when the ligand binds to the receptor.

Most Siglecs trigger the inhibition of cellular biochemical pathways upon ligand binding via their cytoplasmic immunoreceptor tyrosine-based inhibitory motifs (ITIMs). But there are also some Siglecs that activate cellular signals (such as Siglecs14, 15 and 16), as they associate with the immunoreceptor tyrosine-based activation motif (ITAM) adaptor DAP12 via a positively charged amino acid in their transmembrane region. Furthermore, some of those activating receptors as Siglec14 and 16 possess extracellular regions identical to inhibitory Siglecs (Siglec5 and 11), indicating that they may have evolved as a defence mechanism against pathogens that take advantage of inhibitory pathways to infect cells (Cao, *et al.*, 2008; Cao and Crocker, 2011; Fong, *et al.*, 2014).

1.3 Sialic Acids

The native ligands for Siglecs are Sialic acids (Sia), sugars with a nine-carbon backbone as shown in Figure 1.6.

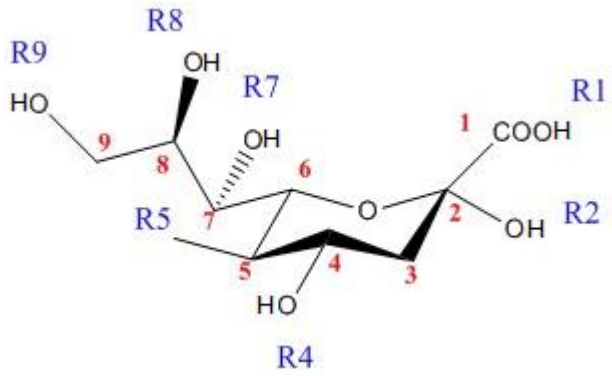


Figure 1.6. Backbone of Sialic Acids.

They are relatively strong acids (pKa ~2.6) due to the C1 carboxylate adjacent to the C2 anomeric carbon. Over 50 forms of sialic acid have been identified to date, due to the number of locations available for structural modification as well as the core variation. The hydroxyl groups located at carbons 4, 7, 8 and 9 may undergo modifications such as acetylation, phosphorylation, sulfation or methylation. Furthermore, different substitutions may be made at carbon 5. The most common sialic acid in human body is the N-acetylneuraminic acid (NeuAc), shown in Figure 1.7. Its exocyclic glycerol chain (C7, C8 and C9) provides opportunity for extensive hydrogen bonding and its N-acetyl group can participate in hydrophobic interactions. Thus, the different moieties and chemical groups play an important role in Sialic Acids binding specificity and function (Schauer, 2004).

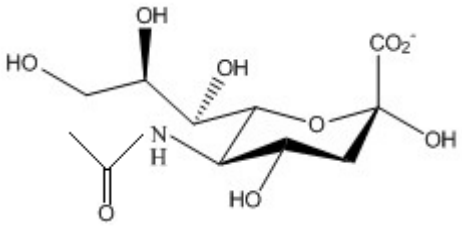


Figure 1.7. N-acetylneuraminic acid structure (NeuAc).

1.4 Sialic acids' role and location in the human body.

All eukaryotic and several prokaryotic systems express Sialic Acids. In humans, they are normally part of glycan oligomers and appear most frequently as

terminal molecules of N-glycans, O-glycans and glycosylphosphatidylinositol anchored proteins (GPI), being able to take part in many intermolecular and intercellular reactions because of their acidic negatively charged nature. However, Sialic Acids are not commonly used as energy resources (Varki and Angata, 2006) and normally play an important role as ligands of protein receptors. Thus, through their very specific interactions they take part in vertebrates' innate immune system signalling among other key physiological processes such as development, cellular recognition, cell-cell attachment and cell function regulation.

Changes in cellular sialylation have been also associated with cancer, infection, diabetes and other illnesses, but many details of the events in which they take part in as well as their pathological states remain still ambiguous. There is a fast growing of knowledge of Sialic Acids and molecules that specifically bind to them and it supposes a research field of huge interest (Crocker, 2002; Varki and Angata, 2006; Crocker, *et al.*, 2007).

For example, several bacteria, parasites and viruses like influenza possess receptors which recognize specifically cell surface Sialic Acids as ligands. Thus, pathogens can adhere to cells and infect them (Schauer, 2004; Crocker, *et al.*, 2007). As a defence mechanism against pathogen attacks, Siglecs can bind to sialylated glycans displayed in the same cell surface, leading to *Cis* interactions (Macauley, *et al.*, 2014). By attaching to these nearby endogenous sialic acids, they reduce the likelihood of viral attachment, protecting the sialylated glycans from external attachments to viral hemagglutinins. Contrarily, the loss of terminal human Sialic Acids through sialidase cleavage unmasks the Siglecs and makes them available for *Trans* binding to another host cell or to a pathogen that has a stronger affinity for the Siglec binding site (Varki and Angata, 2006).

Cis interactions also enable identifying cells as 'self' to the immune system and preventing autoimmune attacks from macrophages and other immune system cells (Macauley, *et al.*, 2014).

1.5 Sialic Acids Biosynthesis

Sialic Acid synthesis in animals takes place mainly in the cellular cytosol and consists in a five-step process which involves four enzymes. NeuNAc is synthesised and activated in the nucleus; and subsequently transferred to the Golgi Apparatus where the molecule is attached to larger glycans. Then, glycans are transported to the location where they carry out their biological function (soluble or attached to glycoproteins). They also can undergo different modifications when substitutions are added or relocated (Altheide, *et al.*, 2006).

Figure 1.8 summarizes the cycle of Sialic Acids biosynthesis in the cellular cytosol starting by their precursor 5'-diphosphate-N-acetyl-D-glucosamine (UDPGlcNAc) which is transformed into N-acetyl-D-mannosamine (ManNAc) with removal of the UDP moiety and epimerization of the carbohydrate by means of the dual activity enzyme UDP-GlcNAc 2-epimerase/ManNAc-6-kinase. Then, the kinase function of the same enzyme phosphorylates the sugar to produce N-acetyl-d-mannose-6-phosphate (ManNAc-6-P) which is transformed into N-acetylneuraminic acid 9-phosphate (Neu5Ac-9-P) by means of a condensation reaction with phosphoenolpyruvate (PEP) performed by the enzyme NeuAc-9-P-synthetase. Finally, this precursor is dephosphorylated by Neu5Ac-9-P-phosphatase to produce the Neu5Ac. After biosynthesis, Neu5Ac is transported to the cell nucleus and attached to cytidine 5'-triphosphate (CTP) to produce the donor substrate cytidine-5'-monophospho-N-acetylneuraminic acid (CMP-Neu5Ac) for subsequent transfer to an oligosaccharide structure after the loss of pyrophosphate via catalysis by the enzyme CMP-NeuAc synthetase. The CMP-Neu5Ac substrate is transported back to the cytoplasm and then to the Golgi where the syaliltransferases attach Neu5Ac residues to oligosaccharides using CMP-Neu5Ac as a sugar donor. These transferases produce specific glycosidic linkages for sialic acids (α2→3, α2→6 or α2→8) determined by their preference for specific monosaccharide acceptors. Those specificities are indicated by the enzyme's nomenclature. However, a lack of substrate selectivity was observed in some *in vitro* experiments using unnatural ManNAc structures. Figure 1.2 shows again the different linkages which Sialic Acids can form with other glycans (Altheide, *et al.*, 2006).

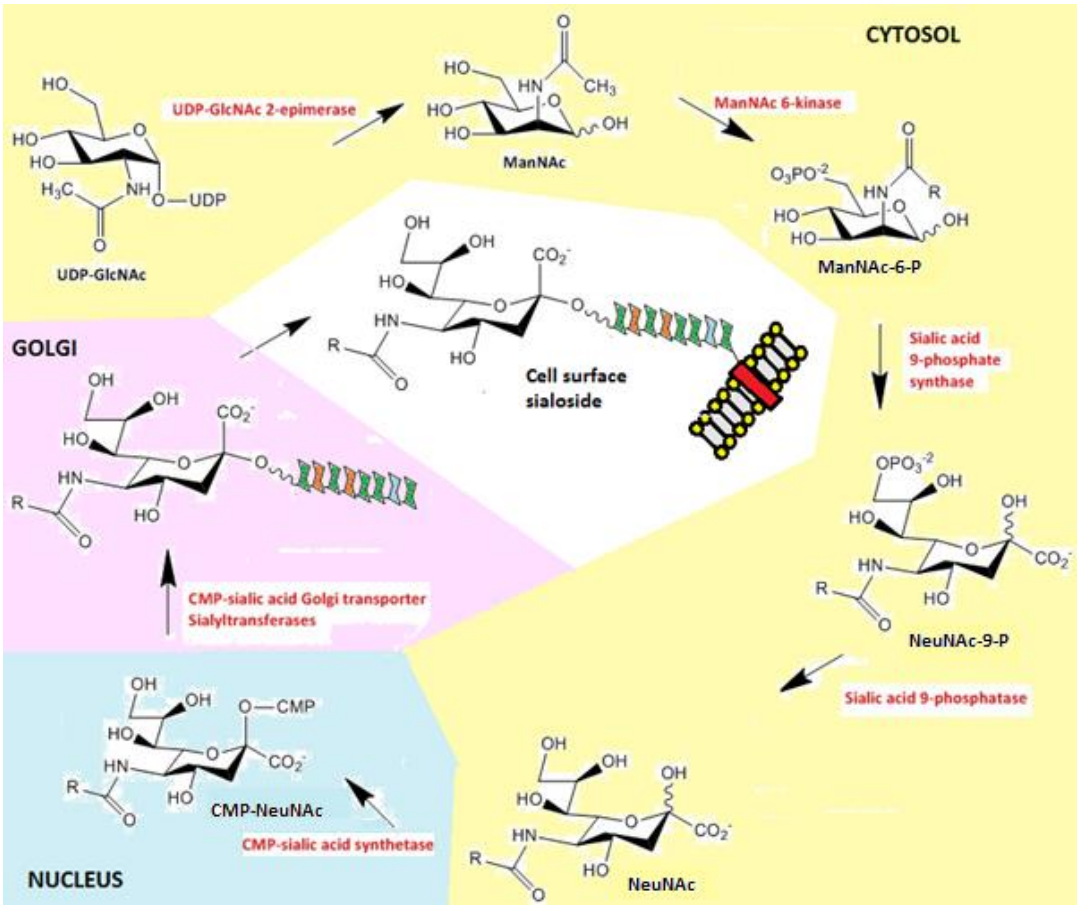


Figure 1.8. Summary of Sialic Acids synthesis, activation, transfer and transport (Altheide, *et al.*, 2006).

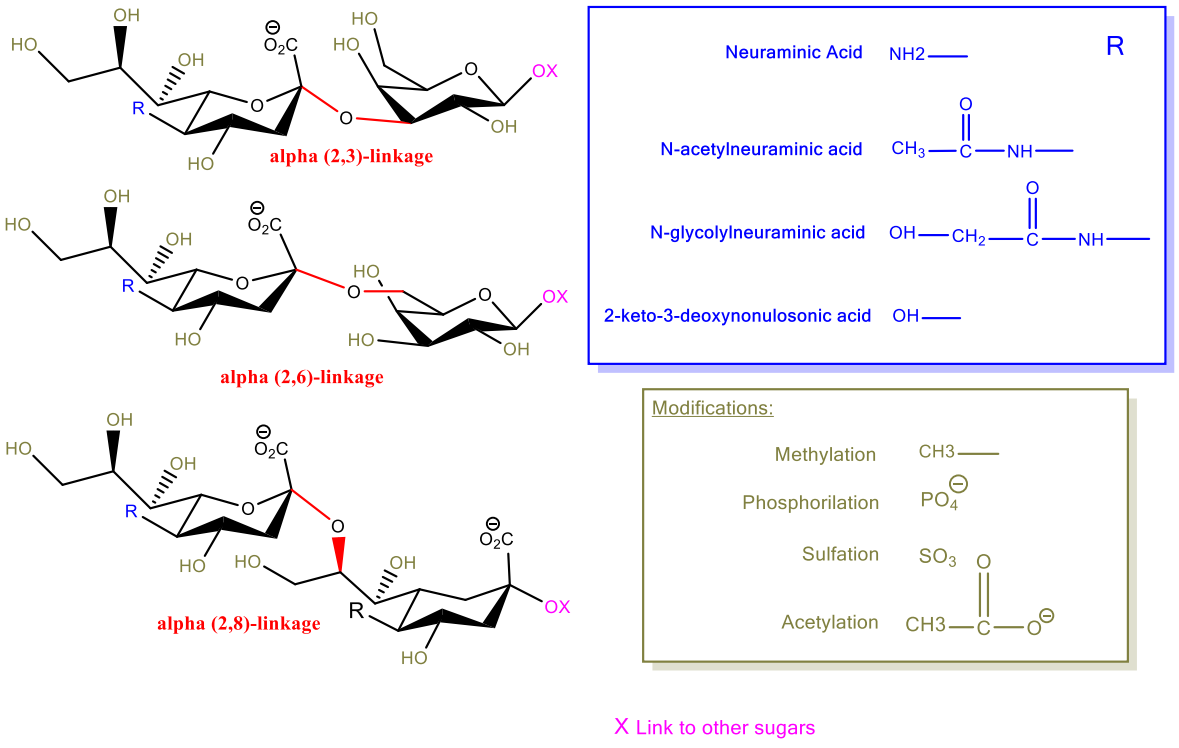


Figure 1.2. Neu5Ac and the possible linkages that it forms with other glycans (Schauer, 2004).

1.6 Types and Effects of Siglec-ligand Interactions

Each Siglec family member has specificity for different types of sialylated glycans either found on the surface of mammalian cells or soluble in extracellular fluids. Thus, we talk about *Cis* interactions when both ligand and receptor are expressed on the surface of the same cell, and *Trans* interactions when ligand and receptor belong to different cell, or the ligand is a soluble glycoconjugate, both types of interactions are represented in Figure 1.9 (Crocker, *et al.*, 2007; Pillai, *et al.*, 2012; Macauley, *et al.*, 2014). The binding can have a variety of effects on immune system cell functions such as the downregulation of cell signals triggered by other receptor-ligand interactions, the induction of apoptosis and cell migration, or the internalization of a ligand into cellular cytoplasmic vesicles. Siglecs may also be sequestered by other non-activator carbohydrate ligands which can block their active site and avoid the interaction with sialic acids, influencing over cellular signalling (Macauley, *et al.*, 2014). All types of possible interactions are summarized in Figure 1.9.

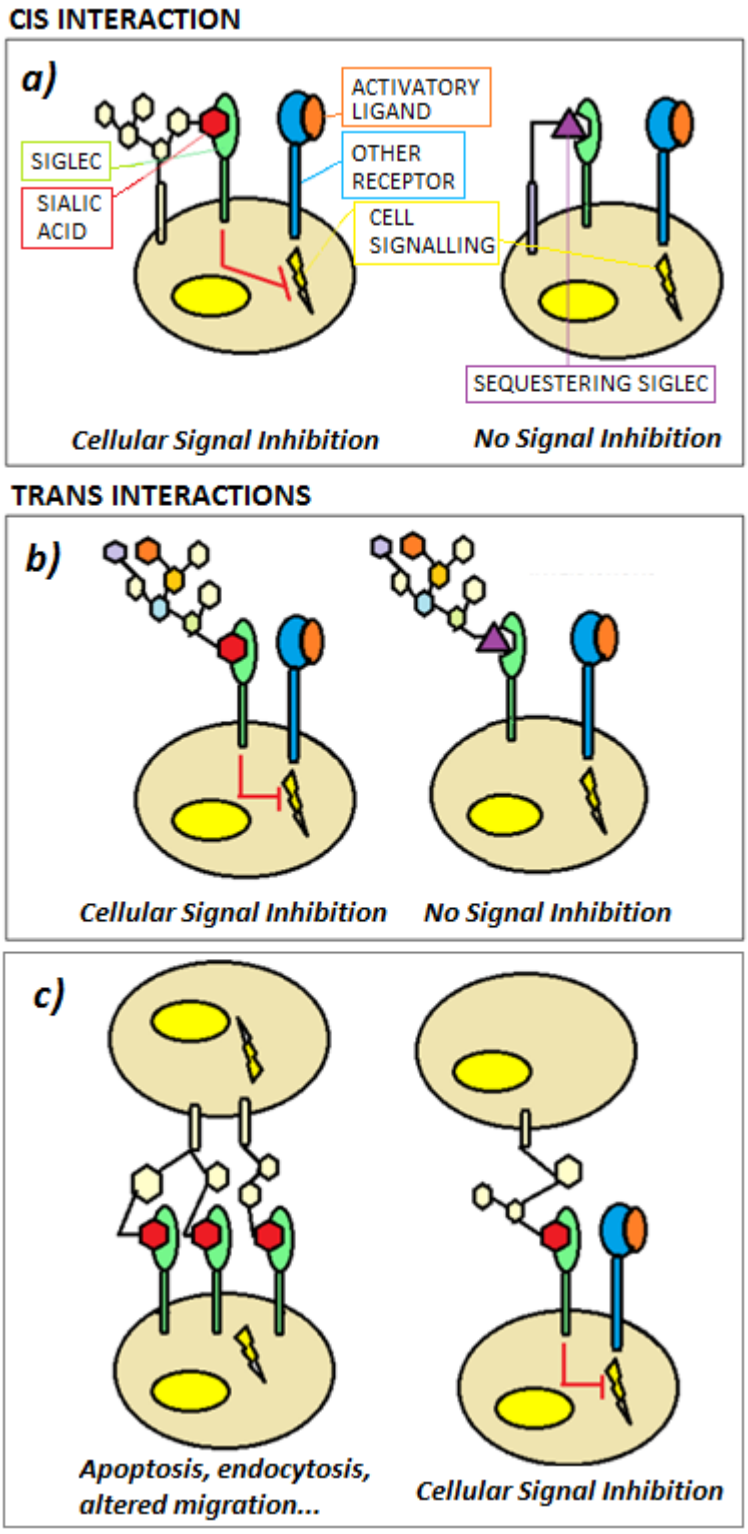


Figure 1.9. Representation of all possible types of interactions that Siglecs can carry out according to the location of their sialylated ligand. Panel a): *Cis* interactions occur when the ligand or the inhibitor are attached to a glycoprotein expressed by the same cell, and lead to cellular signal modulation. Panel b): *Trans* interactions can happen with free ligands or inhibitors, leading as well to cellular signal modulation. Panel c): *Trans* interactions with ligands or inhibitors expressed by nearby cells leading to cellular events such as migration, apoptosis or endocytosis (Macauley, *et al.*, 2014).

1.7 Siglecs and Human Disease

Siglecs have been associated with different human diseases such as bacterial and viral infections, lung inflammation, autoimmune disease and cancer (Macauley, *et al.*, 2014). Indeed, many studies have identified polymorphisms and mutations in Siglec genes linked to diseases: some Siglec3 (CD33) genetic variants are associated with a major risk factor for Alzheimer's disease (Bertram, *et al.*, 2008; Hollingworth, *et al.*, 2011; Naj, *et al.*, 2011) and Siglec8 polymorphisms are related to a higher susceptibility to asthma (Gao, *et al.*, 2010). Furthermore, several human bacterial pathogens can display sialic acid-based ligands that bind Siglecs and suppress immune response, leading to pathogen survival (Crocker, *et al.*, 2007; Pillai, *et al.*, 2012; Macauley, *et al.*, 2014), i.e. Group B Streptococcus (GBS) is classified into nine serotypes that display a capsular polysaccharide (CPS) and have in common a terminal sialic acid residue. Some studies hypothesise that GBS evolved to display that CPS Sia as a form of molecular mimicry to limit the activation of an effective innate immune response by interacting with certain human Siglecs expressed on neutrophils and monocytes, as Siglec5 and Siglec9 (Carlin, *et al.*, 2007). Table 1.1 (showed before) summarizes all known Siglec family members as well as the cell types in which they are expressed, the number of extracellular Ig-like domains and cytoplasmic motifs they contain, and the human disease to which they are linked (Macauley, *et al.*, 2014).

1.8 Siglecs as Interesting Pharmaceutical Targets

Siglecs have been highlighted as interesting targets for the development of immunotherapeutics, due to their highly restricted expression in immune system cells and differential expression in specific cellular types which relate them to very specific diseases (Crocker, *et al.*, 2007; Jellusova and Nitschke, 2012; Kawasaki, *et al.*, 2013; Kiwamoto, *et al.*, 2013; Muller and Nitschke, 2014). A good example of the successful development of an immunotherapeutic targeting a Siglec receptor is the humanized anti-CD33/Siglec3 monoclonal antibody covalently attached to a cytotoxic antitumor drug from Pfizer (commercial name Mylotarg®) for the treatment of acute myeloid leukaemia. The anti-tumour mechanism of Mylotarg® is represented in Figure 1.10: The non-cytotoxic antibody binds to the Siglec3 antigen of normal and leukemic myeloid cells (but not to their hematopoietic precursors) and is internalized into

lysosomes where acidification activates the cytotoxic conjugated drug calicheamicin by a spontaneous reaction with reduced glutathione (GSH). Then, calicheamicin targets and destroys DNA strains, killing the tumour cells (Laszlo, *et al.*, 2014; Loke, *et al.*, 2015; Garfin and Feldman, 2016).

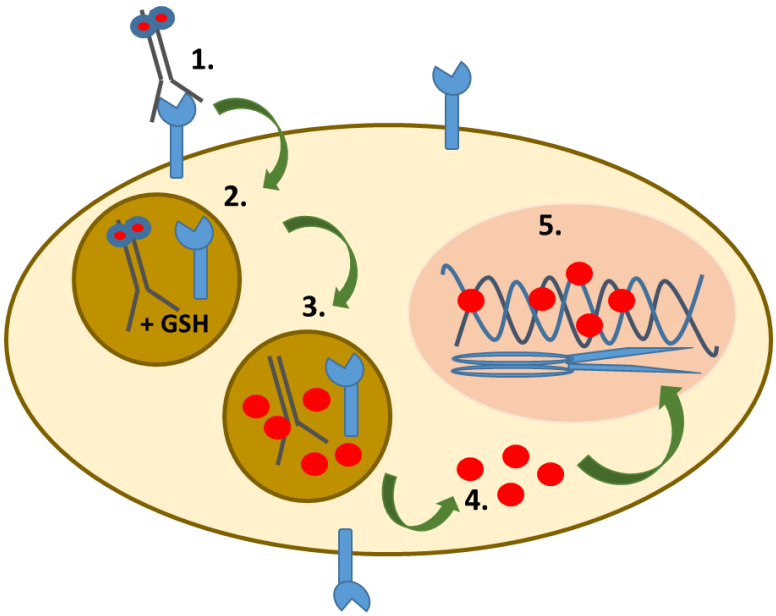


Figure 1.10. Summary of action mechanism of Mylotarg® to treat acute myeloid leukaemia. **1.** The non-cytotoxic conjugated antibody binds specifically to Siglec3. **2.** Antibody and receptor are internalised into lysosomes by endocytosis. **3.** The acid pH activates the cytotoxic conjugated drug calicheamicin by spontaneous reaction with GSH **4.** Calicheamicin is released into the cytoplasm. **5.** Calicheamicin gets into the nucleus and destroys DNA, killing the leukemic cell.

This drug was approved by the US Food and Drug Administration in 2000, but it was linked to a serious liver condition and the beneficial effects of its administration in combination with chemotherapy in acute myeloid leukaemia patients was not demonstrated in clinical trials. Mylotarg® was voluntarily withdrawn from the market by Pfizer in mid-2010. However, some regulatory authorities did not agree with that decision and more clinical trials have been performed showing Mylotarg benefits. Finally, in early 2017 Pfizer reapplied for US and EU approval for the treatment of myeloid acute leukaemia (Laszlo, *et al.*, 2014). There are also clinical studies investigating antibody-drug conjugates that target CD-22 (Siglec2) to treat B-cell malignancies (Shor, *et al.*, 2015; Jahn, *et al.*, 2016).

1.9 Siglec8 as a Therapeutic Target to Treat Asthma

Siglec8 is a transmembrane receptor (Floyd, *et al.*, 2000; Foussias, *et al.*, 2000) exclusively expressed on human eosinophils, basophils and mast cells, where it regulates their function and survival (Macauley, *et al.*, 2014). It has also been associated with chronic respiratory diseases such as asthma. Asthma is a chronic inflammatory disease of the respiratory airways characterized by variable and recurring symptoms such as wheezing, coughing, chest tightness or shortness of breath, joined by reversible airflow obstruction and bronchospasm (Bochner, *et al.*, 2013). Siglec8 is a promising pharmaceutical target to treat asthma since eosinophils are important effectors in this human allergic disease (Kiwamoto, *et al.*, 2013; Ilmarinen and Kankaanranta, 2014a; Kano, *et al.*, 2014).

Eosinophil numbers generally increase in tissues, blood and bone marrow in association with allergic reactions. There is a correlation between eosinophil levels and disease severity as they orchestrate inflammation, tissue damage and airway obstruction during an asthma crisis by releasing proteases (Ilmarinen and Kankaanranta, 2014a). Several studies have reported the induction of cytokine-activated eosinophil apoptosis (programmed cell death) when Siglec8 is activated by its ligands. This would decrease protease production and lead to a reversion of inflammation and tissue damage during an asthma crisis. Thus, Siglec8 has therapeutic potential in treatment of allergic asthma (Farid, *et al.*, 2012; Feng and Mao, 2012; Kiwamoto, *et al.*, 2012a).

Although the mechanisms by which Siglec8 receptor activation induces apoptosis still have not been completely elucidated, there are several studies that have contributed some information. Siglec8 has an intracellular immunoreceptor tyrosine-based inhibitory motif (ITIM) and an immunoreceptor tyrosine-based switch motif (ITSM) responsible for signal transduction through association with the SHP-2 protein tyrosine phosphatase (Nutku, *et al.*, 2005; Kano, *et al.*, 2013; Janevska, *et al.*, 2015; Kano, *et al.*, 2017). The studies conclude that the main causes of eosinophils apoptosis are increased mitochondrial damage and reactive oxygen species (ROS) production. Both seem to be mediated by the MAPK (mitogen-activated protein kinase)/ERK (extracellular receptor-stimulated kinase) pathway, a chain of cytoplasmic and nuclear

protein kinases that communicate the signal from the receptor on the cell surface to the DNA in the nucleus (Kano, *et al.*, 2017).

Genetic studies showed a significant association between genetic sequence variants or polymorphisms in the SIGLEC8 gene and susceptibility to asthma in diverse populations, supporting the role of Siglec8 in the disease (Gao, *et al.*, 2010). Furthermore, Siglec8 could be a potential human serum biomarker to diagnose certain hyper eosinophilic diseases which are not tissue localized (Ilmarinen and Kankaanranta, 2014a). Figure 1.11 summarizes the relevance of Siglec8 in lung inflammatory diseases and the effect that a Siglec8 agonist could perform during a disease exacerbation.

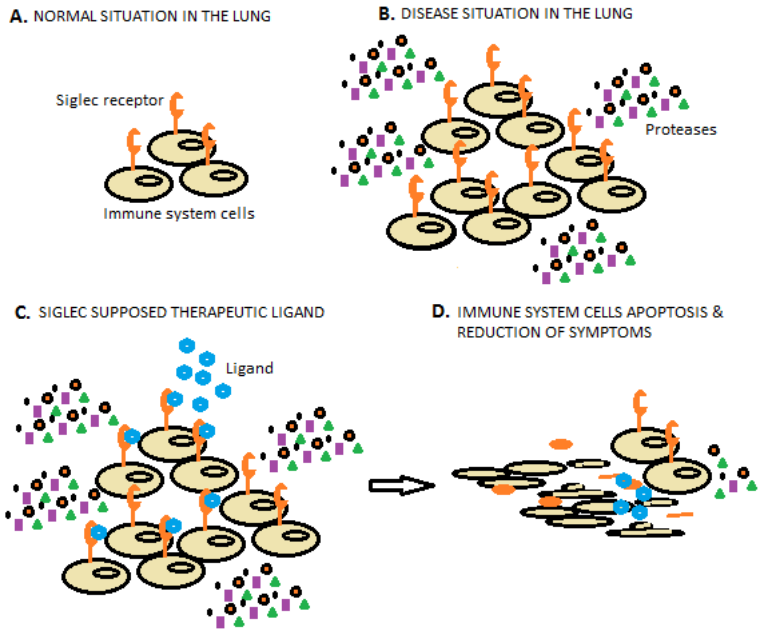


Figure 1.11. Summary of a normal and asthma situation in the lung and the effect that a Siglec8 agonist could make to treat the asthma. **A.** In a normal situation, there will be a basal concentration of immune system cells in the lung or eosinophils in this case. **B.** When an asthma crisis happens, the levels of eosinophils increase, and they release proteases that damage and produce tissue inflammation. **C. and D.** A Siglec8 agonist would induce eosinophils apoptosis and thus, decrease the proteases concentration and their effects on the lung tissue relieving the asthma symptoms.

1.10 Identification of Siglec8 Sialylated Ligands

Glycan microarrays are based on the simultaneous presentation of a library of known glycans arranged on a surface in a resolvable pattern for defining binding specificities of glycan binding proteins (Song, *et al.*, 2014). The specificity of the interactions is determined by a quantitative signal that enables to compare binding of

a protein to all glycans in the microarray, including those cases where there is no interaction, and detecting the ones that associate with the highest affinity.

It has been estimated that the human glycome is composed of more than 7,000 glycan structures, encompassing penta- and hexa-saccharide sequences and their chemical modifications. Therefore, the full potential of defined glycan microarrays will only be covered when the libraries of glycans represent the complete glycome of an organism, tissue or cell (Song, *et al.*, 2014).

Glycan microarray analyses were employed to identify the unique known weak binders for Siglec8 (Bochner, *et al.*, 2005; Campanero-Rhodes, *et al.*, 2006). They revealed specific and low affinity binding for only one single sialylated glycan: NeuAcα2-3(6-O-Sulfo) Galβ1-4[Fucα1-3] GlcNAc or 6'-Sulpho-sialyl-Le^X, whose structure is shown in Figure 1.12.

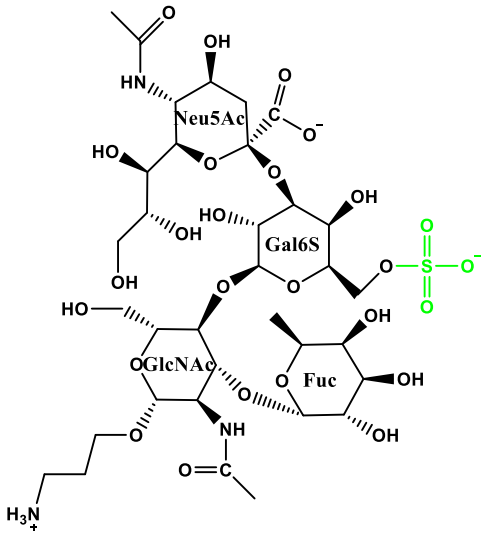


Figure 1.12. NeuAcα2-3(6-O-Sulpho) Galβ1-4[Fucα1-3] GlcNAc or 6'-Sulpho-Sialyl-Le^X structure.

Initially, Floyd and co-workers showed Siglec8 binding to Sialic Acid forms linked to Galβ1-4GlcNAc through α2-3 or α2-6 bonds (Floyd, *et al.*, 2000), but a subsequent study performed with 28 different carbohydrate molecules could not determine a clear binding preference of these molecules for Siglec8 (Blixt, *et al.*, 2003). Later, Siglec8 binding was tested again against 172 different glycan structures, 40 of which were sialylated, resulting in the detection of the 6'-Sulpho-Sialyl-LewisX glycan Figure 1.12) as a selective Siglec8 ligand with high affinity for the receptor (Bochner, *et al.*, 2005). In the same study, neither the Sialyl-LewisX nor the 6-Sulpho-Sialyl-LewisX (sulphated on the 6-position of the GlcNAc) (Figure 1.13) showed

binding, indicating that the Sialyl-LewisX 6'-sulphation is important for the interaction (Bochner, *et al.*, 2005). This result was corroborated in a different study and extended to identify another low affinity binder for Siglec8: the disulphated 6, 6'-Sulpho-Sialyl-LewisX (Figure 1.13) (Campanero-Rhodes, *et al.*, 2006).

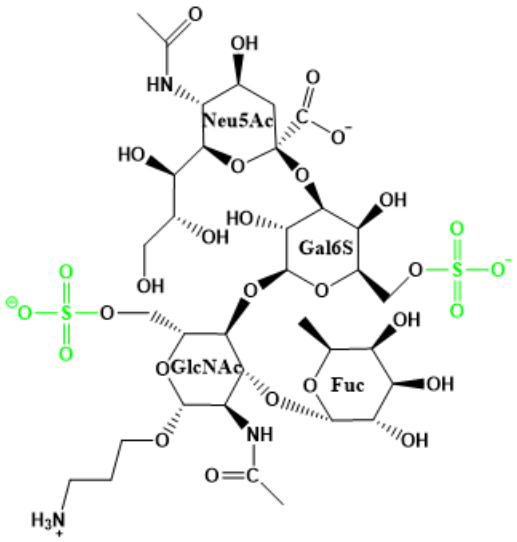


Figure 1.13. 6, 6'-DiSulpho-Sialyl-Le^X structure.

1.11 Precedents in Structural Characterization of Siglec-ligand Interactions

There are several studies in which the 3D structures of the Ig-like V-type or carbohydrate binding domains of human Siglecs have been characterized alone and/or in complex with ligands. In 2003 the Ig-like V-type N-terminal domain of the Natural Killer cell inhibitory receptor Siglec7 was resolved for the first time by means of X-ray Crystallography alone and in complex with the ligand (2,8)-linked disialic acid (pdb IDs 1O7S and 1O7V) (Alphey, *et al.*, 2003). Siglec7 structure was resolved again by means of the same technique in 2004 (pdb ID 1NKO), (Dimasi, *et al.*, 2004) and finally three more Siglec7 X-ray resolved crystal structures in complex with other three different ligands (α(2,3)-α(2,6)-Disialyl lactotetraosyl 2-(trimethylsilyl)ethyl, methyl-9-(aminooxalyl-amino)-9-deoxyNeu5Ac (oxamido-Neu5Ac) and the α(2,8)-disialylated ganglioside GT1b) were published in 2006 (pdb IDs 2DF3, 2G5R and 2HRL) (Attrill, *et al.*, 2006b; Attrill, *et al.*, 2006c). Figure 1.14 shows the structure of the Siglec7 Ig-like V-type carbohydrate binding domain in complex with the ligand α(2,3)-α(2,6)-Disialyl-lactotetraosyl-2-(trimethylsilyl) ethyl resolved by X-ray Crystallography as an example (pdb ID 2DF3).

Figure 1.14. Siglec7 Ig-like V-type domain in complex with α (2, 3)-α (2, 6)-Disialyl lactotetraosyl 2-(trimethylsilyl) ethyl resolved by X-ray Crystallography (pdb ID 2ZG1).

Later, in 2008, the structure of the two N-terminal domains of Siglec5 alone (pdb ID 2ZG2), and in complex with 6'-sialyllactose (pdb ID 2ZG1) and 3'-sialyllactose (pdb ID 2ZG3) were resolved by means of X-ray Crystallography (Zhuravleva, *et al.*, 2008a). Figure 1.15 shows the structure of the Siglec5 Ig-like V-type carbohydrate binding domain in complex with 6'-Sialylactose resolved by X-ray Crystallography.

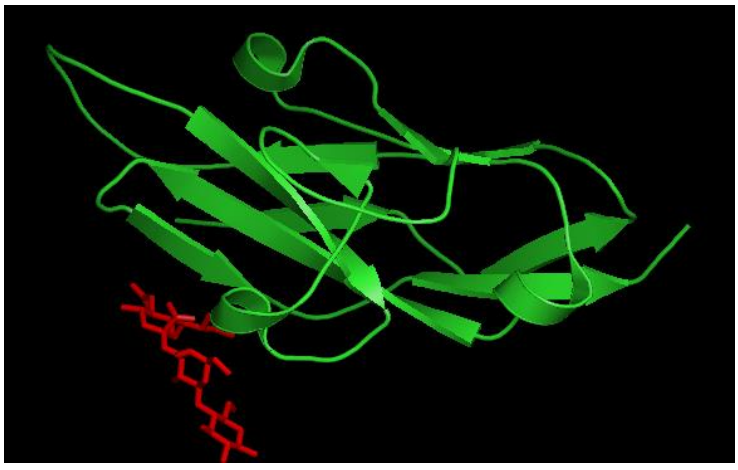


Figure 1.15. Siglec5 Ig-like V-type domain in complex with 6'-Sialyllactose resolved by X-ray Crystallography (pdb ID 2ZG1).

In July of 2016, the Siglec8 carbohydrate-binding domain structure was resolved in complex with the 6'-Sulpho-Sialyl-LewisX ligand by means of solution state NMR spectroscopy (Propster, *et al.*, 2016) (Figure 1.16).

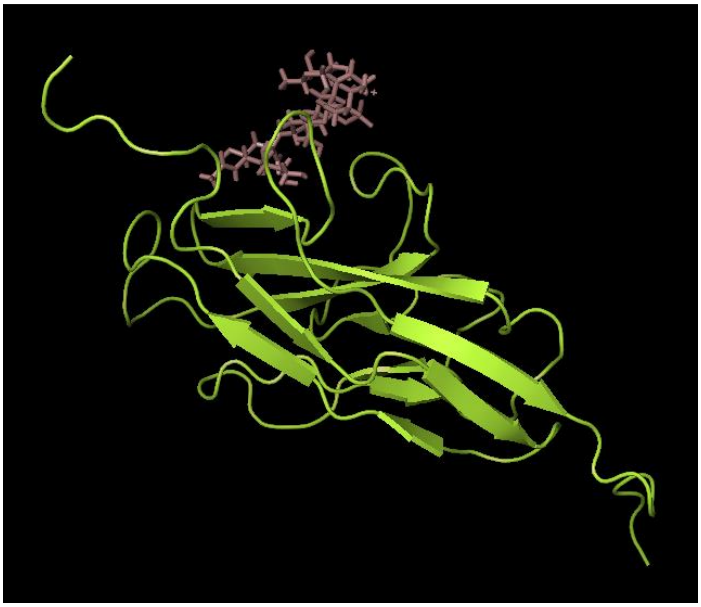


Figure 1.16. Siglec8 carbohydrate binding domain in complex with its ligand 6'-Sulpho-Sialyl-LewisX resolved by solution state NMR spectroscopy (Propster, *et al.*, 2016).

In that work, the Siglec8 Ig-like V-type carbohydrate binding domain was cloned, soluble expressed and purified (Propster, *et al.*, 2015). Subsequently the dissociation constants (K_D) for the interaction between the Siglec8 Ig-like V-type domain and four variants of the sialic acid Sialyl LewisX (SulphoSialyl-LewisX, 6'-SulphoSialyl-LewisX, 6-SulphoSialyl-LewisX and 6, 6'-DisulphoSialyl-LewisX) were determined by solution state NMR. This work confirmed 6, 6'-DisulphoSyalil-LewisX as the known weak ligand with strongest affinity (K_D = 183±18 μM) followed by 6'-SulphoSialyl-Lewis X (K_D = 295±26 μM), and distantly by 6-SulphoSialyl-LewisX (K_D = 2.7±0.8 mM) and Sialyl LewisX (K_D = 8.3±1.9 mM) as summarized in Figure 1.17.

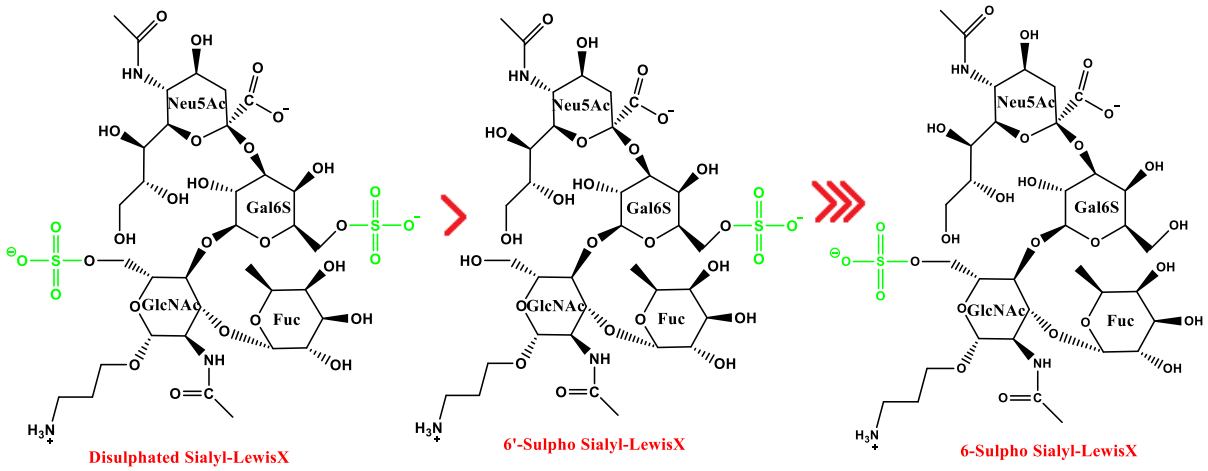


Figure 1.17. From left to right, structures of the Disulphated Sialyl LewisX which binds Siglec8 with stronger affinity than the monosulphated 6'-Sulpho-Sialyl LewisX and 6-Sulpho-Sialyl LewisX which did not show any binding to Siglec8 in the glycan microarrays (Campanero-Rhodes, *et al.*, 2006).

The molecular model proposed in this study is displayed in Figure 1.18, where the Siglec8 amino acids interacting with the 6'-Sulpho-Sialyl-LewisX as well as the nature of the interactions are indicated (Propster, *et al.*, 2016). The model describes the essential arginine (Arg109) which is conserved in all Siglecs and plays an important role in their ligand binding, anchoring the carboxyl group of the terminal sialic acid inside the binding pocket. Mutations of this residue have been shown to result in loss of binding of most Siglecs including Siglec8 (Propster, *et al.*, 2016).

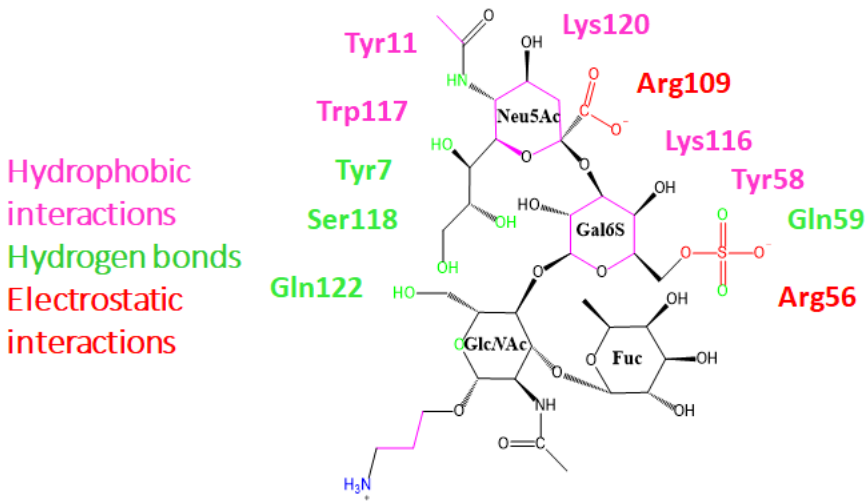


Figure 1.18. 6'-Sulpho-Sialyl-Lewis X and Siglec8 amino acids interacting with it. Amino acids and chemical groups carrying out hydrophobic interactions are represented in pink meanwhile the ones carrying out ionic interactions are shown in red and the ones forming hydrogen bonds in green (Propster, *et al.*, 2016).

1.12 Investigation of Siglecs ligands.

Similarly to many lectins, the literature describes the binding between Siglecs and sialylated oligosaccharides as low affinity interactions, with dissociation constant values between 0.1 and 3 mM (Attrill, *et al.*, 2006a), i.e. Siglec5 recognizes both α(2,3)- and α(2,6)-linked sialic acids with dissociation constants of K_D = 8.7 mM and K_D = 8.0 mM, respectively (Zhuravleva, *et al.*, 2008b). Those dissociation constant values are high in comparison to other lectins such as the C-type lectin DC-SIGN which has been found to bind mannose and fucose containing glycans with a range of affinities between 3 and 48 nM (van Liempt, *et al.*, 2006). Therefore, Siglecs high affinity binding relies on multivalent clustering (Attrill, *et al.*, 2006a).

Although weak Siglec8 binders (6,6'-Disulpho-Sialyl-LewisX and the 6'-sulpho-Sialyl-LewisX) have been found, the identity of the endogenous Siglec8 ligand/s or any other stronger binders remains unknown.

The discovery of those weak Siglec8 ligands as well as the molecular model which describes the chemical nature of the Siglec8 amino acids-ligand interactions provide the basis for the investigation and design of molecules with stronger affinity for the receptor Siglec8 with therapeutic proposes. The design of therapeutic molecules targeting Siglecs has already been done for the receptor Siglec7, expressed on the surface of natural killer (NK) cells which are able to directly lyse tumour cells. In that work, high affinity low molecular weight Siglec7 ligands (low μM K_D) were synthetized to interfere with cancer cell immune system evasion mechanisms. The mission of those molecules is to block the NK Siglec7 receptor avoiding that malignant cells camouflage themselves by displaying Siglec7 ligands on their surface and are recognised by the NK cells as 'self' (Prescher, *et al.*, 2017).

1.13 Rational and aims of this project

Apart from the 6,6'-Disulpho-Sialyl-LewisX and the 6'-sulpho-Sialyl-LewisX weak binders, at present, no more Siglec8 ligands are known. The aim of this project is the rational designing of molecules that can bind to the carbohydrate binding domain of the protein membrane receptor Siglec8 with stronger affinity, and test the binding using biophysical methods. The interest of investigating new Siglec8 agonists is to induce eosinophils apoptosis to treat hyper-eosinophilic diseases such as some subtypes of asthma (Nutku, *et al.*, 2003; Farid, *et al.*, 2012; Ilmarinen and Kankaanranta, 2014b). Furthermore, the involvement of eosinophils in another chronic lung disease, the Chronic Obstructive Pulmonary Disease (COPD), has been recently described in the literature. Those studies found increased levels of these cells in tissues and blood circulation during both stable and disease exacerbations in patients with COPD, indicating that Siglec8 may be a good biomarker for this disease (Bafadhel, *et al.*, 2017).

Asthma and COPD are phenotypically complex diseases in which different cases vary in the underlying pathogenesis. However, approximately half of the patients with asthma seem to belong to a phenotype associated with predominant eosinophilic

inflammation known as allergic asthma, where eosinophils are the main immune system effectors responsible for the symptoms (Woodruff, *et al.*, 2009). Asthma is a long-term condition that affects the respiratory airways. It is triggered when the immune system, mainly the eosinophils, react in an anomalous way against one or several stimuli such as pollen, stress, pollution… This induces inflammation, irritation and obstruction of the airways, and complicates the breathing (AsthmaUK). Chronic Obstructive Pulmonary Disease (COPD) is a chronic progressive lung disease characterised by a non-reversible airflow obstruction. COPD exacerbations are a key cause of morbidity and mortality (Dretzke, *et al.*, 2015). In England, both diseases suppose a huge economic burden for the NHS with the cost of COPD rising to £1.9 billion, and £3 billion in case of asthma each year (British Lung Foundation, 2016). In some cases, especially in older people, asthma and COPD may overlap and converge in the asthma-chronic obstructive pulmonary disease overlap syndrome (ACOS) which shares some characteristics with asthma and COPD (Papaiwannou, *et al.*, 2014).

Currently, a lot of effort is put into the investigation of new therapies to treat those cases of chronic lung diseases where the symptoms cannot be kept under control with the usual preventing and relieving treatments. The development of Siglec8 agonists which induce eosinophils apoptosis has been highlighted as a promising strategy to treat asthma (Ilmarinen and Kankaanranta, 2014b).

No Siglec8 agonists have been developed before, and there are not more clues available in the literature about Siglec8 binders than its recently resolved structure in complex with the ligand 6'-Sulpho-Syalil Lewis X. Thus, that study provided some chemical basis for the investigation and designing of higher affinity molecules targeting Siglec8.

In this work, several constructs of the Siglec8 carbohydrate binding protein domain were cloned, expressed and purified in a soluble way. Subsequently, the soluble domains were characterized by means of different biophysical techniques. Several low molecular weight carbohydrate and peptide ligands were designed using as a basis the solution state NMR Siglec8 resolved structure in complex with 6'-Sulpho-Syalil Lewis X. Finally, the binding of those molecules to the Siglec8 Ig-like V-type domain was tested in solution, and dissociation constant (K_D) values were estimated for some of the Siglec8-ligand interactions. The discovery of small molecules that bind the Siglec8 Ig-like V-type domains may lead to development of

stronger binders with therapeutic activity to treat asthma or other chronic diseases characterised by an anomalous eosinophilic proliferation.

Apart from their well-known implication in asthma, eosinophils have recently been found in the airways, tissues, and circulation of patients with COPD, during both stable and disease exacerbations, and been pointed out as biomarkers and mediators of the COPD (Bafadhel, *et al.*, 2017).

This PhD project also encompassed an industrial collaboration with Mologic Ltd. (Bedford, UK). The protocol to produce and purify Siglec8 recombinant protein domains was developed at Warwick in collaboration with Mologic. Subsequently some of the material produced was used to generate antibodies anti-Siglec8 by boosting the immune response of immunised animals. Thus, those antibodies may be employed in the development of a diagnostic test for chronic lung disease exacerbations by detecting increased levels of eosinophils in patient's body fluids.

Chapter 2.
MATERIALS AND METHODS.

2. Materials and methods.

2.1 Cloning, expression and purification of Siglec8 Ig-like V-type domains.

Strain, genotype, use and origin of all *Escherichia coli* bacterial strains are summarized in Table 2.1.

Table 2.1. Description of all *Escherichia coli* strains used in this work.

Strain	Genotype	Uses	Reference	Origin
DH5alpha	F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d**lacZ**ΔM15 Δ(**lacZYA-argF**)U169, hsdR17(r_K^- m_K^+), λ–	General cloning and storage of common plasmids	http://blog.addgene.org/plasmids-101-common-lab-e-coli-strains	Laboratory stock (stored at -80°C)
BL21 (DE3) PlySs	F- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pLysS (Cam^R)	BL21 host expression strain that expresses T7 RNA polymerase and also encode T7 lysozyme that suppresses basal expression of toxic target proteins prior to induction	http://www.merckmillipore.com/	Laboratory stock (stored at -80°C)
Rosetta gami B (DE3)	F– *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm lacY1 ahpC* (DE3) *gor522*::Tn*10 trxB* pRARE (Cam^R, Kan^R, Tet^R)	Rosetta-gami B strains combine the key features of BL21, Origami, and Rosetta to enhance both the expression of eukaryotic proteins and the formation of target protein disulfide bonds	http://www.merckmillipore.com/	Purchased from Novagen (USA)

A description of all cloning and expression vectors used in this work is included in Table 2.2.

Table 2.2. Description of all cloning and expression vectors used in this work.

Plasmid	Features	Reference	Origin
pUC-57	General cloning and storage of genes vector in *E. coli*. It was purchased with the codon optimised Siglec8 synthetic gene.	http://www.genscript.com/	Purchased from GeneScript (USA)
pET-28a(+)	T7 bacteriophage transcription system and expression induced by providing a source of T7 RNA polymerase in the host cell. N-terminal His-Tag/thrombin cleavage site to enable removal of the tag after protein purification/T7-Tag configuration plus an optional C-terminal His-Tag sequence/kanamycin resistance	http://www.merckmillipore.com/	Laboratory stock (stored at -80°C)
pET-43.1a(+)	T7 bacteriophage transcription system and expression induced by providing a source of T7 RNA polymerase in the host cell. Optional N-terminal or C-terminal His-Tag/optional Nus-Tag for expression of peptide sequences fused/ampicillin resistance.	http://www.merckmillipore.com/	Purchased from Novagen (USA)

pET151/D-TOPO T	T7 bacteriophage transcription system and expression induced by providing a source of T7 RNA polymerase in the host cell. N-terminal tag containing the V5 epitope and a 6xHis tag/TEV cleavage site to enable removal of the tag after protein purification/ampicillin resistance	https://tools.thermofisher.com/content/sfs/manuals/pettopo_man.pdf	Purchased from Invitrogen (USA)

All primers were supplied by Sigma Aldrich (UK) and are listed in Table 2.3, including the Siglec8 fragments that were amplified as well as the name of the restriction enzyme specific recognition sequences incorporated to ligate the PCR products into expression vectors. Those nucleotides complementary to the Siglec8 gene are coloured in red, specific restriction enzyme recognition sequences and TOPO cloning sequences in fuchsia, spacers in grey, stop codons in black, TEV (tobacco etch virus protease) cleavage site in orange and non-specific extra final nucleotides in blue which are required for enzymatic digestion. The table also includes primers complementary to the T7 promoter and terminator sequences, the universal RNA polymerase starting and terminating sequences found in many plasmids and vectors.

Table 2.3. All primers sequences used in this work, including incorporated restriction sites and amplified Siglec8 fragments. Those nucleotides complementary to the Siglec8 gene are coloured in red, restriction enzymes specific recognition sequences and TOPO-cloning sequences in fuchsia, spacers in grey, stop codons in black, TEV cleavage site in orange and non-specific extra ending nucleotides in blue.

Siglec8 fragment	Forward Primer (5'-3')	Forward restriction site	Reverse primer (5'-3')	Reverse restriction site
Gly40-Phe123	GATATCCATATGGGC CTGTGCGTCCATGT	*Nde* I	CTAGCTGGATCCTTA AAAGTAAGAGCCTTT GTCGCG	*Bam* HI
Met17-His155	GATATCCATATGATG GAAGGTGACCGCCAG TAT	*Nde* I	ATCTAACTCGAGGCC CTGAAAATACAGGTT TTCGCTGCCATGGGT CAGGGCCGTGA	*Xho* I
Met17-His155 (Cys42>Ser42) mutagenesis	GAAGGCCTGTCCGTC CATGTG	-----	CTGAACGGTCACCAG TTC	-----
Met17-His155 (Cys42>Ser42) For TOPO cloning	CACCATGGAAGGTGA CCGCCAGTATGGT	TOPO cloning sticky end	ATGGGTCAGGGCCGT GACGAA / TTAATGGGTCAGGGC CGTGACGAA	-----
T7 Primers	TAATACGACTCACTA TAGGG		GCTAGTTATTGCTCA GCGG	

2.1.1. DNA manipulation and cloning techniques:

Siglec 8 gene

A codon optimised gene for expression in *E. coli* encoding full-length human Siglec8 within the cloning vector pUC57, was purchased from GenScript (USA), the protein sequence is shown in Figure 2.1. The gene was used as a template to amplify different versions of the Ig-like V-type domain by means of PCR.

```
        10         20         30         40         50         60         70         80         90        100
MLLLLLLLPL LWGTKGMEGD RQYGDGYLLQ VQELVTVQEG LCVHVPCSFS YPQDGWTDSD PVHGYWFRAG DRPYQDAPVA TNNPDREVQA ETQGRFQLLG
       110        120        130        140        150        160        170        180        190        200
DIWSNDCSLS IRDARKRDKG SYFFRLERGS MKWSYKSQLN YKTKQLSVFV TALTHRPDIL ILGTLESGHS RNLTCSVPWA CKQGTPPMIS WIGASVSSPG
       210        220        230        240        250        260        270        280        290        300
PTTARSSVLT LTPKPQDHGT SLTCQVTLPG TGVTTTSTVR LDVSYPPWNL TMTVFQGDAT ASTALGNGSS LSVLEGQSLR LVCAVNSNPP ARLSWTRGSL
       310        320        330        340        350        360        370        380        390        400
TLCPSRSSNP GLLELPRVHV RDEGEFTCRA QNAQGSQHIS LSLSLQNEGT GTSRPVSQVT LAAVGGAGAT ALAFLSFCII FIIVRSCRKK SARPAAGVGD
       410        420        430        440        450        460        470        480        490
TGMEDAKAIR GSASQGPLTE SWKDGNPLKK PPPAVAPSSG EEGELHYATL SFHKVKPQDP QGQEATDSEY SEIKIHKRET AETQACLRNH NPSSKEVRG
```

Figure 2.1. Siglec8 full-length protein sequence.

Preparation of Siglec 8 DNA stocks

E. coli DH5 alpha cells were transformed with the vector pUC57 containing the Siglec 8 gene. Plasmid DNA was isolated from 5 mL *E. coli* DH5 alpha strain overnight cultures using the geneJET Plasmid Miniprep kit (Thermo Scientific, Lithuania). The kit was used according to the manufacturer's instructions and plasmid DNA was recovered in 50 µL of Elution Buffer and stored at -20°C.

PCR

All PCR reactions were performed by mixing 10 µL of Phusion Buffer GC 5x, 1 µL of 10 mM dNTPs mix and 0.5 µL of Phusion High-Fidelity DNA Polymerase (2,000 U/mL), all purchased from New England Biolabs (UK). Subsequently, 2.5 µL of both forward and reverse primers (shown in table 2.3) at a concentration of 10 mM, up to 10 ng of DNA plasmid template and deionised H_2O were added until a total volume of 50 µL was reached. All reactions were performed in a Biometra T3 Thermocycler with the following cycling conditions (Table 2.4):

Table 2.4. PCR programme used to amplify the Siglec8 Methionine 17-Histidine 155 fragment.

PCR programme	
Initial Denaturalization	98 °C, 30''
32 cycles	98 °C, 10''
	Primers annealing T^{re}, 30''
	72 °C, 30''
Final extension	72 °C, 10'
Hold	4 °C

Agarose gel electrophoresis

Agarose gels were prepared by dissolving 1% (w/v) agarose in TBE buffer (89 mM Tris, 89 mM H_3BO_3, 20 mM EDTA, pH 8.0). 1:10,000 dilution of Safe DNA Gel Stain from Invitrogen (USA) was added to allow visualization of DNA. DNA samples were prepared by mixing 2 μL of Gel Loading Dye Purple (6x) from New England Biolabs (USA) and 10 μL of DNA solution. Gels were submerged in TBE buffer and electrophoresis was performed at 100-150 V until the dye front was at least half way through the gel. DNA was visualized using a UV transilluminator and photographs of gels were taken.

Purification of DNA by gel extraction.

DNA was excised from agarose gels and extracted using a QIAprep Gel Extraction kit from Qiagen (Germany). The kit was used according to the manufacturer's instructions and DNA was recovered in 30 μL of elution buffer.

Restriction endonuclease digestion of DNA

Restriction endonuclease digestion of DNA was performed according to the enzyme manufacturer's instructions. Reactions were typically incubated at 37 °C for 1 hour before being subjected to gel electrophoresis and purified with the DNA gel extraction kit described above.

Ligation of fragments

Ligation reactions were performed using T4 DNA ligase according to the manufacturer's instructions. Vector and insert were mixed in a 1:3 ratio, and typically incubated overnight into an ice bucket at room temperature, such that the temperature gradually increased from 0 degrees to room temperature. For the fragment cloned in

the TOPO cloning vector pET-151 (Invitrogen, USA), the TOPO cloning kit manufacturer's instructions were followed.

Sequencing of plasmid DNA

Samples were sent for sequencing to GATC Biotech (Konstanz, Germany) using T7 promotor primers (shown in table 2.3).

Mutagenesis of Siglec 8 domain Cys42 > Ser42

The substitution of the Siglec 8 Cys42 residue with serine (Ser42) was performed by making a one nucleotide mutation using the Q5 Site-Directed Mutagenesis Kit from New England Biolabs (USA) according to the manufacturer's instructions. Primers used to introduce that mutation are shown in table 2.3.

2.1.2. Growth and maintenance of *E. coli*:

Media

Luria-Bertami (LB), 2x Yeast extract and Triptone (2xYT) and minimal 9 (M9) media were used for the aerobic growth of *E. coli* in liquid media.

LB (10 g/L bacto-tryptone, 5 g/L yeast extract, 10 g/L sodium chloride, pH 7.4) was used as a standard media and 2xYT (16 g/L tryptone, 10 g/L yeast extract, 5 g/L sodium chloride, pH 7.4) as well as M9 media (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl, 2.5 g/L glucose, 240 mg/L $MgSO_4$, 11.2 mg/L $CaCl_2$, 15 mg/L $ZnCl_2$, 15 mg/L $FeCl_2$, 200 mg/L yeast extract and 10 mL of BME Vitamins Solution 100x from Sigma (USA), pH 7.4) were used to optimise the bacterial growing conditions and protein expression levels. LB medium with 16 % (w/v) agar was used for the aerobic growth of *E. coli* on plates. Antibiotics were added, when necessary, to the following concentrations: Ampicillin 100 μg/mL, Kanamycin 50 μg/mL and Chloramphenicol (35 μg/mL).

Maintenance

For long term storage of *E. coli* strains and constructs generated in this study, glycerol stocks were prepared by adding 1 mL of 50 % glycerol to 1 mL of overnight culture. These were frozen immediately on dry ice and stored at -80°C.

Preparation of E. coli competent cells

A starter culture was prepared by inoculating 5 mL of LB with *E. coli* cells and incubating overnight at 37°C with shaking at 180 rpm. The next morning, 1 mL of the overnight culture was used to inoculate 100 mL of LB and was incubated for 1.5-3 hours at 37°C with 180 rpm shaking. Subsequently cells were put on ice for 10 minutes, collected by centrifugation for 3 minutes at 2,500 x g re-suspended in 10 mL of cold 0.1 M $CaCl_2$ and left on ice for 20 minutes. Another centrifugation step was carried out for 10 minutes at 2,500 x g and the pellet was re-suspended again in 5 mL of cold 0.1 $CaCl_2$ containing 15% glycerol, aliquoted in cryotubes (100 µL/tube) and frozen at -80°C.

Transformation of competent E. coli cells

1µL of DNA plasmid solution (80-100 ng/µL) was added to 100 µL of competent cells and incubated on ice for 30 minutes. Subsequently a heat shock was applied by leaving them in a 42°C water bath for 45 seconds followed by 2 minutes on ice again. Then they were re-suspended in 1 mL of LB media and incubated for 1 hour at 37°C with shaking at 180 rpm. Finally, cells were spun down, re-suspended in 200 µL of LB, and spread on LB-agar plates supplemented with the corresponding selective antibiotic and incubated overnight at 37°C.

2.1.3. Protein preparation for characterization and binding studies:

Expression of non-isotopically labelled Siglec8 Ig-like V-type domains

When LB media was used, 2 liters were inoculated with a 10 mL LB overnight culture of either *E. coli* BL21 PlySs or Rosetta gami B (DE3) and incubated at 37°C and 180 rpm shaking until an OD_{600} of 0.6 was reached. When 2xYT or M9 media were used, 2 liters were inoculated with a 10 mL LB overnight culture and incubated at 37°C and 180 rpm shaking until an OD_{600} of 2 was reached. At this stage, expression was induced by adding β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1.0 mM and cells were grown for a further 40 hours at an incubation temperature of 15°C. Cultures were taken then and centrifuged at 14,000 x g. 500 µL samples were removed before and after induction to monitor the expression by means of SDS-PAGE electrophoresis.

Optical density of culture samples

Absorbance of cell culture samples was measured at 600 nm using a UV-vis spectrometer. If required, samples were diluted to an OD_{600} between 0.1 and 0.8 with the same media used for growth. Media was used as a blank for all measurements.

Cell lysis by cell disruption

After 40 hours of expression, cell cultures were centrifuged for 15 minutes at 4,000xg, re-suspended in 5 mL of BugBuster Protein Extraction Reagent (Merck Millipore, USA) per gram of wet cells supplemented with EDTA-free protease cocktail inhibitor (Roche, Germany) at the manufacturer's recommended concentration, and frozen at - 20°C. The same day or days after, the cell suspension was thawed at room temperature and passed 3 times through a cell disruptor (One Shot, Constant Systems Ltd., UK) at 24 poises pressure. Subsequently soluble and insoluble material were separated by centrifugation at 30,000 x g for 30 minutes, and samples of both were removed in order to determine the solubility of the protein by running an SDS-PAGE gel.

SDS poly-acrylamide gel electrophoresis (SDS-PAGE)

Acrylamide sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was used for the separation of proteins based on their molecular weight and charge. All SDS-PAGE gels were made of one stacking and one resolving layer which contained 4 and 12 % (v/v) of acrylamide respectively and were prepared using the standard Laemmli Method (Laemmli 1970). Standard running buffer (3 g/L Tris Base, 14.4 g/L glycine and 1 g/L SDS in H_2O) provided by the Chemical Biology Research Facility was used. Gels were run at 125 V for 15 minutes and later at 200 V for 35 minutes.

SDS-PAGE sample preparation

All cell culture samples were normalised with respect to the concentration of cells in the pre-induction sample according to their OD_{600} by applying the following relationship:

$$Volume\ of\ SDS-SDSPAGE\ loading\ buffer = \frac{Volume\ of\ sample - OD}{6}$$

Thus, samples were diluted in the suitable volume of SDS-PAGE loading buffer (50 mM Tris-HCl, 100 mM dithiothreitol (DTT), 2 % (w/v) SDS, 0.1 % bromophenol blue and 10 % (v/v) glycerol, pH 6.8), boiled at 90°C for 5 minutes, and 10 μL loaded per well in the gels. All cellular protein samples were prepared by means of the same procedure, mixing 5 μL of loading buffer and 15 μL of sample.

SDS-PAGE gel staining

Coomassie stained gels were immersed in 20 mL of fixing solution for 10 minutes (50% (v/v) MeOH, 10% (v/v) Acetic Acid) and then stained in 20 mL of Coomassie solution (0.025% (w/v) Coomassie blue R250, 10% (v/v) MeOH) for 2 to 4 hours. Subsequently, they were destained in a 10% (v/v) acetic acid solution until the bands were clearly visible. Alternatively, gels were immersed for 1 hour in 10 mL of the pre-made dye Instant Blue from Expedeon (UK) and distained in water.

Western Blotting

SDS-PAGE gels were electro-blotted onto a 0.45 nm pore size nitrocellulose membrane immersed in transfer buffer (25 mM Tris Base, 190 mM glycine and 20% methanol, pH 8.3) at 25 V for 1 hour and 30 minutes. After the transfer, the membrane was blocked in 5% (w/v) powdered skimmed milk dissolved in Tris-Buffered Saline and Tween-20 (TBST) (30mM Tris-HCl, 140mM NaCl, 3mM KCl and 0.1 % Tween-20, pH 7.4) for 30 minutes at room temperature and subsequently washed several times with TBST. Later, the membrane was incubated overnight at 4°C in the same milk solution supplemented with 2 μL of mouse monoclonal anti-6x C-terminal His Tag antibody (Abcam, UK), washed again the next morning and incubated for 1 hour at room temperature in the milk solution supplemented with 2 μL of Goat Anti-Mouse IgG2c antibody with pre-adsorbed alkaline phosphatase. After several washes the membrane was incubated for a maximum of 15 minutes with Novex Alkaline Phosphatase Chromogenic Substrate (Invitrogen, USA).

Insoluble material treatment and protein preparation for refolding

Insoluble material was washed several times to isolate the protein inclusion bodies from other cellular material. Three repeated washes were carried out by re-suspending the insoluble material in 5 mL of wash buffer (100 mM Tris-HCl, 5 mM EDTA, 5 mM DTT, 2 M urea and 2 % (w/v) Triton X-100, pH 7.0) per gram of original

wet cells and centrifuging for 20 minutes at 30,000 x g. The fourth wash was performed with the same buffer prepared without Triton X-100. At this point, inclusion bodies were aliquoted in individual tubes containing 10 mg of dry material which were freeze-dried and kept at – 80 °C for further refolding.

Refolding of Siglec8 Ig-like V-type protein domains using a drop-wise method

Two aliquots of 10 mg washed inclusion bodies were defrosted and solubilized in 10 mL of protein extraction buffer (50 mM Tris-HCl, 8 M guanidine hydrochloride, 5 mM DTT, 5 mM EDTA, pH 7.5), giving a concentration of 2 mg/mL. Subsequently, the solution was added drop by drop to 1 L of refolding buffer (6.05 g/L TRIS-HCl, 43 g/L guanidine hydrochloride, 0.5 g/L reduced glutathione and 0.5 g/L oxidized glutathione, pH 8.5) and kept 2 hours under slow shaking with a magnetic stirrer at room temperature. After 2 hours, the solution was transferred to the cold room and kept overnight under slow shaking at 4°C.

Purification of refolded Siglec 8 Ig-like V-type protein domains

After overnight refolding, the 1 L of refolding solution was centrifuged at 2,500 x g for 10 minutes to remove any unfolded protein and other insoluble material. The soluble fraction was passed overnight through a 1 mL HisTrap HP nickel column (GE Healthcare Life Sciences, UK) using a peristaltic pump. The next day, the column was washed with 5 column volumes of 20 mM NaPi, 500 mM NaCl, 20 mM imidazole, pH 7.5 buffer, and the protein was eluted in 5 mL of 20 mM NaPi, 500 mM NaCl, 500 mM imidazole, pH 7.5 buffer. Subsequently, the protein solution was washed with a 20x volume excess of 20 mM NaPi, 500 mM NaCl, pH 7.5 buffer using 10 kDa MWCO centrifugal concentrators (Merckmillipore, UK) to get rid of the imidazole.

2.2 Biophysical characterization of the Siglec 8 Ig-like V-type domain.

Siglec8 construct composition and physicochemical characteristics

The Siglec8 amino acid sequence in FASTA format was introduced as an input in the online software ProtParam from ExPASy (SIB). The analysis provided domain's amino acid composition, molecular weight, isoelectric point, extinction coefficient and hydrophobicity.

The Siglec8 3D-structure illustrating the distribution of different amino acid types was downloaded from the Protein Data Bank (Propster, *et al.*, 2016) and loaded into Pymol (Pymol), where hydrophobic, hydrophilic and charged amino acids where manually highlighted in different colours.

Siglec8 identity confirmation by means of tandem mass spectrometry (nano-LC-ESI-MS/MS)

A SDS-PAGE gel with 10 µL of pure Siglec 8 protein was run and the protein band was excised and sent to the Proteomics facility in the School of Life Sciences at Warwick for analysis. The protein was extracted from the poly-acrylamide gel band and enzymatically digested with trypsin. An aliquot containing 20 µL of extracted peptides (total sample volume 50 µl) was analysed by means of nano LC-ESI-MS/MS with the Ultimate 3000/Orbitrap Fusion instrumentation (Thermo Scientific) using a 30-minute LC separation on a 25-cm column. The data were used to interrogate the Homo sapiens and Escherichia coli databases and the common contaminant database from MaxQuant. MaxQuant software was used for protein identification and Scaffold software for data analysis and visualise the results.

Quantification of Siglec 8 Ig-like V-type domain concentration by UV-vis absorbance

After refolding and purification, the UV-vis absorbance of 50 µL of the resulting solution of Siglec 8 Ig-like V-type constructs in 20 mM NaPi, 500 mM NaCl, pH 7.5 buffer was measured between 230 and 330 nm using a Jasco V-660 spectrophotometer. After subtracting the buffer absorbance from the protein solution, the intensity at 280 nm was used to calculate the domain concentration by means of the Beer-Lambert law using a calculated extinction coefficient of 35,535 $L \cdot mol^{-1} \cdot cm^{-1}$.

Siglec 8 Ig-like V-type domain fold monitored by Circular Dichroism (CD)

A CD spectrum was recorded for a 200 µL fresh Siglec 8 Ig-like V-type domain sample after refolding, purification and buffer exchange into 50 mM Tris, pH 7.5, between 180 and 300 nm. The protein concentration was approximately 1 mg/mL. The instrument was a Jasco J-1500 CD Spectrometer and the measurement parameters

were: data pitch 0.2 nm, CD and FL scales 200 mdeg/1.0 dOD, D.I.T. 2 s, bandwidth 2.0 nm, scanning speed 100 nm/min and 32 accumulations.

Siglec 8 Ig-like V-type domain folding monitored by means of fluorescence

1 mg/mL Siglec 8 Ig-like V-type domains and lysozyme samples in 20 mM NaPi, 500 mM, and pH 7.5 buffer were solubilised with increasing concentrations of denaturing agent GdnHCl (between 0 and 2.0 M for Siglec 8 and between 0 and 3.5 M for lysozyme). Samples were left to equilibrate for 24 h at 4°C, and the next day fluorescence measurements were performed for all samples using a JASCO FP-6500 Spectro fluorometer. Samples were excited at 280 nm (Trp and Tyr excitation wavelength) and the emission spectra were collected between 320 and 380 nm.

Siglec8 Ig-like V-type domain size characterization by small angle X-Ray scattering (SAXS)

Small-angle X-ray scattering (SAXS) measurements were made using a Xenocs Xeuss 2.0 equipped with a micro-focus Cu Kα source collimated with Scatterless slits. The scattering was measured using a Pilatus 300k detector with a pixel size of 0.172 μm x 0.172 μm. The distance between the detector and the sample was calibrated using silver behenate ($AgC_{22}H_{43}O_2$), giving a value of 1.195(5) m. The magnitude of the scattering vector (q) is given by $q=4\pi\sin\theta\lambda$, where 2θ is the angle between the incident and scattered X-rays and λ is the wavelength of the incident X-rays. This gave a q range for the detector of 0.01 Å-1 and 0.33 Å-1. Samples were injected into a Xenocs Low Noise Flow Cell giving a sample thickness of 1.1 mm. A radial integration of the 2D scattering profile was performed and the resulting data corrected for the absorption, sample thickness and background from the sample holder. Finally, the scattering intensity was then rescaled to absolute intensity using glassy carbon as a standard.

The radius of gyration prediction was performed using the downloaded Siglec 8 Ig-like V-type solution state NMR resolved structure from the Protein Data Bank (pdb ID 7N2A) and the software Gromacs (Gromacs). An energy minimisation setting water as the solvent, followed by two equilibrium simulations and a 1 ns molecular

dynamics simulation without position restraints were run. Subsequently, the radius gyration (Rg) was measured across a 1 ns trajectory.

2.3 Exploring Siglec 8 as a therapeutic and diagnostic target.

Design of small molecules targeting Siglec 8 Ig-like V-type domains

Small molecules targeting Siglec 8 Ig-like V-type domains were designed using the chemical model of a Siglec 8-binding glycan described in the literature as a starting point (Propster, *et al.*, 2016). Dry de-salted di-peptides Glu-Val, Glu-Asp, Glu-Leu and Glu-Phe and Glu-Pro were synthesised using FMOC chemistry and purified by Insight Biotechnology (UK). Both L-Phospho-Tyrosine and N-Acetylneuraminic Acid were purchased from Sigma Aldrich (UK). Glucuronic Acid and D-Galactose-6-Sulphate were purchased from Carbosynth (UK).

Siglec 8 Ig-like V-type domain biotinylation

Siglec 8 Ig-like V-type domain biotinylation was performed using an EZ-LinkSulfo-NHS-LC-Biotinylation Kit with a 1:1 ratio, according to Thermo Fisher Scientific instructions. The total length added to the protein chain (Biotin + spacer arm) was 22.4 Å.

Siglec 8-small molecule binding measurements using the Octet technology

Ligands were solubilised in 20 mM NaPi, 500 mM NaCl, pH 7.5 and different dilutions were made to achieve a range of experimental concentrations. The concentration of biotinylated Siglec 8 protein domain was maintained at a constant 20 μM in the same buffer. A total volume of 250 μL per well of either protein or ligand were added to the Octet plate according to the experimental design described in Chapter 5 and the experiment was carried out using Dip and Read Streptavidin Biosensors (Pall ForteBio, Europe).

The experiment was programmed setting the following times according to ForteBio instructions: 60 s for the baseline, 60 s of loading, 60 s of custom (washing) step, 60 s for a second baseline, between 300 and 600 s for association and between 300 and 600 s for dissociation. Data were processed by subtracting the blank spectra (experiment without protein and experiment without ligand) from the sample spectra

according to manufacturer's instructions and curves were fitted using a heterogeneous, local and partial model.

Siglec 8-small molecule binding measurements using equilibrium dialysis.

One equilibrium dialysis device was used to test the protein-ligand binding at each ligand concentration. Each device has two chambers separated by a dialysis membrane of 5,000 kDa MWCO which allows the ligand molecules to pass from one compartment to the other but not the protein molecules. Ligands were solubilised in 20 mM NaPi, 500 mM NaCl, pH 7.5 and different dilutions were made to achieve a range of experimental concentrations between 50 and 500 μM. The concentration of Siglec8 protein domains in the same buffer was maintained at a constant 50 μM concentration. 50 μL of 50 μM protein solution were added to one of the equilibrium dialysis device chambers and 50 μL of a specific concentration of ligand solution to the other. Each device was left to equilibrate for 2 hours at room temperature under gentle shaking. After equilibration, 50 μL of solution from the ligand were recovered to measure the ligand concentration. For each ligand concentration, another experiment with protein but no ligand was set up as a control. Each experiment was repeated for each ligand concentration at a fixed protein concentration plus a blank.

Quantitative solution state NMR after Siglec8-Phospho-Tyrosine equilibrium dialysis

A volume of 50 μL was taken from each equilibrium dialysis device's ligand chamber and diluted with 85 μL of HPLC quality H_2O and 15 μL of D_2O (10% v/v) to reach a volume of 150 μL for solution state NMR measurements. The DDS (4,4-dimethyl-4-silapentane-1-sulfonic acid) standard was prepared in the same way by diluting 50 μL of a 100 μM DDS solution with 85 μL of HPLC quality H_2O and 15 μL of D_2O (10% v/v).

One dimensional ^1H NMR spectra of all samples were measured in the 700 MHz Bruker Avance NMR spectrometer at the University of Warwick. All spectra were measured using an offset (O1) of 4.719 ppm, a spectral window of 14 ppm, a receiver gain of 912, a 90-degree pulse of 9.00 μs, 4 dummy scans and 256 scans.

Spectra were processed using the Topspin 3.2 software. First, a baseline correction was applied, subsequently all peaks were integrated using identical integral

regions to compare their intensities. For each concentration, the intensity of the sample was subtracted from the intensity of the control, ΔI ($I_{control} - I_{sample}$). The ′sample′ was the ligand solution taken from the equilibrium dialysis experiment with protein and the ′control′ the ligand solution at the same concentration as the sample taken from the equilibrium dialysis experiment without protein. The integral of the peak belonging to the three DSS methyl groups was used as standard to convert ΔI into ΔC ($C_{control} - C_{sample}$). ΔC was used to calculate the amount of ligand bound to the protein receptor ($C_{bound} = (C_{control} - C_{sample}) *2$) and a dissociation curve was built by representing the initial concentration of ligand added to the chamber versus the fraction of ligand bound to the protein receptor ($B = C_{bound} / C_{receptor}$).

Different values of the fraction of Phospho-Tyrosine bound to Siglec8 were represented against the different initial concentrations of Phospho-Tyrosine added to the equilibrium dialysis device and curves were fitted according to the Hill equation using the software Origin (Origin) and information about binding cooperativity and affinity were extracted from the obtained equation.

Quantitative fluorescence after Siglec8-Phospho-Tyrosine equilibrium dialysis

A volume of 50 µL was taken from each equilibrium dialysis device's ligand chamber and a fluorescence spectrum measured between 250 and 500 nm using a JASCO FP-6500 Spectro fluorometer. Samples were excited at 268 nm, which is the maximum absorbance wavelength of Phospho-Tyrosine. The maximum emission fluorescence wavelength was 297 nm and the values of fluorescence intensity at this wavelength were used to calculate the difference in fluorescence emission between sample and control ($F_{control} - F_{sample}$). A standard solution of Phospho-Tyrosine was used to convert ΔF into ΔC ($C_{control} - C_{sample}$). ΔC was used to calculate the amount of ligand bound to the protein receptor ($C_{bound} = (C_{control} - C_{sample}) *2$) and a dissociation curve was built by representing the initial concentration of ligand added to the chamber versus the fraction of ligand bound to the protein receptor ($B = C_{bound} / C_{receptor}$). Data were fitted to the Hill equation using the software Origin.

Chapter 3.

CLONING, EXPRESSION AND PURIFICATION OF SIGLEC8 IG-LIKE V-TYPE DOMAINS

3. Cloning, expression and purification of Siglec8 Ig-like V-type domains.

3.1 Introduction to recombinant protein expression.

A recombinant protein is a protein encoded by a gene that has been cloned in a cell (bacterial, yeast, mammalian cell) that supports expression of the gene, and translation of messenger RNA. The production of recombinant proteins in heterologous organisms has represented a huge advancement to produce desired proteins of medical or industrial interest, avoiding use of large amounts of animal or plant tissue otherwise required to obtain sufficient quantities of protein. Production of recombinant proteins first involves (Rosano and Ceccarelli, 2014; Lyons, *et al.*, 2016) cloning of the gene of interest into an expression vector. A vector is a fragment of circular DNA provided with specific regulatory sequences that act as enhancer and promoter regions, leading to efficient production of the protein encoded by the inserted gene. This is achieved by the production of significant amount of stable messenger RNA, which can be translated into the protein. The vector is introduced into a suitable expression host, which can be prokaryotic such as bacteria, or eukaryotic as mammalian cells. Prokaryotic cells have a simpler structure than eukaryotic cells; their main differences reside in the presence of membrane-bound organelles, cell wall, DNA structure and the energy producer organelles, chloroplasts and mitochondria. The host cell is then grown to cell densities of 0.6-0.8, at which point protein expression is induced with a molecule that triggers transcription, typically Isopropyl β-D-1-thiogalactopyranoside (IPTG) if a gene is under the control of the lac operator. This compound mimics allolactose, a lactose metabolite which binds to the lac repressor sequence, taking its place and triggering gene transcription. Furthermore, unlike allolactose, IPTG cannot be hydrolysed by the cell and the expression of lac-controlled genes is not inhibited during the experiment. Once expressed, the cells are harvested, and the desired protein is purified. Protein purification is required to get as much soluble, folded and functional protein as possible. The complexity of this step varies depending on whether high amounts of soluble protein are produced in the cell cytoplasm, if the protein levels are low, or if the protein is insoluble and aggregates forming inclusion bodies.

Many challenges can occur throughout the different steps of recombinant protein production, and therefore it is very important to choose the appropriate combination of host, vector, media and culture conditions. One of the first considerations is whether there are any features of the protein sequence and structure that could adversely affect expression and solubility (Lyons, *et al.*, 2016). This could include post translational modifications, enzymatic modification of proteins during or after protein biosynthesis which modify an existing functional group or introduce a new one. Some modifications such as phosphorylation commonly regulate the activity of enzymes, and others like glycosylation can promote protein folding and improve stability. Prokaryotic cells lack the required machineries to perform them. Another modification is formation of disulphide bonds, established through the thiol groups of two cysteines, which are key protein folding features. The presence of signal peptide sequences should also be considered. These sequences are usually 16-30 amino acids in length and are present at the N-terminus of newly synthesized proteins. In eukaryotes, they direct the insertion of proteins into the membrane of the endoplasmic reticulum and subsequently, they are cleaved off. When expressing eukaryotic proteins in bacteria, they must not be included as they cannot be recognized and lead to insolubility.

Careful consideration must also be given to the choice of expression host. Prokaryotic hosts (bacteria) and eukaryotic hosts (e.g. yeast, unicellular algae, filamentous fungi, or tumoral mammalian cell lines) have strengths and weaknesses. Bacteria are a very convenient option because of their easy production of large amounts of protein in very short times using relatively cheap and simple media, molecular tools and reagents. However, prokaryotes lack the enzymatic machinery to carry out post-translational modifications or molecular folding, thus multi-domain eukaryotic proteins expressed in bacteria often are non-functional or become insoluble as inclusion bodies. Inclusion bodies are difficult to recover and require several extraction steps with denaturing agents followed by protein refolding methods that aim to yield protein in its native conformation (often at much lower yield). Therefore, in those cases, yeast may be a better choice as expression hosts. Moreover, when human-like splicing of mRNA is required, insect or mammalian cell lines are the best alternative (Lyons, *et al.*, 2016).

3.2 Cloning and expression of protein domains

A protein domain is a conserved part of a given protein sequence which has a tertiary structure that can evolve, function and exist independently of the rest of the protein chain (Lyons, *et al.*, 2016). Each domain forms a compact three-dimensional structure that often can be independently stable and folded. The expression of a single protein domain of interest as opposed to the larger, full-length protein has the advantages of simplified data interpretation and (potentially) simpler expression. However, the protein sequence must be carefully analysed for the presence of post-translational modifications as well as inter-domain disulphide bonds. In addition, it is important to carefully select the sequence to be cloned to ensure that the entire domain is present, with no residues missing and no extra amino acids included that belong to the next domain. In this work the Immunoglobulin-like (Ig-like) V-type carbohydrate binding domain of the human membrane protein receptor Siglec8 was amplified by PCR using the full length SIGLEC8 synthetic gene, and subsequently cloned and expressed to perform ligand binding studies.

3.3 Analysis of Siglec8 structural features

The protein sequence and reported Siglec8 receptor structural features were found in the protein database 'Uniprot'. All well delimited and characterized domains are shown in Figure 3.1. The receptor contains three extracellular immunoglobulin-like domains and one cytoplasmic immunoreceptor tyrosine-based inhibitor involved in modulation of cellular responses (Foussias, *et al.*, 2000). The Ig-like, V-type carbohydrate binding domain spans from glycine 40 to phenylalanine 123 and forms an inter-domain disulphide bond between the cysteines 42 and 181, and an intra-domain disulphide bond between the cysteines 47 and 107 (Foussias, *et al.*, 2000).

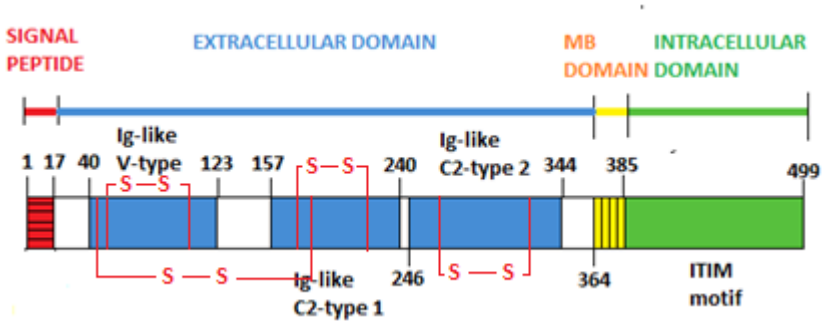


Figure 3.1. Representation of Siglec8 amino acid sequence, highlighting the location of the different domains and disulphide bonds (Foussias, *et al.*, 2000).

Classical Ig-like domains are well conserved, but they can be composed of 7-10 β strands and differ significantly in size due to the high variability in the length of their loops (Smith and Xue, 1997). Thus, Ig-like domains are classified according to their number of β strands: C2-type domains are composed of two sheets with three and four antiparallel β-strands and V-type is formed of two sheets with four and five antiparallel β-strands (Smith and Xue, 1997) (Figure 3.2).

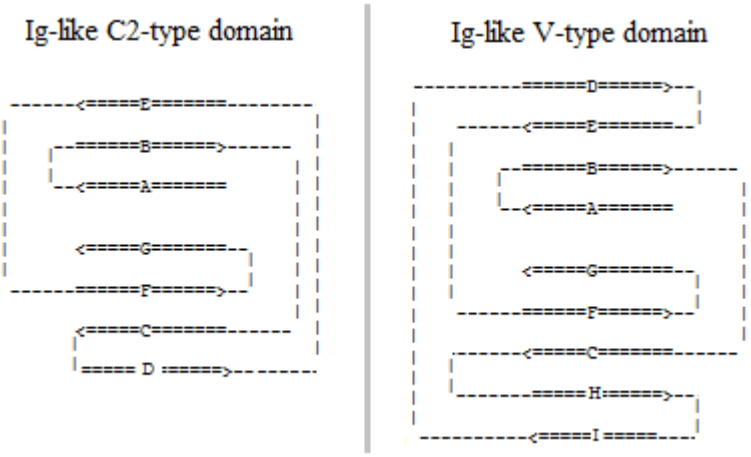


Figure 3.2. Ig-like C2-type and Ig-like V-type domains schematic structure (Smith and Xue, 1997).

Regarding post translational modifications, the Siglec8 receptor is glycosylated at asparagines 172, 249 and 267. Figure 3.3 shows a schematic representation of the primary, secondary and tertiary structure of Siglec8, as well as disulphide bonds and glycosylation positions.

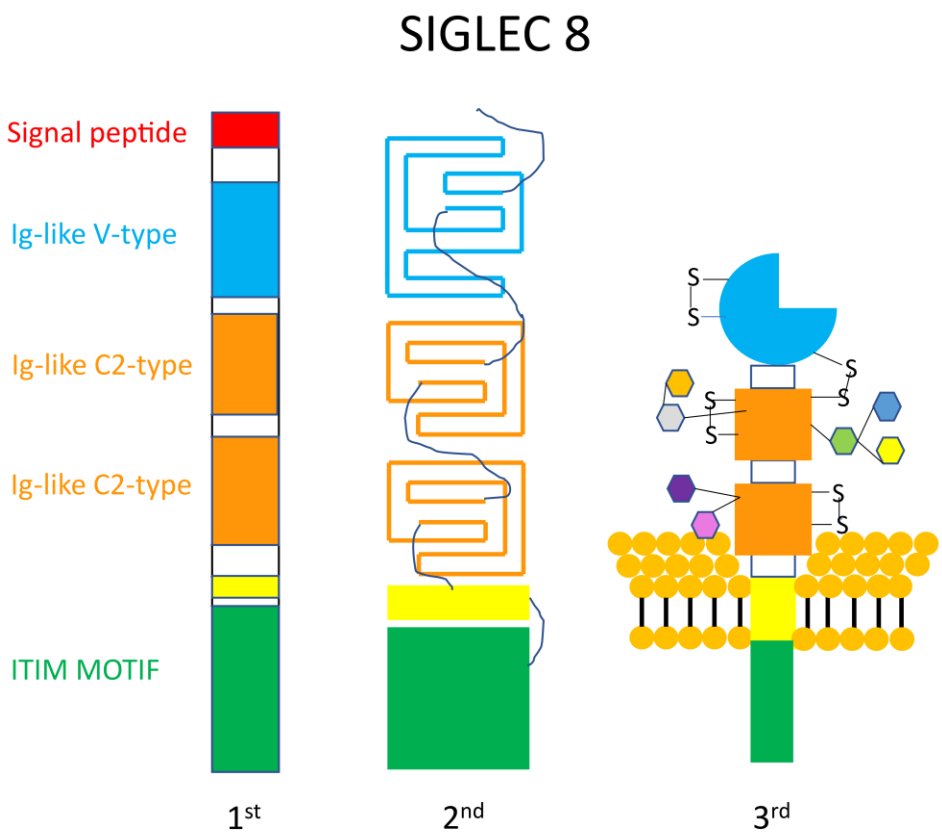


Figure 3.3. Schematic representation of the primary, secondary and tertiary structure of Siglec8 (Foussias, *et al.*, 2000).

3.4 Optimization of the production of soluble Siglec8 Ig-like V-type domain

A codon optimised gene encoding full-length human Siglec8 protein for expression in *E. coli* within the cloning vector pUC57 was purchased, and complementary primers were designed to amplify the fragment corresponding to the Ig-like V-type ligand binding domain by means of PCR. Codon optimization is required to improve the efficiency of the protein translation as not all tRNAs (adaptor molecules that serves as physical link between mRNA and the amino acid sequence) coding for a specific amino acid are expressed at the same level across species. However, some *E. coli* expression strains are available with extra tRNAs that the wild-type strains do not produce abundantly, and in some cases the optimization may not be required.

In this work, three different vectors were used in order to try to improve the solubility and expression levels of the Siglec8 Ig-like V-type protein domain: pET-28a(+), pET-43a(+) and pET151/D-TOPO T (pictures of the vectors sequences in Materials and Methods). For pET-28a(+) and pET-43a(+), primers were designed with

hanging non-complementary 5'-ends which introduced appropriate restriction sites and a stop codon (TAA), or the TEV protease cut site (ENLYFQG) when necessary. Figure 3.4 summarizes the Siglec8 Ig-like V-type domain cloning steps for pET vectors which do not belong to the TOPO cloning technology.

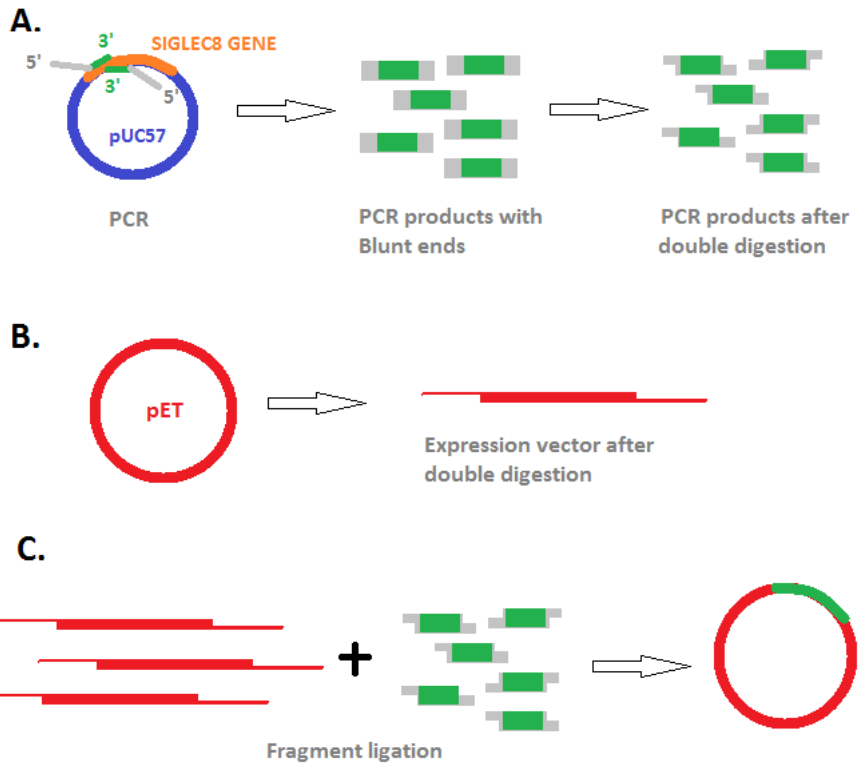


Figure 3.4. Siglec8 Ig-like V-type domain cloning steps. **A.** Amplification of Siglec8 Ig-like V-type domains by PCR followed by double digestion with restriction enzymes. **B.** Double digestion of the expression vector pET with the same enzymes. **C.** Ligation reaction of the pET expression vector and the PCR products coding for the Ig-like V-type Siglec8 domain.

As mentioned above, different constructs of the Siglec8 Ig-like V-type domain were cloned in three different vectors in order to try to optimise expression levels and/or solubility. Initially, the amino acid sequence encoding only the Siglec8 Ig-like V-type domain from Gly40 to Phe123 was amplified and cloned into pET-28(a+), but no overexpression was achieved. A literature search of all members of the human Siglec family whose Ig-like V-type domains have been overexpressed in bacteria (Siglecs5, 7, 8 and 9) revealed that all constructs encompassed an equivalent region of the receptors and included some extra inter-domain amino acids (Barb, *et al.*, 2013; Propster, *et al.*, 2015). A further BLAST sequence alignment of Siglecs5, 7, 8 and 9 revealed high amino acid sequence conservation (Figure 3.5).

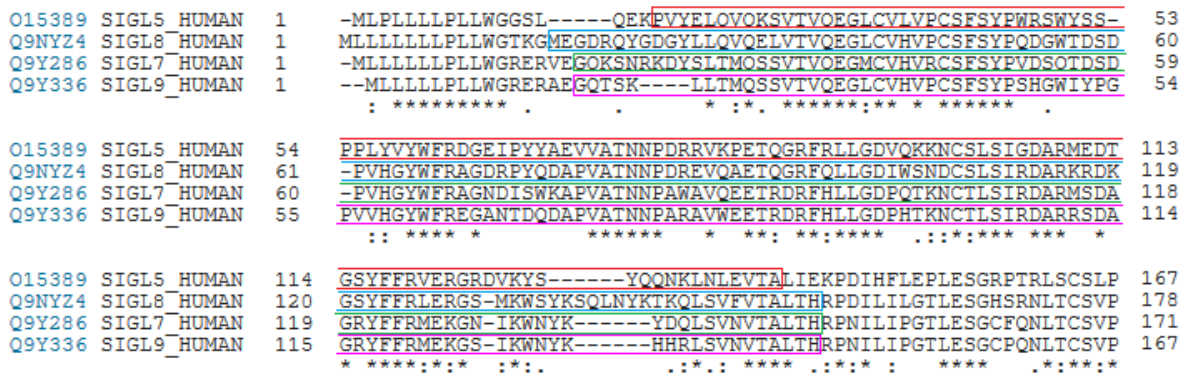


Figure 3.5. Siglecs5, 7, 8 and 9 BLAST alignments of the previously cloned and expressed domains highlighted in different colours. The BLAST was performed with UniProt (Barb, et al., 2013; Propster, et al., 2015).

Therefore, the region encoding Siglec8 Ig-like V-type domain residues from Met17 to His155, the Siglec8 region that was successfully expressed by Propster and co-workers (Propster, *et al.*, 2015), was amplified and cloned into pET-43(a+) and pET-151 TOPO. Different combinations of *E. coli* bacterial strains, culture media, IPTG concentrations, additives, temperatures and growth times were tested to try to optimise expression levels and/or solubility of the different constructs.

An overview of all trials and optimization steps, in chronological order, is summarized below in Table 3.1, and discussed in more detail in the following sections.

Table 3.1. Different steps followed to clone and express Siglec8 Ig-like V-type domain chronologically ordered.

Siglec8 fragment	Vector	Experiment	Outcome
Gly40-Phe123	pET-28(a+)	Cloning	Domain was cloned.
Gly40-Phe123	pET-28(a+)	Expression	Domain did not express.
Met17-His155	pET-43(a+)	Cloning	Domain was cloned.
Met17-His155	pET-43(a+)	Expression and solubility	Domain was expressed at low levels. It was soluble but formed oligomers.
Met17-His155	pET-43(a+)	Mutagenesis	Unpaired Cys42 was substituted by Ser42 to avoid oligomerization.
Met17-His155 (Cys42>Ser42)	pET-43(a+)	Expression and solubility	Domain was expressed at low levels. It became insoluble.
Met17-His155 (Cys42>Ser42)	pET-43(a+)	Solubility optimization	Domain remained been insoluble.
Met17-His155 (Cys42>Ser42)	pET151/D-TOPO T	Cloning	Domain was cloned.
Met17-His155 (Cys42>Ser42)	pET151/D-TOPO T	Expression and solubility	Domain was expressed at high levels with some extra amino acids at the end (a Cys included). It was still insoluble.
Met17-His155 (Cys42>Ser42)	pET151/D-TOPO T	Mutagenesis	Stop codon (TAA) was introduced after His155 to remove extra amino acids.
Met17-His155 (Cys42>Ser42)	pET151/D-TOPO T	Expression and solubility	Domain was expressed at high levels. It was still insoluble.

Siglec 8 Gly40-Phe123 cloning into pET-28(a+)

Figure 3.6 shows the agarose gel (1% w/v) for the amplified DNA fragment encoding residues Gly40 to Phe123. Lane 1 contains the standard DNA ladder, lanes 2 and 3 contain the 279bps blunt ended DNA fragment, and lanes 4 and 5 contain the same fragment after *Bam* HI and *Nde* I double digestion (265 bps), before and after purification ready for cloning.

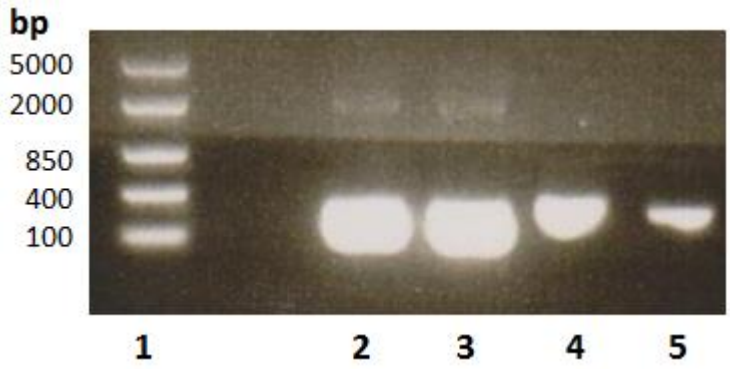


Figure 3.6. Amplified 279 bps blunt ended DNA fragment encoding Gly40-Phe123 (lanes 2 and 3) and same fragment of 265 bps after double digestion with *Nde* I and *Bam* HI (lanes 4 and 5). The first lane corresponds to the ladder Middle Range (Life Technologies) with 5000-2000-850-400-100 bps long fragments from top to bottom.

Figure 3.7 displays the 1% (w/v) agarose gel of the pET-28(a+) vector (6,829 bps long) after single and double digestion. Lane 3 corresponds to the 6,829bps linearized vector after single digestion with *Bam* HI. Lane 2 shows the two fragments of 1,500 and 5,329 bps with cohesive ends after *Nde* I and *Bam* HI double digestion. The longer fragment (5,329 bps) with all sequences required for expression was subsequently purified before insertion of the Siglec8 domain.

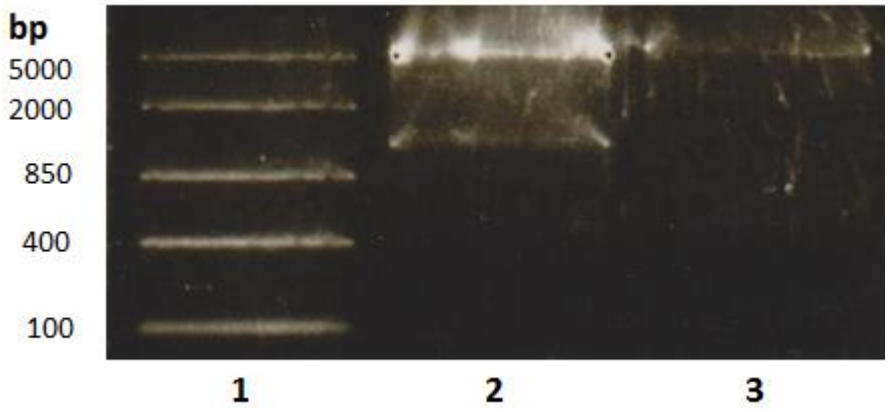


Figure 3.7. 1% (w/v) agarose gel demonstrating single (*Bam* HI, lane 3) and double digestion (*Bam* HI, *Nde* I, lane 2) of pET-28(a+). The first lane corresponds to the ladder Middle Range (Life Technologies) with 5000-2000-850-400-100 bps long fragments from top to bottom.

The gene encoding the Siglec8 fragment Gly40 – Phe123 (279 bps) was ligated into pET-28(a+) and resulting colonies were screened for the presence of the inserted gene by means of colony PCR. Figure 3.8 shows the agarose gel containing the different PCR reactions after ligation into pET-28(a+). Lanes 2, 4, 6, 7 and 14 confirmed the insertion of the 279bps domain into the plasmid pET-28(a+). Those plasmids which gave a positive result in the colony PCR were sequenced to check that frame, sequence and genetic construction were all correct. All PCR positive plasmids contained the correct genetic construction. An example of the sequencing data is shown in Figure 3.9. The fragment contained an N-terminal 6 x His-tag followed by a thrombin protease site. The starting methionine was introduced by the *Nde* I restriction site and the stop codon was introduced with the PCR primer immediately following the last amino acid Phe123.

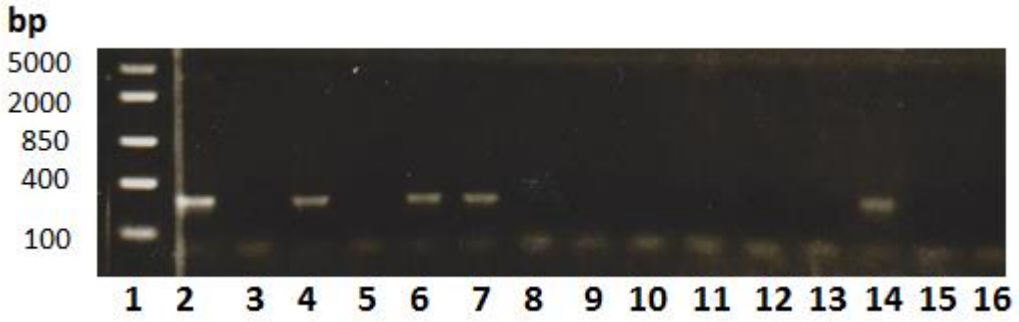


Figure 3.8. PCR colonies. The first lane corresponds to the ladder Middle Range (Life Technologies) with 5000-2000-850-400-100 bps long fragments from top to bottom. The rest are PCRs carried out using plasmid preps which came from different bacterial colonies transformed with ligation reaction of pET-28(a+) and the fragment Gly40-Phe123. Colonies from lane 2, 4, 6 and 14 were positive for the 279 bps insert.

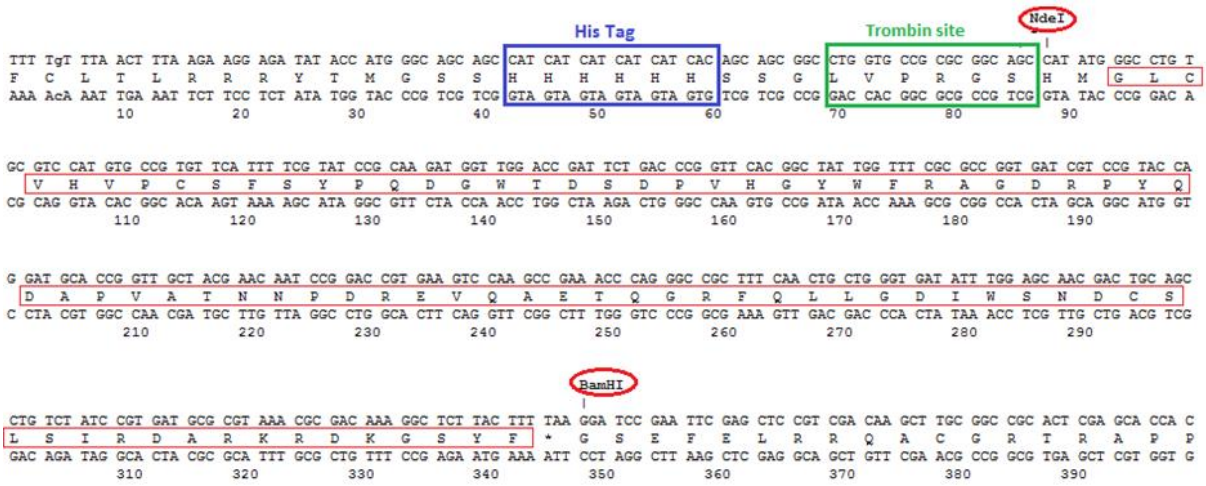


Figure 3.9. Sequencing result for one of the positive PCR colonies of Siglec8 Gly40-Phe123 cloned into pET-28(a+). The N-terminal 6xHis-Tag and the thrombin site as well as the *Nde* I and *Bam* HI cloning sites are highlighted.

Expression of Siglec8 Gly40 - Phe123 cloned into pET-28(a+)

Figure 3.10 shows the SDS-PAGE analyses of *E. coli* BL21 PlySs cell cultures transformed with the Siglec8 Gly40-Phe123 Ig-like V-type domain in the vector pET-28(a+). Four different IPTG concentrations and growth temperatures were tested: 37°C and 0.1 mM IPTG, 37°C and 0.5 mM IPTG, 15°C and 0.1 mM IPTG; and 15°C and 0.5 mM IPTG; samples were collected immediately prior to induction and 2, 4, 5, 22 and 24 hours after induction. No expression of the 11.86 kDa Ig-like V-type domain was observed at any of those conditions.

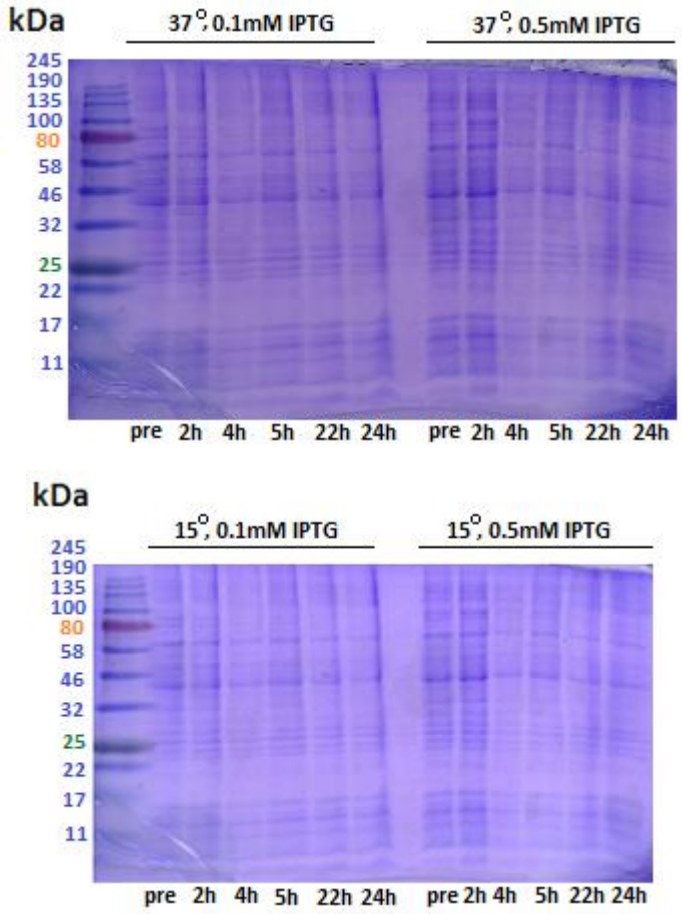


Figure 3.10. SDS-PAGE gels belonging to the BL21 PlySs cultures grown in LB and transformed with pET-28a (+) and the insert Siglec8 Gly40-Phe123. From left to right cultures were performed at 37°C induced with 0.1mM IPTG, 37°C induced with 0.5mM IPTG, 15°C induced with 0.1mM IPTG and 15°C induced with 0.5mM IPTG. For each condition samples were taken out before and 2, 4, 5, 22 and 24 hours after induction. All samples contained 100mM DTT.

At this stage, the expression and purification of the Siglec 8 Ig-like V-type domain was published (Propster, *et al.*, 2015). Based on that paper, together with a literature search and a sequence BLAST of all Siglec Ig-like V-type domains cloned and expressed in bacteria until that time (5, 7, 8 and 9), I decided to clone the same fragment and use the same vector as the authors used. An alignment of the Siglec5, 7,

8 and 9 Ig-like V-type domains sequences is shown in Figure 3.9 and revealed high amino acid sequence conservation and that all authors produced equivalent regions of the domain (Barb, *et al.*, 2013; Propster, *et al.*, 2015)Therefore, this new study (Propster, *et al.*, 2015) was taken as reference, and the amino acid sequence encoding the Siglec8 Ig-like V-type domain from Met17 to His155 was amplified and cloned into pET-43(a+).

Siglec8 Met17-His155 cloning into pET-43(a+)

Figure 3.11 shows the agarose gel (1% w/v) for the amplified, blunt ended 468bps DNA fragment which encodes for the Siglec8 Ig-like V-type domain from Met17 to His155 (lane 2). The second band in lane 2 below 100 bps corresponds to the primers used for PCR amplification. The same fragment of 454 bps with cohesive ends after *Nde* I and *Xho* I double digestion is shown in lane 3 and after purification ready for cloning in lane 4.

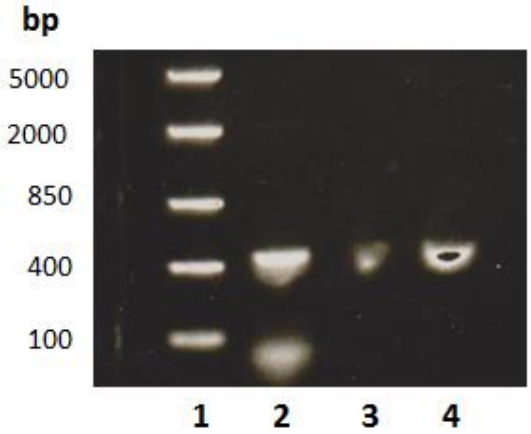


Figure 3.11. Amplification of DNA fragment encoding Met17-His155, and subsequent double digestion. Lane 1: Middle Range DNA ladder (Life Technologies) whose fragments of known sizes are 5000-2000-850-400-100 base pairs. Lane 2: PCR amplified blunt ended product, with a theoretical size of 468 bps. Lanes 3 and 4: product after double digestion with *Nde* I and *Xho* I restriction enzymes (theoretical size of 454 bps).

The 7,275bps pET-43(a+) vector singly digested with *Nde* I (Figure 3.12, lane 1) and doubly digested with *Nde* I and *Xho* I (lanes 2 and 3) giving rise to two fragments of 1,787 and 5,488 bps and cohesive ends. The longer fragment was subsequently purified ready for ligation of the gene encoding the Siglec8 domain.

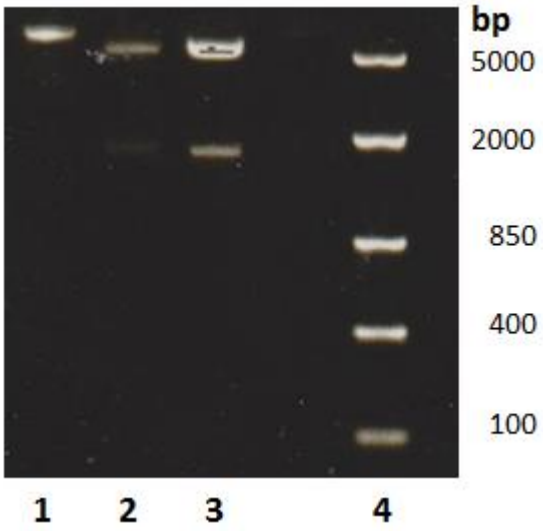


Figure 3.12. 1% agarose gel demonstrating pET-43(a+) single *Nde* I digestion of 7,275 bps long (lane 1) and the two fragments of 1,787 and 5,488 bps long after double digestion with *Nde* I and *Xho* I (Lanes 2 and 3).

The gene encoding the Siglec8 fragment Met17-His155 (454 bps) was ligated into pET-43(a+) and resulting colonies were screened for the presence of the inserted gene by means of colony PCR. Figure 3.13 shows the agarose gel containing the different PCR reactions after ligation into pET-43(a+). Lanes 8, 11 and 13 confirmed the insertion of the 454bps fragment into the plasmid pET-43(a+). Those plasmids which gave a positive result in the colony PCR were sequenced to check that frame, sequence and genetic construction were all correct. All positive plasmids contained the correct genetic construction. An example of the sequencing data is shown in Figure 3.14. The fragment contained a C-terminal 6 x His-tag preceded by a TEV protease site.

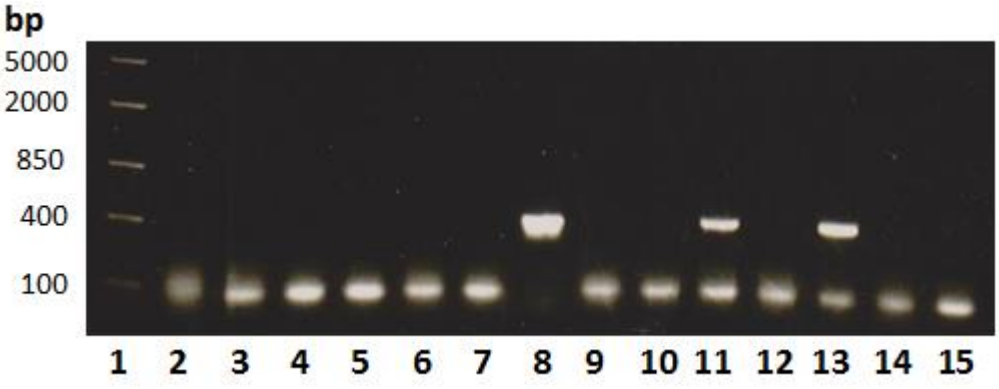


Figure 3.13. PCR colonies. From left to right the first lane corresponds to the Middle Range ladder and the rest are PCRs carried out using plasmid preps which came from different bacterial colonies transformed with ligation reaction of pET-43 a(+) and the fragment Met17-His155. Colonies from lanes 8, 11 and 13 showed the presence of the 444bps fragment.

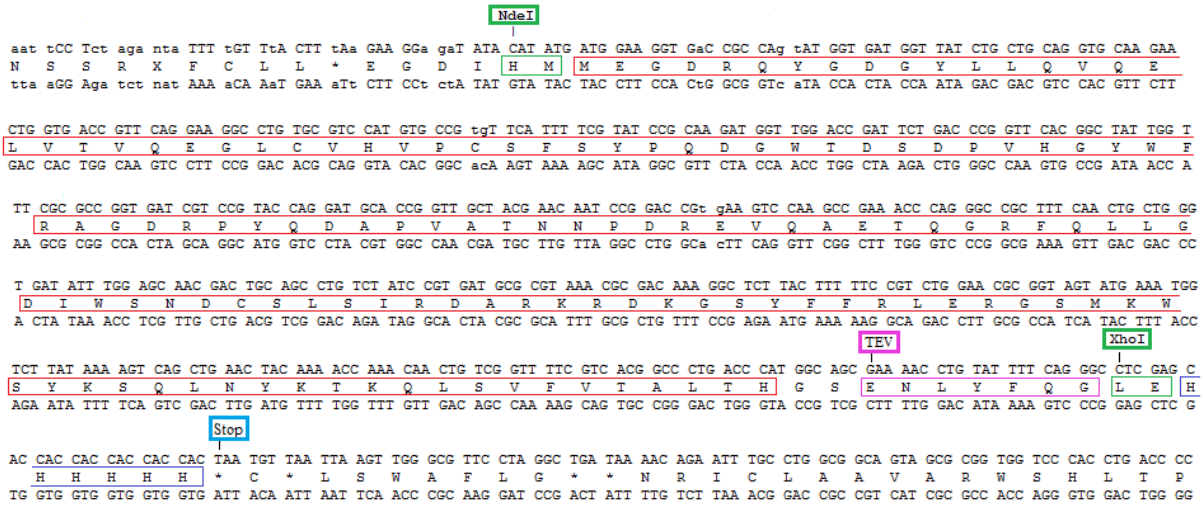


Figure 3.14. Sequencing results pET-43(a+) with Met17-His155 Siglec8 insert and TEV site between *Nde* I and *Xho* I.

Expression and solubility of Siglec8 Met17-His155 cloned in pET-43(a+)

Figures 3.15, 3.16, 3.17 and 3.18 show the SDS-PAGE and Western Blot analyses of *E. coli* BL21 PlySs cell cultures transformed with the Siglec8 Met17-His155 Ig-like V-type domain in the vector pET-43(a+). Western Blot was carried out in two steps using a mouse anti-C-terminal 6x-HisTag primary monoclonal antibody and a goat anti-mouse IgG2c (immunoglobulin G2c) secondary antibody with the alkaline phosphatase (AP) enzyme pre-adsorbed. The presence of antibody attached to the protein was revealed using an AP chromogenic substrate.

Four different IPTG concentrations and growth temperatures were tested: 37°C and 0.1 mM IPTG, 37°C and 0.5 mM IPTG, 15°C and 0.1 mM IPTG; and 15°C and 0.5 mM IPTG; and samples were removed before, 2, 4 and 6 hours after induction when cultures were performed at 37°C, and before, 2, 4, 7, 22, 30 and 34 hours after induction when cultures were performed at 15°C.

At 37°C and 2 hours post-induction, a band migrating at an approximate molecular weight of 72 kDa was observed in both SDS-PAGE and Western Blot (Figures 3.15 and 3.16). However, this size was much larger than the expected 18.24 kDa molecular weight, suggesting strong aggregation into tetramers or even larger aggregates, despite of the addition of 100 mM DTT to the SDS-PAGE sample. At 15°C and 2 hours post-induction with 0.5 mM IPTG, a faint band migrating between the 17 and 20 kDa markers was observed (Figure 3.17). This is in good agreement

with the expected molecular weight of 18.24 kDa for the Ig-like V-type domain, and the band became more intense at longer expression times but was still very faint after 34 hours of expression, and the band was not observed in the Western Blot (Figure 3.18).

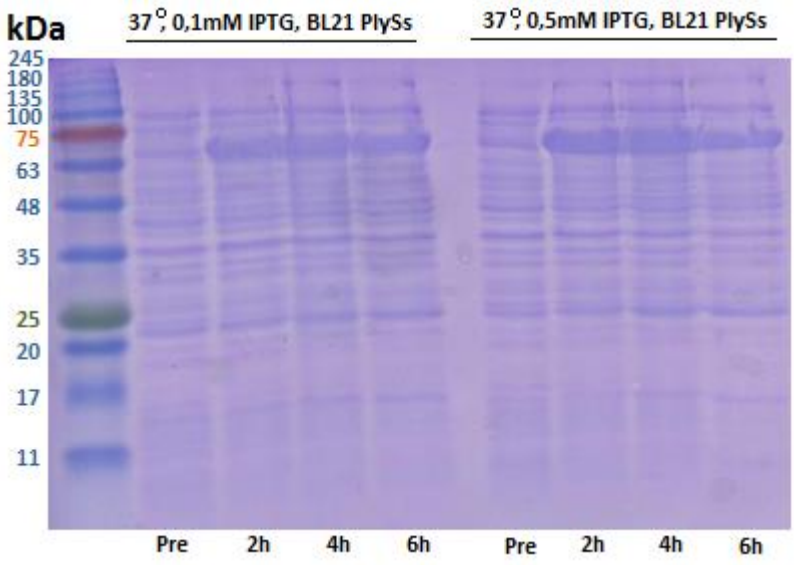


Figure 3.15. SDS-PAGE gels belonging to the BL21 PlySs cultures grown in LB and transformed with pET-43a (+) and the insert Siglec8 Met17-His155. From left to right cultures were performed at 37°C induced with 0.1mM IPTG and 37°C induced with 0.5mM IPTG. For each condition samples were taken out before and 2, 4 and 6 hours after induction. All samples contained 100mM DTT.

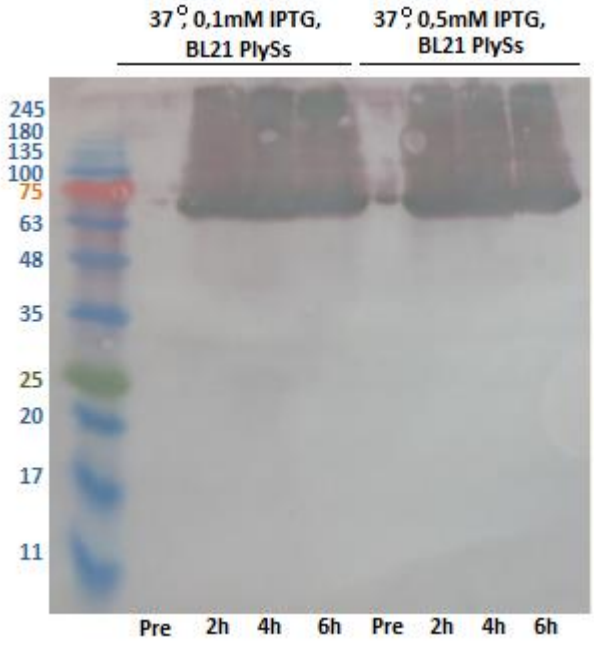


Figure 3.16. Western Blot belonging to the BL21 PlySs cultures grown in LB and transformed with pET-43a (+) and the insert Siglec8 Met17-His155. From left to right cultures were performed at 37°C induced with 0.1mM IPTG and 37°C induced with 0.5mM IPTG. For each condition samples were taken out before and 2, 4 and 6 hours after induction. All samples contained 100mM DTT.

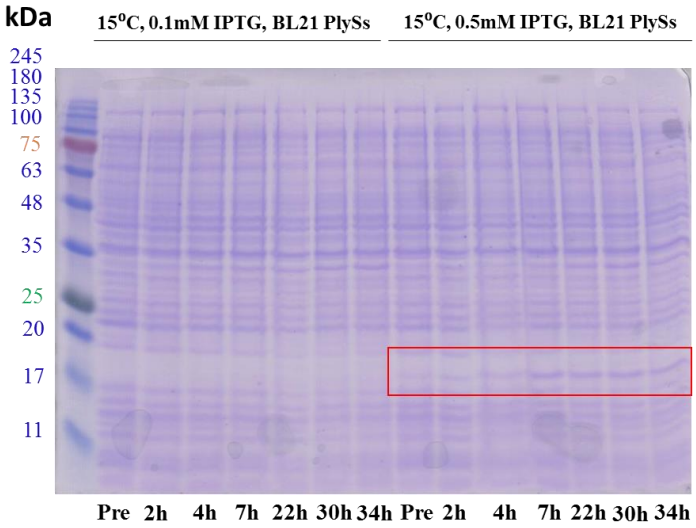


Figure 3.17. SDS-PAGE gels belonging to the BL21 PlySs cultures grown in LB and transformed with pET-43a (+) and the insert Siglec8 Met17-His155. From left to right cultures were performed at 15°C induced with 0.1mM IPTG and 15°C induced with 0.5mM IPTG. For each condition samples were taken out before and 2, 4, 7, 22, 30 and 34 hours after induction. All samples contained 100mM DTT.

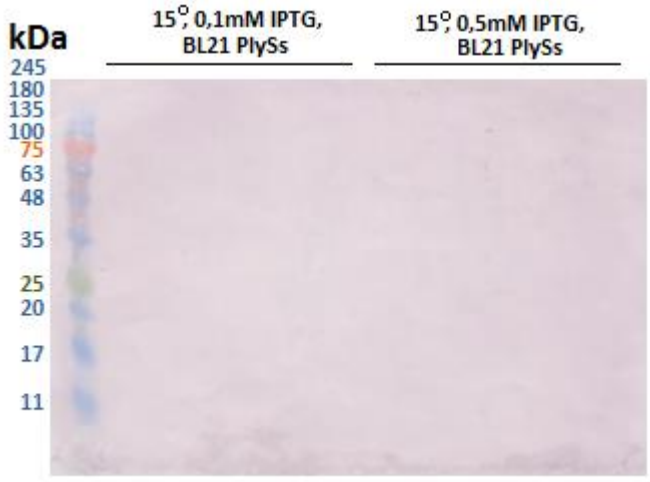


Figure 3.18. Western Blot belonging to the BL21 PlySs cultures grown in LB and transformed with pET-43a (+) and the insert Siglec8 Met17-His155. From left to right cultures were performed at 15°C induced with 0.1mM IPTG and 15°C induced with 0.5mM IPTG. For each condition samples were taken out before and 2, 4, 7, 22, 30 and 34 hours after induction. All samples contained 100mM DTT.

At this stage, the strain *E. coli* Rosetta gami B (DE3) was purchased taking reference from previous published work (Propster, *et al.*, 2015). The strain's genotype is described in detail in the Bacterial strains section (Materials and Methods). In brief, mutations in enzymes Glutathione and Thioredoxin reductases facilitate the formation of disulphide bonds between cysteines in the bacterial cytoplasm and thus, protein folding. Furthermore, the strain is provided with those genes encoding *E. coli* rare codon tRNAs, which enhances translation of mRNA transcribed from genes belonging to other species.

Figures 3.19 and 3.20 show the SDS-PAGE gels analyses performed using *E. coli* Rosetta gami B (DE3) cell cultures transformed with the Siglec8 Met17-His155 Ig-like V-type domain cloned in the vector pET-43(a+). Four different IPTG concentrations and growth temperatures were tested in LB media: 37°C and 0.1 mM IPTG, 37°C and 0.5 mM IPTG, 15°C and 0.1 mM IPTG; and 15°C and 0.5 mM IPTG. Samples were collected before and 2, 4, 24, 30 and 38 hours after induction. In all cases, a band migrating between the 16.6 and 25 kDa markers was observed, in good agreement with the expected mass of 18.24 kDa for the Siglec8 Ig-like V-type domain. However, the band was still very faint, suggesting low expression levels.

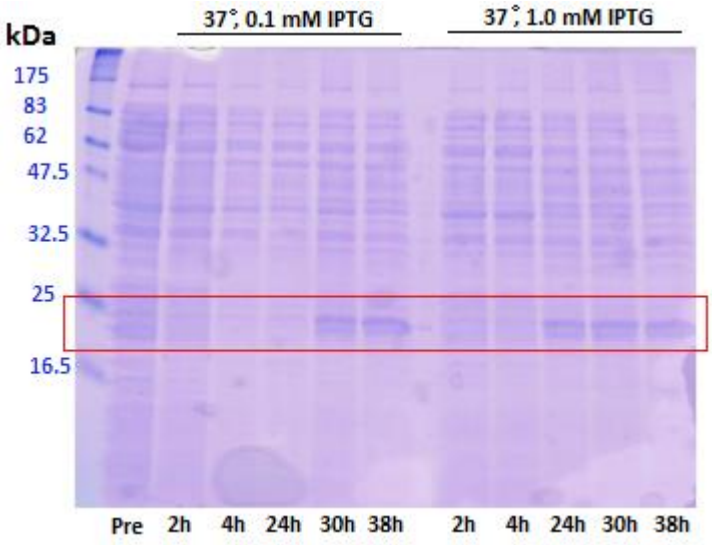


Figure 3.19. SDS-PAGE gels belonging to the *E. coli* Rosetta gami B (DE3) cultures grown in LB and transformed with pET-43a (+) and the insert Siglec8 Met17-His155. From left to right cultures were performed at 37°C induced with 0.1mM IPTG and 37°C induced with 1.0mM IPTG. For each condition samples were taken out before and 2, 4, 24, 30 and 38 hours after induction. All samples contained 100mM DTT.

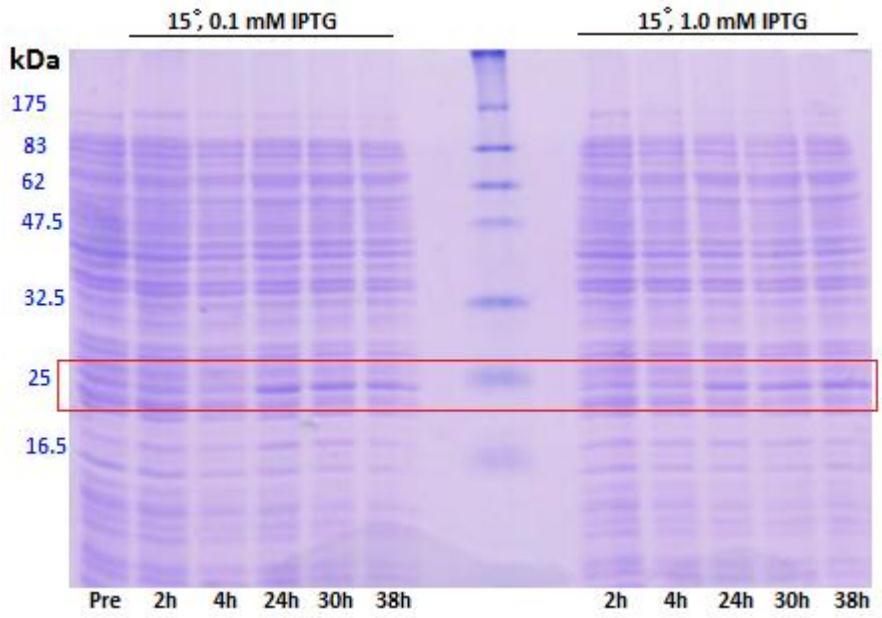


Figure 3.20. SDS-PAGE gels belonging to the *E. coli* Rosetta gami B (DE3) cultures grown in LB and transformed with pET-43a(+) and the insert Siglec8 Met17-His155. From left to right cultures were performed at 15°C induced with 0.1mM IPTG and 15°C induced with 1.0mM IPTG. For each condition samples were taken out before and 2, 4, 24, 30 and 38 hours after induction. All samples contained 100mM DTT.

Since NMR studies require large quantities of stable ^{15}N, ^{13}C and/or ^{2}H isotope labelled proteins, expression in minimal media (M9) was also checked and four different 1 L cultures of *E. coli* Rosetta gami B (DE3) transformed with the construct were grown under a variety of conditions: 37°C and 0.1 mM IPTG; 37°C and 1.0 mM IPTG; 15°C and 0.1 mM IPTG; and 15°C and 1.0 mM IPTG. Both SDS-PAGE and Western Blot analyses were carried out with samples removed before and 2, 4 and 24 hours after induction, and these results are shown in Figures 3.21 and 3.22. In all cases apart from the culture grown at 15°C and induced with 0.1 mM IPTG, a band migrating between the 17 and 25 kDa standards was observed in both SDS-PAGE and in the Western blot strongly suggesting this band belonged to the 18.24 kDa Siglec8 Ig-like V-type domain. Low expression levels were still observed and indicated that the yield is similar in both LB and M9 media.

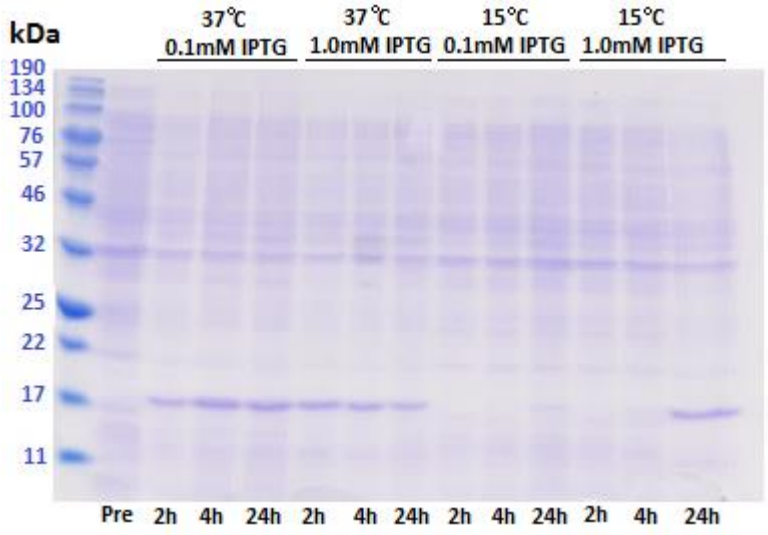


Figure 3.21. SDS-PAGE gels belonging to the *E. coli* Rosetta gami B (DE3) cultures grown in M9 media and transformed with pET-43a(+) and the insert Siglec8 Met17-His155. From left to right cultures were performed at 37°C induced with 0.1mM IPTG, 37°C induced with 1.0mM IPTG, 15°C induced with 0.1mM IPTG and 15°C induced with 1.0mM IPTG. For each condition samples were taken out before and 2, 4, and 24, 30 hours after induction. All samples contained 100mM DTT.

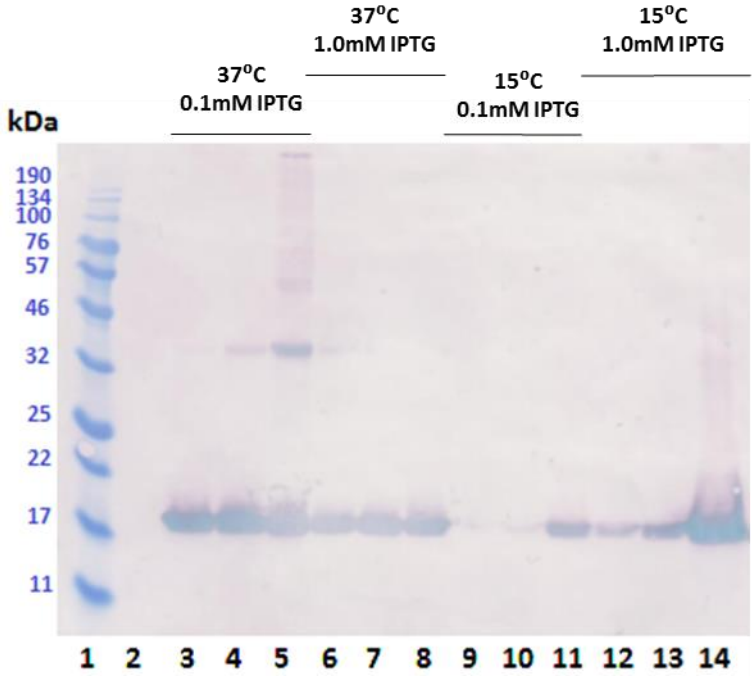


Figure 3.22. Western Blot belonging to the E. coli Rosetta gami B (DE3) cultures grown in M9 media and transformed with pET-43a(+) and the insert Siglec8 Met17-His155. From left to right cultures were performed at 37°C induced with 0.1mM IPTG, 37°C induced with 1.0mM IPTG, 15°C induced with 0.1mM IPTG and 15°C induced with 1.0mM IPTG. For each condition samples were taken out before and 2, 4, and 24, 30 hours after induction. All samples contained 100mM DTT.

After 30 hours, the culture grown at 15°C and induced with 1.0 mM IPTG was spun down and cells were lysed using a cell disruptor. Soluble and insoluble materials

were separated by centrifugation, and the soluble fraction was passed through a Nickel NTA column. Figure 3.23 shows the SDS-PAGE and the Western Blot analyses of the different elution fractions collected from the Nickel column (flow through, washes and elution steps with increasing concentration of imidazole). A very faint band above 46 kDa was observed in the elution steps containing 332-500 mM imidazole (lanes 9 to 14), indicating a low concentration of soluble protein and oligomerization of the domains, probably into trimers (54.72 kDa).

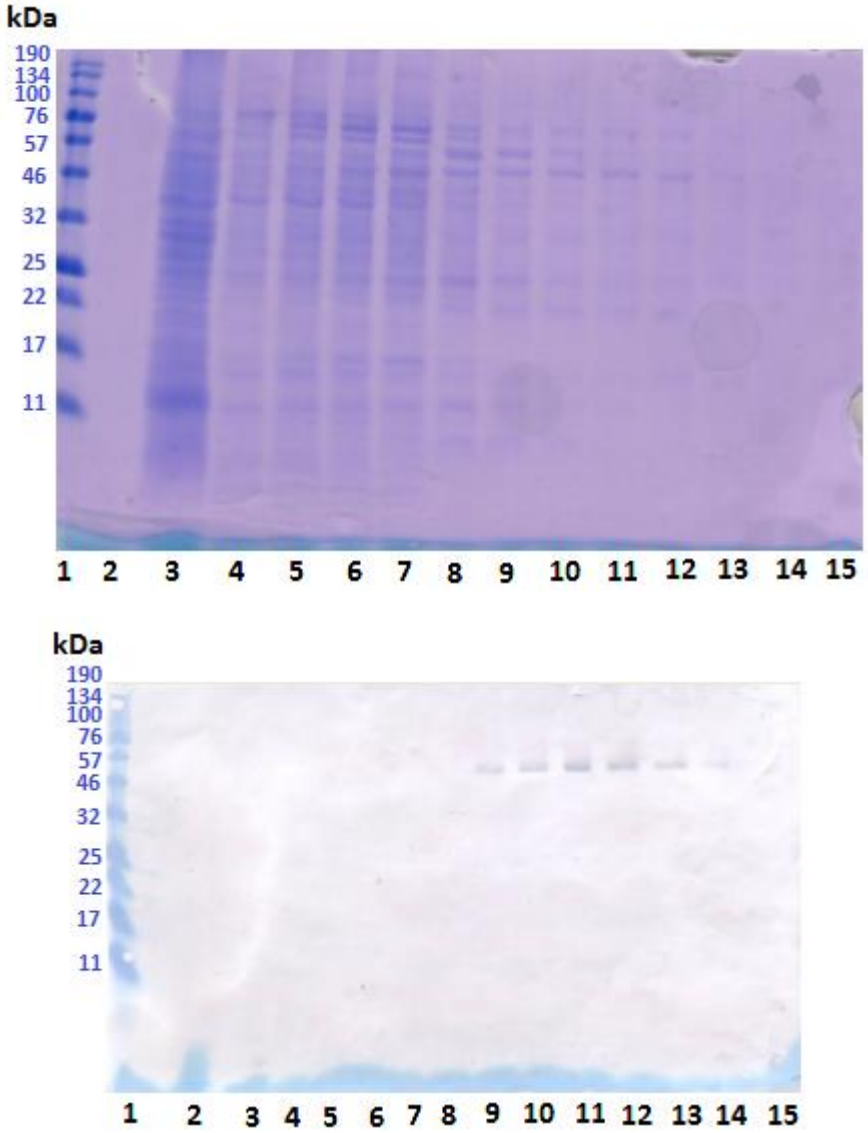


Figure 3.23. Lane 1 belongs to the protein ladder. Lane 2 consists in just binding buffer. Lanes 3 and 4 belong to washes with binding buffer. Lanes from 5 to 25 belong to the flow through with a linear increasing imidazole concentration: 20, 44, 68, 92, 116, 140, 164, 188, 212, 236, 260, 284, 308, 332, 356, 380, 404, 428, 452, 476 and 500 mM respectively.

To eliminate oligomer formation, the Siglec8 Cysteine residue located in position 42 (Cys42) which forms an inter-domain disulphide bond with Cys181 (not

included in the cloned construct), was substituted by a Serine (Ser42) in order to avoid the formation of incorrect disulphide bonds and/or oligomers via this unpaired Cys (Propster, *et al.*, 2015).

Siglec8 Met17-His155 (Cys42 > Ser42) mutagenesis

Primers were designed to mutate Cys42 to a serine within the Siglec8 Met17-His155 Ig-like V-type domain using a Site-Directed Mutagenesis Kit. The pET-43 (a+) plasmid containing the wild-type domain sequence was used as a template. Figure 3.24 shows the sequencing results for the Siglec8 Met17-His155 domain mutated at position 42 with the substitution Cys42>Ser42. The fragment still contained the same features: A C-terminal 6 x His-tag coming after a TEV protease cutting site.

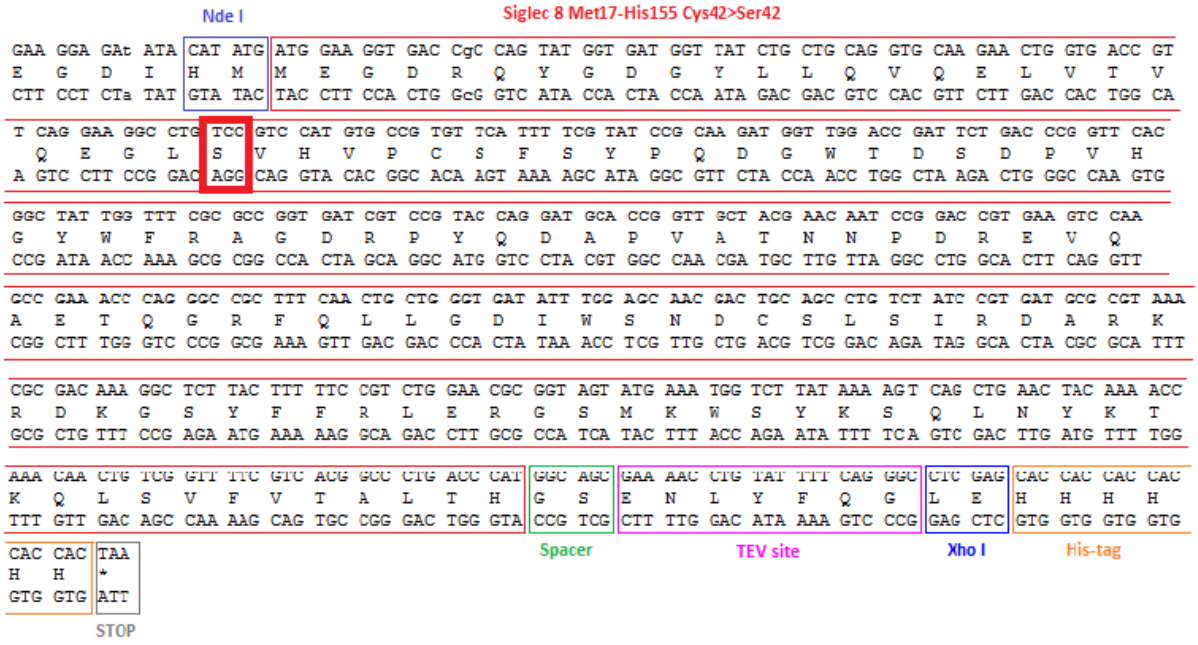


Figure 3.24. Sequencing results for the Siglec8 Met17-His155 mutated domain with a Cys42>Ser42 substitution cloned into pET-43(a+).

Expression and solubility of Siglec8 Met17-His155 (Cys42>Ser42) mutated domain cloned into pET-43(a+)

Figures 3.25 and 3.26 show the SDS-PAGE and Western Blot performed for four different M9 cultures of *E. coli* Rosetta gami B (DE3) transformed with the construct pET-43(a+), Siglec8 Met17-His155 (Cys42>Ser42). Samples were removed before and 2 and 7 hours after induction in case of both cultures grown at 37ºC and induced with 0.1 and 1.0 mM IPTG; samples were removed before, 2 and 24 hours after induction for cultures grown at 15 ºC and induced with 0.1 and 1.0 mM IPTG. A

band migrating just above the 17 kDa marker, belonging to the 18.19 kDa mutated protein domain, was observed in all cases, even in the pre-induction samples. However, the expression levels still were not high.

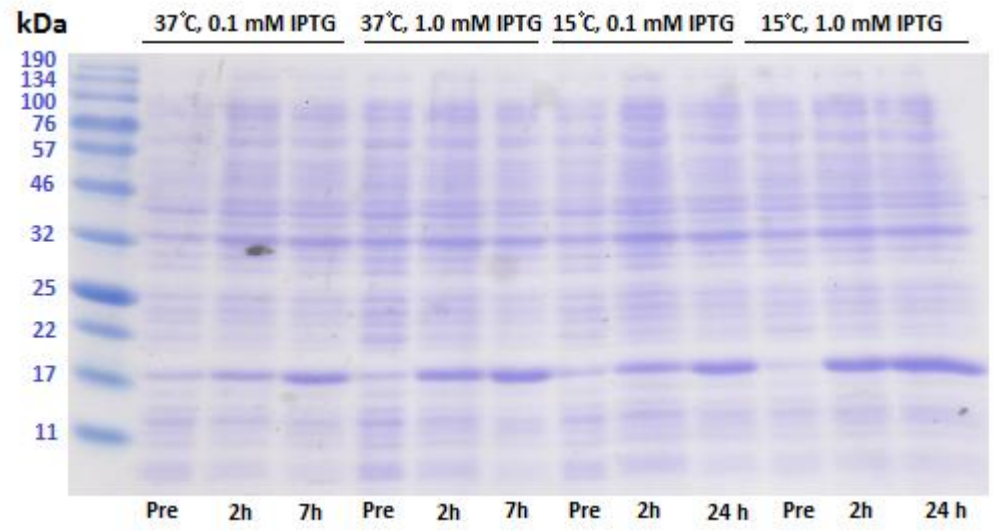


Figure 3.25. SDS-PAGE gel of samples from four *E. coli* Rosetta gami B (DE3) cultures performed in M9 media at 37ºC and induced with 0.1 mM IPTG, 37ºC and 1.0 mM IPTG, 15ºC and 0.1 mM IPTG; and 15ºC and 1.0 mM IPTG. Samples were taken out before and at different times after induction. All of them contained 100 mM DTT.

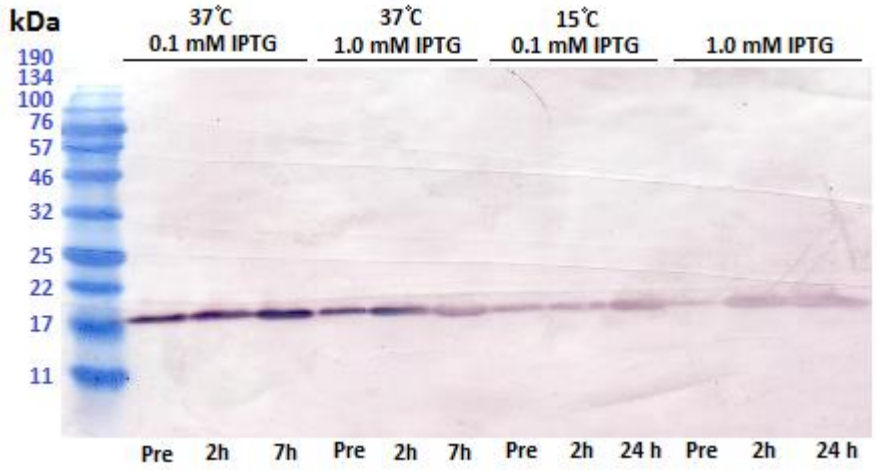


Figure 3.26. Western Blot of samples from four *E. coli* Rosetta gami B (DE3) cultures performed in M9 media at 37ºC and induced with 0.1 mM IPTG, 37ºC and 1.0 mM IPTG, 15ºC and 0.1 mM IPTG; and 15ºC and 1.0 mM IPTG. Samples were taken out before and at different times after induction. All of them contained 100 mM DTT.

After expression, cell cultures were spun down, re-suspended in lysis buffer and lysed using a cell disruptor to break the cells and release all their contents. Protease cocktail inhibitor was added to minimise protein degradation. Subsequently, the suspension was spun down to separate soluble and insoluble material. The lanes

labelled as “Sup” and “Pellet”, correspond to the soluble and insoluble material respectively, obtained after cell disruption and centrifugation. Soluble proteins stayed solubilised in the supernatant and the insoluble proteins formed part of the pellet. Figure 3.27 shows the SDS-PAGE and Western Blot analyses for *E. coli* Rosetta gami B (DE3) cells grown in LB at 15°C and induced with 1.0 mM IPTG. Along with the pre-induction and post-induction samples, supernatant and pellet samples were included and showed that most of the protein was in the pellet and insoluble, likely forming inclusion bodies. A very small amount of protein was soluble but could not be seen in the SDS-PAGE gel. The Western Blot also revealed some basal expression before induction, which could be interfering with the production of soluble post-induction protein.

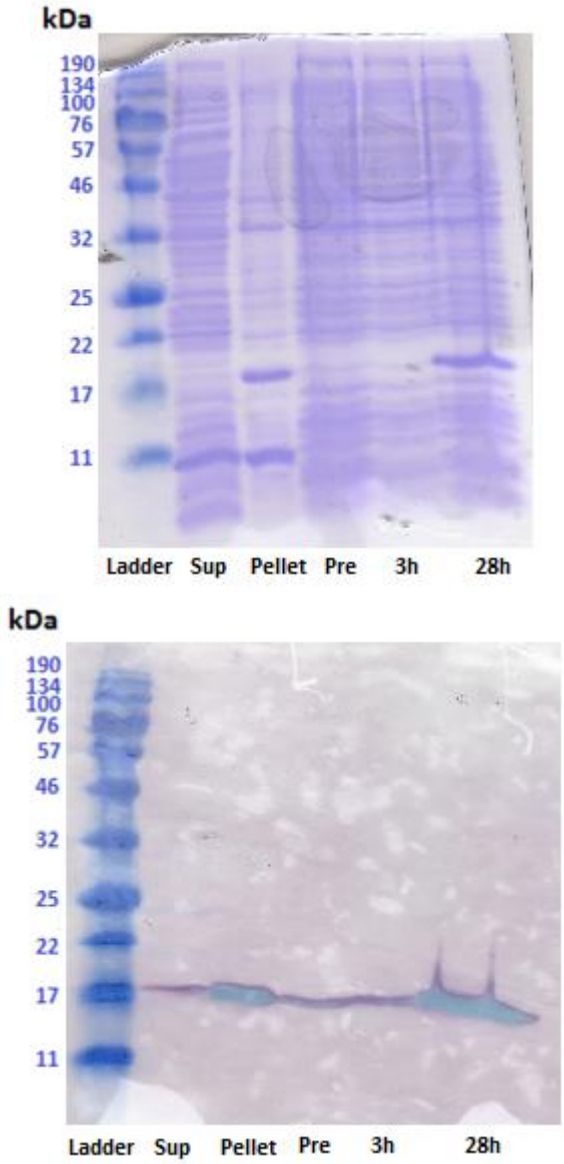


Figure 3.27. SDS-PAGE gel and Western Blot of samples from an *E. coli* Rosetta gami B (DE3) LB culture overexpressing Siglec8 Met17-His155 (Cys42 > Ser42) in pET-43(a+) performed at 15ºC and induced with 1.0 mM IPTG. The first three lanes correspond to the Protein Ladder, supernatant and pellet respectively, after expression and disruption. The last three samples were taken out before and at different times after induction. All of them contained 100 mM DTT.

Optimization of the solubility of Siglec8 Met17-His155 (Cys42>Ser42) mutated domain cloned into pET-43(a+)

The addition of 1% (w/v) of glucose sometimes can eliminate the basal expression before induction, by lowering the cellular cAMP levels, which stimulates the lacUV5 promoter in DE3 strains. Thus, cellular collapse due to accumulation of protein in the cytoplasm is avoided. To see if this had an impact on expression levels, three cultures were grown at 15, 25 and 37°C with the addition of 1% (w/v) glucose to the LB before transferring the cells to M9 media and inducing expression. Figure

3.28 displays the SDS-PAGE gel and Western Blot analyses of the soluble and insoluble material from those cultures. Although the basal expression in the pre-induction sample was eliminated, the Siglec8 Ig-like V-type domain was still in the pellet after expression and cell disruption. In the Western Blot, only at 15°C, a very faded band could be observed in the lane belonging to the supernatant, suggesting a very small improvement in the solubility. This improvement may be due to the fact that, at low temperatures, the speed of cell metabolism is slower and protein aggregation is less likely. Although the expression levels improved, the protein domain solubility remained very low.

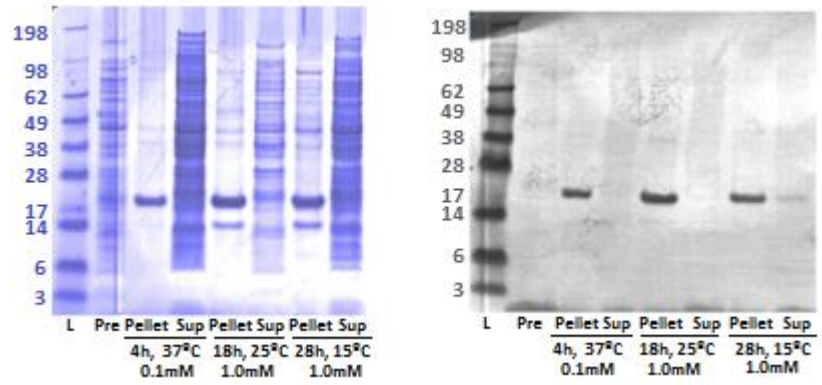


Figure 3.28. SDS-PAGE gel and Western Blot of pellet and supernatant samples from three *E. coli* Rosetta gami B (DE3) M9 cultures overexpressing Siglec8 Met17-His155 (Cys42 > Ser42) domain performed at 15, 25 and 37C°. All cultures were supplemented with 1% (w/v) of glucose before expression induction.

A second attempt to improve the domain solubility was carried out using the GroEL/ES chaperone system (Houry, 2001). Chaperones are protein systems which assist the proper folding of target proteins, when they are synthesized on the ribosome, as well as the maintenance of pre-existing ones in a stable conformation. They are necessary for cell viability under both normal and stress conditions. The *E. coli* Hsp 60 chaperone system assists in the folding of folding intermediates with exposed hydrophobic surfaces, avoiding their aggregation and the formation of inclusion bodies. It consists of GroEL and its cofactor GroES. GroEL is an oligomer composed of two rings with seven 57 kDa subunits each, which has an inner cavity were the protein intermediates are folded. GroES is a complex of seven units of a 10 kDa monomer. The system is used in molecular biology to assist folding of recombinant proteins by co-expressing both GroEL/ES system and the recombinant protein. Figure 3.29 summarizes the GroEL/ES reaction cycle, where initially one GroEL ring in

'open' or *trans* conformation, captures a non-native substrate protein. Subsequently, ATP binding induces that GroEL ring conformational change into *cis*, enabling GroES to bind. This releases the substrate into the cavity, and the ATP molecule is hydrolysed leading to the binding of another ATP to the other unoccupied, *trans* GroEL ring. Finally, the binding induces the releasing of GroES from the complex, and the substrate into the cell.

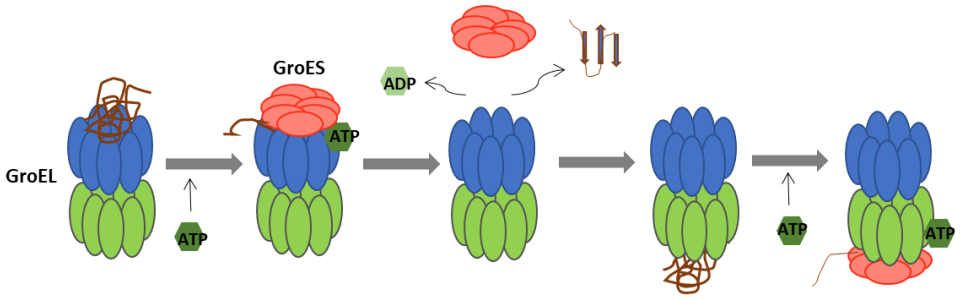


Figure 3.29. Schematic summary of GroEL/ES reaction cycle, assisting protein folding.

Two *E. coli* Rosetta gami B (DE3) 1L LB cultures were transformed with both Siglec8 domain Met17-His155 (Cys42>Ser42) cloned into pET-43(a+), and GroEL/ES system cloned into pET-28(a). All gene expression was under the control of the *lac* operon. Cultures were grown at 37ºC and 15 ºC respectively, and both induced with 1.0 mM IPTG. Samples were removed before and 4 hours after induction for cells grown at 37ºC, and before, 4 and 40 hours after induction for cells grown at 15ºC. Cell cultures were spun down, cells were lysed, and soluble and insoluble material was separated by centrifugation. Figures 3.30 and 3.31 show the SDS-PAGE gel and Western Blot analyses corresponding to those samples. In the SDS-PAGE gel, insoluble GroEL 57 kDa subunits (highlighted with a red arrow) were seen in the pellet, and soluble GroES 10 kDa subunits were observed in the supernatant. However, neither the SDS-PAGE gel nor the Western Blot show any trace of Siglec8 Met17-His155 (Cys42>Ser42) expression.

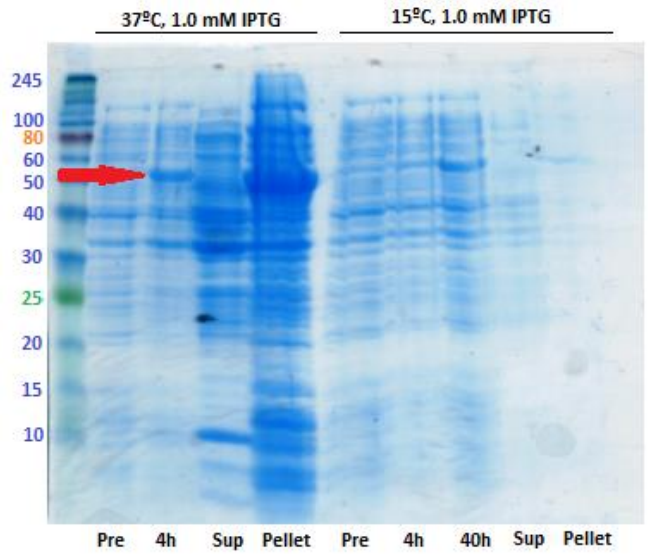


Figure 3.30. SDS-PAGE gel of *E. coli* Rosetta gami B (DE3) cell cultures transformed with both pET43-a(+) vector with the Siglec8 Ig-like V-type domain (18.10 kDa) and GroEL/ES (57 and 10 kDa subunits) cloned into pET28(a). Cells were grown in 1L of LB at 15 and 37ºC respectively, and expression was induced with 1.0 mM IPTG. Samples were taken out before, 4 (and 40) hours after expression induction; soluble and insoluble materials were also included. 100 mM DTT was added to all samples.

Figure 3.31. Western Blot of *E. coli* Rosetta gami B (DE3) cell cultures transformed with both pET43-a(+) vector with the Siglec8 Ig-like V-type domain (18.10 kDa) and GroEL/ES (57 and 10 kDa subunits) cloned into pET28(a). Cells were grown in 1L of LB at 15 and 37ºC respectively, and expression was induced with 1.0 mM IPTG. Samples were taken out before, 4 (and 40) hours after expression induction; soluble and insoluble materials were also included. 100 mM DTT was added to all samples.

The same mutated domain, Siglec8 Met17-His155 (Cys42 > Ser42), was cloned into a new vector, pET151/D-TOPO T, aiming to improve both yield and solubility.

Siglec8 Met17-His155 cloning (Cys42>Ser42) mutated domain cloning into pET151/D-TOPO

Figure 3.32 shows the agarose gel (1% w/v) for the amplified DNA fragment which encodes the blunt ended 421 bps Siglec8 Ig-like V-type domain from Met17 to His155 (Cys42>Ser42) (lanes 1, 2 and 3), ready to clone using the pET151/ Directional TOPO cloning vector. The domain was amplified by means of PCR using the pET-43a (+) vector as template and complementary primers.

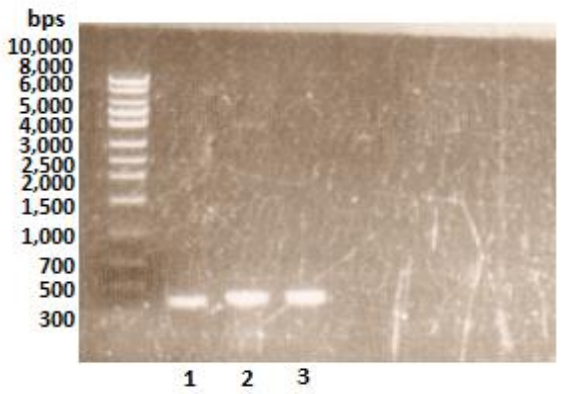


Figure 3.32. Siglec8 Ig-like V-type domain from Met17 to His155 blunt ended DNA fragment amplified by PCR ready to clone into the pET151/D-TOPO T kit.

Figure 3.33 shows the sequencing results for the domain cloned into the pET151/D-TOPO vector. The fragment contained an N-terminal 6 × His-Tag followed by a TEV site and eight extra amino acids at the C-terminus: KGELRSGC. The C-terminal amino acid was a cysteine which could be used for protein immobilization by means of thiol coupling. However, in this case it may take part in the formation of undesirable disulphide bonds and enhance protein domain aggregation.

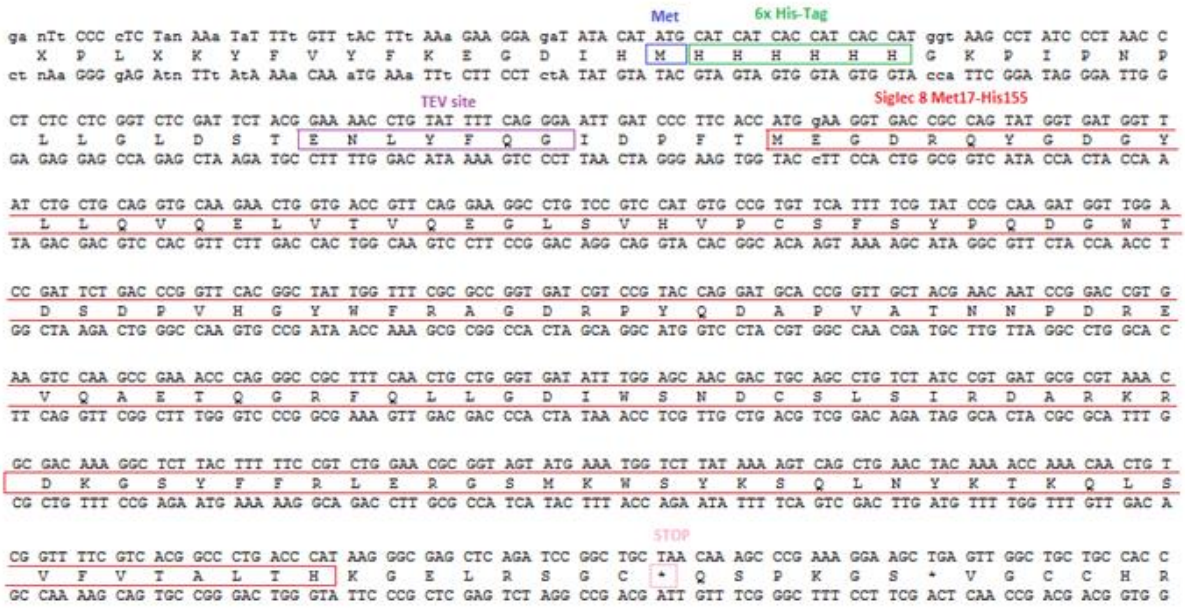


Figure 3.33. Sequencing results for the Siglec8 Met17-His155 Cys42>Ser42 domain cloned into the pET151/D-TOPO vector with some extra amino acids at the end (KGELRSGC).

Expression and solubility of Siglec 8 Met17-His155 (Cys42>Ser42) mutated domain cloned into pET151/D-TOPO T with the extra C-terminal amino acids (KGELRSGC)

Figure 3.34 shows the SDS-PAGE gel performed to two different 1 L *E. coli* Rosetta gami B (DE3) cultures transformed with the construct pET151/D-TOPO T, Siglec8 Met17-His155 (Cys42>Ser42) with extra C-terminal amino acids (KGELRSGC). Both were performed at 15° C and induced with 1.0 mM IPTG. The first one was performed in 2xYT media, and the second one in M9 media. Samples were taken out before, and 40 hours after expression induction. At 40 hours post-induction, a thick band migrating between the 15 and 25 kDa markers was observed and is in good agreement with the mass of the new construct (20.61 kDa). The expression levels of the domain cloned in the pET151/D-TOPO T vector improved significantly in both media in comparison with the other constructs. However, the protein domain was still insoluble.

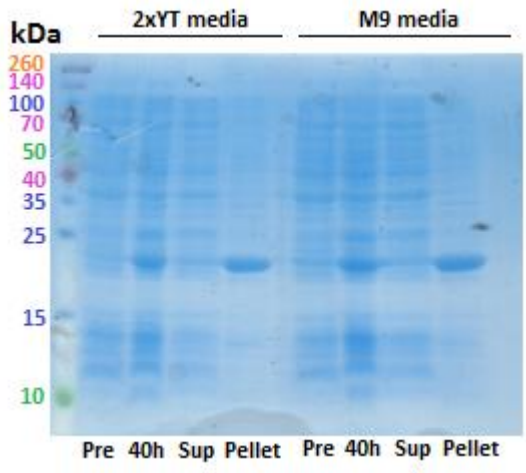


Figure 3.34. SDS-PAGE performed to *E. coli* Rosetta gami B (DE3) cell cultures grown in M9 and 2xYT media at 15°C for 40 hours. Samples were taken out before expression induction, 40 hours after and to both soluble and insoluble fractions after cell disruption and centrifugation.

Considering the possibility that the non-native, C-terminal cysteine could affect the domain folding and/or aggregation, a STOP codon was introduced immediately following His155.

Introduction of a STOP codon (TAA) following His155 of Siglec8 Met17-His155 (Cys42>Ser42) cloned into pET151/D-TOPO

Figure 3.35 shows the sequencing results for the domain cloned into the pET151/D-TOPO vector with a STOP codon introduced after His155 using the Q5 Site-Directed Mutagenesis Kit. The fragment still contained an N-terminal 6 × His-Tag followed by a TEV site.

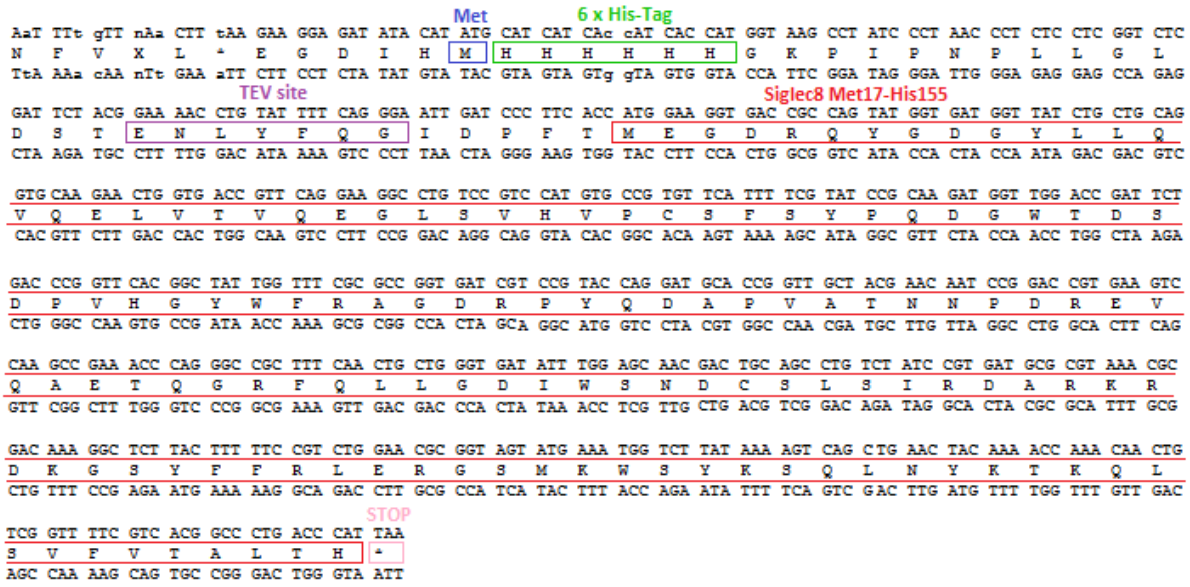


Figure 3.35. Sequencing results for the Siglec8 Met17-His155 Cys42>Ser42 domain cloned into the pET151/D-TOPO vector.

Expression and solubility of Siglec8 Met17-His155 (Cys42>Ser42) domain cloned into pET151/D-TOPO T

Figure 3.36 shows the SDS-PAGE analysis of a 1 L *E. coli* Rosetta gami B (DE3) LB culture, grown at 15° C and induced with 1.0 mM IPTG. Samples were taken out before, 40 hours after expression induction and from different dilutions of soluble and insoluble material. At 40 hours, a thick band migrating just below the 20 kDa standard was clearly distinguished, and likely to belong to the 19.82 kDa Met17-His155 (Cys42>Ser42) construct. The expression levels of the Siglec8 Ig-like V-type Met17-His155 (Cys42>Ser42) domain remained constant after the deletion of the eight extra amino acids (KGELRSGC), and there was no improvement in the domain's solubility.

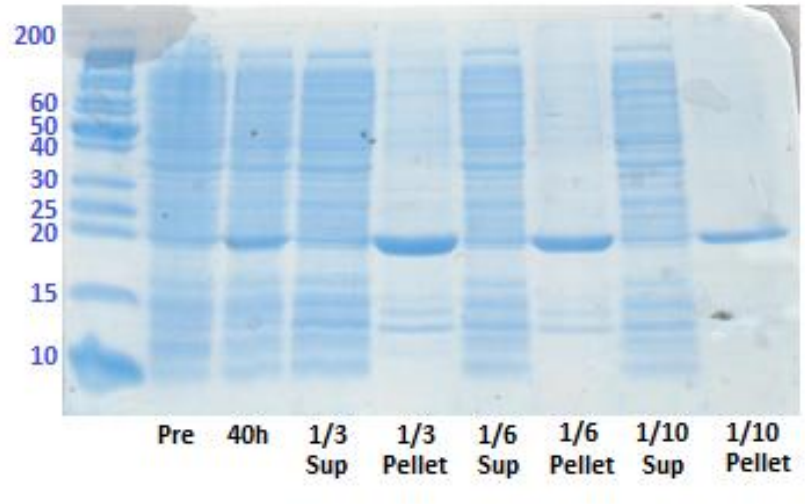


Figure 3.36. SDS-PAGE performed to *E. coli* Rosetta gami B (DE3) cell cultures grown in LB media at 15°C for 40 hours. Samples were taken out before expression induction, 40 hours after and to both soluble and insoluble fractions after cell disruption and centrifugation.

After all the above attempts to produce soluble protein had shown little success, it was decided to pursue a strategy of refolding Siglec8 Ig-like V-type domain from the insoluble fraction (i.e. from the inclusion bodies).

3.5 Refolding insoluble proteins from inclusion bodies.

The choice between working with insoluble recombinant protein or optimising the expression conditions to try to increase solubility relies on the nature of the protein and the amount required to perform an analysis. Strategies such as changing expression vector, fusion tag, host strain, growth conditions or co-expression with chaperones can really make a difference in achieving expression of soluble protein molecules. In contrast, a small protein (10-17 kDa) containing one – two cysteine residues might be expected to re-fold in a reasonable yield from inclusion bodies, while larger proteins with many Cys residues are the most problematic to refold (Palmer and Wingfield, 2012). In our case, the 19.82 kDa protein domain containing a single disulphide bond is a reasonable candidate for refolding.

Inclusion bodies are non-crystalline, amorphous structures formed by a mixture of unfolded and partially folded proteins, and perhaps native secondary structures densely packed (Palmer and Wingfield, 2012).

In vitro folding studies with a mixture of proteins suggested that aggregation occurs between folded subdomains or intermediates which associate with complementary binding surfaces from other molecules to form oligomers and

eventually aggregates, therefore those interactions are not random or unspecific (Palmer and Wingfield, 2012). The propensity of a protein to form inclusion bodies does not seem to be directly related to the presence of cysteines and their sulfhydryl groups, as proteins without cysteines have been found to form inclusion bodies. However, the reducing environment of the *E. coli* cytoplasm may contribute to the instability and low solubility of these types of protein. Some other factors that may contribute to aggregate formation are the lack of post-translational modifications as glycosylation, and chaperones or enzymes catalysing folding such as the cis-trans isomerise. Furthermore, the high levels of protein concentration together with the limited solubility of their intermediates can also contribute (Palmer and Wingfield, 2012).

3.6 Refolding methods

Refolding methods require the disruption of pure inclusion bodies into individual unfolded protein molecules in solution (Yamaguchi and Miyazaki, 2014). Initially, several washes must be carried out to remove other components in the pellet such as unbroken cells, membrane and cell material, or any other sources of contamination. Wash buffers contain a mixture of the non-ionic chaotropic agent urea at low concentrations, a reducing agent as DTT, detergent to help to remove cell membranes, and protease inhibitors to avoid proteolysis. Washes must be performed with the same buffer free of detergent to remove any excess which may influence the folding later. Once inclusion bodies are clean, they must be mechanically disrupted with extraction buffer containing high concentrations (5-8 M) of the strong chaotropic agent guanidine hydrochloride (GdnHCl) combined with a reducing agent (DTT). Thus, individual unfolded molecules are solubilised and ready for refolding (Yamaguchi and Miyazaki, 2014).

The procedure for the removal of the denaturant from denatured proteins is a key step in their efficient recovery and refolding (Yamaguchi and Miyazaki, 2014**).** There are several approaches that have been reported to refold an inactive protein into an active one such as size-exclusion chromatography, reversed micelles, zeolite absorbing systems and the natural GroEL-GroES chaperones. However, in most cases there is a significant amount of protein precipitation due to the presence of high numbers of aromatic and aliphatic residues which decrease the solubility of proteins

during refolding and contribute to the aggregation of protein folding intermediates (Yamaguchi and Miyazaki, 2014).

3.7 Description of the refolding methods tested in this work

The most conventional refolding methods are dialysis and dilution, which can yield as much as 40% and 80% refolding of the total protein, respectively, at concentrations of 1-10 mg/mL (Yamaguchi and Miyazaki, 2014). Dialysis is based on the removal of denaturant by buffer exchange through a membrane with a molecular weight cut off (MWCO) of up to half of the protein molecular weight, present in refolding buffer of at least 50x the volume of the protein solution. Additives that inhibit aggregation are typically included when the target protein is refolded at intermediate to high concentration (0.1 to 10 mg/mL) (Yamaguchi and Miyazaki, 2014). Figure 3.37 summarizes the basis of dialysis, where the membrane's molecular weight cut off enables the small molecules, but not the protein, to pass through the membrane led by osmosis.

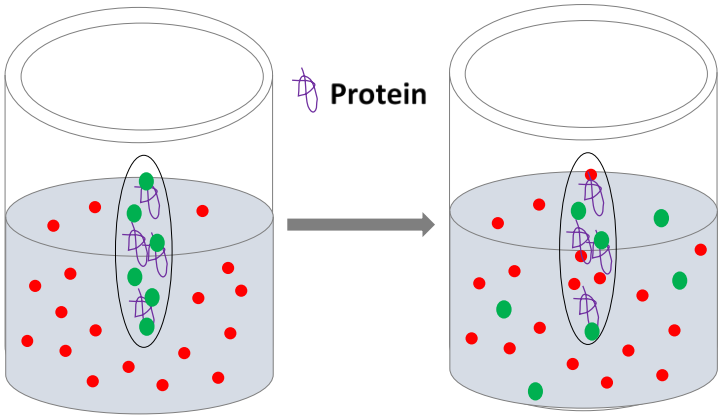


Figure 3.37. Schematic representation of a dialysis, where the membrane's molecular weight cut off enables the small molecules to pass through the membrane but not the protein molecules.

However, the rapid decrease in denaturant concentration can initiate the re-formation of aggregates, possibly because of the contact between exposed hydrophobic surfaces. The dialysis method can be optimised by removing the denaturant in a step-wise fashion, in which the protein is brought to equilibrium with gradually decreasing concentrations of denaturant (Yamaguchi and Miyazaki, 2014). This strategy helps to avoid aggregation of improperly folded protein and denaturing of properly folded molecules. Figure 3.38 summarizes both types of dialysis:

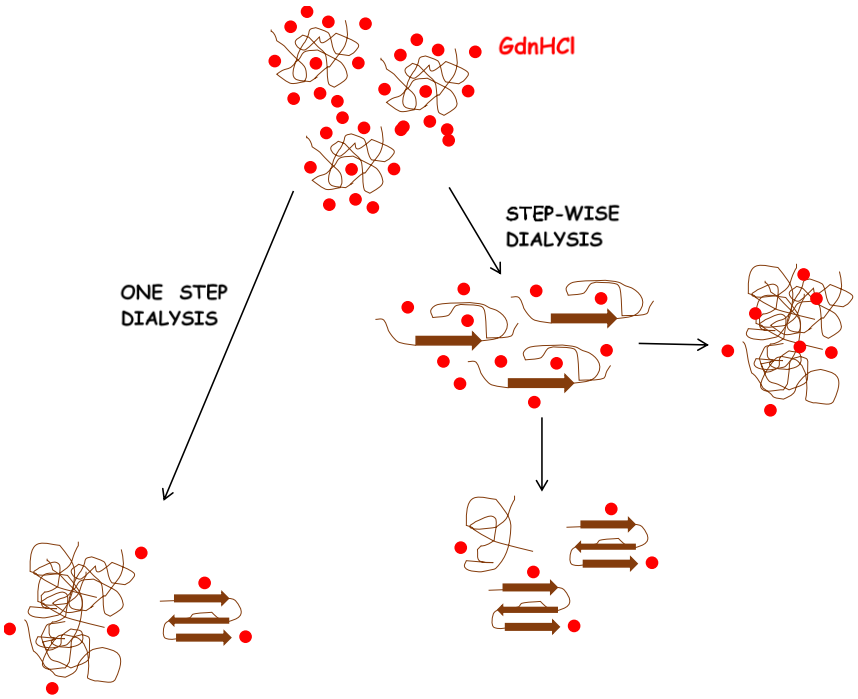


Figure 3.38. The two types of dialysis: in one step dialysis (left), the concentration of denaturing agent is reduced all at once. Meanwhile in the step-wise dialysis (right), it is reduced gradually.

The Dilution method involves diluting the unfolded protein solution by adding it dropwise directly into a 100× volume of refolding buffer (Yamaguchi and Miyazaki, 2014). The protein concentration is so low when a drop is added that protein molecules should not aggregate and refold. When the next drop is added, the refolded molecules should (theoretically) not participate in aggregation. The disadvantage of this method is that it requires concentration of a large volume of refolding buffer after the refolding (Yamaguchi and Miyazaki, 2014). Figure 3.39 summarizes the drop-wise refolding method:

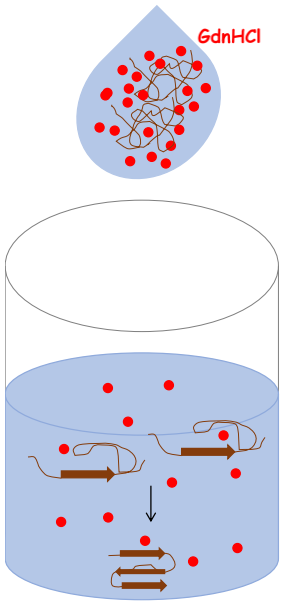


Figure 3.39. Drop-wise refolding method, where denatured protein molecules in high concentrations of denaturing agent, are added drop-wise to a 100x volume with a low concentration of denaturing agent.

3.8 Refolding buffer composition according to protein properties

The composition of the refolding buffer is critical and must be determined according to the protein properties. A wide variety of compounds have been demonstrated to enhance the refolding process for specific proteins, including amino acids, mild detergents, polymers, prosthetic groups, sugars, ligands and mixed micelles (Yamaguchi and Miyazaki, 2014). Those additives are thought to decrease the tendency of folding intermediates or mis-folded proteins to aggregate, stabilize the folded proteins and inhibit non-productive refolding pathways, or a mixture of all of these. However, there is no “one size fits all” formulation. To determine optimal refolding conditions one must still test many different additives and conditions (Yamaguchi and Miyazaki, 2014).

One key protein property that influences the refolding conditions is the presence of disulphide bonds. When disulphide bonds are required for correct protein folding, a reducing agent like DTT must be included in the solubilisation step and must be maintained during the initial refolding to prevent the formation of intermolecular disulphide cross-linking. In addition, thiol reagents such as the pair glutathione reduced (GSH) - glutathione oxidized (GSSG). These compounds facilitate the formation of disulphide bonds through a twostep process, where the protein forms a

mixed disulphide intermediate with the thiol, followed by a second exchange reaction with a second thiol to form the intra-molecular disulphide bond. Thus, the reversibility of any mis-paired disulphide bonds in folding intermediates' is allowed. The reducing agent is typically added in a five to ten-fold molar excess over the oxidant i.e. 1 mM GSH-0.2 mM GSSG.

Another protein property that influences the refolding conditions is the hydrophobicity. The presence of high numbers of aromatic and aliphatic residues (Ala, Cys, Val, Ile, Leu, Met, Phe and Trp) is believed to decrease solubility of proteins during refolding and contribute to aggregation. The easiest way to minimise hydrophobic aggregation and allow the recovery of monomeric active protein is to add a mild detergent like N-lauroylsarcosine to the refolding buffer.

Knowledge of any essential prosthetic groups is also important. For example, bacterial alkaline phosphatase has been reported to require Zn^{2+} for proper refolding.

Finally, the net charge must also be considered. The net charge of a protein is calculated by summing the basic amino acid residues and subtracting the acidic amino acids. The pI of a protein is the pH at which it has a net charge of zero, and therefore its solubility is minimal. By shifting the pH away from neutral, greater solubility and recovery of functional protein is possible. It is recommended working at a pH at least one unit higher or lower than the protein's pI (Yamaguchi and Miyazaki, 2014).

3.9 Analysis of Siglec8 Ig-like V-type domain Met17-His155 (Cys42>Ser42) relevant characteristics for refolding

Table 3.2 shows the analysis of the Siglec8 Ig-like V-type domain, using the ProtParam online tool. The presence of two cysteines require the usage of DTT in the washing and first refolding stages, as well as the redox pair GSH-GSSG in the last stages. Furthermore, the pI value 6.11 was considered when selecting the pH of the refolding buffer (pH 7.5 or higher).

Table 3.2. Siglec8 Ig-like V-type (Cys42>Ser42) domain relevant parameters for refolding (ProtParam).

	Number of amino acids (aa')	MW (kDa)	Cysteines	Hydrophobic aa'	Prosthetic groups	pI
Siglec8 Met17-His155 (Cys42>Ser42)	172	19.82	2	67 (39%)	None	6.11

3.10 Comparison between one-step dialysis, step-wise dialysis and drop-wise methods to refold the Siglec8 Ig-like V-type domain

Cell pellets from 500 mL cell cultures were washed and extracted according to the protocol described in Materials and Methods. Subsequently, the concentration in extraction buffer was adjusted to 2 mg/mL by comparison of the size and intensity of SDS-PAGE bands with a sample of 2 mg/mL of Chicken Lysozyme, as shown in Figure 3.40, as well as measuring the absorbance.

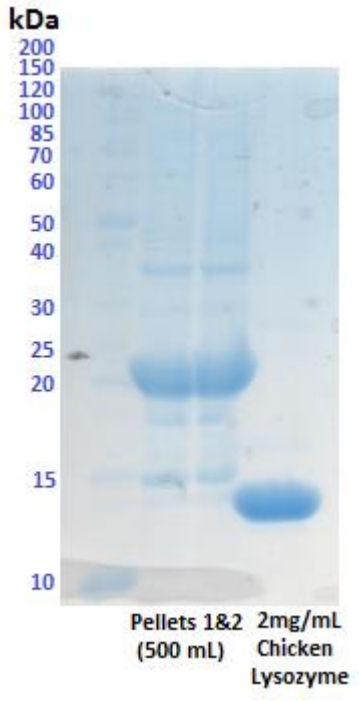


Figure 3.40. SDS-PAGE gel made to samples of unfolded protein in extraction buffer and to a 2 mg/mL solution of Chicken Lysozyme to adjust the concentration of the extracted Siglec8 Ig-like V-type protein domains to 2 mg/mL.

The three refolding methods, one-step and step-wise dialysis, and drop-wise dilution, were tested with the initial same amount of unfolded protein (2 mg/mL), solubilised in 2 mL of extraction buffer, following the protocols described in Materials and Methods. Subsequently, the three samples belonging to the three different refolding methods were purified using a 1mL HisTrap HP column, and buffer was exchanged into 500 μL of 20mM NaPi, 500mM NaCl, pH 7.5. Figure 3.41 shows the SDS-PAGE analyses of three 10 μL samples, belonging to samples refolded using the three different protocols: from left to right one-step dialysis, step-wise dialysis and drop-wise refolding. Drop-wise refolding provided turned out to be the method that provided a higher yield of protein after purification.

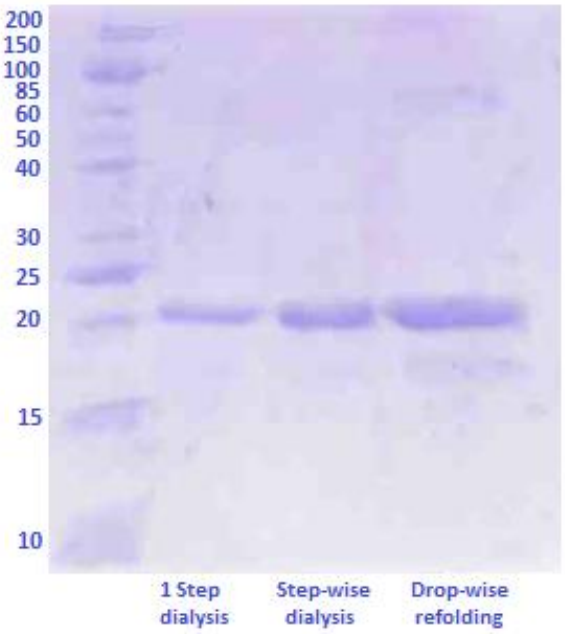


Figure 3.41. SDS-PAGE performed to three 10µL samples belonging to the three protein solutions refolded by means of three different methods (one step dialysis, step-wise dialysis and drop-wise).

Based on all cloning, expression, refolding and purification trials, I kept working with the Siglec8 Ig-like V-type Met17-His155 (Cys42>Ser42) construct cloned into the vector pET151/D-TOPO T, expressed in the *E. coli* strain Rosetta gami B (DE3), refolded using the drop-wise method and purified by means of Niquel affinity column into 20 mM NaPi, 500 mM NaCl, pH 7.5 buffer.

3.11 Industrial collaboration with Mologic Ltd. Role of eosinophils in Chronic Obstructive Pulmonary Disease (COPD) and diagnostic through Siglec8.

Apart from their well-known implication in asthma, eosinophils have recently been found in the airways, tissues, and circulation of patients with COPD, during both stable and disease exacerbations (Bafadhel, *et al.*, 2017). Furthermore, epidemiological studies showed an association between the blood eosinophil count and the risk of COPD exacerbations, mortality and response to corticosteroid treatment. (Casanova, *et al.*, 2017; Prins, *et al.*, 2017; Zeiger, *et al.*, 2017). Thus, eosinophils have been pointed out as biomarkers and mediators of the COPD, but their contribution to the mechanism of this disease remains to be elucidated (Bafadhel, *et al.*, 2017).

This PhD project also encompassed an industrial collaboration with Mologic Ltd. (Bedford, UK). Mologic is an independent SME (Small and medium-sized enterprises) company with approximately 60 employees at two sites in the UK. Mologic develops diagnostic technologies using pure, well characterised, biologically relevant samples of the biomarkers of interest. Those biomarkers are usually proteins or peptides used to generate proprietary antibodies and serve as calibration standards. The antibodies are used to develop immunochromatographic diagnostic devices as lateral flow strips for the biomarkers of interest. This was the case of the Siglec8 human receptor, an excellent biomarker to predict COPD exacerbations. The protocol to produce and purify Siglec8 recombinant protein domains was developed at Warwick in collaboration with Mologic. Subsequently some of the material produced was used to generate antibodies anti-Siglec8 by boosting the immune response of immunised animals. Thus, those antibodies may be employed in the development of a diagnostic test for chronic lung disease exacerbations by detecting increased levels of eosinophils in patient's body fluids.

Mologic conducts exploratory research in biotechnology and healthcare diagnostics to develop novel assay systems, combinations of biomarkers and approaches to in-home testing and disease management. The company recently obtained a $4.8 M grant for the Bill and Melinda Gates Foundation to establish the Centre for Advance Rapid Diagnostics (CARD) at its Bedford site. The challenge of CARD is to create a step-change in simple, rapid diagnostic test technology for the detection of protein biomarkers at the point of care. The new, ultra-sensitive technology platform is being developed with a focus on manufacturing costs and ease-of-use to enable deployment throughout the developing world. While new immunodiagnostic technology carried out in the laboratory is achieving unprecedent sensitivity and accuracy, rapid diagnostic tests, which are well suited for use at the point of care and in the field, have changed little since their invention in the mid-1980s. The grant to establish CARD as the means to deliver a new ultra-sensitive point of care technology platform, is aligned with the Gates Foundation's strategy to eradicate poverty through enhancing healthcare and bring the diagnostic standards and capability of developed world care to those who need it most within low resource countries.

The Mologic team includes synthetic organic/peptide chemists, immunoassay experts, biochemists, microbiologists and immunologists, working in up-to-date laboratories. Mologic had a long-standing interest in developing diagnostics for markers of infection and inflammation in human disease. To understand how Mologic operates and the kind of products developed, it is helpful to look at the two first commercialised products as illustrative examples:

- PERiPLEX®. Mologic's first diagnostic product was CE Marked in December 2016. It is a rapid point-of-care test for the detection of infection in patients on Peritoneal Dialysis (PD). There are over 200,000 patients being treated with PD globally, and it is widely accepted as the preferred method for managing severe chronic kidney disease (CDK) or end stage renal disease (ESRD). PD is a type of dialysis that uses the peritoneum in a patient's abdomen as the membrane through which fluid and dissolved substances are exchanged with the blood. It is used to remove toxins and excess fluid, and correct electrolyte problems is those with kidney failure. However, PD is known to have a high rate of associated infections, which can lead to peritonitis. Using current methods, diagnosis of infection can take more than 24 hours. This delay risks damage to the peritoneal membrane or the formation of scar tissue, avoiding the long-term use of PD. PERiPLEX detects two critical biomarkers of infection in PD waste fluid using a lateral flow immunoassay system, and provides results within 10 minutes. A positive result can give PD patients warning of an infection, prompting them to contact their healthcare provider to initiate antibiotic therapy. Thus, earlier identification of infection would enable preventive use of available antibiotics, resulting in improved management of infection and prevention of infection-related damage to the peritoneal membrane.
- BVPro® is a simple, rapid, point-of-care diagnostic intended for professional use as an aid in the diagnosis of bacterial vaginosis (BV). The test detects sialidase enzyme activity in vaginal fluid, using a lateral flow immunoassay system which gives results in 15 minutes. Sialidase is an enzyme produced by bacterial pathogens in the vagina. Untreated BV is associated with a slightly increased risk of developing complications during pregnancy, inclusing miscarriage or preterm birth. It also increases the chance of developing an infection after uterus operations and the risk of contracting STIs. BV can usually be successfully treated

with antibiotics, but for a number of woman, the first course of treatment is not effective with the symptoms returning within three months. In these cases, a different form of treatment may be prescribed. BVPro® uses patented technology to detect sialidase activity. Sialidase is an enzyme that degrades sugars and has been reported to weaken the natural mucosal barriers in the vagina, allowing anaerobic bacteria to colonise and out-compete healthy lactobacilli. BVPro® detects this activity by supplying the enzyme with an artificial substrate. Thus, if sialidase is present, it will modify the substrate which is then captured on the test line using specific antibodies. Where no enzyme is present, the substrate remains unchanged and will not be captured. A procedural control line is included to indicate that the assay has been performed correctly. The presence of a red line in addition to the control line indicates a positive result.

Chapter 4.

BIOPHYSICAL CHARACTIZATION OF SIGLEC8 IG-LIKE V-TYPE DOMAINS

4. Biophysical characterization of Siglec8 Ig-like V-type domains.

4.1 Application of biophysical techniques to protein characterization

Biophysics is defined as an interdisciplinary science that applies physical techniques to study and characterize biological systems. In this chapter bioinformatics and different biophysical techniques such as protein fluorescence, circular dichroism (CD), UV-vis absorbance, single X-ray scattering (SAXS) and mass spectrometry (MS) were employed to verify the identity, quantify and characterize the secondary structure of the Siglec8 Ig-like V-type carbohydrate binding domains after refolding and purification.

The biophysical characterization of a recombinant protein is critical to determine its suitability for further ligand-binding studies. Those characterizations may reveal features such as the formation of protein aggregates in solution or the adoption of the correct protein three-dimensional conformation also known as native structure, which is essential for protein function.

Protein folding is the process by which a protein structure assumes its functional shape or conformation, so it is able to perform its biological function. The main stages of the folding process are described and represented below using the NMR solution state resolved structure of the SUMO-1 protein (pdb ID 2ASQ) as an example in Figure 4.1.

The linear amino-acid sequence of a protein is known as its primary structure. The formation of intramolecular hydrogen bonds between the carbonyl O of one amino acid and the amino H of another leads to regions in which chains are organised into secondary regular structures. The most common ones are known as alpha-helices and beta sheets. Alpha helices form a spiral shape and the beta sheets' backbone can bend over itself in a parallel or an antiparallel way. Subsequently, the protein folds into its tertiary structure so that the hydrophilic sides remain in contact with the aqueous environment surrounding the protein and the hydrophobic patches face the core. There may be also covalent disulphide bridges formed between cysteine residues stabilizing

the protein. In some cases, the assembly of several subunits that have already folded may interact to form a fully functional quaternary protein.

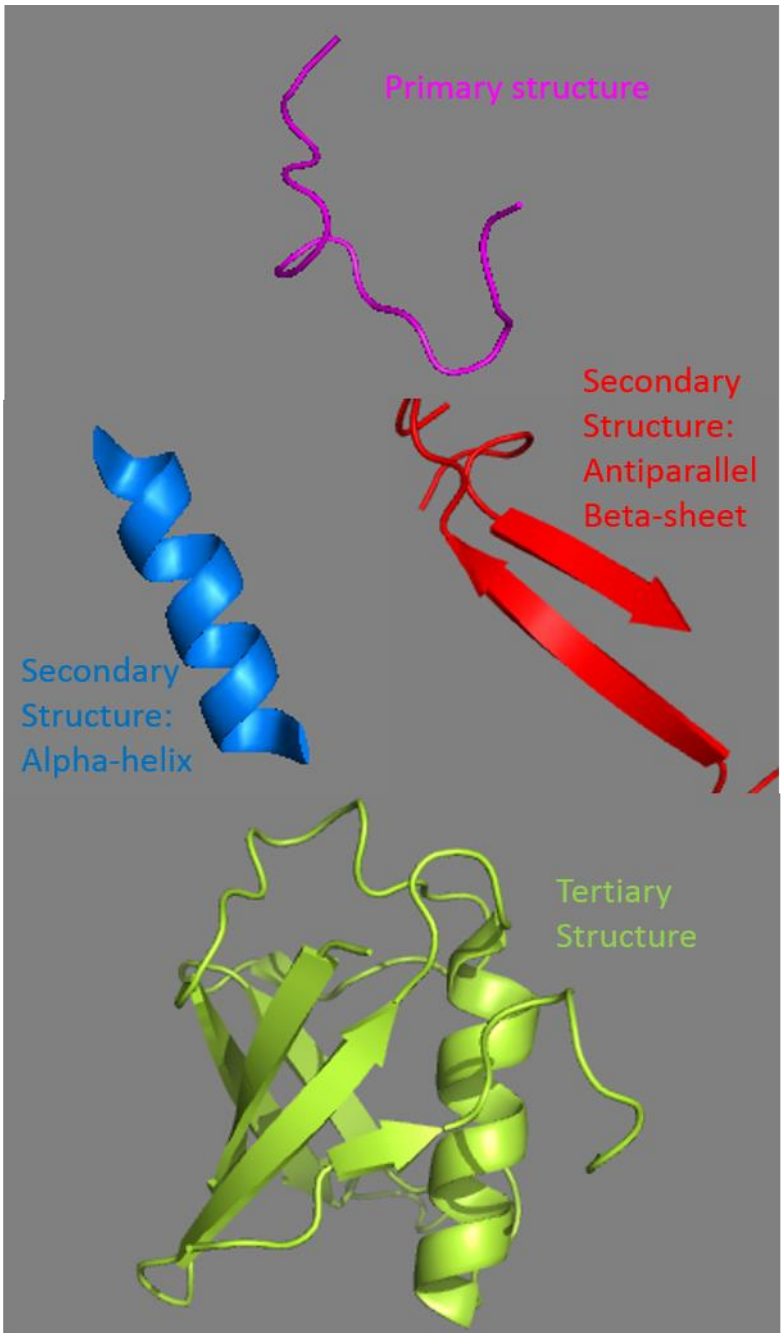


Figure 4.1. Representation of the main stages of the protein folding process using the solution state NMR resolved structure of the SUMO-1 protein (pdb ID 2ASQ) as an example.

The aim of this chapter is to characterize the soluble Siglec8 Ig-like V-type carbohydrate binding domains in 20 mM NaPi, 500 mM NaCl, pH 7.5 to determine if they are suitable to perform further ligand binding studies.

4.2 Siglec8 Ig-like V-type domain sequence, composition and structural analysis

The analysis of the amino acid sequence reveals important protein features and physicochemical parameters such as molecular weight (MW), isoelectric point (pI), extinction coefficient (ε) or hydrophobicity. These are required to manipulate the protein in the laboratory, optimise the composition of buffers and perform biophysical characterizations.

The online software ProtParam from ExPASy (SIB) was used to calculate the number of amino acids, molecular weight (MW), isoelectric point (pI) and the amino acid composition of the Siglec8 Ig-like V-type domain (Table 4.1). The isoelectric point is the pH at which a protein carries no net electrical charge. It is an important property because it is at this pH that the protein is least soluble, and therefore unstable (Kozlowski, 2017). Thus, it is convenient to keep the pH of buffers at least one unit above or below this value when working with proteins.

Table 4.1. Number of amino acids, molecular weight and pI of Siglec8 Ig-like V-type construct

	Siglec8 Ig-like V-type construct
Number of amino acids	172
Molecular weight (Da)	19,819.01
pI	6.11

The amino acid composition for the Siglec8 Ig-like V-type construct is represented in Figure 4.2, where the 20 naturally-occurring amino acids are classified into three groups according to the chemical nature of their side chain: non-polar, polar and charged. All 20 amino acids are present between 1.2 and 8.7%.

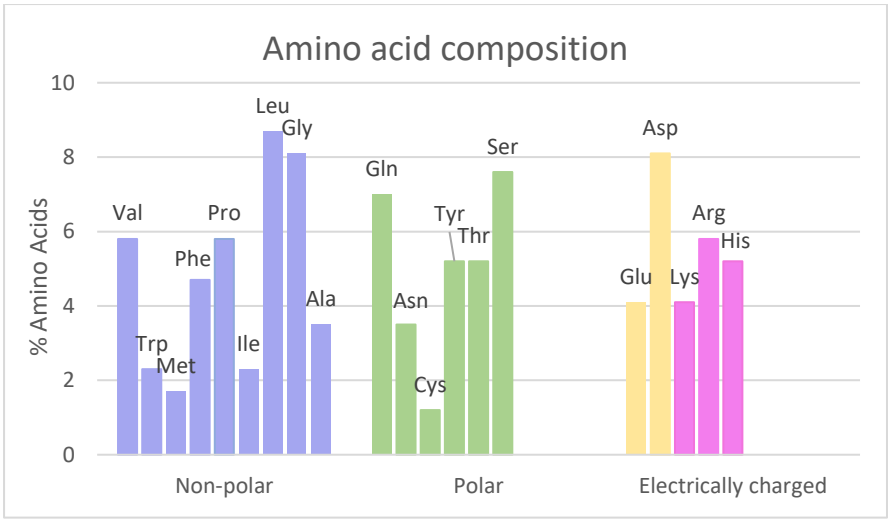


Figure 4.2. Proportion of each amino acid present in the Siglec8 Ig-like V-type construct.

The amino acid spatial distribution may be more relevant than the simple composition to determine any hydrophobic or charged pockets and patches in the protein's surface. Figure 4.3 presents the three-dimensional structure of Siglec8 Ig-like V-type domain resolved by solution state NMR (Propster, et al., 2016) with the non-polar amino acids highlighted in blue, the polar residues in green, basic residues in pink and acidic ones in yellow. The three-dimensional structure reveals a homogeneous distribution of the different subgroups of amino acids with a higher proportion of hydrophobic patches across the domain.

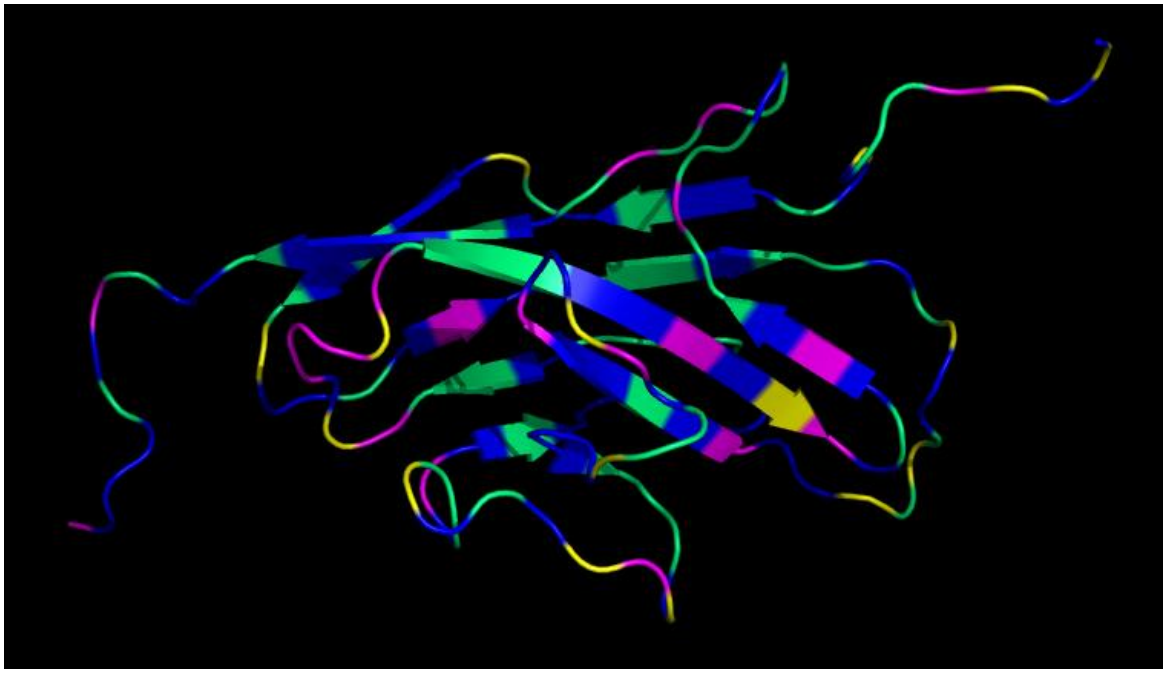


Figure 4.3. Siglec8 Ig-like V-type domain resolved by ssNMR with the different types of amino acids highlighted in different colours: pink and yellow for positive and negative charged ones respectively, green for the polar ones and blue for the apolar ones (Propster, *et al.*, 2016), PDB ID 2N7B.

4.3 Identity confirmation of the recombinant Siglec8 Ig-like V-type domain by Tandem Mass Spectrometry (nano-LC-ESI-MS/MS)

Tandem Mass Spectrometry (MS) was used to confirm the identity of the bacterially-expressed Siglec8 Ig-like V-type protein domain after purification. A SDS-PAGE gel with 10μL of pure protein was run and the protein band was excised and sent to the Proteomics facility in the School of Life Sciences at Warwick for analysis. The protein was extracted from the poly-acrylamide gel band and enzymatically digested with trypsin. Trypsin is an enzyme that cleaves proteins at the carboxyl side of the amino acids lysin or arginine, except when either is followed by proline, generating a mixture of peptides. 20μL of this peptide digest was injected in the nano-LC-ESI-MS/MS system (nano-liquid chromatography, electrospray, tandem mass spectrometry).

The resulting peptides were fractionated using liquid chromatography and directly introduced to the mass spectrometer through electrospray ionization. The mass-to-charge ratio of the ionized peptides was measured by tandem MS and the results were compared with Homo sapiens and Escherichia coli databases from MaxQuant (Stone, et al., 1998). MaxQuant software was used for protein identification and Scaffold software for data analysis and results visualization. Table 4.2 shows a Table with the results. The identity of the protein sample as Siglec8 human receptor was confirmed with more than 95% probability and 14 out of 16 peptides yielded more than 98% sequence match with the receptor.

Table 4.2. Identity confirmation of LC-ESI-MS/MS results of Siglec8 Ig-like V-type domain after comparison with MaxQuant database.

Peptide	Prob	Exclusive To
AGDRPYQDAPVATNNPDR	100%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8
AGDRPYQDAPVATNNPDREV	99%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8
DKGSYFFR	99%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8
DKGSYFFRLER	100%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8
FQLLGDIWSNDCSLSIR	100%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8
FQLLGDIWSNDCSLSIRDAR	93%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8

FQLLGDIWSNDCSLSIRDARK	80%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8
GSMKWSYK	99%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8
GSMKWSYKSQLNYK	100%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8
GSYFFRLER	95%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8
LERGSMKWSYK	100%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8
PYQDAPVATNNPDR	100%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8
PYQDAPVATNNPDREVQAET	99%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8
RDKGSYFFR	100%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8
WSYKSQLNYK	100%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8
WSYKSQLNYKTK	98%	Sialic acid-binding Ig-like lectin 8-Homo sapiens-SIGLEC8

4.4 Quantification of Siglec8 Ig-like V-type domain concentration by UV-vis absorbance

The aromatic amino acids tryptophan (Trp), tyrosine (Tyr) and phenylalanine (Phe), and the disulphide linkages between cysteines (S-S) constitute chromophores that absorb in the near UV region. Absorption of near UV radiation by proteins is usually monitored at 280 nm (A_{280}) due to the high absorption by Trp and Tyr at this wavelength. Thus, if the molar absorption coefficient of the protein and the path length of sample cell are known, the protein concentration (C) can directly be determined using the Beer-Lambert law (WADDELL, 1956):

$$A_{280} = C \cdot \varepsilon_{280} \cdot l$$

Where ε_{280} is the molar absorption coefficient and λ is the path length. The molar extinction coefficient can be easily determined if the composition of the protein is known. Table 4.3 shows the Trp, Tyr and disulphide absorption coefficients at 280 nm. They vary depending on the polarity of the media in the case of Trp and Tyr, and on the oxidative state for the sulphide groups. Values differ for folded and unfolded proteins (WADDELL, 1956; Sapan and Lundblad, 2015):

Table 4.3. Molar extinction coefficients of Tryptophan, Tyrosine and Disulphide bonds at 280 nm.

	ε280 (M⁻1 cm⁻1)		
	Trp	**Tyr**	**(S-S/SH SH)**
Average value in folded proteins	5,500	1,490	125
Value in unfolded proteins	5,690	1,280	120

The calculation for an unfolded protein would be:

ε_{280} = (5,690 · number of Trp) + (1,280 · number of Tyr) + (120 · number of reduced SH-SH)

There are several online software programmes that calculate protein's physicochemical parameters such as ε_{280}. The molar absorption coefficient for the Siglec8 Ig-like V-type construct was determined from the sequence of amino acids, using the online tool ProtParam from ExPASy (SIB), as ε_{280ss} 35,535 M-1·cm-1 and $\varepsilon_{280SHSH}$ 35,410 M-1·cm-1 for the folded and the unfolded domain respectively.

Figure 4.4 displays the absorption spectra between 230 and 330 nm for the three fresh refolded Siglec8 Ig-like V-type domain samples by means of three different methods: one-step dialysis (blue) and step-wise dialysis (green) dialysis; and drop-wise refolding (red). Absorption was measured to 2 mL samples after refolding and buffer exchange into 20 mM NaPi 500 mM NaCl, pH 7.5 buffer. The protein concentration in all samples was determined, by applying the Beer-Lambert law as described above, as 0.06 mg/mL (2.81 μM) for the one-step dialysed sample, 0.27 mg/mL (13.7 μM) for the step-wise dialysed sample and 0.5 mg/mL (25.3 μM) for the drop-wise refolded sample. Refolding and purification yields were 3, 13.5 and 25 % respectively.

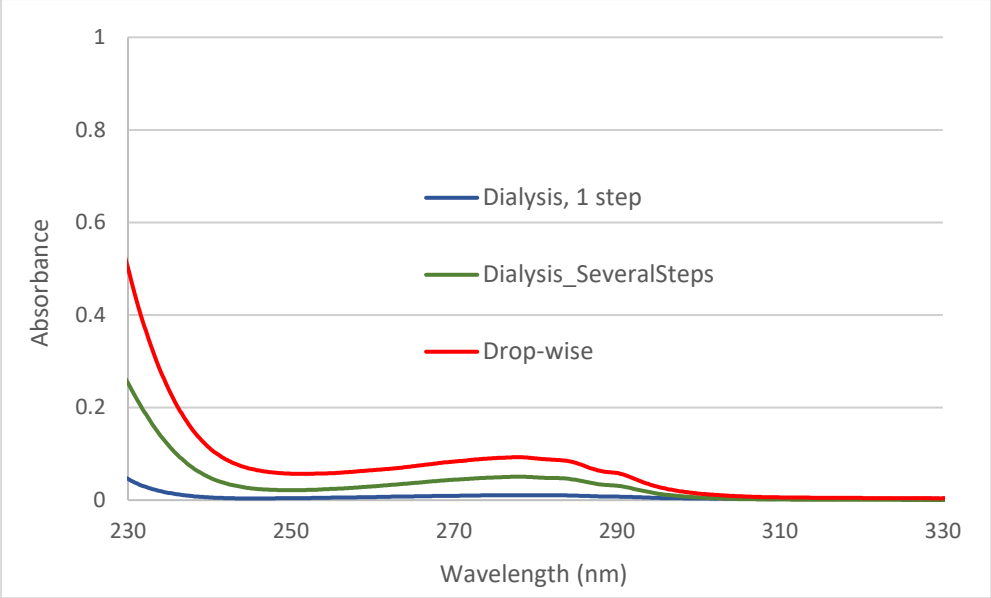


Figure 4.4. Absorption spectra of three Siglec8 Ig-like V-type domain samples refolded by means of one-step dialysis (blue), step-wise dialysis (green) and drop-wise (red).

Figure 4.5 shows the absorption spectrum between 230 and 330 nm for the drop-wise refolded protein after being concentrated down to a volume of 500 µL reaching a concentration of 1 mg/mL (50.6 µM).

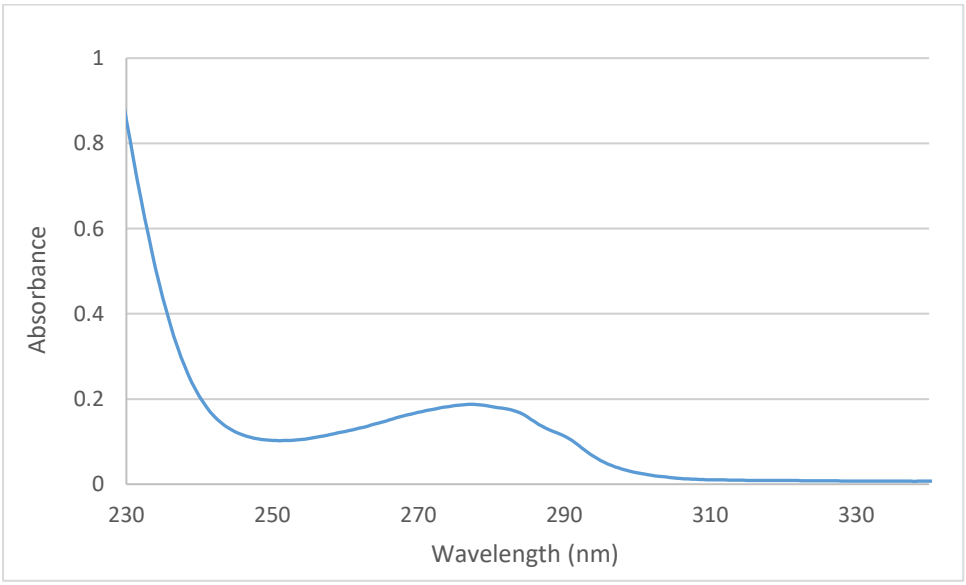


Figure 4.5. Absorption spectrum between 330 and 230 nm of Siglec8 Ig-like V-type refolded domains into 20mM NaPi 500mM NaCl, pH 7.5 buffer.

4.5 Siglec8 Ig-like V-type domain fold monitored by Circular Dichroism (CD)

Circular Dichroism (CD) is the difference in the absorption of left-handed circularly polarised light (L-CPL) and right-handed circularly polarised light (R-CPL), which occurs when a molecule contains one or more light-absorbing groups. In CD spectroscopy, the CD of molecules is measured over a range of wavelengths. The signal can be positive or negative, depending on whether L-CPL is absorbed to a greater extent than R-CPL (positive CD signal) or to a lesser extent (negative CD signal) (Micsonai, et al., 2015).

One of the most important applications of CD is the analysis of protein secondary structures. Protein CD spectra are usually recorded between 180 and 260 nm and can be used as input for software such as Dichroweb (London, 2010-2017) that compare the obtained spectrum to a database of previously obtained spectra for α-helical, β-sheet and random coil proteins to yield an estimate of secondary structure content (Micsonai, et al., 2015).

The typical model CD spectra belonging to the different pure secondary structures are represented in Figure 4.6 (CorrÊa and Ramos, 2009). α-helical structures' spectra (in blue) are positive below 200 nm and show two negative minima around 210 nm and above 220 nm. β-sheet spectra (in red) are also positive below 200 nm and display one less sharp negative minimum around 218 nm. Random coil spectra (in green) are flat or slightly positive above 210 nm and display a sharp minimum below 200 nm (CorrÊa and Ramos, 2009).

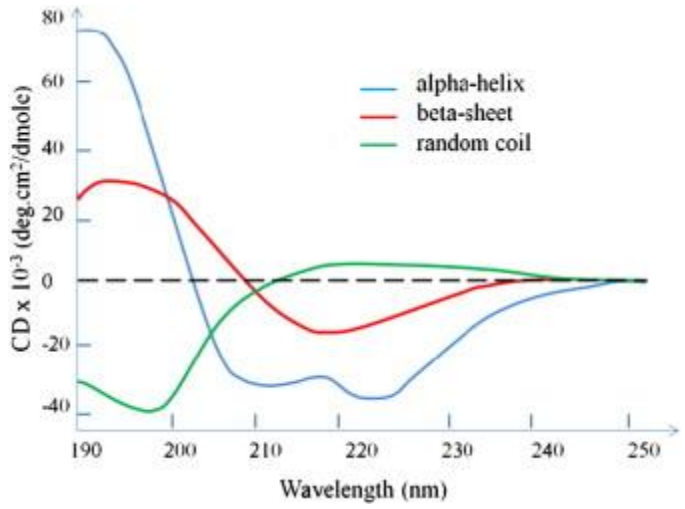


Figure 4.6. Model spectra for α-helix, β-sheet and random coil pure secondary structures (CorrÊa and Ramos, 2009).

A CD spectrum was recorded for a fresh Siglec8 Ig-like V-type domain sample after refolding, purification and buffer exchange into 50 mM Tris, pH 7.5. The concentration was approximately ~ 1 mg/mL (Figure 4.7). The spectrum shows two negative minima at 232 and 208 nm and one positive maximum at 199 nm and could not be fitted to any specific secondary structure by the Dichroweb algorithms (London, 2010-2017). This was unsurprising as β-sheet rich proteins, such as Siglecs, have been described in the literature as problematic when fitting CD spectra (Micsonai, et al., 2015). This is due to the fact that they can arrange in parallel or antiparallel orientations and twist, displaying a huge morphological a spectral variety of β structures.

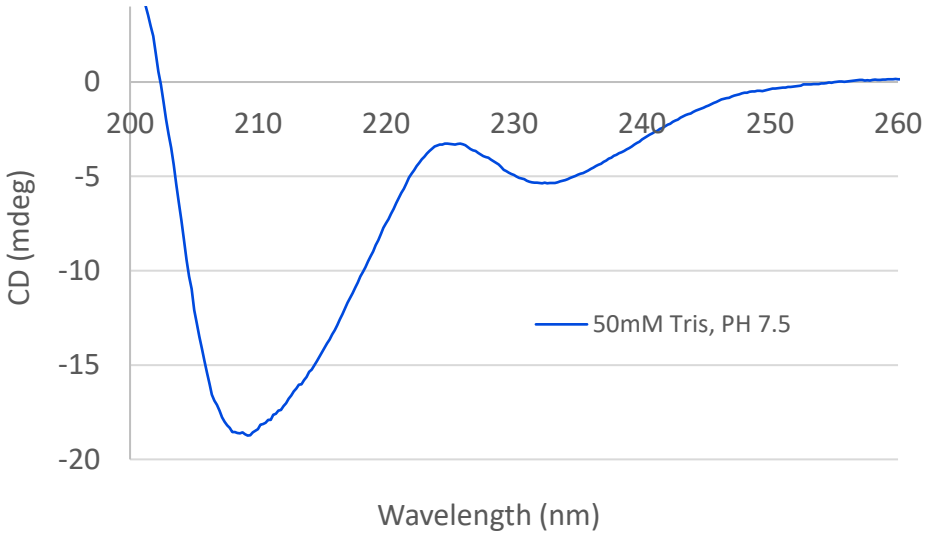


Figure 4.7. CD spectrum of Siglec8 Ig-like V-type domain in 50 mM Tris, pH 7.5 at a concentration of 1 mg/mL.

The spectra obtained for the Siglec8 Ig-like V-type domain is in good agreement with spectra belonging to a set of β-rich proteins described in the literature, the βI-proteins (Manavalan and Johnson, 1983). The characteristic spectrum for βI-proteins is well defined in this published work (Sreerama and Woody, 2003), where the CD spectra of 16 β-rich proteins were measured and they were classified in two different subsets: βI or βII-proteins. βII-proteins have a characteristic negative band around 200 nm and some of them show a small positive band around 190 nm while others display a positive band or a negative shoulder around 220 nm. In contrast, βI-proteins have a stronger positive band around 190 nm and a comparable negative band in the 210-220 nm region. Both subsets can also show CD bands above 225 nm due to aromatic and disulphide groups. In the same work, the crystal structures of those

proteins were analysed to determine the secondary structures that give rise to the basic features of the CD spectra of both subsets of β-proteins. The analysis pointed out that the relative compositions of β- and P2-helix (poly(Pro)II-helix) in β-rich proteins determine the type of a β-protein CD spectrum, as the βII-proteins generally have a larger fraction of residues in P2 structure than the βI ones, and the βI-proteins have a larger β-sheet content than the βII proteins.

Thus, the Siglec8 Ig-like V-type domain CD spectrum agrees well with spectra for βI-proteins, with the second negative peak around 230 nm probably accounted for its aromatic amino acids and disulphide bond. This is spectrum, shown in Figure 4.8, is also in excellent agreement with the CD spectrum (Propster, et al., 2015) and solution state NMR structure published for Siglec8 in 2016 (Propster, et al., 2016).

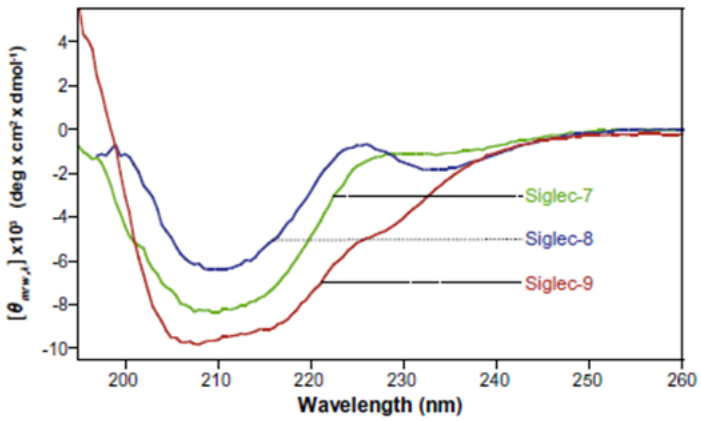


Figure 4.8. CD spectra for Siglec7 (green), Siglec8 (blue) and Siglec9 (red) published (Propster, et al., 2015).

4.6 Siglec8 Ig-like V-type domain folding monitored by means of Fluorescence.

This experiment aims to obtain information about how the unfolding process of the Siglec8 Ig-like V-type domains happens in presence of increasing concentrations of the denaturing agent Guanidine Hydrochloride (GdnHCl). The unfolding process of the protein Lysozyme from chicken egg white, which has already been characterized (Chen, et al., 2014) was reproduced and used as a model to compare.

There are three amino acids with intrinsic fluorescent properties: phenylalanine (Phe), tyrosine (Tyr) and tryptophan (Trp), but only tryptophan and tyrosine have high enough quantum yields (ratio of photons absorbed to photons emitted though fluorescence) to give a high fluorescence signal. Both amino acids can

be used to follow protein folding, as their fluorescent properties (quantum yields) are sensitive to the chemical environment which changes when the protein folds/unfolds as described in Figure 4.9 (Royer, 2006).

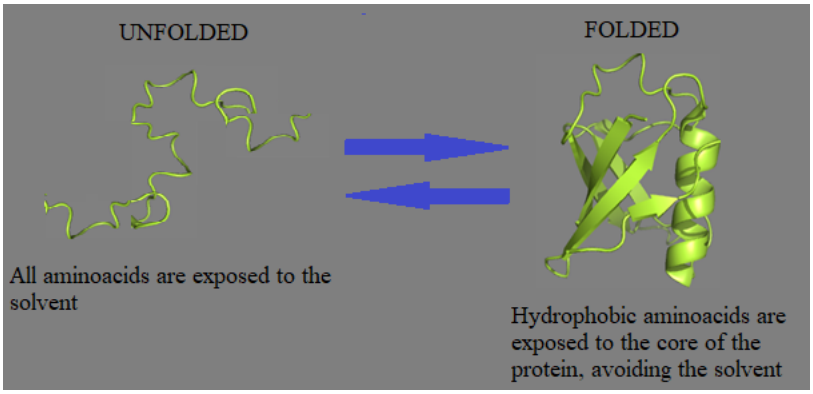


Figure 4.9. Example of the unfolded (left) versus folded (right) structure of a protein. For the unfolded form, all amino acids are exposed to the solvent, while when the protein is folded the hydrophobic amino acids are exposed to the protein's core avoiding the solvent.

At 280 nm both Trp and Tyr are excited, and at 295 nm tryptophan is selectively excited. In a folded protein, tryptophan and tyrosine are usually located within the core, surrounded by a hydrophobic environment where typically they have lower quantum yields and the wavelength of maximum emission intensity (λmax) is lower. In a partially folded or unfolded state, Trp and Tyr become exposed to the aqueous solvent, emit higher fluorescence intensity and λmax can be shifted to higher wavelengths for Trp (Royer, 2006). However, there are other phenomena that may affect the protein's fluorescence yield, such as the quenching that occurs when there are several aromatic rings in the protein core and there is an energy transfer between them, leading to a decrease of the fluorescence intensity for a folded protein in relation to its unfolded form (Kosinski-Collins, et al., 2004; Royer, 2006).

A good example of how the quenching phenomenon can affect the fluorescence emission of a protein is the work carried out with the β-sheet rich, human γD-crystallin protein which has four tryptophan residues (Kosinski-Collins, et al., 2004). In that study, four different mutants with three out of the four tryptophan residues substituted with phenylalanine were generated. Fluorescence spectra were measured for each of the four mutants as well as the wild type protein in the presence and absence of 5.5 M GdnHCl. The wild-type protein and two of the mutants were significantly more fluorescent in their unfolded state. However, two mutants emitted less fluorescence in their unfolded state than in the native (folded) state, when the

tryptophan was surrounded by a hydrophobic environment. Thus, in the second case, there were quenching phenomena decreasing the fluorescence of the native protein structure. There was also a typical red shift of the maximum emission wavelength after denaturing in all cases.

We monitored the unfolding process of both the standard protein Lysozyme and Siglec8 Ig-like V-type domains at a concentration of 1 mg/mL by measuring the Trp and Tyr fluorescence emission of samples with increasing concentrations of GdnHCl (from 0 M to 3.5 M for lysozyme and from 0 M to 2 M for the Siglec8 Ig-like V-type domains). Lysozyme contains three Tyr and six Trp residues and the Siglec8 Ig-like V-type domain nine Tyr and four Trp residues. After preparation, samples were left overnight to equilibrate and the following day their fluorescence emission spectra were measured between 320 and 380 nm (excitation wavelength 280 nm).

The spectra are shown in Figure 4.10 for Lysozyme and Figure 4.11 for the Siglec8 Ig-like V-type domain. In both cases, as the concentration of GdnHCl was increased, the protein domains unfolded, and the maximum intensity of emission became higher. This suggests the presence of the quenching between aromatic rings in the native proteins' fold which decreased the intensity of emission in comparison with the unfolded form, despite the hydrophobic environment surrounding the fluorescent aminoacids. Lysozyme spectra were very similar for all samples containing 0-2.0 M GdnHCl, indicating that the protein's native structure remained intact. However, with the addition of higher concentrations of the denaturing agent (2.5, 3.0 and 3.5 M GdnHCl), the fluorescence emission intensity increased dramatically. In contrast, the intensity of emission for the Siglec8 Ig-like V-type domain increased gradually with the addition of higher concentrations of GdnHCl, suggesting a step-wise unfolding process. Siglec8 spectra also showed the typical red shift of the maximum emission wavelength as the protein lost its native fold and fluorophores became exposed to the solvent.

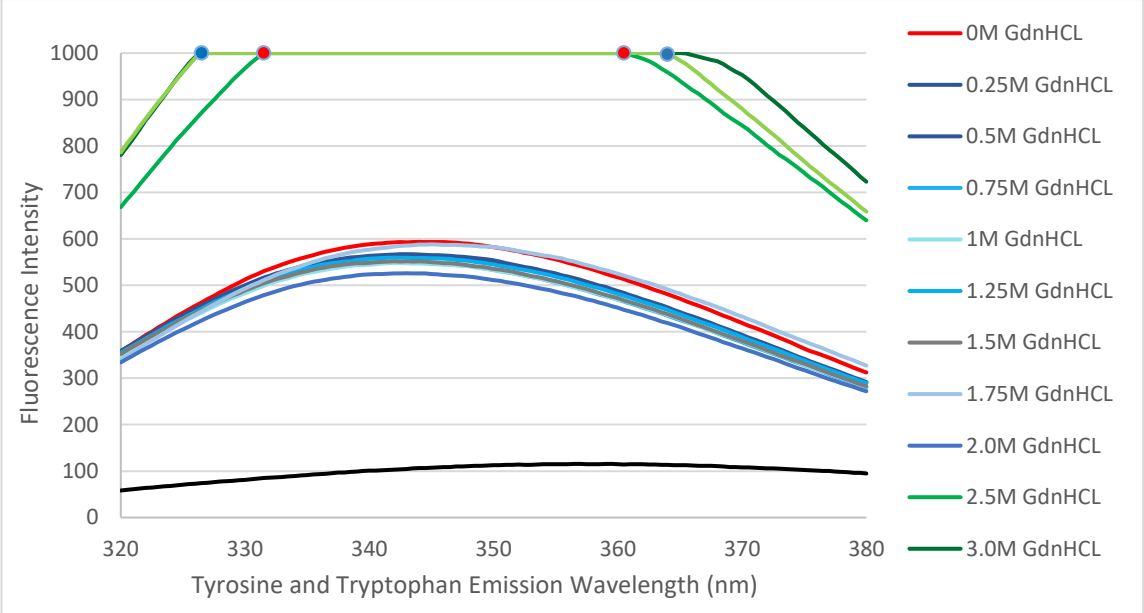


Figure 4.10. Fluorescence emission spectra for Lysozyme samples with increasing concentrations of denaturing agent GdnHCl (0, 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.5, 3.0 and 3.5 M) and a blank with buffer and 4 M GdnHCl. Samples were excited at 280 nm and the emission spectra were measured between 380 and 320 nm.

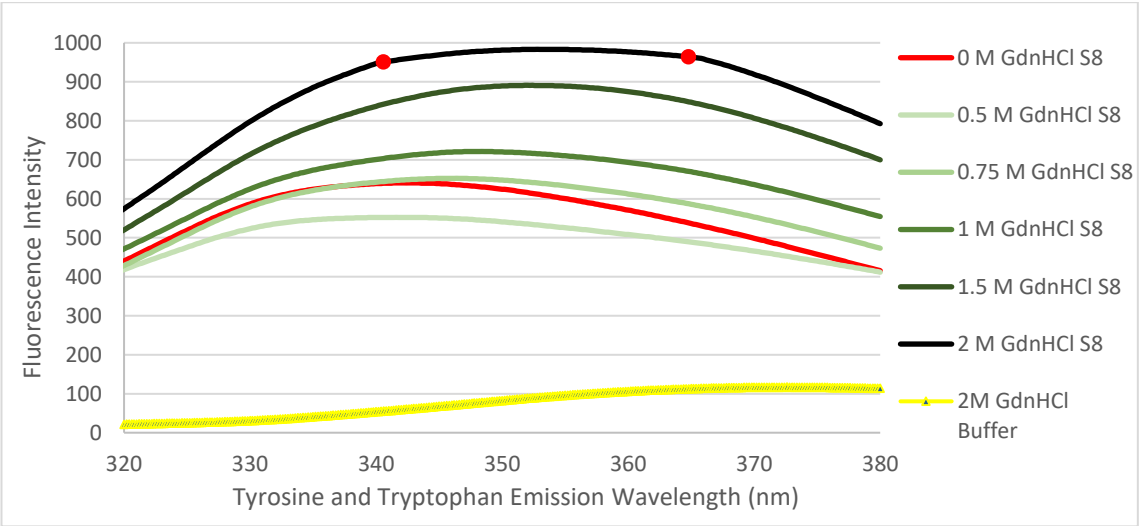


Figure 4.11. Fluorescence emission spectra for Siglec8 Ig-like V-type domain samples with increasing concentrations of denaturing agent GdnHCl (0, 0.5, 0.75, 1.0, 1.5 and 2.0 M) and a blank with buffer and 2M GdnHCl. Samples were excited at 280 nm and the emission spectra were measured between 380 and 320 nm.

Figure 4.12 and Figure 4.13 show a plot of fluorescence intensity and λmax versus GdnHCl concentration for Lysozyme and Siglec8 Ig-like V-type domains, respectively.

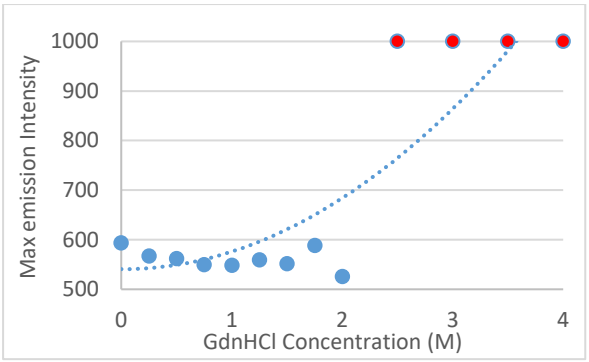


Figure 4.12. Maximum fluorescence intensity for Lysozyme samples with addition of GdnHCl from 0 to 3.5 M.

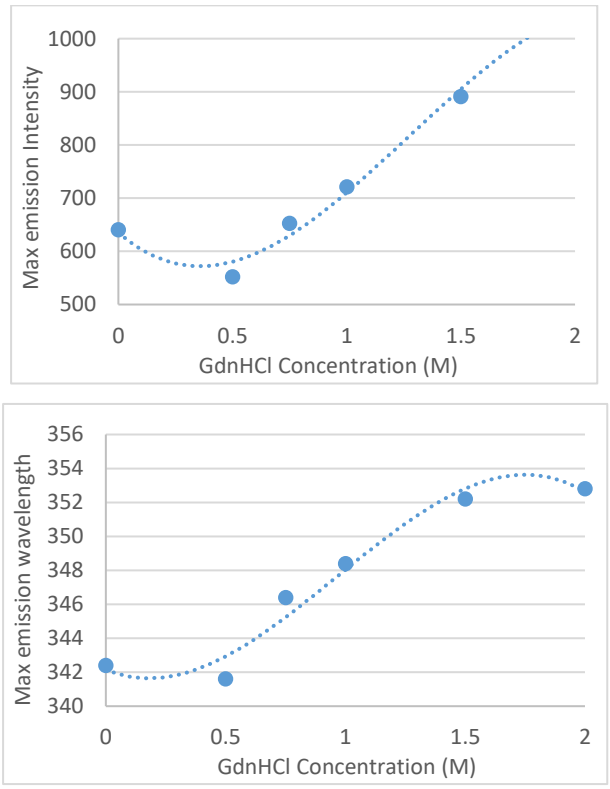


Figure 4.13. Maximum fluorescence intensity (left) and wavelength of maximum emission (right) for Siglec8 Ig-like V-type domain samples with addition of GdnHCl from 0 to 2M.

4.7 Siglec8 Ig-like V-type domain size characterization by Small Angle X-Ray Scattering (SAXS).

The average diameter of Siglec8 Ig-like V-type domain particles in 20mM NaPi 500mM NaCl, pH 7.5 buffer has been measured using SAXS. In a SAXS experiment a solution of particles is illuminated by a collimated monochromatic X-ray beam. The intensity of the scattered X-rays is recorded by an X-ray detector. Likewise, the scattering pattern of the pure solvent is also collected and subtracted from the sample solution. As molecules have different orientations in solution, the

obtained signal is related to the overall shape and size of the particles (a practical guide).

Figure 4.14 shows the SAXS plot of the logarithmic scattering intensity in arbitrary units versus the inverse of the size (q) in $Å^{-1}$. Data were fitted to calculate the average protein domain's size. Extreme values of q do not fit the model as they belong to a small fraction of protein domains which are forming aggregates in solution (lowest q values) and buffer background (highest q values), respectively.

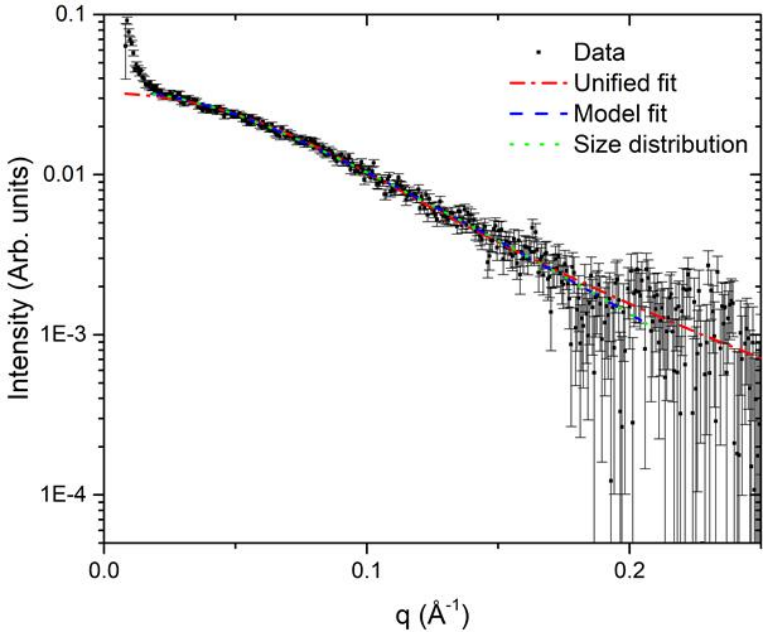


Figure 4.14. SAXS plot of Scattering Intensity versus the inverse of particle size (q) fitted.

Figure 4.15 displays the estimated radio according to a spherical model and Figure 4.16 the average value of the diameter of the Siglec8 Ig-like V-type domain particles in solution. Both calculated from the fitted data. The estimated radio for the protein domains supposing they had spherical shape was approximately 13 Å. However, as the protein domain structure has been previously resolved by means of solution state NMR, we know it does not have a spherical shape (Propster et al).

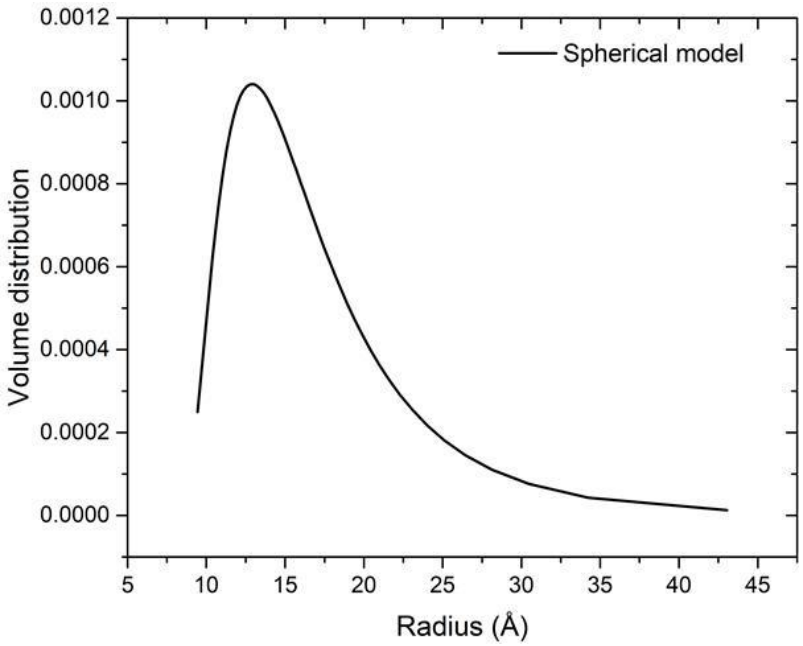


Figure 4.15. SAXS estimated radio for the Siglec8 Ig-like V-type domains according to a spherical model.

The calculated protein domain average diameter was approximately 32 Å (Figure 4.16). This value is in good agreement with the estimation of a 16.2 Å radius gyration (Rg), using the Siglec8 Ig-like V-type domain solution state resolved structure (pdb ID 2N7A) (Propster, et al., 2016) and the software Gromacs, see Figure 4.17.

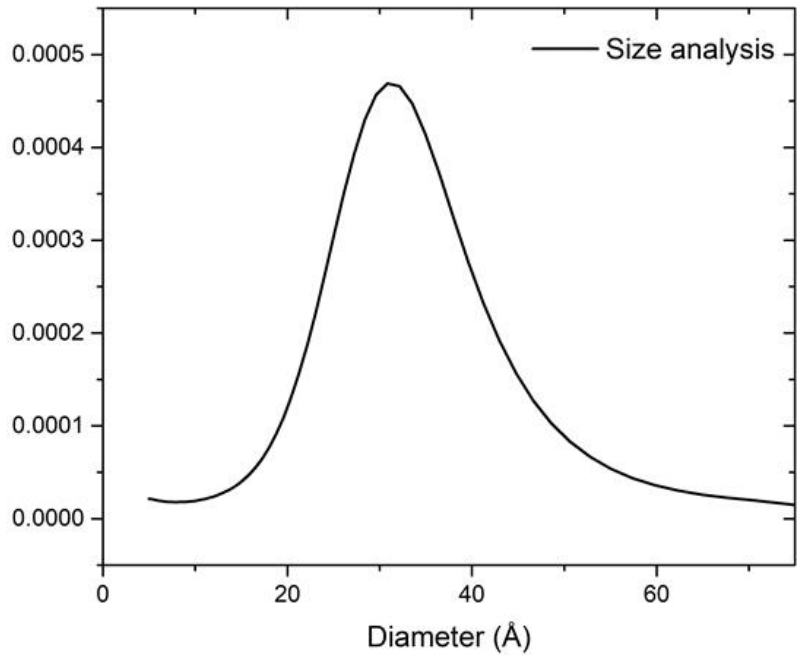


Figure 4.16. SAXS estimated average diameter for the Siglec8 Ig-like V-type domains.

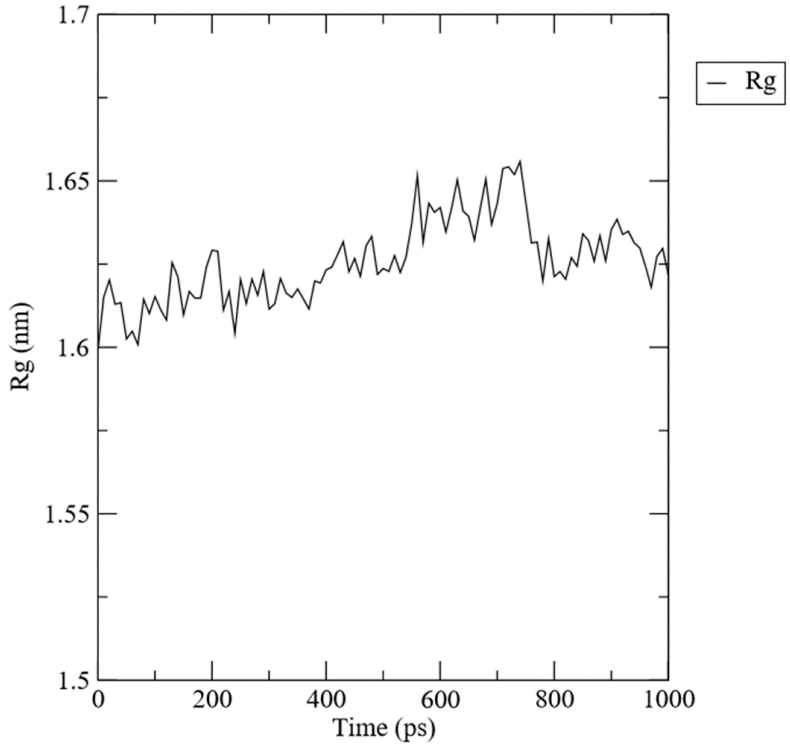


Figure 4.17. Estimation of the Siglec8 Ig-like V-type resolved structure's radius gyration during a 1 ns simulation using the software Gromacs and the pdb file 7N2A as input.

The 3D-structural model obtained from the SAXS data belonging to our soluble Siglec8 Ig-like V-type domains (above) in comparison with the Siglec8 Ig-like V-type domains solution state NMR resolved structure (below) (Propster, *et al.*, 2016) are shown in figure 4.18. Both domains encompass the same Siglec8 sequence but differ slightly in the initial and terminal sequences. The model indicates that we produced globular and folded Siglec8 carbohydrate binding domains which also share the same shape as the solution state NMR Siglec8 Ig-like V-type resolved domains downloaded from the Protein Data Bank. Both figures were visualised using the software Pymol (Pymol).

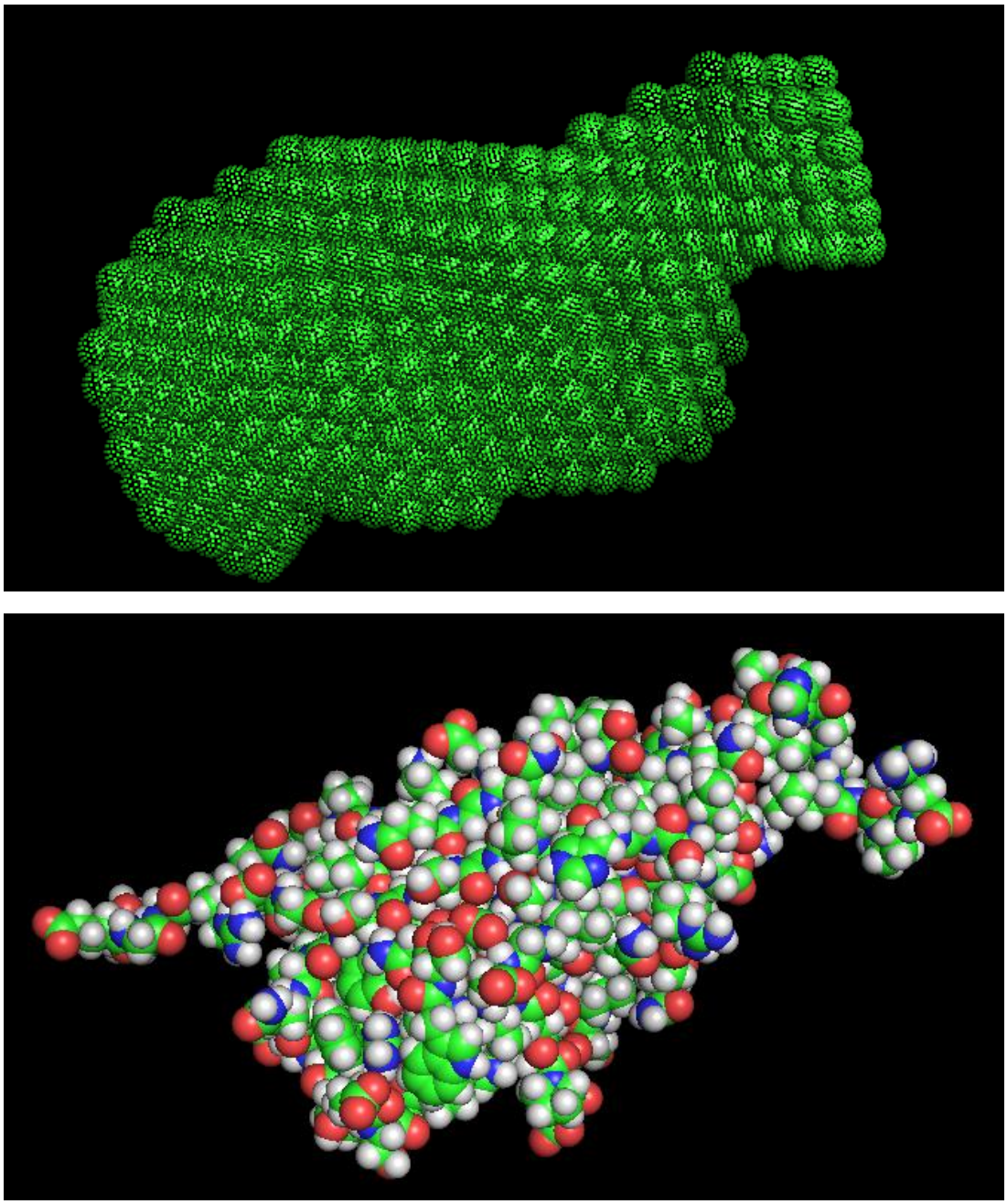


Figure 4.18. 3D-model of Siglec8 carbohydrate binding domains produced in this work by means of SAXS (above). Siglec8 carbohydrate binding domains resolved by means of solution state NMR (Propster, *et al.*, 2016) (below).

Chapter 5.

EXPLORING SIGLEC8 AS A THERAPEUTIC TARGET.

5. Exploring Siglec8 as a therapeutic target.

5.1. An overview of current asthma treatments

The treatments prescribed for asthma vary as a function of the severity of the disease. The most common type of therapeutic according to Asthma UK (AsthmaUK) is use of short-acting reliever inhalers. Inhalers such as Ventolin® are prescribed for every asthmatic person and must be taken when an exacerbation happens. They act locally and induce bronchial smooth muscle relaxation, opening the airways and facilitating the breathing. Daily preventer inhalers are also a common therapeutic approach. Small daily doses of steroids as Turbohaler® reduce patient's sensitivity to triggers, as they inhibit the immune system response locally. They are prescribed to patients who need to use the short-acting reliever inhaler three or more times a week. Long-acting reliever inhalers are prescribed in addition to the short-acting relievers and higher doses of daily preventer inhalers to those patients who still cannot keep their asthma symptoms under control. They induce bronchial smooth muscle relaxation, opening the airways and facilitating breathing. But their effect takes longer in beginning to work and lasts longer than the short-acting reliever inhalers' action. The inhaler Spiriva® Respimat® is an example. Add-on therapies are required when the other treatments are not enough to keep the disease under control. These include steroid tablets which help to reduce the inflammation during acute asthma attacks when the preventer inhaler is not enough, or leukotriene receptor antagonists that reduce the body's response to certain allergens and situations triggering asthma attacks.

There are cases of severe asthma in which the disease cannot be kept under control with the medicines mentioned above, and higher doses of daily preventer inhalers and steroid tablets are prescribed (AsthmaUK). Oral steroid tablets are usually administered during short occasional courses and come with possible side effects such as high blood pressure or lower resistance to infections. For this reason, they are only prescribed when the benefits outweigh the risks (AsthmaUK). Current research efforts in asthma therapies are focused on the development of new therapeutics that enable to keep the patients' severe asthma symptoms under control, minimizing emergency hospital admissions as well as side effects (AsthmaUK).

5.2 Role of eosinophils in asthma and targetability through Siglec8.

Eosinophils' enhanced survival and delayed apoptosis in tissues and circulation seem to be the cause of the accumulation of these cells in the lungs of asthmatic individuals (Simon, *et al.*, 1997; Kankaanranta, *et al.*, 2000). Biochemically, there is evidence of eosinophil survival-prolonging cytokines such as GM-CSF, IL-5, IL-3, and other life-supporting molecules being strong contributing factors towards initiation and maintenance of eosinophilic airway inflammation in asthma (Tai, *et al.*, 1991; Yamaguchi, *et al.*, 1991; Stern, *et al.*, 1992). At present, three drugs whose mode of action induces eosinophil apoptosis are being used to treat asthma: glucocorticoids, theophylline (both steroids) and leukotriene modifiers. Furthermore, potential novel anti-asthmatic drugs with eosinophilic apoptotic effects such as inhibitors of histone deacetylases (HDAC), antibodies targeting the IL-5 receptor alpha, cyclin-dependent kinase (CDK) inhibitors and Siglec8 activators are being investigated (Ilmarinen and Kankaanranta, 2014b).

There is evidence that the *in vitro* activation of Siglec8 induces apoptosis of eosinophils (Ilmarinen and Kankaanranta, 2014b). The mechanisms of regulation and signalling of this apoptotic effect have not been completely elucidated, but it is known that they may vary under different chemical environments (Ilmarinen and Kankaanranta, 2014b). Thus, in resting eosinophils, Siglec8-induced apoptosis is dependent on reactive oxygen species (ROS) and caspases, a family of endoproteases whose activation results in a cascade of signalling events that lead to the controlled demolition of eosinophils' cellular components, apoptosis (McIlwain, *et al.*, 2015). Conversely, eosinophils activated with the growth factors IL-5 or IL-33 appeared to be more susceptible to death via Siglec8 activation by means of necrosis or autolysis instead of apoptosis. Production and accumulation of reactive oxygen species (ROS), and activation of the MAPK/ERK pathway have been demonstrated to play an essential role in mediating Siglec8-induced cell death (Kano, *et al.*, 2014; Kano, *et al.*, 2017). The MAPK (mitogen-activated protein kinase) / ERK (extracellular receptor-stimulated kinase) pathway is made up of a collection of cytosolic and nuclear enzymes that communicate a signal from a receptor on the cell's surface to the DNA in the nucleus of the cell, involving phosphorylation of protein kinases belonging to the Src family (non-receptor tyrosine kinases) (Kano, *et al.*, 2017).

The exclusive expression of the Siglec8 receptor in human eosinophils and mast cells makes Siglec8 a good target to specifically inhibit or deplete those cellular types in eosinophil and mast cell-related diseases. The challenge is to selectively deliver a therapeutic agent (Kiwamoto, *et al.*, 2012b).

5.3. Rational drug design in the pharmaceutical industry

Structure-based or rational drug design is a method used by the pharmaceutical industry to find new active molecules based on the structural knowledge of a biological target. It involves the design of small molecules or peptides that are complementary in shape and charge to the target, which is normally a bio-molecule such a protein. The small molecules must specifically activate or inhibit the target's function, triggering a therapeutic effect (Todd, *et al.*, 2009). Figure 5.1 schematically shows a protein target and a series of different small molecules that are complementary to the target in charge and shape.

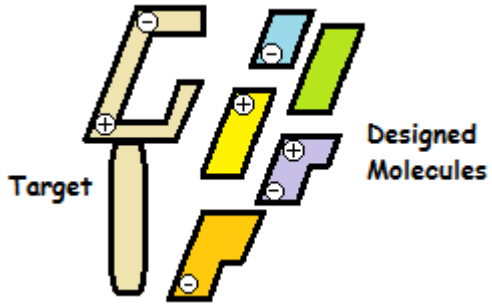


Figure 5.1. Schematic representation of a target and five designed small molecules with complementary shape and charge.

Rational drug design starts with the search for compounds that fulfil specific requirements such as size complementarity, charge complementarity or the addition of a donor, an acceptor or a hydrophobic ring in a specific position. Subsequently these molecules are synthesized and tested for activity. All activity assays will provide information about those chemical groups that are more likely to be active and the molecules can be redesigned in order to improve the activity over the target (Todd, *et al.*, 2009).

However, bringing a pharmaceutical drug to the market is a very long-term process whose current costs have been estimated at approximately 2.6 billion dollars. The probability of failure in the drug discovery and development pipeline is high, with 90% of the drugs tested in clinical trials failing to obtain FDA approval and get into the market (Leelananda and Lindert, 2016). At present, pharma and biotech companies

use computer-aided drug discovery (CADD) and high throughput screening (HTS) technologies to optimise and accelerate the drug discovery process (Leelananda and Lindert, 2016). HTS technologies enable the activity of thousands of molecules to be tested against biological targets utilising robotic automation, miniaturized assays and large-scale data analysis, but these technologies are expensive and require resources of targets and ligands (Mayr and Bojanic, 2009; Leelananda and Lindert, 2016). CADD technologies are powerful tools that can reduce the number of ligands that need to be screened in experimental assays. They not only reduce costs by ensuring that the best possible lead compound enters animal studies, but also decrease the time it takes for a drug to reach the market. CADD tools can identify lead drug molecules for testing, predict effectiveness and possible side effects, and assist in improving bioavailability of possible candidates (Clark, *et al.*, 2016). CADD methods can be generally classified into two groups: structure-based (SB) and ligand-based (LB). The method used depends on the availability of target structural information (Leelananda and Lindert, 2016).

Structure-based methods are based on the knowledge of the 3D structure of the biological target and its active site. Sometimes the presence of reactive groups in the active site is also known. Target information is usually obtained experimentally by X-ray crystallography or NMR spectroscopy, but computational methods that predict the three-dimensional structures of targets may be used as well. This structural information can be used to virtually evaluate the affinity of molecules to targets, as well as estimate their binding free energies. A final filtering and optimization lead to those molecules whose activity will be tested *in vitro* (Todd, *et al.*, 2009; Leelananda and Lindert, 2016).

When the target's structure has not been experimentally determined, ligand-based CADD methods are the alternative. They are based on the knowledge of active binders to the target. Thus, the features of those active small molecules can be compared with the inactive ones to generate a structure-activity hypothesis about which chemical groups (pharmacophores) are relevant to trigger biological activity (Leelananda and Lindert, 2016).

5.4 Rational drug design of molecules targeting Siglec8

In this work, several low molecular weight molecules targeting the Siglec8 Ig-like V-type domain were rationally designed using the high-resolution structure reported for the Siglec8 Ig-like V-type domain in complex with 6'-Sulpho-Sialyl-LewisX (Propster, *et al.*, 2016). A schematic of the amino acids in Siglec8 that interact with 6'-Sulpho-Sialyl-LewisX is shown in Figure 5.2 and summarised in Table 5.1, where the nature of the interactions is indicated (Propster, *et al.*, 2016).

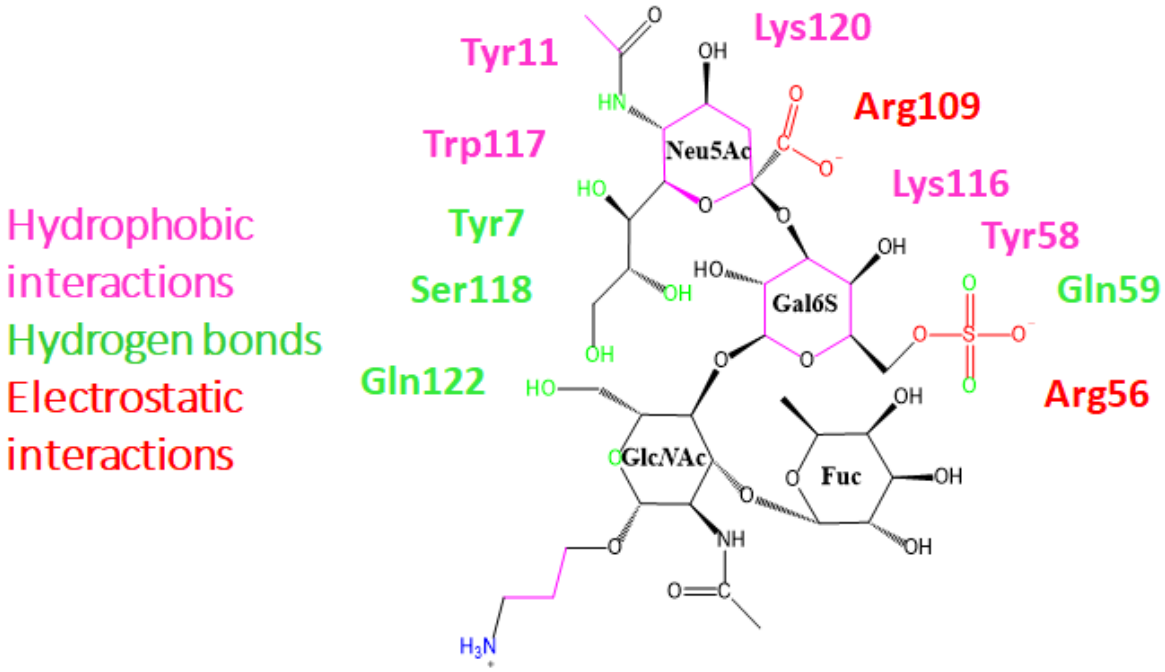
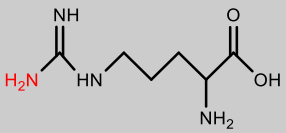
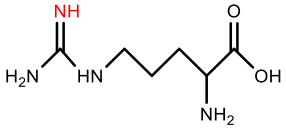
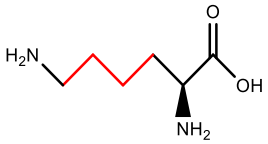
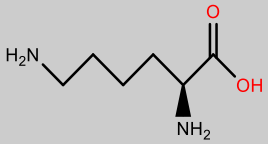
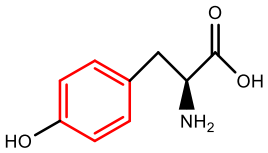
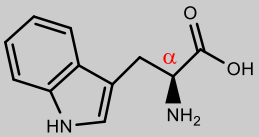
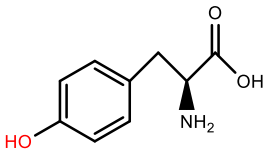
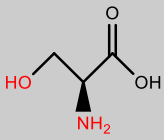
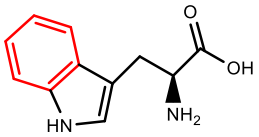
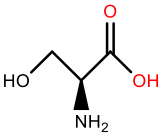
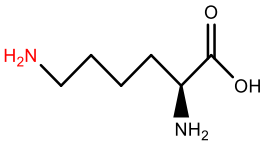
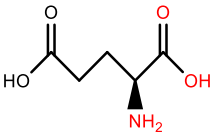
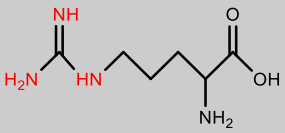
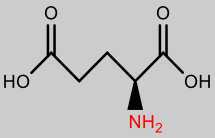
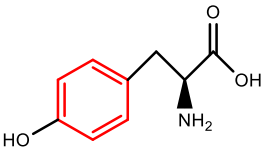
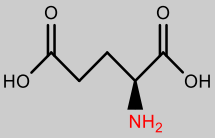
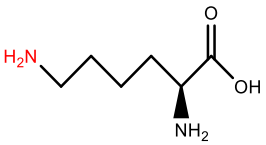
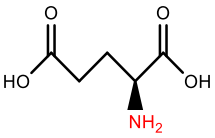


Figure 5.2. 6'-Sulpho-Sialyl-Lewis X and Siglec8 amino acids interacting with it. Amino acids and chemical groups carrying out hydrophobic interactions are represented in pink meanwhile the ones carrying out ionic interactions are shown in red and the ones forming hydrogen bonds in green (Propster, et al., 2016).

Table 5.1. Summary of the Siglec8 amino acids interacting with 6'Sulpho-Sialyl LewisX and the types of interactions (Propster, et al., 2016).

6'Sulpho-Sialyl LewisX chemical group	**Siglec8 amino acid**		**Chemical nature of the interaction**
Neu5Ac (COO^-) O1A/O1B	**Arg109 $N^{\eta 1}$**		Salt Bridge
Neu5Ac (COO^-) O1A/O1B	Arg109 $N^{\eta 2}$		Salt Bridge

6'Sulpho-Sialyl LewisX chemical group	Siglec8 amino acid		Chemical nature of the interaction
Neu5Ac H3/H4	**Lys116**		**Hydrophobic Contacts**
Neu5Ac N5	Lys116 O		**Hydrogen Bonds**
Neu5Ac CH_3	**Tyr11**		**Hydrophobic Contacts**
Neu5Ac H6	**Trp117 H^{α}**		**Hydrophobic Contacts**
Neu5Ac O7	**Tyr7 O^{γ}**		**Hydrogen Bonds**
Neu5Ac O8	**Ser118 N** Ser118 O^{γ}		**Hydrogen Bonds** Hydrogen Bonds
Neu5Ac H7, H91, H92	**Trp117**		**CH/π interactions**

6'Sulpho-Sialyl LewisX chemical group	Siglec8 amino acid		Chemical nature of the interaction
Neu5Ac O9	**Ser118 O**		**Hydrogen Bonds**
	Lys120 N^{ζ}		Hydrogen Bonds
	Gln $O^{\varepsilon 1}$		**Hydrogen Bonds** Hydrogen Bonds
	Gln $N^{\varepsilon 2}$		
Gal6S SO^{-}_{3}	**Arg56 $N^{\eta 1}$**		**Salt Bridge**
	Arg56 $N^{\eta 2}$		Salt Bridge
	Arg56 N^{ε}		Salt Bridge
	Gln59 $N^{\varepsilon 2}$		Hydrogen Bonds
Gal6S H3, H4, H5	**Tyr58**		**CH/π interactions**
GlcNAc O5	**Gln122 $N^{\varepsilon 2}$**		**Hydrogen Bonds**
GlcNAc O6	Lys120 N^{ζ}		**Hydrogen Bonds**
	Gln122 $N^{\varepsilon 2}$		Hydrogen Bonds

The model highlights the interactions of an essential arginine (Arg109) which is conserved in all Siglecs and plays an important role in their ligand binding by anchoring the carboxyl group of the terminal sialic acid inside the binding pocket. Mutations of this residue have been shown to result in loss of binding of most Siglecs, including Siglec8 (Propster, *et al.*, 2016). In addition, this work reveals the importance of both sulphate groups attached to the D-Galactose 6-Sulphate (Gal6S) and the N-Acetyl-D-Glucosamine (GlcNAc) as pharmacophores that enhance the affinity of the interaction. The dissociation constant of the 6,6'-Disulphated Sialyl-LewisX (K_D = 183±18 μM) was the lowest of all ligands tested in the study, followed by the 6'-Sulpho Sialyl-LewisX (K_D = 295±26 μM) whose sulphate group is attached to the Gal6S (see Figure 5.3). The 6-Sulpho Sialyl LewisX (sulphate attached to GlcNAc) showed lower affinity for the Siglec8 Ig-like V-type domain (K_D = 2.7±0.8 mM) and finally, Sialyl LewisX (K_D = 8.3±1.9 mM) had the lowest affinity (Propster, *et al.*, 2016). Therefore, the sulphate group attached to Gal6S plays a key role in the binding through its interaction with arginine 56 (Arg56), and the addition of another sulphate group to the GlcNAc enhanced the affinity of the ligand for the protein domain as represented in Figure 5.3. Furthermore, the model showed a lack of contacts between the α1,3-linked Fucose (Fuc) and Siglec8. However, the Fucose takes part stabilizing the LewisX trisaccharide conformation for binding with an inter-carbohydrate hydrogen bond between the Fucose C5-H5 and the O5 pyranose ring oxygen of Gal6S, and it may also contribute to enhance the affinity.

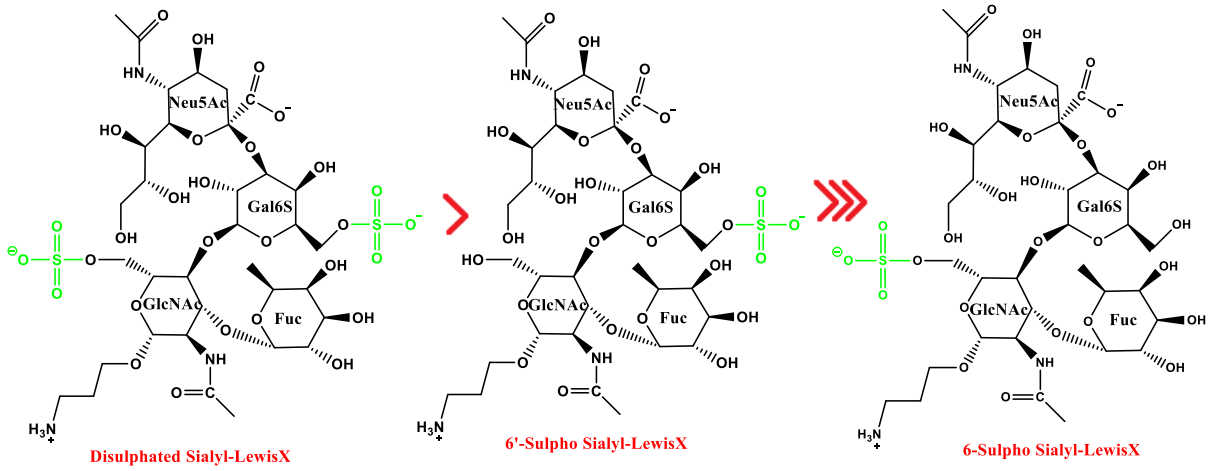


Figure 5.3. From left to right, structures of the Disulphated Sialyl LewisX which binds Siglec8 with stronger affinity than the monosulphated 6'-Sulpho-Sialyl LewisX and 6-Sulpho-Sialyl LewisX which did not show any binding to Siglec8 in the glycan microarrays (Campanero-Rhodes, et al., 2006).

Using the molecular model described above as a basis, we designed a library of six peptide-based and three glycan-based commercially available small molecules as possible Siglec8 Ig-like V-type domain binders. The approach taken for the glycan ligands was fragment based, encompassing the N-acetylneuraminic acid (NeuAc) and the D-Galactose 6-Sulphate (GalS6). The D-Glucuronic Acid (GluA) was also included due to its hexacyclic structure with a carboxylic group that provides it with a negative charge (Table 5.2). All peptide-based ligands included at least one negative charge and either carbohydrate chains or aromatic rings or a pentacyclic structure in an attempt to partially imitate the chemical environment of the large 6'-Sulpho Sialyl LewisX (Table 5.3). Ligands' structures and molecular weights are presented in Tables 5.2 and 5.3.

Table 5.2. Glycan based designed library attempting to imitate the chemical environment of the 6'-Sialyl LewisX ligand.

Glycan based library

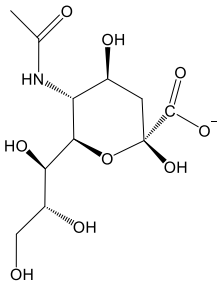
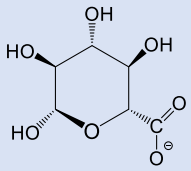
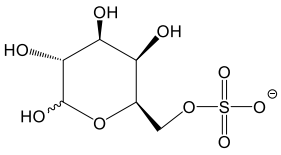
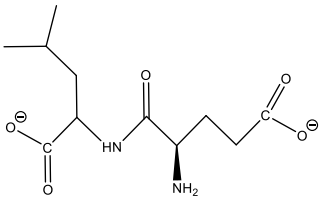
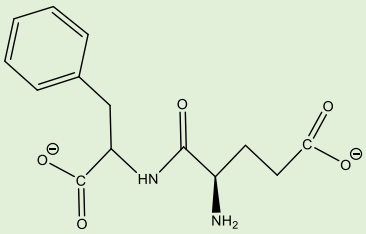
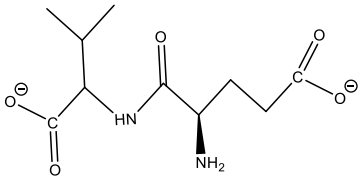
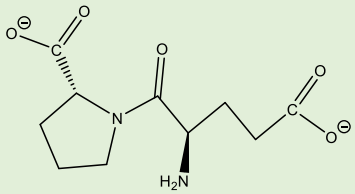
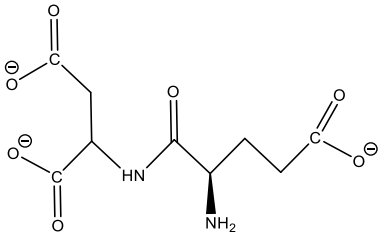
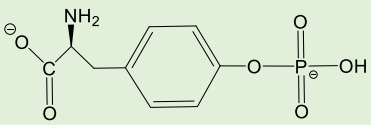
Glycan Compound	Structure	Molecular Weight (MW) (g/mol)
N-acetylneuraminic acid (NeuAc)		309.273
D-Glucuronic Acid (GluA)		194.139
D-Galactose 6-Sulphate (GalS6)		260.213

Table 5.3. Peptide based designed library attempting to imitate the chemical environment of the 6'-Sialyl LewisX ligand.

Peptide based library

Peptide	Structure	Molecular Weight (MW) (g/mol)
Glutamic Acid-Leucine (Glu-Leu)		260.29
Glutamic Acid-Phenylalanine (Glu-Phe)		294.3
Glutamic Acid-Valine (Glu-Val)		246.26
Glutamic Acid-Proline (Glu-Pro)		244.24
Glutamic Acid-Aspartic Acid (Glu-Asp)		262.22
O-Phospho-L-Tyrosine (P-Tyr)		261.17

5.5 Binding of molecular libraries designed to target Siglec8 using the Octet

The Octet system enables high throughput, real-time and label-free kinetic and affinity characterization of biomolecular interactions, including binding between proteins and small molecules (Fortebio). Here we analysed Siglec8-small molecule interactions. The experiment involved immobilization of the Siglec8 Ig-like V-type domain onto the surface of a sensor, followed by its immersion in a series of solutions (including buffer and ligand solutions), emulating the protein-ligand association and dissociation processes. Inside the instrument, a white light is emitted and passed through the biosensor tip. Subsequently, the detector analyses the interference pattern of white light reflected from two surfaces: the layer of immobilized protein on the biosensor tip and an internal reference layer. Thus, when a ligand is interacting with the protein, there is an increase in optical thickness at the biosensor tip, which results in a wavelength shift, $\Delta\lambda$ (Fortebio). Only molecules binding to or dissociating from the biosensor can shift the interference pattern and generate a response profile in the Octet System. By exposing the protein to a buffer solution containing ligand, and subsequently washing it with buffer, rates of association k_a (on-rate) and dissociation k_d (off-rate) as well as affinity constants K_D can be monitored (Fortebio). The basis of the Octet instrument's working principles is represented in Figure 5.4.

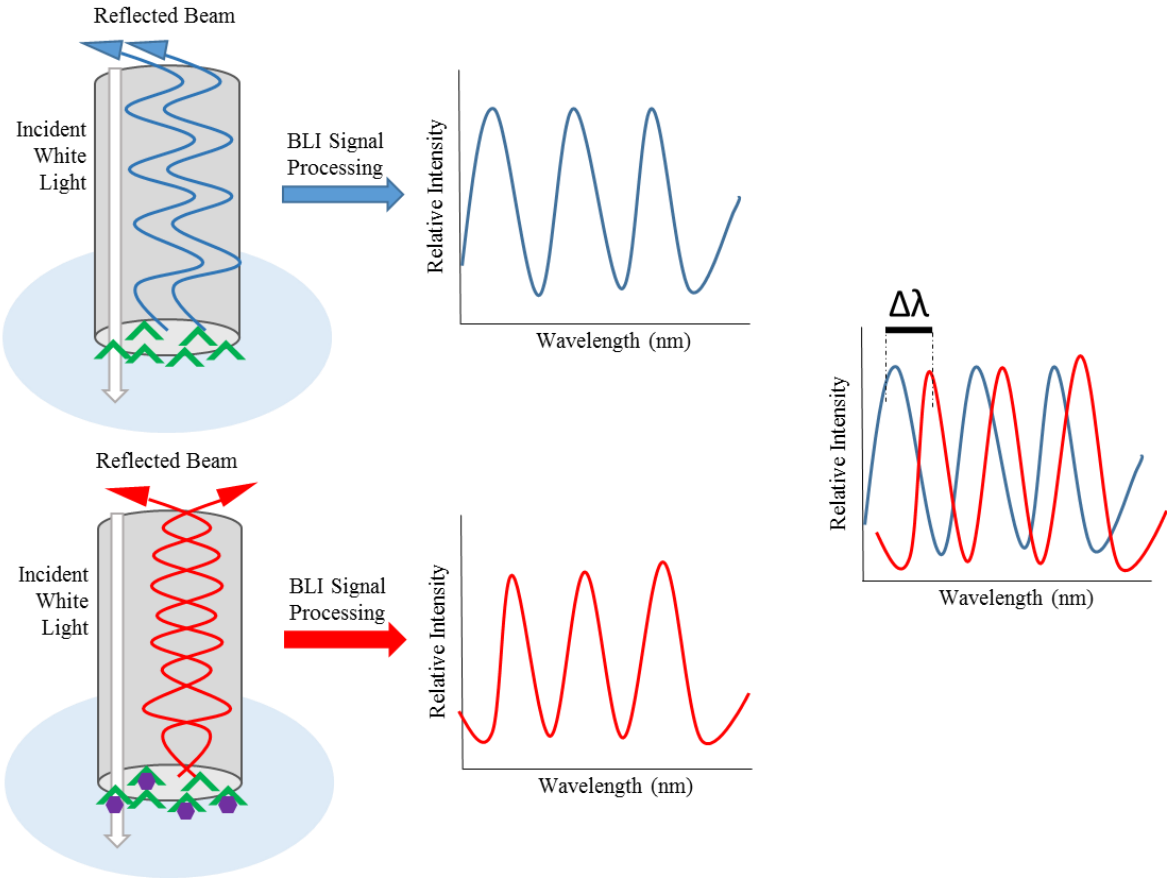


Figure 5.4. Representation of the Octet instrument working principle. Protein is immobilised in the biosensor surface (green). When there is not ligand attached (above), the incident white light is reflected according to a pattern which changes when there is ligand (purple) bound to the protein (below), leading to a shift in the phase of reflected light (Δλ) (Fortebio).

The initial binding screen was carried out for each of the nine ligands at a constant concentration of 10 mM. First, the Siglec8 Ig-like V-type domains were biotinylated, by covalent attachment of biotin to the protein through its free primary amine groups (N-terminus or lysines side chain epsilon amines). This does not usually perturb the protein's natural function, due to the small size of biotin (MW = 244.31 g/mol). A lysine biotinylation reaction is shown in Figure 5.5 below.

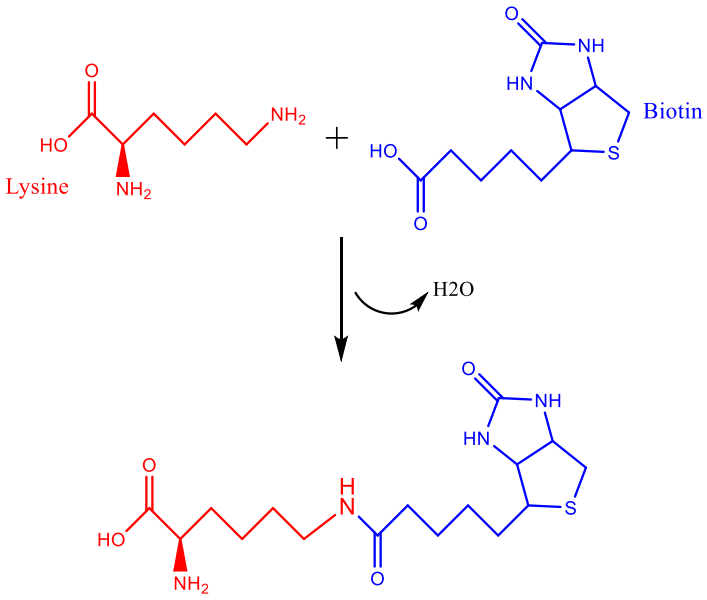


Figure 5.5. Protein biotinylation reaction. Condensation between a lysine's epsilon side chain amine and biotin with loss of a water molecule.

The distribution of seven lysine residues across the primary sequence of the Siglec8 Ig-like V-type domain used in this work is shown below (see circled residues in Figure 5.6). The biotynilation reaction was set up to achieve a yield of one biotin molecule reacting with either a random lysine or the N-terminal amine per protein domain, leading to a random distribution of eight different orientations when the protein domain is attached to the Octet sensor.

Figure 5.6. Distribution of Lysine amino acids across the Siglec8 Ig-like V-type domain sequence.

After biotinylating, the Octet experiment was set up as shown in Figure 5.7. Eleven sensors (one per ligand plus a reference without protein and a reference without ligand) were sequentially immersed into five different wells from left to right, following the five steps described below:

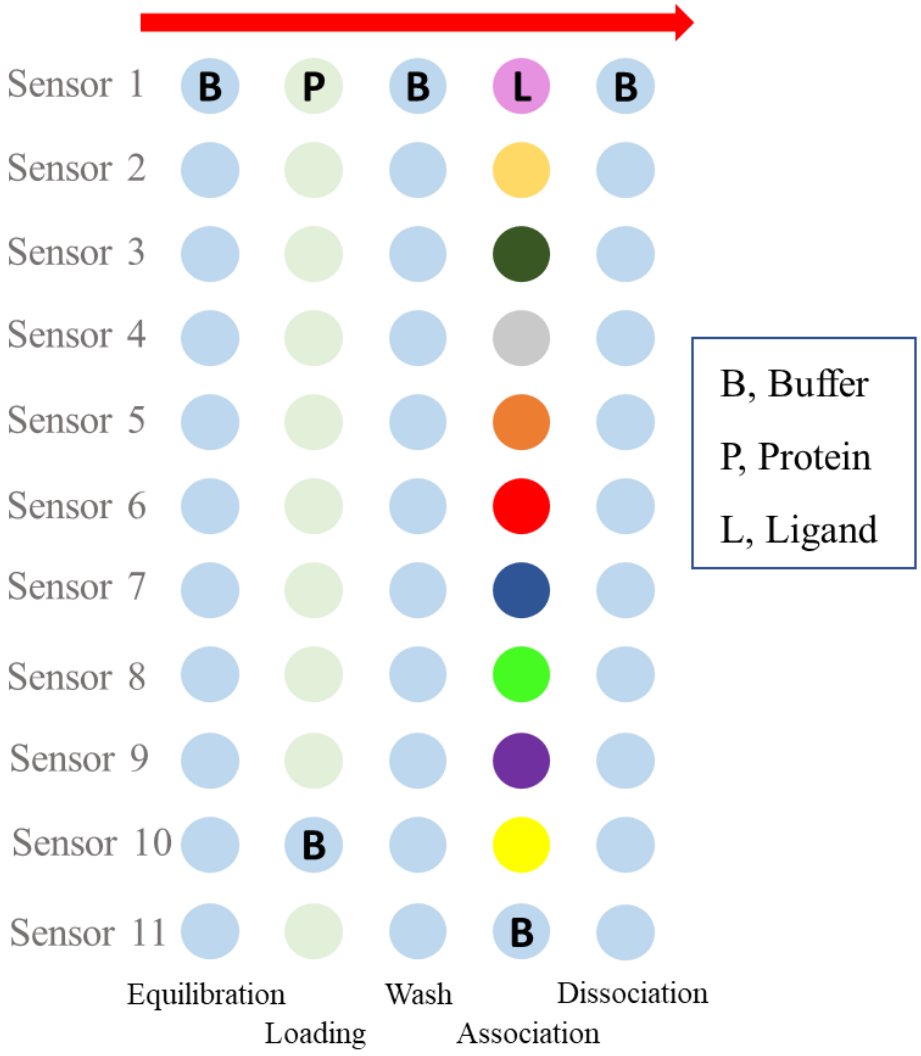


Figure 5.7. Representation of the Siglec8 ligand binding screening Octet experimental set up.

1. Equilibration.

   A sensor coated with the protein Streptavidin is immersed in 20 mM NaPi, 500 mM NaCl, pH 7.0 buffer to establish a constant chemical environment.

2. Protein loading.

   The sensor is then immersed in a solution of 20 µM biotinylated Siglec8 Ig-like V-type domain in matching buffer. Biotin binds to Streptavidin with an extremely high affinity, fast on-rate, and high specificity. Furthermore, biotin binding to streptavidin is resistant to heat, pH and proteolysis, making the capture of

biotinylated molecules possible in a wide variety of environments. Thus, the sensor surface was covered with Siglec8 Ig-like V-type domains randomly orientated depending on the site of biotin attachment (i.e. attachment to one of six lysines or to the N-terminus). In this step, one well contained no protein as a negative control, so one of the sensors was not loaded with protein.

3. Wash.

   The sensor is dipped in the matching buffer to get rid of any excess of protein.

4. Association.

   Each sensor is immersed in a specific concentration of ligand solution in buffer, enabling protein-ligand association. For the reference experiment, one of the sensors loaded with protein was immersed in buffer, as a second negative control.

5. Dissociation.

   The sensor is dipped in buffer again enabling the ligand dissociation process.

The reflected light signal is collected at each of the five steps for each sensor. Figure 5.8 displays idealised raw data for the binding of an immobilised protein and two different ligands (graphic on the left) and the data after subtracting the reference experiments (graphic on the right). The processed data give information about the protein-ligand association and dissociation rates as well as the magnitude of the interaction. Thus, for the same concentration of ligands 1 and 2, the association to the protein occurs more rapidly for ligand 1 (red curve) than for ligand 2 (green curve) ($k_{a1} > k_{a2}$), but ligand 2 dissociates more slowly from the protein than ligand 1 ($k_{d1} > k_{d2}$). Ligand 1 also produces a higher response than ligand 2, meaning that more molecules of ligand 1 bind to the protein.

However, affinity information cannot be obtained until more experiments are carried out for each ligand at different concentrations, and all data are fitted according to the suitable model. That will provide a dissociation constant ($K_D=k_d/k_a$) for each ligand. The comparison of association and dissociation rates, the magnitude of the binding response as well as the dissociation constants between ligands is relevant to select the best candidates.

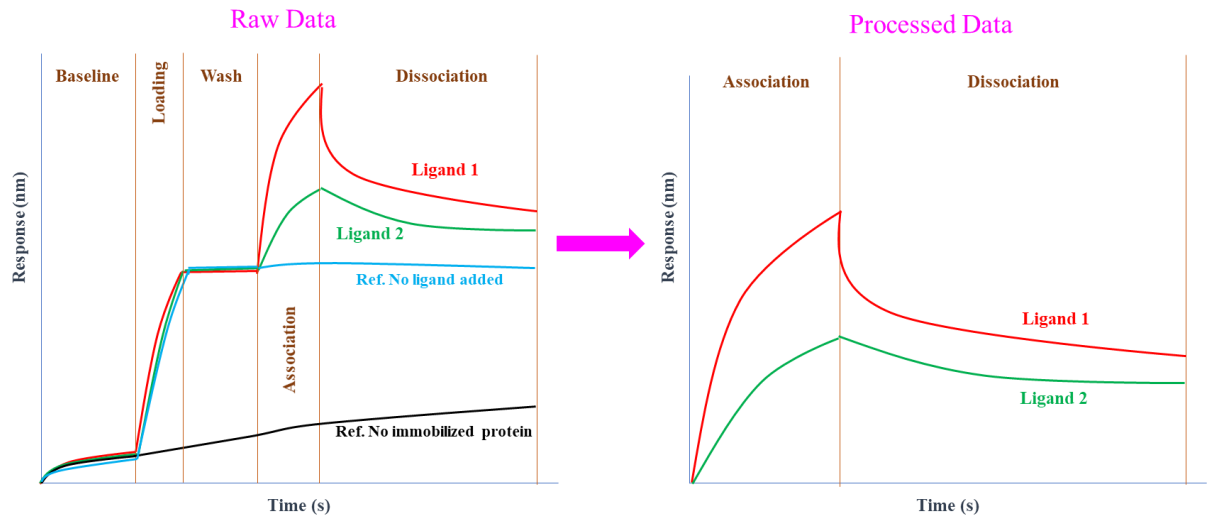


Figure 5.8. Hypothetical Octet data for the binding of two different ligands to an immobilised protein (ligand 1 in red and ligand 2 in green). Raw data are shown in the graphic on the left, while the processed data after the subtraction of references are represented in the graphic on the right.

The first Siglec8 ligand binding screen was performed at a constant ligand concentration of 10 mM for each one of the nine ligand molecules. Figures 5.9 and 5.10 display the processed Octet data for this Siglec8 binding screen for all ligands. Curves are labelled with the name of each ligand. In Figure 5.10, there are also curves belonging to experiments performed for the ligand L-Phosphotyrosine at three additional concentrations (5, 2.5 and 1.25 mM).

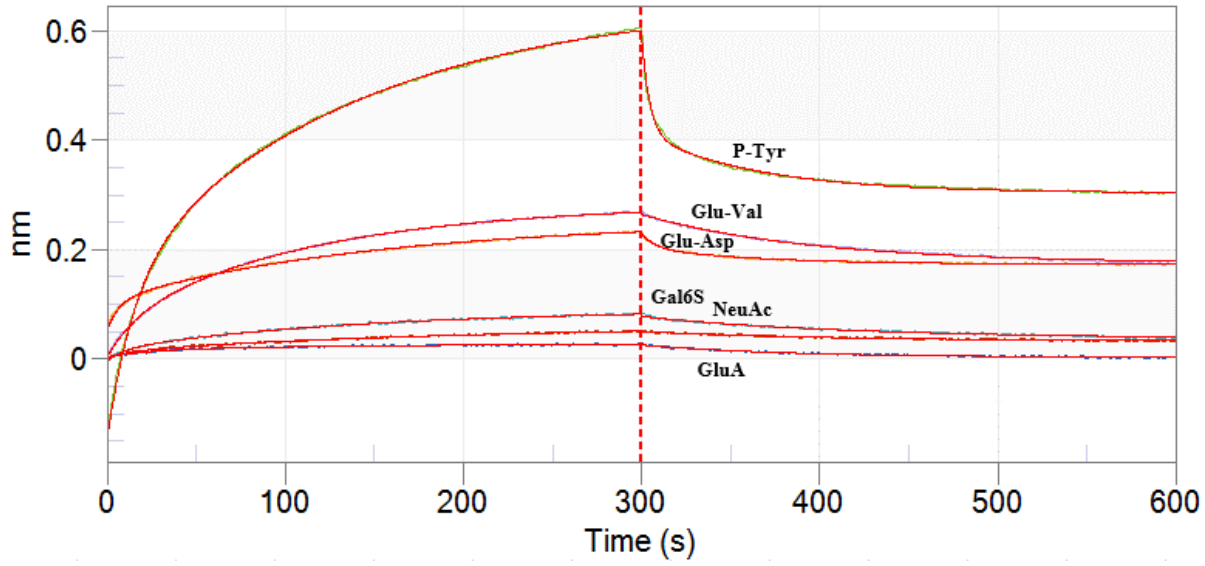


Figure 5.9. Six Octet Siglec8 binding experiments performed with a 10 mM concentration of Phospho L-Tyrosine, Glu-Val, Glu-Asp, D-Galactose-6-Sulphate, N-Acetyl Neuraminic Acid and Glucuronic Acid.

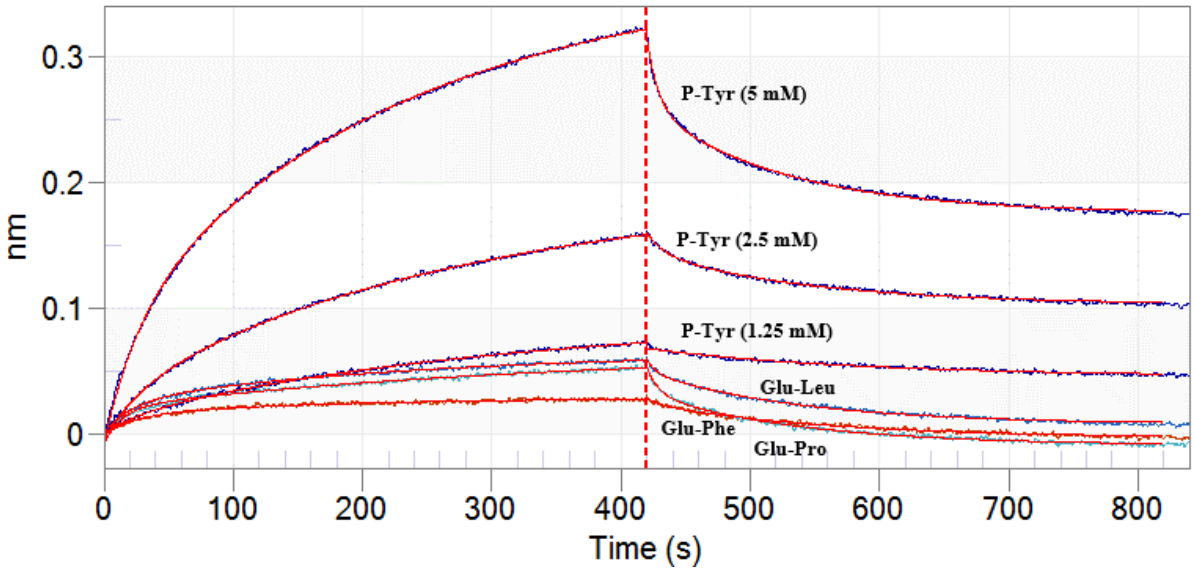


Figure 5.10. Three Octet Siglec8 binding experiments performed with a 10 mM concentration of Glu-Leu, Glu-Phe and Glu-Pro. Three extra experiments with 5, 2.5 and 1.25 mM Phospho-L-Tyrosine are also included.

L-Phospho-Tyrosine showed by far the highest Siglec8 binding response at a concentration of 10 mM (0.602 nm), followed by Glu-Val (0.269 nm), Glu-Asp (0.233), D-Gal-6-O-Sulphate (0.0827) and Glu-Leu (0.056).

Further Siglec8 Octet binding experiments were performed with the four ligands that gave the highest binding response (L-Phospho-Tyrosine, Glu-Asp, Glu-Val and Glu-Leu) at different concentrations in the millimolar range. The aim was to obtain information about affinity, association and dissociation rates. Table 5.4 includes all concentrations tested for each ligand.

Table 5.4. Different concentrations of each ligand tested with the Octet to estimate their affinity for Siglec8.

Ligand	L-Phospho-Tyrosine	Glu-Asp	Glu-Val	Glu-Leu
Concentrations (mM)	20	30	22.5	25
	8	15	18	12.5
	4	6	9	7
	0.4	3	0.9	5
	0.2	0.3	0.45	2.5
	0.04	0.15	0.09	0.5
		0.03		

Figures 5.11 to 5.14 display the Octet ligand binding curves for L-Phospho-Tyrosine, Glu-Asp, Glu-Val and Glu-Leu, respectively. As before, L-Phospho-Tyrosine was the ligand demonstrating the highest response, followed by Glu-Leu and

Glu-Asp. The results for Glu-Val showed a very low binding response, inconsistent with the first 10 mM screening.

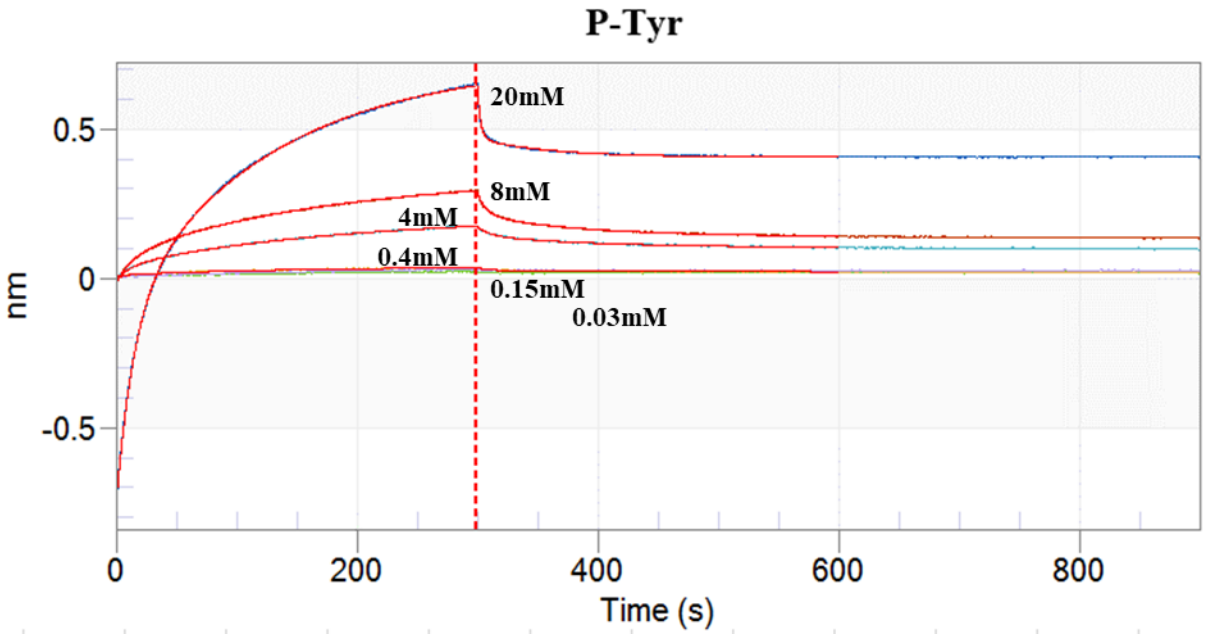


Figure 5.11. Octet experiment to test the binding between Siglec8 and L-Phospho-Tyrosine at different ligand concentrations.

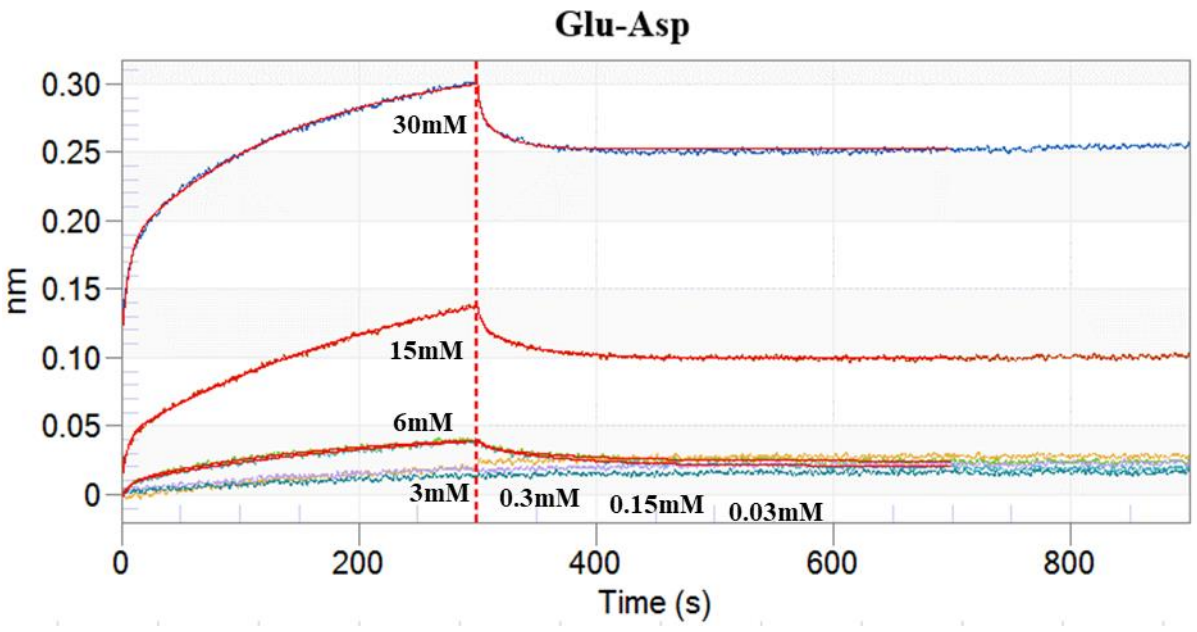


Figure 5.12. Octet experiment to test the binding between Siglec8 and Glu-Asp at different ligand concentrations.

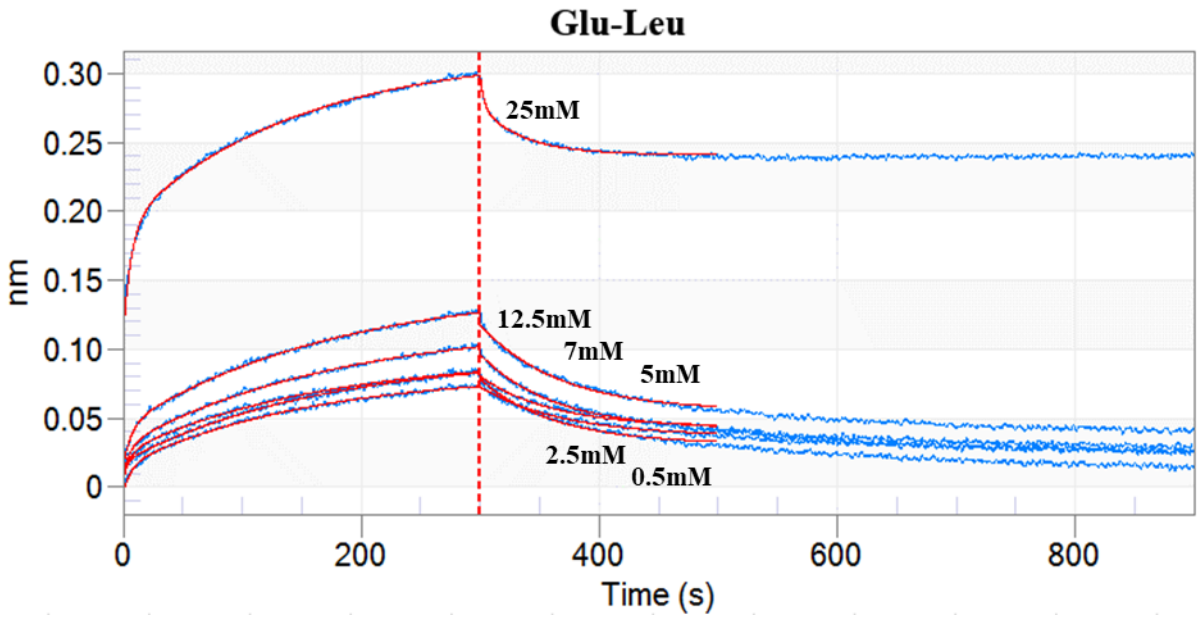


Figure 5.13. Octet experiment to test the binding between Siglec8 and Glu-Leu at different ligand concentrations.

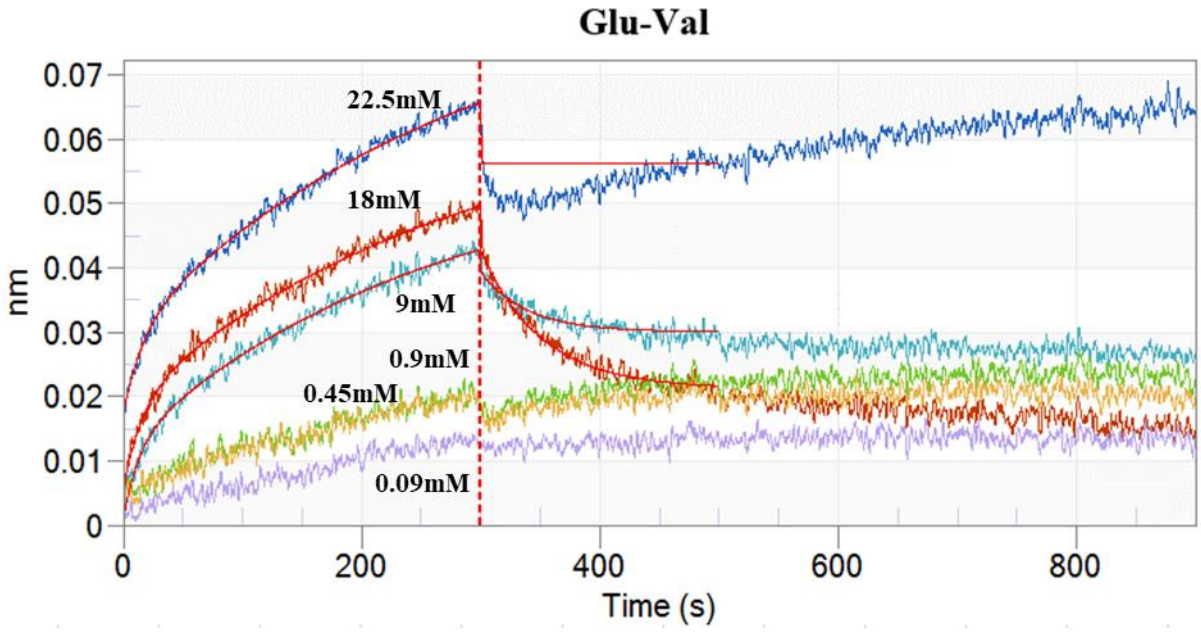


Figure 5.14. Octet experiment to test the binding between Siglec8 and Glu-Val at different ligand concentrations.

After trying different models, the best fitting results were obtained using an heterogeneous (n ligand molecules : 1 protein domain), local and partial model. Heterogeneous modelling is in good agreement with the multivalent ligand binding of other Siglecs described in the literature (Attrill, *et al.*, 2006a). The term local means that every concentration was fitted independently for each ligand, and the term partial indicates that association and dissociation were considered separately. L-Phospho-Tyrosine showed the highest Siglec8 binding response above 4 mM and dissociated

slowly from the protein domain. Glu-Asp bound Siglec8 at concentrations above 6 mM and dissociated slowly from the protein domain. However, the binding response was lower than for L-Phospho-Tyrosine. Glu-Leu bound Siglec8 at concentrations above 0.5 mM and dissociated faster than P-Tyr and Glu-Asp from the protein domain and its binding response was lower than for L-Phospho-Tyrosine too.

5.6 Equilibrium dialysis combined with 1D H^+ NMR to monitor Siglec8-Phospho-Tyrosine binding interactions.

Equilibrium dialysis enables the estimation of protein-ligand binding affinity constants in combination with any suitable ligand quantifying technique (Tipping, *et al.*, 2015). Each equilibrium dialysis device has two compartments separated by a dialysis membrane with smaller pore size than the protein but larger than the ligand. Thus, ligand and salts can diffuse between both chambers, but the protein will remain in the compartment where it was initially put.

The basics of the equilibrium dialysis experiment are illustrated in Figure 5.15. For each tested ligand concentration, two equilibrium dialysis devices were used: a control without Siglec8, and a sample with a fixed Siglec8 concentration in 20 mM NaPi, 500 mM NaCl, pH 7.5 buffer.

In the control device, one chamber was filled with 50 μL of the same buffer, and the other with 50 μL of a solution with a specific Phospho-Tyrosine concentration in buffer. In the sample device, one chamber was filled with 50 μL of a 50 μM concentration solution of Siglec8 domains in buffer, and the other with 50 μL of the same Phospho-Tyrosine solution added to the control device. Thus, in both devices the ligand diffused through the dialysis membrane until it reached an equilibrium concentration. After 2 hours of gently shaking, the chamber with the ligand solution was opened, and the 50 μL recovered from both devices.

Subsequently, the concentration of ligand was measured using 1D H^+ NMR and fluorescence when possible. When there is protein-ligand binding, the concentration of free ligand in the control should be higher than in the sample as some of the ligand would be bound to the protein, shifting the equilibrium and making more ligand to diffuse through the dialysis membrane (Figure 5.15c). Contrarily, in absence of interaction, for both control and sample devices, the same amount of ligand would diffuse through the dialysis membrane leading to an equal ligand concentration in control and sample (Figure 5.15b).

The same experiment with a fixed concentration of 50 µM Siglec8 was repeated at different ligand concentrations to build binding curves for some of the ligands.

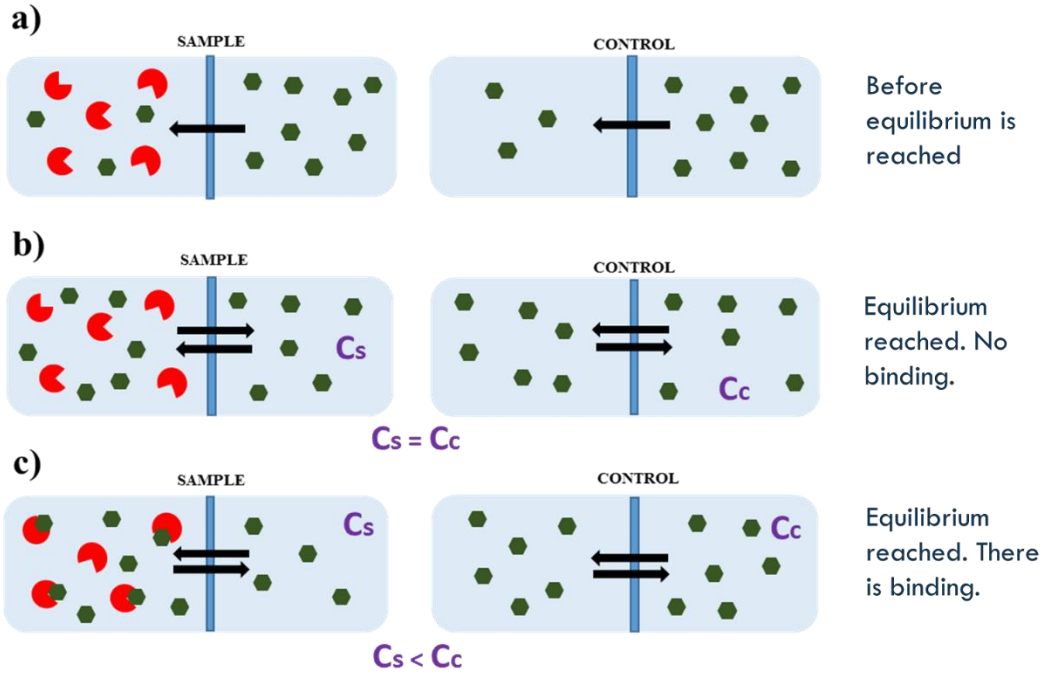


Figure 5.15. Representation for one equilibrium dialysis experiment at one ligand concentration. The picture illustrates both the sample and control devices in different situations: a) Before the equilibrium is reached. b) After equilibrium was reached but there was no protein-ligand interaction. c) After equilibrium was reached but there was protein-ligand interaction.

Estimation of the L-Phospho-Tyrosine fraction bound to Siglec 8 by means of 1D ^{1}H NMR.

An equilibrium dialysis experiment was set up with a constant Siglec8 concentration (50 µM) and different Phospho-Tyrosine concentrations as described above. The aim of this experiment was to estimate a KD value for the interaction between Siglec 8 and this ligand using 1D ^{1}H NMR as a quantitative technique for measurement of ligand concentration. Table 5.5 displays the concentration values of L-Phospho-Tyrosine added to the ligand chamber for all equilibrium dialysis samples:

Table 5.5. Equilibrium dialysis concentration values of Phospho-Tyrosine

C P-Tyr Samples (µM)	C P-Tyr Controls (µM)
50	50
100	100
500	500

After the six equilibrium dialysis chambers reached equilibrium, each chamber with ligand solution was opened, and 50 µL of the resulting ligand solution were recovered from all devices. Samples were then diluted with 85 µL of HPLC quality H_2O and 15 µL of D_2O (10% v/v) to reach a volume of 150 µL for NMR measurements.

All solution state NMR measurements were performed on the 700 MHz Spectrometer using identical acquisition parameters for every sample in order to allow direct comparison of peak integrals in the spectra. An additional sample of 50 µL of 100 µM DDS (4, 4-dimethyl-4-silapentane-1-sulfonic acid) diluted with 85 µL of HPLC quality H_2O and 15 µL of D2O (10% v/v) was prepared and measured in the same way as an external standard of accurately known concentration. The structure of this compound is displayed in Figure 5.16.

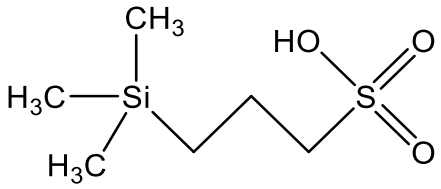


Figure 5.16. Structure of the NMR standard DSS (4, 4-dimethyl-4-silapentane-1-sulfonic acid).

Figure 5.17 shows an assigned L-Phospho-Tyrosine 1D ^1H NMR spectrum where the numbers indicate which peak belongs to each proton or pair of equivalent protons in the molecule. The intensity of an NMR peak is proportional to the number of equivalent protons that give rise to that signal as well as the sample concentration. Therefore, for a specific peak corresponding to two samples with different concentrations of the same ligand, the intensity increments are proportional to the concentration difference between them. The peaks 2 and 3 belonging to the aromatic ring protons were used to compare the intensities of each control-sample pair ($\Delta I = I_{control} - I_{sample}$). Subsequently, the concentration increments ($\Delta C = C_{control} - C_{sample}$) were calculated using the DSS's methyl proton peak at 0 ppm. Since DSS is of accurately known concentration, it can be used as an external reference by normalising all intensities to a one proton intensity value. The spectrum belonging to the DSS sample used as the external standard is shown in Figure 5.18.

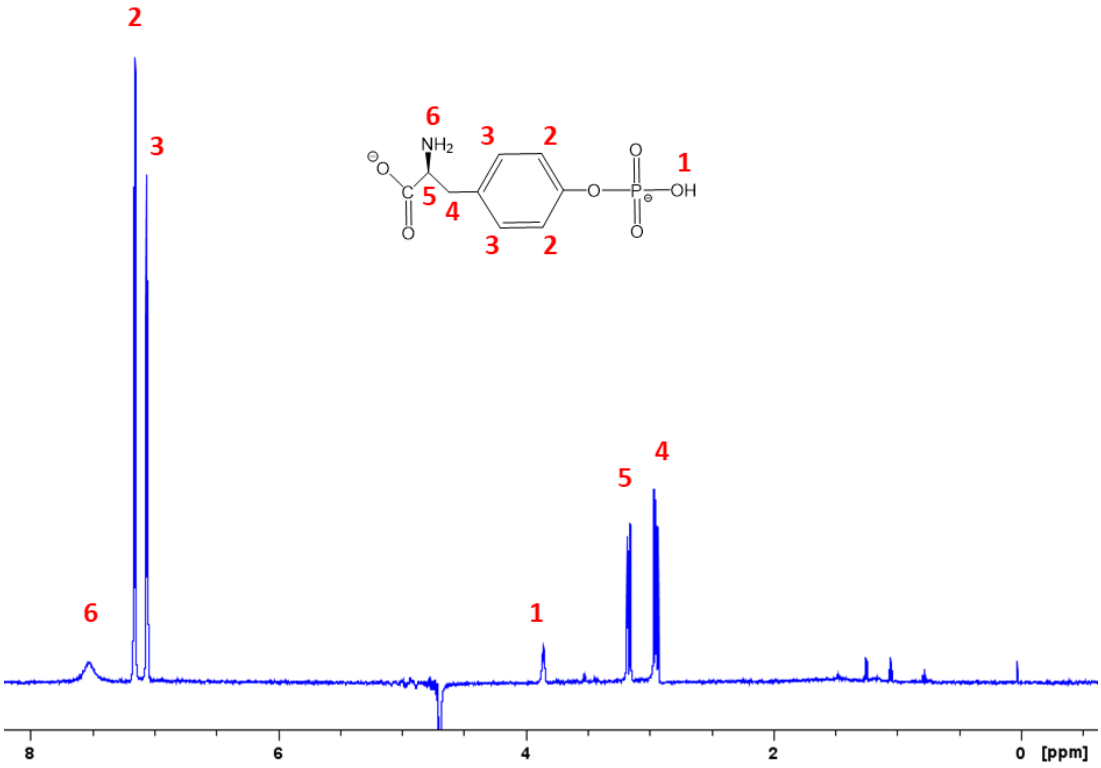


Figure 5.17. L-Phospho-Tyrosine 1D H^+ solution state NMR assigned spectrum.

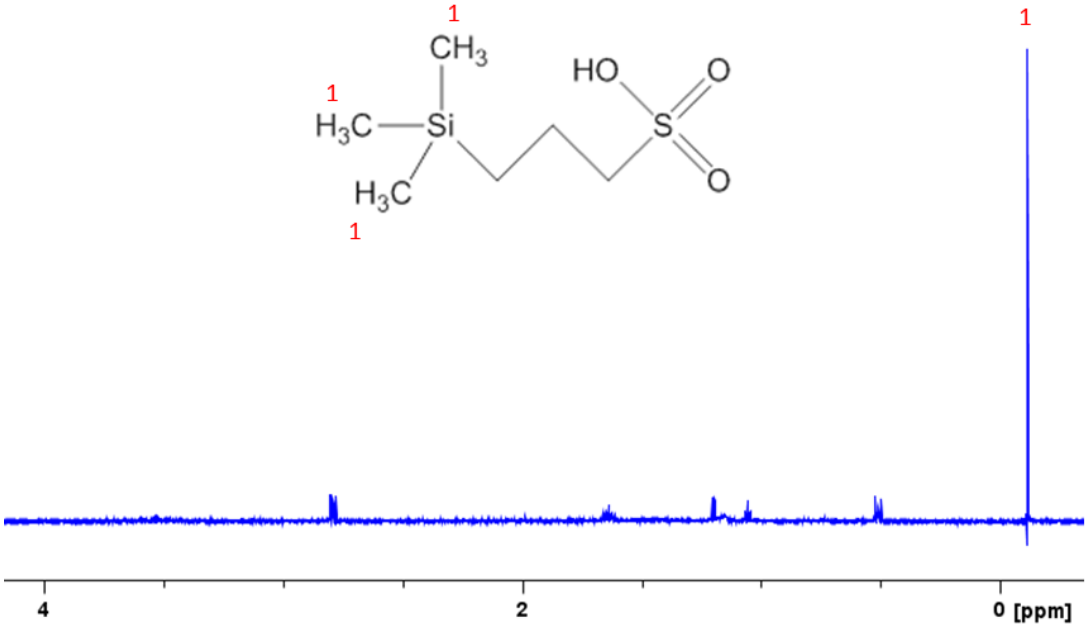


Figure 5.18. DSS solution state NMR 1D H^+ spectrum.

The intensity ($\Delta I = I_{control} - I_{sample}$) and concentration ($\Delta C = C_{control} - C_{sample}$) increments, concentration of ligand bound to the protein ($C_{bound} = (C_{control} - C_{sample})$

*2) and fraction of ligand bound to the protein receptor (B = C_{bound} / $C_{receptor}$) values for peak 2 are shown in Table 5.6 and for peak 3 in Table 5.7.

Table 5.6. Intensity, concentration and fraction of P-Tyr bound to Siglec8 values using the NMR peak 2.

Peak 2						
L (µM)	**$I_{control}$ $1H^+$**	**I_{sample} $1H^+$**	**($I_{control}$ - I_{sample}) $1H^+$**	**I_{DSS} $1H^+$**	**(C_{bound}) (µM)**	**B = $C_{bound}/C_{receptor}$**
50	0.2709	0.1715	0.09935	-	29.81	0.5961
100	0.5	0.2505	0.2495	0.111	74.85	1.497
500	0.9122	0.2339	0.6783	-	203.5	4.070

Table 5.7. Intensity, concentration and fraction of P-Tyr bound to Siglec8 values using the NMR peak 3.

Peak 3						
L (µM)	**$I_{control}$ $1H^+$**	**I_{sample} $1H^+$**	**($I_{control}$ - I_{sample}) $1H^+$**	**I_{DSS} $1H^+$**	**(C_{bound}) (µM)**	**B = $C_{bound}/C_{receptor}$**
50	0.2472	0.1503	0.09695	-	29.09	0.5817
100	0.4848	0.2302	0.2546	0.111	76.38	1.528
500	0.8464	0.2188	0.6276	-	188.3	3.765

The fraction of ligand bound to the protein receptor B ($C_{bound}/C_{receptor}$) is also the fraction of receptor occupied by ligand when the stoichiometry of a protein-receptor interaction is 1:1. Thus, for a 1:1 stoichiometry the expected values of B should be lower or equal to one, meaning that for a 50 µM total concentration of protein receptor, the maximal concentration of ligand that could interact with it would be 50 µM. However, our data show values of B higher than 1 for ligand concentrations higher than 50 µM (Tables 5.6 and 5.7) supporting the heterogeneous binding model indicated by the Octet data, that is a stoichiometry n·ligand molecules :1 protein domain (nL: R).

Estimation of the L-Phospho-Tyrosine fraction bound to Siglec 8 by means of Fluorescence.

Another equilibrium dialysis experiment was set up with a constant Siglec8 concentration (50 µM) and different Phospho-Tyrosine concentrations. The aim of this experiment was to estimate a KD value for the interaction between Siglec 8 and this ligand using Fluorescence as a quantitative technique for measurement of ligand

concentration. Table 5.8 displays the concentration values of L-Phospho-Tyrosine added to the ligand chamber for all equilibrium dialysis samples:

Table 5.8. Equilibrium dialysis concentration values of Phospho-Tyrosine

C P-Tyr Samples (µM)	C P-Tyr Controls (µM)
1	1
10	10
25	25
50	50
100	100

After the ten equilibrium dialysis chambers reached equilibrium, each chamber with ligand solution was opened, and 50 µL of the resulting ligand solution were recovered from all devices. Subsequently, a fluorescence emission spectrum was measured for each sample. Samples were excited at 268 nm, the Phospho-Tyrosine's maximum absorbance wavelength, and fluorescence emission spectra were recorded between 250 and 500 nm, revealing a maximum of emission at 297 nm. The intensity of emission at 297 nm was used to compare sample and control and calculate the fraction of Phospho-Tyrosine bound to the Siglec8 carbohydrate protein domains. The intensity ($\Delta F = F_{control} - F_{sample}$) and concentration ($\Delta C = C_{control} - C_{sample}$) increments, concentration of ligand bound to the protein ($C_{bound} = (C_{control} - C_{sample}) * 2$) and fraction of ligand bound to the protein receptor ($B = C_{bound} / C_{receptor}$) values are shown in Table 5.9.

Table 5.9. Intensity, concentration and fraction of P-Tyr bound to Siglec8 values obtained using fluorescence.

L (µM)	$F_{control}$	F_{sample}	$(F_{control} - F_{sample})$	(C_{bound}) (µM)	$B = C_{bound}/C_{receptor}$
1	30.72	23.99	6.736	0.438	0.00876
10	126.89	109.07	17.823	1.16	0.0232
25	141.54	24.04	117.50	7.64	0.1528
100	560.07	159.40	400.67	26.08	0.5216

5.7 Estimation of Siglec 8-ligand dissociation constant values.

Receptor-ligand (R-L) binding kinetics describes the rate at which a receptor (R) and its ligand (L) bind to each other. When a receptor and a ligand with affinity for it are mixed in solution, k_{on} ($M^{-1} \cdot s^{-1}$) and k_{off} (s^{-1}) are the kinetic constants that provide information about the rate of forward and reverse binding:

$$\text{Receptor (R)} + n\cdot\text{Ligand (nL)} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \text{Receptor-}n\cdot\text{Ligand (R-}n\cdot\text{L)}$$

After equilibrium is reached, the rates of the association and dissociation reactions become equal. This means that free receptor (R) and n ligand (nL) molecules and receptor-ligand complex (R-nL) are being created at the same rate as represented below assuming a stoichiometry 1:1 (n=1):

$$kon \cdot [R] \cdot [L] = koff \cdot [RL]$$

The affinity of the receptor-ligand interaction measures the strengh of the binding. It is commonly measured and reported by the equilibrium dissociation constant K_D (M).

$$KD = \frac{koff}{kon} = \frac{[L] \cdot [R]}{[RL]}$$

$$[RL] = \frac{[L] \cdot [R]}{KD}$$

Other important concepts are the fractional occupancy (F.O.), and the binding occupancy (B). F.O. is expressed as the fraction of all receptors that are bound to the ligand and B as the fraction of all ligands bound to the receptor as shown below:

$$F.O. = \frac{[RL]}{[R] + [RL]}$$

$$B = \frac{[L]bound}{[R]total}$$

Using the definition of the equilibrium dissociation constant K_D, the fractional occupancy can be expressed as follows:

$$F.O. = \frac{[RL]}{[R] + [RL]} = \frac{{}^{[L] \cdot [R]}/_{KD}}{[R] + \left({}^{[L] \cdot [R]}/_{KD}\right)} = \frac{{}^{[L] \cdot [R]}/_{KD}}{{}^{(KD \cdot [R] + [L] \cdot [R])}/_{KD}} = \frac{[L]}{KD + [L]}$$

Assuming a one to one stoichiometry for the ligand-receptor interaction, and introducing the maximal binding occupancy (B_{max}), which represents the maximum

number of receptors that can interact with the ligand, the fractional occupancy may also be writen as:

$$(\text{Bmax} = \frac{[\text{L}]\text{max}}{[\text{R}]\text{total}})$$

$$\text{F.O.} = \frac{[\text{L}-\text{R}]}{[\text{R}]\text{total}} = \frac{[\text{L}]\text{bound}}{[\text{R}]\text{total}} = \frac{[\text{L}]\text{bound}}{[\text{R}]\text{max}} = \frac{[\text{L}]\text{bound}}{[\text{L}]\text{max}} = \frac{\text{B}}{\text{Bmax}}$$

Making the last two fractional occupancy expressions equal, the following equation enables to fit a binding curve to determine B_{max} and K_D, by representing different concentrations of ligand for a constant amount of receptor in solution, against the binding occupancy of the receptor at each concentration.

$$\text{B} = \frac{[\text{L}] \cdot \text{Bmax}}{\text{KD} + [\text{L}]}$$

By linearizing the equation, the K_D can be determined as the negative reciprocal of the slope of the line, and B_{max} is given by its X-axis intercept (Figure 5.19):

$$\text{B} \cdot \text{KD} + \text{B} \cdot [\text{L}] = \text{Bmax} \cdot [\text{L}]$$

$$\text{B} \cdot \text{KD} = (\text{Bmax} - \text{B}) \cdot [\text{L}]$$

$$\frac{\text{B}}{[\text{L}]} = (\text{Bmax} - \text{B}) \cdot \frac{1}{\text{KD}}$$

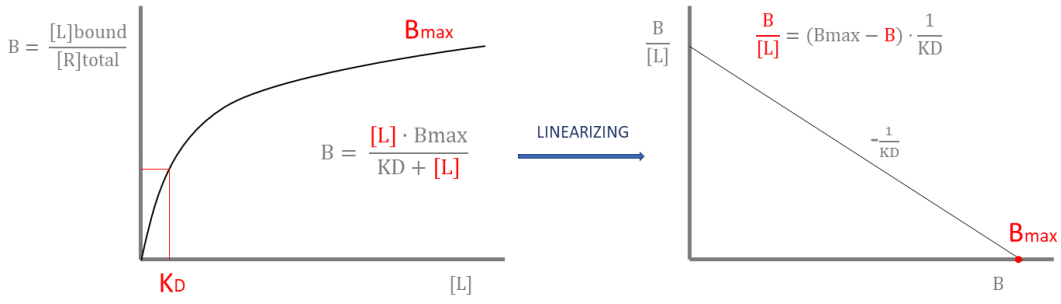


Figure 5.19. Graphical representation of a binding curve and its linearized form.

However, this model is not useful in all situations as it makes the following assumptions:

- The stoichiometry of the interaction is 1:1 (n=1).
- All receptors are equally accessible to ligands.

- There is no state of partial binding, so receptors are either free or bound to the ligand.
- Binding does not alter the ligand or receptor as enzymes do.
- Binding is reversible.

The model described above was applied to the data obtained from the equilibrium dyalisis-1D ^{1}H NMR experiment for the interaction between Siglec8 and P-Tyr but the data do not fit the model which requires B values lower than 1 (B < 1). This is in good agreement with the heterogeneous binding model used to fit the Octet data where the number of ligand molecules interacting with each protein domain is higher than 1 (n >1) and therefore B is also higher than 1 for some of the data points (B > 1).

The best fit was obtained using a variant of the Hill equation. The Hill equation was originally formulated in 1910 to describe the sigmoidal O2 binding curve of haemoglobin or the fraction of protein saturated (Goutelle, *et al.*, 2008) by ligand as a function of the ligand concentration or cooperativity:

$$B' = \frac{[L]^{\eta}}{KD + [L]^{\eta}}$$

In the Hill equation B' is the fraction of protein receptor occupied by ligand ($B'=R_{bound}/R_{total}$), L is the total ligand concentration, K_D is the dissociation constant and n is the Hill coefficient. The Hill coefficient describes the cooperativity of the binding, where $\eta>1$ indicates a positively cooperative interaction, meaning that the binding of one ligand molecule enhances the affinity for other ligand molecules and $\eta<1$ indicates the opposite case. In addition, $\eta=1$ means non-cooperative binding, where the affinity of the protein for a ligand molecule is not dependent on other ligand molecules.

Our experimental data did not allow us to calculate the value of B' or the fraction of occupied Siglec 8 domains, due to the lack of information about the stoichiometry of the Siglec 8-P-Tyr interaction. B' (fraction of occupied receptor) is different from B (fraction of bound ligand) when the stoichiometry of the interaction is nL:R and n is higher than 1 (n>1). However, we represented the fraction of ligand bound to the receptor (B) against the total ligand concentration (L).

Figures 5.20 and 5.21 display the Hill equation fitting for the Phospho-Tyrosine fraction bound to Siglec 8 (B) versus the initial concentration of ligand used for the equilibrium dialysis experiment (L) according to the NMR 1D ^{1}H spectrum peaks 2 and 3, using the software Origin. The estimated values for each of the Hill equation components are represented below with an extra constant that we have called A. Thus, we hypothesized that the meaning of this constant is related to the number of Phospho-Tyrosine molecules interacting with each Siglec 8 domain, and B' may be calculated as the quotient between B and A (B'=B/A). The K_D, A and ɳ values resulting from the Hill equation fitting for the NMR 1D H^+ spectrum peaks 2 and 3 are represented in Tables 5.10 and 5.11.

$$B = \frac{A \cdot [L]^{\eta}}{KD + [L]^{\eta}} \longrightarrow B' = \frac{B}{A} = \frac{[L]^{\eta}}{KD + [L]^{\eta}}$$

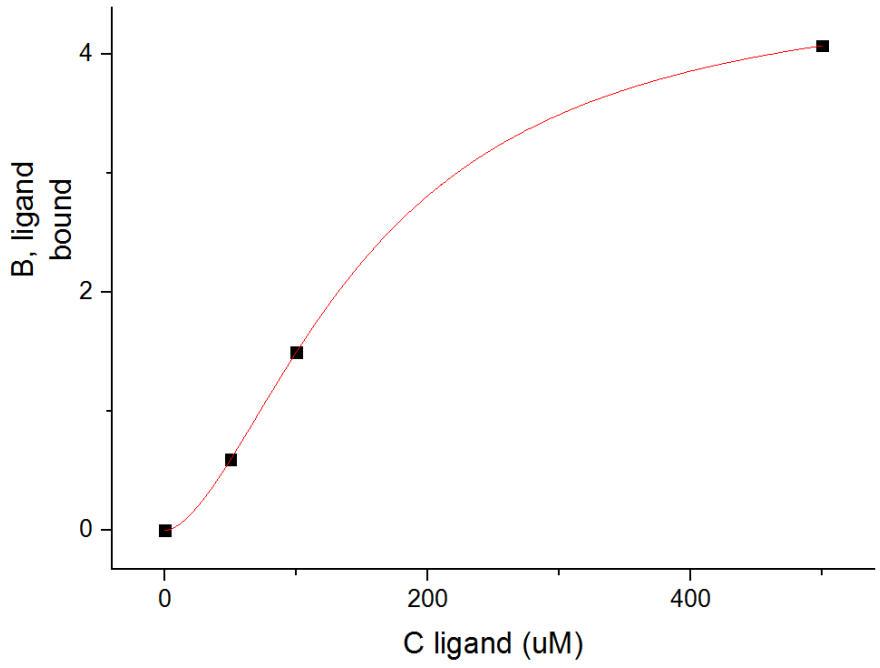


Figure 5.20. Representation of Phospho-Tyrosine concentration against the fraction of Phospho-Tyrosine bound calculated from the 1D H+ NMR peak 2 and fitting according to the Hill equation.

Table 5.10. Hill equation parameters for the Siglec8 and P-Tyr interaction according to the NMR peak 2.

Hill equation parameters for the Siglec8 and Phospho-Tyrosine interaction. NMR Peak 2

A	4.63	± 1.12157E-15
KD (µM)	154.58	± 8.08213E-14
η	1.69	± 8.83144E-16

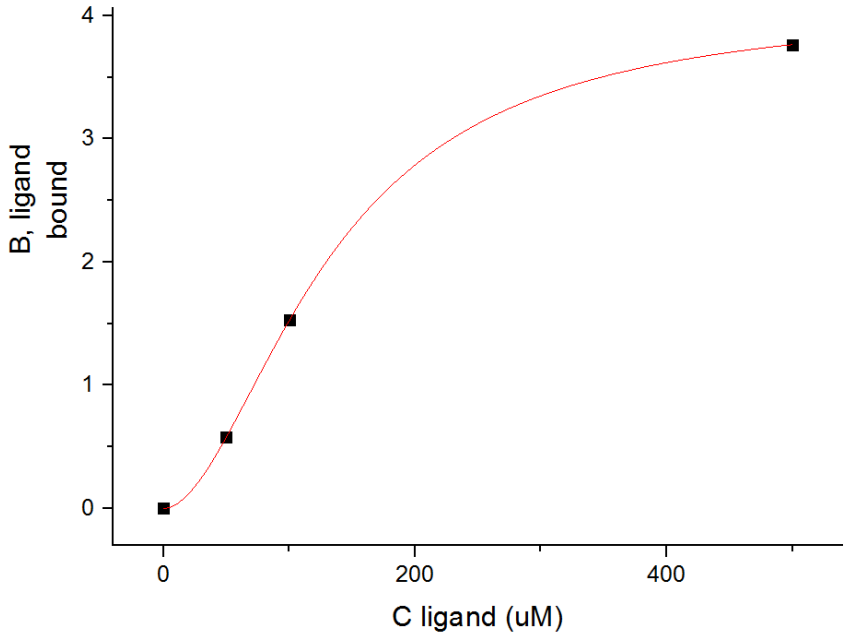


Figure 5.21. Representation of Phospho-Tyrosine concentration against the fraction of Phospho-Tyrosine bound calculated from the 1D H+ NMR peak 3 and fitting according to the Hill equation.

Table 5.11. Hill equation parameters for the Siglec8 and P-Tyr interaction according to the NMR peak 3.

Hill equation parameters for the Siglec8 and Phospho-Tyrosine interaction. NMR Peak 3

A	4.09	± 6.76153E-16
KD (μM)	132.29	± 4.80302E-14
η	1.85	± 8.9904E-16

Figure 5.22 display the Hill equation fitting for the Phospho-Tyrosine fraction bound to Siglec 8 (B) versus the initial concentration of ligand used for the equilibrium dialysis experiment (L) according to Fluorescence measurements, using the software Origin. The estimated values for each of the Hill equation components are represented in table 5.12.

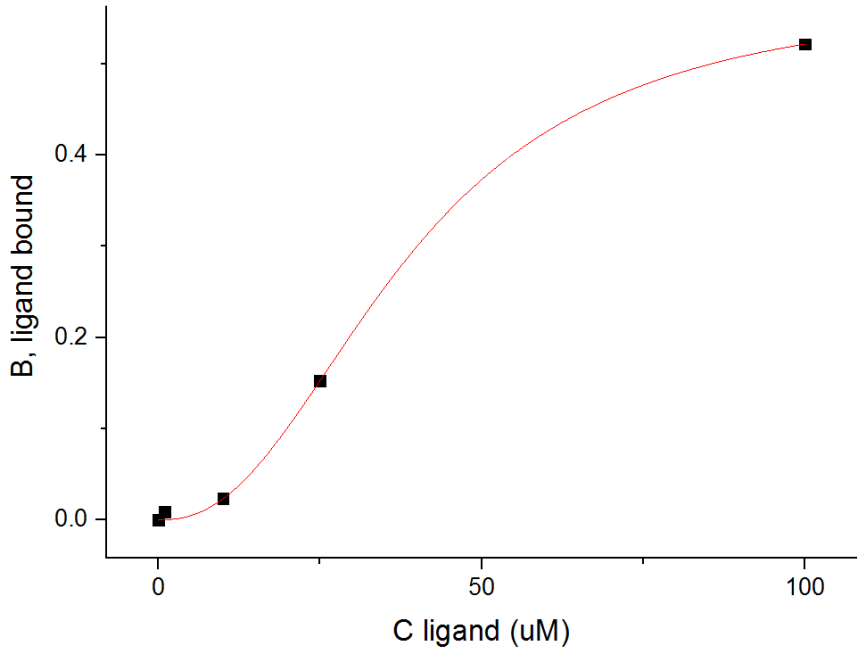


Figure 5.22. Representation of Phospho-Tyrosine concentration against the fraction of Phospho-Tyrosine bound calculated from fluorescence measurements and fitting according to the Hill equation.

Table 5.12. Hill equation parameters for the Siglec8 and P-Tyr interaction according to the NMR peak 3.

Hill equation parameters for the Siglec8 and Phospho-Tyrosine interaction. Fluorescence		
A	0.57832	± 0.03121
KD (μM)	38.73537	± 3.83112
η	2.33946	± 0.32154

In summary, this chapter provided information about the interaction between Siglec8 and nine different small mimetic molecules, with four of them (Glu-Asp, Glu-Leu, Glu-Val and Phospho-Tyr) showing a significant binding response in the Octet experiments. Phospho-Tyrosine was the ligand which demonstrated the highest binding response to our soluble Siglec8 carbohydrate binding domains. All binding curves were fitted according to a heterogeneous, local and partial binding model, suggesting a stoichiometry of more than 1 ligand molecule interacting with 1 protein receptor. Further equilibrium dialysis together with 1D H^+ NMR and fluorescence measurements for the Siglec8-Phospho-Tyrosine interaction supported the heterogeneous binding interaction model. Those data fitted the Hill equation and revealed a positive cooperative binding between Siglec8 and Phospho-Tyrosine with an affinity constant value around 140 μM according to NMR and around 40 μM according to fluorescence measurements. Table 5.13 summarizes all dissociation

constant values obtained for Siglec8-Phospho-Tyrosine interaction by means of the two different techniques.

Table 5.13. Summary of the dissociation constant values obtained by means of equilibrium dialysis plus NMR 1D $^{+}$H and fluorescence.

Technique used to complement Equilibrium Dialysis	Dissociation constant obtained value (μM)
NMR 1D $^{+}$H Peak 2	154.58
NMR 1D $^{+}$H Peak 3	132.29
Fluorescence	38.735

Chapter 6.
DISCUSSION AND FUTURE PERSPECTIVES.

6. Discussion and future perspectives.

In this work, the Siglec8 Ig-like V-type domain has been expressed in *E. coli*, refolded from inclusion bodies and purified to test interaction with small sialic acid mimetic molecules in solution with therapeutic purposes.

To produce this domain, production of which has only been reported by one other group (Propster, *et al.*, 2015), several different *E. coli* strains, expression vectors, chaperones, media additives and culture conditions were screened to achieve prokaryotic soluble expression. In some 2 L cultures grown at 15°C, very low levels of soluble expression could be detected, but the concentration was not high enough to obtain a pure sample due to the high concentration of bacterial soluble proteins present in the culture's soluble fraction. Although there are not many examples of Siglec soluble expression in the literature, previous works reflect the difficulty of producing human Siglecs' carbohydrate binding domains in prokaryotic expression systems. Human Siglec7 Ig-like V-type domains were produced in Chinese Hamster ovary Lec1 cells for X-ray structure resolution (Alphey, *et al.*, 2003) and human Siglec5 extracellular domains were expressed in *E. coli* and refolded from inclusion bodies with a final yield of 2 mg/mL for the same purpose (Barb, *et al.*, 2013). The presence of disulphide bonds, a secondary structure based on β-sheets together with the Asn-linked glycosylation of Siglecs' extracellular domains may be the main reasons for their challenging expression in bacteria. However, Siglec8 Ig-like V-type domains were solubly expressed in *E. coli* Rosetta gami B (DE3) cells in milligram quantities (Propster, *et al.*, 2015) for multidimensional solution state NMR structural characterization (Propster, *et al.*, 2016). We attempted to replicate the cloning construct and expression conditions used in that work, where the construct Siglec8 Ig-like V-type Met17-His155 (Cys42>Ser42) was expressed in *E. coli* Rosetta gami B (DE3) cells using the vector pET-43(a+) but could not obtain the soluble expression levels reported in this work. This is probably due to a lack of means in our chemical biology facility to achieve their culture shaking speed, 250 rpm. Shaking speed is crucial to deliver oxygen to cells in liquid cultures, with the rate of oxygen demand increasing as the cell mass increases and it affects the yields and solubility of overexpressed recombinant proteins. After many attempts to improve our Siglec8 Ig-like V-type domain's solubility, a protocol for refolding from inclusion bodies was

developed in collaboration with Mologic Ltd., and pure samples were obtained at an approximate concentration of 1 mg/mL. This concentration was sufficient to perform a biophysical characterization of the domain as well as protein-ligand binding studies. Likewise, Mologic made use of those samples to produce antibodies anti-Siglec8 for diagnostic purposes.

The circular dichroism spectrum of our soluble Siglec8 Ig-like V-type domains (Figure 4.7) confirmed the production of folded domains with a predominant β-sheet composition and suggested the presence of the disulphide bond. CD data also matched the published spectrum for the soluble Siglec 8 domains expressed in previous work (Propster, *et al.*, 2015). In addition, SAXS data (Figure 4.16) confirmed the production of well-folded soluble Siglec8 Ig-like V-type domain with an average diameter of 32 Å, which is in good agreement with the estimation of a 16.2 Å radius of gyration for the Siglec8 Ig-like V-type domain structure resolved by means of solution state NMR (Propster, *et al.*, 2016).

Due to its restrictive expression in eosinophils and mast cells, Siglec8 has been pointed out as a promising pharmaceutical target to treat allergic diseases such as asthma where those inflammatory cells are involved. Siglec8 agonists have been demonstrated to inhibit cell function and survival, a desirable effect to modulate the excessive eosinophilic response that occurs in the airways of a high proportion of asthmatic patients (Feng and Mao, 2012; Kiwamoto, *et al.*, 2012; Kiwamoto, *et al.*, 2013; Gangwar, *et al.*, 2017). However, the identity of Siglec8's physiological sialylated ligands still remains unknown, with all Siglec8 binding information coming from glycan array data and biophysical studies (Campanero-Rhodes, *et al.*, 2006; Propster, *et al.*, 2016) which revealed 6,6'- Sulpho-Sialyl-Lewis X as the best candidate (estimated K_D = 183±18 μM). In this work, that information was employed to design nine glycan and peptide-based 6,6'- Sulpho-Sialyl-Lewis X mimetic small molecule candidates. Octet and equilibrium dialysis together with either 1D ^{1}H NMR or fluorescence binding experiments revealed the ligand L-Phospho-Tyrosine as our best Siglec8 ligand candidate, with an estimated dissociation constant between 130 - 150 μM in case of NMR and around 40 μM using fluorescence. With this study, we aim to provide some insight and basis for the development of highly-specific ligands with a suitable affinity for the receptor Siglec8. Good candidates could include larger

molecules than L-Phospho-Tyrosine which may contain one or more sulphate or other negatively charged groups together with aromatic rings, as they seem to be important pharmacophores for the interaction.

In our body, drugs can produce dangerous side effects through non-specific interactions with non-target proteins if they are not designed to interact specifically with a particular target protein. Likewise, it is very important to be aware of the important self-defence role that immune system cells play and find a balance between highly specific Siglec8 ligands with the suitable affinity, so they do not induce apoptosis of all eosinophils and affect the immune system physiological response. Thus, ligands must have enough affinity for the receptor to reduce the levels of eosinophils in the patient's airways during an asthmatic crisis without depleting or affecting the immune system responses in a physiological environment.

Siglecs carry out critical immune system regulatory functions in our body and this may be the reason why only low-affinity Siglec ligands with lower than millimolar K_D values have been previously reported. In a physiological situation where the immune system cells release hundreds of enzymes, cytokines and cytotoxic molecules, it is likely that the same balance between high-specificity and the suitable affinity is required to regulate in the right proportion the immune system cellular activity through Siglecs.

Despite the huge advance in the investigation of Siglecs since this important protein family was described ten years ago (Crocker, *et al.*, 2007), there is still much about the molecular structures and ligand interactions that remains unknown. Probably the most important key to moving this field forward is the collaboration between professionals from different disciplines including immunology, glycobiology, synthetic and computational chemistry, biochemistry, cellular biology and biophysics.

Chapter 7.
BIBLIOGRAPHY.

7. Bibliography.

Alphey, M.S., Attrill, H., Crocker, P.R. and van Aalten, D.M.F. (2003). High resolution crystal structures of Siglec-7 - Insights into ligand specificity in the Siglec family. *Journal of Biological Chemistry* **278**(5), pp. 3372-3377.

Altheide, T.K., Hayakawa, T., Mikkelsen, T.S., Diaz, S., Varki, N. and Varki, A. (2006). System-wide genomic and biochemical comparisons of sialic acid biology among primates and rodents - Evidence for two modes of rapid evolution. *Journal of Biological Chemistry* **281**(35), pp. 25689-25702.

AsthmaUK. https://www.asthma.org.uk/advice/inhalers-medicines-treatments/how/. [Online]. Asthma UK:

AsthmaUK. https://www.asthma.org.uk/advice/inhalers-medicines-treatments/steroids/.

AsthmaUK. https://www.asthma.org.uk/advice/triggers/understanding/.

AsthmaUK. https://www.asthma.org.uk/research/future/.

Attrill, H., Imamura, A., Sharma, R.S., Kiso, M., Crocker, P.R. and van Aalten, D.M. (2006a). Siglec-7 undergoes a major conformational change when complexed with the alpha(2,8)-disialylganglioside GT1b. *J Biol Chem* **281**(43), pp. 32774-32783.

Attrill, H., Imamura, A., Sharma, R.S., Kiso, M., Crocker, P.R. and van Aalten, D.M.F. (2006b). Siglec-7 undergoes a major conformational change when complexed with the alpha(2,8)-disialylganglioside GT1b. *Journal of Biological Chemistry* **281**(43), pp. 32774-32783.

Attrill, H., Takazawa, H., Witt, S., Kelm, S., Isecke, R., Brossmer, R., Ando, T., Ishida, H., Kiso, M., Crocker, P.R. and van Aalten, D.M.F. (2006c). The structure of siglec-7 in complex with sialosides: leads for rational structure-based inhibitor design. *Biochemical Journal* **397**pp. 271-278.

Bafadhel, M., Pavord, I.D. and Russell, R.E.K. (2017). Eosinophils in COPD: just another biomarker? *Lancet Respir Med* **5**(9), pp. 747-759.

Barb, A.W., Wang, X. and Prestegard, J.H. (2013). Refolded recombinant Siglec5 for NMR investigation of complex carbohydrate binding. *Protein Expression and Purification* **88**(2), pp. 183-189.

Bertram, L., Lange, C., Mullin, K., Parkinson, M., Hsiao, M., Hogan, M.F., Schjeide, B.M.M., Hooli, B., DiVito, J., Ionita, L., Jiang, H.Y., Laird, N., Moscarillo, T., Ohlsen, K.L., Elliott, K., Wang, X., Hu-Lince, D., Ryder, M., Murphy, A., Wagner, S.L., Blacker, D., Becker, K.D. and Tanzi, R.E. (2008). Genome-wide Association Analysis Reveals Putative Alzheimer's Disease Susceptibility Loci in Addition to APOE. *American Journal of Human Genetics* **83**(5), pp. 623-632.

Blixt, O., Collins, B.E., van den Nieuwenhof, I.M., Crocker, P.R. and Paulson, J.C. (2003). Sialoside specificity of the siglec family assessed using novel multivalent probes - Identification of potent inhibitors of myelin-associated glycoprotein. *Journal of Biological Chemistry* **278**(33), pp. 31007-31019.

Bochner, B.S., Alvarez, R.A., Mehta, P., Bovin, N.V., Blixt, O., White, J.R. and Schnaar, R.L. (2005). Glycan array screening reveals a candidate ligand for Siglec-8. *Journal of Biological Chemistry* **280**(6), pp. 4307-4312.

Bochner, B.S., Rothenberg, M.E., Boyce, J.A. and Finkelman, F. (2013). Advances in mechanisms of allergy and clinical immunology in 2012. *Journal of Allergy and Clinical Immunology* **131**(3), pp. 661-667.

British Lung Foundation, U. (2016). Estimating the economic burden of respiratory illness in the UK.

Campanero-Rhodes, M.A., Childs, R.A., Kiso, M., Komba, S., Le Narvor, C., Warren, J., Otto, D., Crocker, P.R. and Feizi, T. (2006). Carbohydrate microarrays reveal sulphation as a modulator of siglec binding. *Biochemical and Biophysical Research Communications* **344**(4), pp. 1141-1146.

Cao, H., Lakner, U., Bono, B., Traherne, J.A., Trowsdale, J. and Barrow, A.D. (2008). SIGLEC16 encodes a DAP12-associated receptor expressed in macrophages that evolved from its inhibitory counterpart SIGLEC11 and has functional and non-functional alleles in humans. *European Journal of Immunology* **38**(8), pp. 2303-2315.

Cao, H.A. and Crocker, P.R. (2011). Evolution of CD33-related siglecs: regulating host immune functions and escaping pathogen exploitation? *Immunology* **132**(1), pp. 18-26.

Carlin, A.F., Lewis, A.L., Varki, A. and Nizet, V. (2007). Group B streptococcal capsular sialic acids interact with siglecs (immunoglobulin-like lectins) on human leukocytes. *Journal of Bacteriology* **189**(4), pp. 1231-1237.

Chen, B., Zhang, H., Xi, W., Zhao, L., Liang, L. and Chen, Y. (2014). Unfolding mechanism of lysozyme in various urea solutions: Insights from fluorescence spectroscopy. p. 524-528, vol. 1076.

Clark, A.J., Tiwary, P., Borrelli, K., Feng, S., Miller, E.B., Abel, R., Friesner, R.A. and Berne, B.J. (2016). Prediction of Protein-Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations. *J Chem Theory Comput* **12**(6), pp. 2990-2998.

CorrÊa, D. and Ramos, C. (2009). The use of circular dichroism spectroscopy to study protein folding, form and function. p. 164-173, vol. 3 (5). African Journal of Biochemistry Research.

Crocker, P.R. (2002). Siglecs: sialic-acid-binding immunoglobulin-like lectins in cell-cell interactions and signalling. *Current Opinion in Structural Biology* **12**(5), pp. 609-615.

Crocker, P.R., Paulson, J.C. and Varki, A. (2007). Siglecs and their roles in the immune system. *Nature Reviews Immunology* **7**(4), pp. 255-266.

Dimasi, N., Attrill, H., van Aalten, D.M.F., Moretta, A., Moretta, L., Biassoni, R. and Mariuzza, R.A. (2004). Structure of the saccharide-binding domain of the human natural killer cell inhibitory receptor p75/AIRM1 (vol D60, pg 401, 2004). *Acta Crystallographica Section D-Biological Crystallography* **60**pp. 620-620.

Dretzke, J., Blissett, D., Dave, C., Mukherjee, R., Price, M., Bayliss, S., Wu, X., Jordan, R., Jowett, S., Turner, A.M. and Moore, D. (2015). The cost-effectiveness of domiciliary non-invasive ventilation in patients with end-stage chronic obstructive pulmonary disease: a systematic review and economic evaluation. *Health Technol Assess* **19**(81), pp. 1-246.

Farid, S.S., Mirshafiey, A. and Razavi, A. (2012). Siglec-8 and Siglec-F, the new therapeutic targets in asthma. *Immunopharmacology and Immunotoxicology* **34**(5), pp. 721-726.

Feng, Y.H. and Mao, H. (2012). Specific regulator of eosinophil apoptosis: Siglec-8-new hope for bronchial asthma treatment. *Chinese Medical Journal* **125**(11), pp. 2048-2052.

Floyd, H., Ni, J., Cornish, A.L., Zeng, Z.Z., Liu, D., Carter, K.C., Steel, J. and Crocker, P.R. (2000). Siglec-8 - A novel eosinophil-specific member of the immunoglobulin superfamily. *Journal of Biological Chemistry* **275**(2), pp. 861-866.

Fong, J., Deng, L.W., Varki, N., Nizet, V. and Varki, A. (2014). Discovering a sialic acid independent ligand for paired receptors Siglec-5 and-14. *Faseb Journal* **28**(1), pp.

Fortebio. *https://www.fortebio.com/bli-technology.html*. [Online].

Foussias, G., Yousef, G.M. and Diamandis, E.P. (2000). Molecular characterization of a Siglec8 variant containing cytoplasmic tyrosine-based motifs, and mapping of the Siglec8 gene. *Biochemical and Biophysical Research Communications* **278**(3), pp. 775-781.

Gao, P.S., Shimizu, K., Grant, A.V., Rafaels, N., Zhou, L.F., Hudson, S.A., Konno, S., Zimmermann, N., Araujo, M.I., Ponte, E.V., Cruz, A.A., Nishimura, M., Su, S.N., Hizawa, N., Beaty, T.H., Mathias, R.A., Rothenberg, M.E., Barnes, K.C. and Bochner, B.S. (2010). Polymorphisms in the sialic acid-binding immunoglobulin-like lectin-8 (Siglec-8) gene are associated with susceptibility to asthma. *European Journal of Human Genetics* **18**(6), pp. 713-719.

Garfin, P.M. and Feldman, E.J. (2016). Antibody-Based Treatment of Acute Myeloid Leukemia. *Current Hematologic Malignancy Reports* **11**(6), pp. 545-552.

Goutelle, S., Maurin, M., Rougier, F., Barbaut, X., Bourguignon, L., Ducher, M. and Maire, P. (2008). The Hill equation: a review of its capabilities in pharmacological modelling. *Fundam Clin Pharmacol* **22**(6), pp. 633-648.

Gromacs. http://www.gromacs.org/.

Hollingworth, P. Harold, D. Sims, R. Gerrish, A. Lambert, J.C. Carrasquillo, M.M. Abraham, R. Hamshere, M.L. Pahwa, J.S. Moskvina, V. Dowzell, K. Jones, N. Stretton, A. Thomas, C. Richards, A. Ivanov, D. Widdowson, C. Chapman, J. Lovestone, S. Powell, J. Proitsi, P. Lupton, M.K. Brayne, C. Rubinsztein, D.C. Gill, M. Lawlor, B. Lynch, A. Brown, K.S. Passmore, P.A. Craig, D. McGuinness, B. Todd, S. Holmes, C. Mann, D. Smith, A.D. Beaumont, H. Warden, D. Wilcock, G. Love, S. Kehoe, P.G. Hooper, N.M. Vardy, E. Hardy, J. Mead, S. Fox, N.C. Rossor, M. Collinge, J. Maier, W. Jessen, F. Ruther, E. Schurmann, B. Heun, R. Kolsch, H. van den Bussche, H. Heuser, I. Kornhuber, J. Wiltfang, J. Dichgans, M. Frolich, L. Hampel, H. Gallacher, J. Hull, M. Rujescu, D. Giegling, I. Goate, A.M. Kauwe, J.S.K. Cruchaga, C. Nowotny, P. Morris, J.C. Mayo, K. Sleegers, K. Bettens, K. Engelborghs, S. De Deyn, P.P. Van Broeckhoven, C. Livingston, G. Bass, N.J. Gurling, H. McQuillin, A. Gwilliam, R. Deloukas, P. Al-Chalabi, A. Shaw, C.E. Tsolaki, M. Singleton, A.B. Guerreiro, R. Muhleisen, T.W. Nothen, M.M. Moebus, S. Jockel, K.H. Klopp, N. Wichmann, H.E. Pankratz, V.S. Sando, S.B. Aasly, J.O. Barcikowska, M. Wszolek, Z.K. Dickson, D.W. Graff-Radford, N.R. Petersen, R.C. van Duijn, C.M. Breteler, M.M.B. Ikram, M.A. DeStefano, A.L. Fitzpatrick, A.L. Lopez, O. Launer, L.J. Seshadri, S. Berr, C. Campion, D. Epelbaum, J. Dartigues, J.F. Tzourio, C. Alperovitch, A. Lathrop, M. Feulner, T.M. Friedrich, P. Riehle, C. Krawczak, M. Schreiber, S. Mayhaus, M. Nicolhaus, S. Wagenpfeil, S. Steinberg, S. Stefansson, H. Stefansson, K. Snaedal, J. Bjornsson, S. Jonsson, P.V. Chouraki, V. Genier-Boley, B. Hiltunen, M. Soininen, H. Combarros, O. Zelenika, D. Delepine, M. Bullido, M.J. Pasquier, F. Mateo, I. Frank-Garcia, A. Porcellini, E. Hanon, O. Coto, E. Alvarez, V. Bosco, P. Siciliano, G. Mancuso, M. Panza, F. Solfrizzi, V. Nacmias, B. Sorbi, S. Bossu, P. Piccardi, P. Arosio, B. Annoni, G. Seripa, D. Pilotto, A. Scarpini, E. Galimberti, D. Brice, A. Hannequin, D. Licastro, F. Jones, L. Holmans, P.A. Jonsson, T. Riemenschneider, M. Morgan, K. Younkin, S.G. Owen, M.J. O'Donovan, M. Amouyel, P. Williams, J. Alzheimer's Dis, N. Consortium, C. and Consortium, E. (2011). Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nature Genetics* **43**(5), pp. 429-+.

Houry, W.A. (2001). Chaperone-assisted protein folding in the cell cytoplasm. *Current Protein & Peptide Science* **2**(3), pp. 227-244.

Ilmarinen, P. and Kankaanranta, H. (2014a). Eosinophil Apoptosis as a Therapeutic Target in Allergic Asthma. *Basic & Clinical Pharmacology & Toxicology* **114**(1), pp. 109-117.

Ilmarinen, P. and Kankaanranta, H. (2014b). Eosinophil Apoptosis as a Therapeutic Target in Allergic Asthma. *Basic & Clinical Pharmacology & Toxicology* **114**(1), pp. 109-117.

Jahn, L., Hagedoorn, R.S., van der Steen, D.M., Hombrink, P., Kester, M.G.D., Schoonakker, M.P., de Ridder, D., van Veelen, P.A., Falkenburg, J.H.F. and Heemskerk, M.H.M. (2016). A CD22-reactive TCR from the T-cell allorepertoire for the treatment of acute lymphoblastic leukemia by TCR gene transfer. *Oncotarget* **7**(44), pp. 71536-71547.

Janevska, D., O'Sullivan, J.A., Cao, Y. and Bochner, B.S. (2015). Specific subsets of kinases mediate Siglec-8 engagement-induced reactive oxygen species (ROS) production and apoptosis in primary human eosinophils. *Glycobiology* **25**(11), pp. 1243-1243.

Jellusova, J. and Nitschke, L. (2012). Regulation of B cell functions by the sialic acid-bincing receptors Siglec-G and CD22. *Frontiers in Immunology* **2**pp. 14.

Kankaanranta, H., Lindsay, M.A., Giembycz, M.A., Zhang, X., Moilanen, E. and Barnes, P.J. (2000). Delayed eosinophil apoptosis in asthma. *J Allergy Clin Immunol* **106**(1 Pt 1), pp. 77-83.

Kano, G., Almanan, M., Bochner, B.S. and Zimmermann, N. (2013). Mechanism of Siglec-8-mediated cell death in IL-5-activated eosinophils: Role for reactive oxygen species-enhanced MEK/ERK activation. *Journal of Allergy and Clinical Immunology* **132**(2), pp. 437-445.

Kano, G., Bochner, B.S. and Zimmermann, N. (2014). Regulation Of Reactive Oxygen Species Production Involving Src Family Kinase In Siglec-8 Induced Eosinophil Cell Death. *Journal of Allergy and Clinical Immunology* **133**(2), pp. AB59-AB59.

Kano, G., Bochner, B.S. and Zimmermann, N. (2017). Regulation of Siglec-8-induced intracellular reactive oxygen species production and eosinophil cell death by Src family kinases. *Immunobiology* **222**(2), pp. 343-349.

Kawasaki, N., Vela, J.L., Nycholat, C.M., Rademacher, C., Khurana, A., van Rooijen, N., Crocker, P.R., Kronenberg, M. and Paulson, J.C. (2013). Targeted delivery of lipid antigen to macrophages via the CD169/sialoadhesin endocytic pathway induces robust invariant natural killer T cell activation. *Proceedings of the National Academy of Sciences of the United States of America* **110**(19), pp. 7826-7831.

Kiwamoto, T., Katoh, T., Tiemeyer, M. and Bochner, B.S. (2013). The role of lung epithelial ligands for Siglec-8 and Siglec-F in eosinophilic inflammation. *Current Opinion in Allergy and Clinical Immunology* **13**(1), pp. 106-111.

Kiwamoto, T., Kawasaki, N., Paulson, J.C. and Bochner, B.S. (2012a). Siglec-8 as a drugable target to treat eosinophil and mast cell-associated conditions. *Pharmacology & Therapeutics* **135**(3), pp. 327-336.

Kiwamoto, T., Kawasaki, N., Paulson, J.C. and Bochner, B.S. (2012b). Siglec-8 as a drugable target to treat eosinophil and mast cell-associated conditions. *Pharmacology & Therapeutics* **135**(3), pp. 327-336.

Kosinski-Collins, M.S., Flaugh, S.L. and King, J. (2004). Probing folding and fluorescence quenching in human gammaD crystallin Greek key domains using triple tryptophan mutant proteins. *Protein Sci* **13**(8), pp. 2223-2235.

Kozlowski, L.P. (2017). Proteome-pI: proteome isoelectric point database. *Nucleic Acids Res* **45**(D1), pp. D1112-D1116.

Laszlo, G.S., Estey, E.H. and Walter, R.B. (2014). The past and future of CD33 as therapeutic target in acute myeloid leukemia. *Blood Reviews* **28**(4), pp. 143-153.

Leelananda, S.P. and Lindert, S. (2016). Computational methods in drug discovery. *Beilstein J Org Chem* **12**pp. 2694-2718.

Loke, J., Khan, J.N., Wilson, J.S., Craddock, C. and Wheatley, K. (2015). Mylotarg has potent anti-leukaemic effect: a systematic review and meta-analysis of anti-CD33 antibody treatment in acute myeloid leukaemia. *Annals of Hematology* **94**(3), pp. 361-373.

London, U.o. (2010-2017). Dichroweb. http://dichroweb.cryst.bbk.ac.uk/html/home.shtml.

Lyons, J.A., Shahsavar, A., Paulsen, P.A., Pedersen, B.P. and Nissen, P. (2016). Expression strategies for structural studies of eukaryotic membrane proteins. *Current Opinion in Structural Biology* **38**pp. 137-144.

Macauley, M.S., Crocker, P.R. and Paulson, J.C. (2014). Siglec-mediated regulation of immune cell function in disease. *Nature Reviews Immunology* **14**(10), pp. 653-666.

Manavalan, P. and Johnson, W.C. (1983). IDENTIFICATION OF TERTIARY STRUCTURAL TYPES ALL-ALPHA, ALL-BETA, ALPHA+BETA AND ALPHA/BETA FROM CIRCULAR-DICHROISM OF PROTEINS. *Biophysical Journal* **41**(2), pp. A267-A267.

Marchetti, R., Perez, S., Arda, A., Imberty, A., Jimenez-Barbero, J., Silipo, A. and Molinaro, A. (2016). "Rules of Engagement" of Protein-Glycoconjugate Interactions: A Molecular View Achievable by using NMR Spectroscopy and Molecular Modeling. *Chemistryopen* **5**(4), pp. 274-296.

Mayr, L.M. and Bojanic, D. (2009). Novel trends in high-throughput screening. *Curr Opin Pharmacol* **9**(5), pp. 580-588.

McIlwain, D.R., Berger, T. and Mak, T.W. (2015). Caspase functions in cell death and disease. *Cold Spring Harb Perspect Biol* **7**(4), pp.

Micsonai, A., Wien, F., Kernya, L., Lee, Y.H., Goto, Y., Réfrégiers, M. and Kardos, J. (2015). Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc Natl Acad Sci U S A* **112**(24), pp. E3095-3103.

Muller, J. and Nitschke, L. (2014). The role of CD22 and Siglec-G in B-cell tolerance and autoimmune disease. *Nature Reviews Rheumatology* **10**(7), pp. 422-428.

Na, H.Y., Liu, X.L., Li, X.M., Zhang, X.S., Wang, Y., Wang, Z.H., Yuan, M.L., Zhang, Y., Ren, S.Y. and Zuo, Y.F. (2017). Novel roles of DC-SIGNR in colon cancer cell adhesion, migration, invasion, and liver metastasis. *Journal of Hematology & Oncology* **10**pp. 18.

Naj, A.C. Jun, G. Beecham, G.W. Wang, L.S. Vardarajan, B.N. Buros, J. Gallins, P.J. Buxbaum, J.D. Jarvik, G.P. Crane, P.K. Larson, E.B. Bird, T.D. Boeve, B.F. Graff-Radford, N.R. De Jager, P.L. Evans, D. Schneider, J.A. Carrasquillo, M.M. Ertekin-

Taner, N. Younkin, S.G. Cruchaga, C. Kauwe, J.S.K. Nowotny, P. Kramer, P. Hardy, J. Huentelman, M.J. Myers, A.J. Barmada, M.M. Demirci, F.Y. Baldwin, C.T. Green, R.C. Rogaeva, E. St George-Hyslop, P. Arnold, S.E. Barber, R. Beach, T. Bigio, E.H. Bowen, J.D. Boxer, A. Burke, J.R. Cairns, N.J. Carlson, C.S. Carney, R.M. Carroll, S.L. Chui, H.C. Clark, D.G. Corneveaux, J. Cotman, C.W. Cummings, J.L. DeCarli, C. DeKosky, S.T. Diaz-Arrastia, R. Dick, M. Dickson, D.W. Ellis, W.G. Faber, K.M. Fallon, K.B. Farlow, M.R. Ferris, S. Frosch, M.P. Galasko, D.R. Ganguli, M. Gearing, M. Geschwind, D.H. Ghetti, B. Gilbert, J.R. Gilman, S. Giordani, B. Glass, J.D. Growdon, J.H. Hamilton, R.L. Harrell, L.E. Head, E. Honig, L.S. Hulette, C.M. Hyman, B.T. Jicha, G.A. Jin, L.W. Johnson, N. Karlawish, J. Karydas, A. Kaye, J.A. Kim, R. Koo, E.H. Kowall, N.W. Lah, J.J. Levey, A.I. Lieberman, A.P. Lopez, O.L. Mack, W.J. Marson, D.C. Martiniuk, F. Mash, D.C. Masliah, E. McCormick, W.C. McCurry, S.M. McDavid, A.N. McKee, A.C. Mesulam, M. Miller, B.L. Miller, C.A. Miller, J.W. Parisi, J.E. Perl, D.P. Peskind, E. Petersen, R.C. Poon, W.W. Quinn, J.F. Rajbhandary, R.A. Raskind, M. Reisberg, B. Ringman, J.M. Roberson, E.D. Rosenberg, R.N. Sano, M. Schneider, L.S. Seeley, W. Shelanski, M.L. Slifer, M.A. Smith, C.D. Sonnen, J.A. Spina, S. Stern, R.A. Tanzi, R.E. Trojanowski, J.Q. Troncoso, J.C. Van Deerlin, V.M. Vinters, H.V. Vonsattel, J.P. Weintraub, S. Welsh-Bohmer, K.A. Williamson, J. Woltjer, R.L. Cantwell, L.B. Dombroski, B.A. Beekly, D. Lunetta, K.L. Martin, E.R. Kamboh, M.I. Saykin, A.J. Reiman, E.M. Bennett, D.A. Morris, J.C. Montine, T.J. Goate, A.M. Blacker, D. Tsuang, D.W. Hakonarson, H. Kukull, W.A. Foroud, T.M. Haines, J.L. Mayeux, R. Pericak-Vance, M.A. Farrer, L.A. and Schellenberg, G.D. (2011). Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nature Genetics* **43**(5), pp. 436-+.

Nieto, L., Canales, A., Gimenez-Gallego, G., Nieto, P.M. and Jimenez-Barbero, J. (2011). Conformational Selection of the AGA*IA(M) Heparin Pentasaccharide when Bound to the Fibroblast Growth Factor Receptor. *Chemistry-a European Journal* **17**(40), pp. 11204-11209.

Nutku, E., Aizawa, H., Hudson, S.A. and Bochner, B.S. (2003). Ligation of Siglec-8: a selective mechanism for induction of human eosinophil apoptosis. *Blood* **101**(12), pp. 5014-5020.

Nutku, E., Hudson, S.A. and Bochner, B.S. (2005). Mechanism of Siglec-8-induced human eosinophil apoptosis: Role of caspases and mitochondrial injury. *Biochemical and Biophysical Research Communications* **336**(3), pp. 918-924.

Origin. https://www.origin.com/gbr/en-us/store/.

Palmer, I. and Wingfield, P.T. (2012). Preparation and extraction of insoluble (inclusion-body) proteins from Escherichia coli. *Curr Protoc Protein Sci* **Chapter 6**pp. Unit6.3.

Papaiwannou, A., Zarogoulidis, P., Porpodis, K., Spyratos, D., Kioumis, I., Pitsiou, G., Pataka, A., Tsakiridis, K., Arikas, S., Mpakas, A., Tsiouda, T., Katsikogiannis, N., Kougioumtzi, I., Machairiotis, N., Siminelakis, S., Kolettas, A., Kessis, G., Beleveslis, T. and Zarogoulidis, K. (2014). Asthma-chronic obstructive pulmonary disease overlap syndrome (ACOS): current literature review. *J Thorac Dis* **6 Suppl 1**pp. S146-151.

Pillai, S., Netravali, I.A., Cariappa, A. and Mattoo, H. (2012). Siglecs and Immune Regulation. In: Paul, W.E. *Annual Review of Immunology, Vol 30*, vol. 30, pp. 357-392. Palo Alto: Annual Reviews.

Prescher, H., Frank, M., Gütgemann, S., Kuhfeldt, E., Schweizer, A., Nitschke, L., Watzl, C. and Brossmer, R. (2017). Design, Synthesis, and Biological Evaluation of Small, High-Affinity Siglec-7 Ligands: Toward Novel Inhibitors of Cancer Immune Evasion. *J Med Chem* **60**(3), pp. 941-956.

Propster, J.M., Yang, F., Ernst, B., Allain, F.H.T. and Schubert, M. (2015). Functional Siglec lectin domains from soluble expression in the cytoplasm of Escherichia coli. *Protein Expression and Purification* **109**pp. 14-22.

Propster, J.M., Yang, F., Rabbani, S., Ernst, B., Allain, F.H.T. and Schubert, M. (2016). Structural basis for sulfation-dependent self-glycan recognition by the human immune-inhibitory receptor Siglec-8. *Proceedings of the National Academy of Sciences of the United States of America* **113**(29), pp. E4170-E4179.

Pymol. https://pymol.org/2/.

Rosano, G.L. and Ceccarelli, E.A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology* **5**pp. 17.

Royer, C.A. (2006). Probing protein folding and conformational transitions with fluorescence. *Chem Rev* **106**(5), pp. 1769-1784.

Sapan, C.V. and Lundblad, R.L. (2015). Review of methods for determination of total protein and peptide concentration in biological samples. *Proteomics Clin Appl* **9**(3-4), pp. 268-276.

Schauer, R. (2004). Victor Ginsburg's influence on my research of the role of sialic acids in biological recognition. *Archives of Biochemistry and Biophysics* **426**(2), pp. 132-141.

Shor, B., Gerber, H.P. and Sapra, P. (2015). Preclinical and clinical development of inotuzumab-ozogamicin in hematological malignancies. *Molecular Immunology* **67**(2), pp. 107-116.

SIB, B.S. *ProtParam Tool*. [Online].

Simon, H.U., Yousefi, S., Schranz, C., Schapowal, A., Bachert, C. and Blaser, K. (1997). Direct demonstration of delayed eosinophil apoptosis as a mechanism causing tissue eosinophilia. *J Immunol* **158**(8), pp. 3902-3908.

Smith, A.E. and Helenius, A. (2004). How viruses enter animal cells. *Science* **304**(5668), pp. 237-242.

Smith, D.K. and Xue, H. (1997). Sequence profiles of immunoglobulin and immunoglobulin-like domains. *Journal of Molecular Biology* **274**(4), pp. 530-545.

Song, X.Z., Heimburg-Molinaro, J., Cummings, R.D. and Smith, D.F. (2014). Chemistry of natural glycan microarrays. *Current Opinion in Chemical Biology* **18**pp. 70-77.

Sreerama, N. and Woody, R.W. (2003). Structural composition of betaI- and betaII-proteins. *Protein Sci* **12**(2), pp. 384-388.

Stern, M., Meagher, L., Savill, J. and Haslett, C. (1992). Apoptosis in human eosinophils. Programmed cell death in the eosinophil leads to phagocytosis by macrophages and is modulated by IL-5. *J Immunol* **148**(11), pp. 3543-3549.

Stone, K.L., DeAngelis, R., LoPresti, M., Jones, J., Papov, V.V. and Williams, K.R. (1998). Use of liquid chromatography-electrospray ionization-tandem mass spectrometry (LC-ESI-MS/MS) for routine identification of enzymatically digested proteins separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Electrophoresis* **19**(6), pp. 1046-1052.

Tai, P.C., Sun, L. and Spry, C.J. (1991). Effects of IL-5, granulocyte/macrophage colony-stimulating factor (GM-CSF) and IL-3 on the survival of human blood eosinophils in vitro. *Clin Exp Immunol* **85**(2), pp. 312-316.

Tipping, W.J., Tshuma, N., Adams, J., Haywood, H.T., Rowedder, J.E., Fray, M.J., McInally, T., Macdonald, S.J. and Oldham, N.J. (2015). Relative binding affinities of integrin antagonists by equilibrium dialysis and liquid chromatography-mass spectrometry. *ACS Med Chem Lett* **6**(2), pp. 221-224.

Todd, A., Anderson, R. and Groundwater, P.W. (2009). Rational drug design. Identifying and characterising a target., The Pharmaceutical Journal.

van Kooyk, Y. and Rabinovich, G.A. (2008). Protein-glycan interactions in the control of innate and adaptive immune responses. *Nature Immunology* **9**(6), pp. 593-601.

van Liempt, E., Bank, C.M., Mehta, P., Garciá-Vallejo, J.J., Kawar, Z.S., Geyer, R., Alvarez, R.A., Cummings, R.D., Kooyk, Y. and van Die, I. (2006). Specificity of DC-SIGN for mannose- and fucose-containing glycans. *FEBS Lett* **580**(26), pp. 6123-6131.

Varki, A. and Angata, T. (2006). Siglecs - the major subfamily of I-type lectins. *Glycobiology* **16**(1), pp. 1R-27R.

WADDELL, W.J. (1956). A simple ultraviolet spectrophotometric method for the determination of protein. *J Lab Clin Med* **48**(2), pp. 311-314.

Woodruff, P.G., Modrek, B., Choy, D.F., Jia, G., Abbas, A.R., Ellwanger, A., Koth, L.L., Arron, J.R. and Fahy, J.V. (2009). T-helper type 2-driven inflammation defines major subphenotypes of asthma. *Am J Respir Crit Care Med* **180**(5), pp. 388-395.

Yamaguchi, H. and Miyazaki, M. (2014). Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules* **4**(1), pp. 235-251.

Yamaguchi, Y., Suda, T., Ohta, S., Tominaga, K., Miura, Y. and Kasahara, T. (1991). Analysis of the survival of mature human eosinophils: interleukin-5 prevents apoptosis in mature human eosinophils. *Blood* **78**(10), pp. 2542-2547.

Zhuravleva, M.A., Trandem, K. and Sun, P.D. (2008a). Structural implications of Siglec-5-mediated sialoglycan recognition. *Journal of Molecular Biology* **375**(2), pp. 437-447.

Zhuravleva, M.A., Trandem, K. and Sun, P.D. (2008b). Structural implications of Siglec-5-mediated sialoglycan recognition. *J Mol Biol* **375**(2), pp. 437-447.