

Biochemical Approaches for Studying the Role of CD33 in Alzheimer's Disease Susceptibility

by

Jaesoo Jung

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry
University of Alberta

© Jaesoo Jung, 2023

Abstract

Alzheimer's disease (AD) is a common neurodegenerative dementia, and its pathological hallmarks include amyloid- β plaque and neurofibrillary tangle formation in the brain. Genome-wide association studies (GWAS) revealed that a rare single nucleotide polymorphism (SNP) in *CD33* is associated with AD susceptibility. This SNP regulates alternative mRNA splicing to generate long (CD33M) and short (CD33m) protein isoforms. These two CD33 isoforms differ in the presence and absence of the V-set glycan-binding domain. Therefore, as higher expression of CD33M correlates with increased AD risk, there is a correlation between loss of glycan binding and AD susceptibility. As the glycan ligands for CD33M are not well understood, the major objective of this thesis is to develop biochemical tools to study the glycan ligands of CD33. To complement studies on CD33M, a second objective is to develop approaches for better understanding alternative mRNA splicing of CD33.

In Chapter 2, a new version of CD33-Fc was developed to characterize CD33 ligands in applications through electrospray ionization mass spectrometry (ESI-MS) and flow cytometry. This new version of CD33-Fc contains a TEV recognition site between the CD33 extracellular domains and Fc constant domains, C-terminal His₆ tag, and Strep-tag II. In a direct ESI-MS application, we found that CD33 recognizes α 2-3 and α 2-6 linked sialosides with similar affinity. In an indirect ESI-MS application, called Catch-and-Release (CaR) ESI-MS, defined human oligosaccharide libraries and a synthetic *N*-glycan library were screened for binding with CD33-Fc. The screening results are consistent with the ability of CD33 to recognize both α 2-3 and α 2-6 sialosides. Analysis of ligands on cells also supports the conclusions that CD33 can recognize both α 2-3 and α 2-6 linkages. These results are significant in two ways: (i) reviews on Siglec specificity had been claiming that CD33 has specificity for α 2-6 sialosides and (ii) the ability of CD33 to

recognizes α 2-3 sialosides has important implications in setting up the findings made in the next Chapter.

In Chapter 3, the role of carbohydrate sulfation was investigated for its ability to enhance the binding of Siglecs, including CD33, to glycan ligands. This was initially accomplished by overexpression of carbohydrate sulfotransferases (CHSTs), probing of cells with Siglec-Fc proteins, and detection by flow cytometry. Many new interactions were discovered, particularly the ability of CHST1 to upregulate Siglec ligands to numerous Siglecs, including CD33. Production of a highly homogeneous CD33 fragment enabled a mass spectrometry-based binding assay to assess binding affinities. Of note, CD33 showed a significant enhancement in affinity (≥ 28 -fold) for a disulfated ligand, Neu5Ac α 2-3(6-*O*-sulfo)Gal β 1-4(6-*O*-sulfo)GlcNAc, compared to its non-sulfated counterpart. It was also revealed that CHST1 is upregulated in numerous cancers and associated with poorer patient outcome. These results provide new insights into carbohydrate sulfation as a general mechanism for fine-tuning Siglec ligands on cells and has implications on the biological ligands of CD33 in the brain.

In Chapter 4, several anti-CD33m monoclonal antibodies were developed to be able to study this elusive protein isoform of CD33, which is implicated in decreased AD susceptibility. It was shown, for the first time at the protein level, that the AD-relevant SNP in *CD33* affects the CD33m protein levels. U937 cells expressing the T allele of this SNP had better uptakes of A β ₁₋₄₂ aggregates, which builds into a larger body of work that CD33m is a gain-of-function protein isoform that enhances phagocytosis. Under endogenous expression levels, CD33m localizes intracellularly, and it was discovered that Cys42 is a key factor for retaining CD33m inside cells. Furthermore, a system for transforming CD33M into CD33m by Cre recombination was devised and validated, and two novel CD33 splicing cell-based reporters were developed for the future

screening of compounds capable of inducing alternative mRNA splicing of CD33. These tools will be valuable in future studies aimed at fully elucidating the AD protective role of CD33.

Preface

This dissertation is submitted for the degree of Doctor of Philosophy at University of Alberta. The research described herein was conducted under the supervision of Professor Matthew S. Macauley in the Department of Chemistry, University of Alberta, between September 2017 and May 2023. This thesis is an original work by Jaesoo Jung. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board.

All MS-based methods in Chapter 2 were performed in Dr. John Klassen's Lab. The protocol for isolating *N*-glycan library from U937 cells was provided by National Center for Functional Glycomics.

Chapter 3 of this thesis has been published as Jung *et al.* *ACS Chemical Biology*, 16(11), 2673-2689. I was responsible for the data collection and analysis as well as the manuscript composition. Jhon R. Enterina assisted with the transcript profiling of CHST1 in Figure 6. Duong T. Bui and Dr. John S. Klassen assisted with the determination of dissociation constant by ESI-MS in Figure 5. Fahima Mozanehe assisted with the sialylation of sulfated LacNAc structures and their NMR analysis. Po-han Lin and Dr. Xuefei Huang assisted with the synthesis of sulfated LacNAc and their NMR analysis in Figure 5. Nitin and Dr. Kay-Hooi Khoo assisted with defining *O*-glycosylation sites on CD33 in Figure 5. Parisa Raeisimakiani and Dr. Lara K. Mahal assisted with analyzing glycomes of CHST overexpression cells by lectin arrays in Figure 1.

Acknowledgements

First and foremost, I am highly grateful to my supervisory committee members, Dr. Matthew S. Macauley, Dr. John S. Klassen and Dr. Ratmir Derda for the supports, encouragement, and advice.

I appreciate Alberta Graduate Excellence Scholarship provided by the Alberta Government. I also wish to acknowledge Glyconet for their support.

I would like to thank Dr. Heajin Park and Duong T. Bui for MS analysis. I thank the Khoo group at the Academia Sinica for defining glycosylation sites on CD33. I thank the Mahal group at the University of Alberta for analyzing cells with lectin array. I thank the Huang group at the Michigan State University for synthesizing sulfated LacNAc structures. I thank for the Zandberg group at the University of British Columbia, Kelowna for characterizing *N*-glycan compositions on CHST overexpressing cell lines. I thank my post and present members of the Macauley group for their help, advice, and friendship, in particular to Jhon R. Enterina, Fahima Mozaneh, Chris D. St. Laurent, Edward Schmidt, Kelli McCord, and Susmita Sarkar.

Thanks also to my friends and to the people I met at the University of Alberta during my time in Edmonton. Last, I would like to express my gratitude to my family for their love, support, and encouragement.

Table of Contents

<i>Abstract</i>	<i>ii</i>
<i>Preface</i>	<i>v</i>
<i>Acknowledgements</i>	<i>vi</i>
<i>Table of Contents</i>	<i>vii</i>
<i>List of Tables</i>	<i>xii</i>
<i>List of Figures</i>	<i>xiii</i>
<i>List of Abbreviations</i>	<i>xviii</i>
<i>Glossary of Terms</i>	<i>xxi</i>
Chapter 1 Introduction	1
1.1 Introduction of Siglecs	2
1.1.1 The functional parts that make up a Siglec.....	2
1.1.2 Ligands for Siglecs	5
1.1.3 Examples of the effects of Siglec-glycan interactions on immune responses	7
1.1.4 The two key types of Siglec-glycan interactions.....	9
1.1.5 Therapeutic aspects of Siglecs in cancer.....	9
1.2 Soluble versions of Siglec to probe their glycan binding	10
1.2.1 Siglec-Fc chimeric fusion proteins	11
1.2.2 The advantages of the Fc chimera	11
1.2.3 Alternatives to Siglec-Fc	12
1.3 Probing the glycan ligands of Siglecs with Siglec-Fc proteins	14
1.3.1 Glycan microarrays	14

1.3.2 Elucidation of Siglec ligands using multivalent ligand presentations	16
1.3.3 Cell- and tissue-based assay	16
1.4 CD33 biology in Late-Onset Alzheimer's Disease	19
1.4.1 Genetics of the rs12459419 SNP in CD33	19
1.4.2 The two isoforms of CD33 and their function.....	22
1.4.3 Ligands for CD33	23
1.4.4 Differences between hCD33 and mCD33	24
1.5 Aims of this Thesis.....	26
<i>Chapter 2 Development of a New Generation of Siglec-Fc for Characterizing Ligands for CD33</i>	
.....	29
2.1 Acknowledgements.....	29
2.2 Introduction	30
2.3 Results.....	32
2.3.1 A new generation of CD33-Fc	32
2.3.2 Validation of Catch-and-Release electrospray ionization mass spectrometry	34
2.3.3 Development of monomeric and homogenous CD33 for an ESI-MS-based binding assay	38
2.3.4 Both α 2-3 or α 2-6 linked sialosides in U937 cells play a role in masking CD33	41
2.3.5 CaR-ESI-MS with CD33-Fc and a natural N-glycan library from U937 cells	45
2.4 Discussion	48
2.5 Conclusions	52
2.6 Methods	53
2.6.1 Neuraminidase-A and -S preparation	53
2.6.2 TEV protease preparation.....	54
2.6.3 Endo H preparation.....	54
2.6.4 CD33-Fc preparation	54

2.6.5 CD33 fragment preparation	55
2.6.6 N-glycan library preparation	55
2.6.7 Flow cytometry with two steps or precomplexing CD33-Fc with Streptactin	56
2.6.8 Formulation of CD33 ligand-targeted liposomes	57
2.6.8 Liposome assay	58
2.6.9 Flow cytometry	58
2.6.10 ESI mass spectrometry	58
2.6.11 CaR-ESI-MS Screening.....	60
<i>Chapter 3 Carbohydrate sulfation as a mechanism for fine-tuning Siglec ligands.....</i>	61
3.1 Acknowledgements.....	61
3.2 Introduction	62
3.3 Results.....	63
3.3.1 Enhanced glycan sulfation by CHST overexpression	63
3.3.2 CHST-dependent binding of Siglecs to U937 cells.....	66
3.3.3 Sialic acid-dependent binding and generality of CHST-dependent Siglec binding	69
3.3.4 Pharmacological perturbation of key cellular glycans	71
3.3.5 Quantifying Siglec binding to sulfated glycans.....	73
3.3.6 Overexpression of both CHST1 and CHST2 in U937 cells greatly enhances Siglec ligands.....	77
3.3.7 Upregulated expression of CHST1 in cancer	80
3.3.8 Disrupting glycan sulfation levels on cancer cells downregulates Siglec ligands	82
3.4 Discussions	87
3.5 Conclusions	90
3.6 Methods and materials.....	91
3.6.1 Cell culture	91
3.6.2 Cloning and transduction of CHSTs.....	91
3.6.3 Construction of the RP173 vector	92

3.6.4 Mutagenesis of the two O-glycosylation sites on CD33.....	92
3.6.5 Staining method for Flow-cytometry assay.....	93
3.6.6 Glycosylation or Sulfation inhibitors	93
3.6.7 Statistical analyses.....	94
3.6.8 Liposome binding assay	94
3.6.9 Measurement of cell viability.....	95
 <i>Chapter 4 A development of a reporter system for screening small molecule drug candidates inducing the expression of hCD33m.....</i>	96
4.1 Acknowledgement	96
4.2 Introduction	97
4.3 Results.....	99
4.3.1 Development of monoclonal antibodies that recognize hCD33m.....	99
4.3.2 CD33 minigene constructs that faithfully mimics differential splicing	102
4.3.3 A role for the free cysteine residue on hCD33m in influencing its cellular localization	104
4.3.4 Development of an artificial system for transforming hCD33M to hCD33m.....	107
4.3.5 Development of cell-based reports of exon-2 skipping.....	109
4.4 Discussion	115
4.5 Conclusions	118
4.6 Methods	119
4.6.1 Generation of hCD33m-specific hybridomas.....	119
4.6.2 Expression, labelling, and isotyping of hCD33m-specific antibodies.....	121
4.6.3 Isolation of adult mouse microglia	121
4.6.4 Mutagenesis on C169A for hCD33M and C42A for hCD33m	122
4.6.5 Cloning of RP172 vectors	122
4.6.6 Lentivirus production	123

4.6.7 Lentiviral transduction of U937 cells	123
4.6.8 Flow cytometry	123
4.6.9 Phagocytosis assessment in U937 cells	123
4.6.10 Intracellular staining	124
4.6.11 Nano-Glo® Luciferase Assay for Construct I	124
4.6.12 Nano-Glo® Dual-Luciferase Assay for Construct II	125
4.6.13 Statistical analyses	125
<i>Chapter 5 Conclusions and Future Directions.....</i>	<i>127</i>
<i>References</i>	<i>132</i>
<i>Appendix A. DNA sequences of all constructs used in this thesis</i>	<i>145</i>

List of Tables

<i>Table 1.1 Discovered sialoside specificity for Siglecs through glycan microarrays.....</i>	<i>8</i>
<i>Table 3.1 The known substrates for each CHST.</i>	<i>64</i>
<i>Table 3.2 RNA expression of CHSTs in the selected cancer cell lines.</i>	<i>84</i>
<i>Table 3.3 Ligands for Siglecs</i>	<i>90</i>

List of Figures

<i>Figure 1.1: The Siglec family and their expression pattern across immune cells.</i>	<i>3</i>
<i>Figure 1.2: Schematic of a Siglec and its different parts.</i>	<i>4</i>
<i>Figure 1.3: The localization of Siglecs is the key for Siglecs to antagonize immune responses. 6</i>	
<i>Figure 1.4: General types of sialosides in mammals.</i>	<i>7</i>
<i>Figure 1.5: The two types of interactions between Siglecs and their glycan ligands.</i>	<i>10</i>
<i>Figure 1.6 Schematic of a Siglec-Fc.</i>	<i>11</i>
<i>Figure 1.7 Alternatives to Siglec-Fc recombinant soluble proteins for probing Siglec-glycan interactions.</i>	<i>13</i>
<i>Figure. 1.8: The key target gene or inhibitors for blocking the three major glycosylation pathways.</i>	<i>18</i>
<i>Figure 1.9: CD33 polymorphisms implicated in AD susceptibility.</i>	<i>20</i>
<i>Figure 1.10: rs12459419 modulates the ratio of the two isoforms of CD33.</i>	<i>21</i>
<i>Figure 1.11: Schematic of the two human isoforms of CD33 implicated in AD susceptibility. 23</i>	
<i>Figure 1.12: The history of the ligand discovery for CD33.</i>	<i>24</i>
<i>Figure 1.13: Schematic showing the major differences between hCD33 and mCD33.</i>	<i>25</i>
<i>Figure 2.1 CaR-ESI-MS for discovering glycan libraries against Siglec-Fc.</i>	<i>32</i>
<i>Figure 2.2 Components of a new versatile Siglec-Fc chimera.</i>	<i>33</i>
<i>Figure 2.3 Preparation of highly pure CD33-Fc.</i>	<i>34</i>

<i>Figure 2.4 Scheme for CaR-ESI-MS</i>	<i>35</i>
<i>Figure 2.5 CaR-ESI-MS with CD22-Fc and a defined glycan library.</i>	<i>36</i>
<i>Figure 2.6 CaR-ESI-MS with CD33-Fc and a defined glycan library.</i>	<i>37</i>
<i>Figure 2.7 The effect of two glycosidases on two CD33-Fc constructs expressed in WT and Lec-1 CHO cells.....</i>	<i>39</i>
<i>Figure 2.8: Generation of a homogeneous CD33 fragment.</i>	<i>40</i>
<i>Figure 2.9 Affinity measurement of CD33 with its high affinity ligand.....</i>	<i>41</i>
<i>Figure 2.10: Affinity measurement of CD33 with its K_d determination of α2-3 or α2-6 linked sialyl LacNAc with the CD33 fragment.</i>	<i>42</i>
<i>Figure 2.11: CD33 is masked by both α2-3 or α2-6 linked sialic acids.</i>	<i>44</i>
<i>Figure 2.12: CD33-Fc binds to U937 cells in NeuA- and NeuS-dependent manners.</i>	<i>45</i>
<i>Figure 2.13: Mass spectrum N-glycans derived from U937 cells.</i>	<i>46</i>
<i>Figure 2.14: CaR-ESI-MS spectra for CD22-Fc against N-glycans from U937 cells.....</i>	<i>47</i>
<i>Figure 2.15: CaR-ESI-MS spectra for CD33-Fc against N-glycans from U937 cells.....</i>	<i>48</i>
<i>Figure 3.1 Characterization of CHST-overexpressing U937 cells.</i>	<i>65</i>
<i>Figure 3.2 Probing CHST-overexpressing U937 cells for Siglec ligands.</i>	<i>67</i>
<i>Figure 3.3 mSiglec binding to the CHST transduced cells.</i>	<i>68</i>
<i>Figure 3.4 Sialic acid-dependent Siglec binding to CHST overexpressing cells.....</i>	<i>70</i>
<i>Figure 3.5. Inhibitors of cellular glycosylation reveal a cell type-specific pattern of Siglec ligands.....</i>	<i>72</i>

<i>Figure 3.6 The effect of Kifunensine or benzyl-α-GalNAc on CD33, Siglec5/14, and Siglec-7 binding to four different cell lines expressing CHST1.....</i>	<i>73</i>
<i>Figure 3.7 Mapping out the novel O-glycosylation sites on the CD33 fragment.....</i>	<i>75</i>
<i>Figure 3.8 Generation of homogeneous CD33 fragment and determination of dissociation constant (K_d).....</i>	<i>76</i>
<i>Figure 3.9 Comparison of the T239A/T260A CD33 fragment with the WT CD33 fragment... </i>	<i>77</i>
<i>Figure 3.10. Evaluation of the level of disulfated sialosides by G270-16 and its effects on Siglec binding.....</i>	<i>78</i>
<i>Figure 3.11 CD33L-liposome binding assay on the CHST1, CHST2, and CHST1/2 overexpressing U937 cells.....</i>	<i>79</i>
<i>Figure 3.12 Upregulated CHST1 expression in solid tumours correlates with poor patient outcome in several cancer types.</i>	<i>80</i>
<i>Figure 3.13 Kaplan-Meier 5-year overall survival curves of 19 cancer types based on CHST1 expression.....</i>	<i>81</i>
<i>Figure 3.14 Inhibiting carbohydrate sulfation on cancer cells decreases Siglec ligands.....</i>	<i>83</i>
<i>Figure 3.15 Assessing cell viability following NaClO₃ treatment.</i>	<i>83</i>
<i>Figure 3.16. Inhibition of endogenous carbohydrate sulfation leads to the decrease of the binding of Siglec-7, 5/14, and 15 in a panel of cancer cell line.....</i>	<i>85</i>
<i>Figure 3.17. Assessing Siglec-Fc binding to cell lines expressing low levels of CHST1 and CHST2.....</i>	<i>86</i>

<i>Figure 3.18. Treatment of cancer cell lines with NaClO₃ does not alter staining with SNA or MAA.....</i>	<i>86</i>
<i>Figure. 3.19 KS can be extended from different types of glycosylation.....</i>	<i>88</i>
<i>Figure. 3.20 The upregulated carbohydrate sialylation and sulfation in cancer may be able to exploit Siglecs for immune-evasion.....</i>	<i>89</i>
<i>Figure 4.1 Antibody staining of the two isoforms of human CD33.....</i>	<i>100</i>
<i>Figure 4.2 Development of two anti-hCD33m monoclonal antibodies.</i>	<i>101</i>
<i>Figure 4.3 Effect of DDT on S503 and S823 staining of hCD33m.</i>	<i>102</i>
<i>Figure 4.4 Intracellular staining of S503 reveals that hCD33m is localized inside mouse microglia.....</i>	<i>102</i>
<i>Figure 4.5 Recapitulation of alterative splicing at the rs12459419 locus with CD33 minigene constructs.....</i>	<i>104</i>
<i>Figure 4.6 The effect of an interdomain disulfide bond on S503 recognition.</i>	<i>105</i>
<i>Figure 4.7 A novel system for studying rs12459419 and evaluation of C42A mutation.....</i>	<i>106</i>
<i>Figure 4.8 3xFLAG-tag helps to evaluate the total expression level of hCD33m.....</i>	<i>107</i>
<i>Figure 4.9 Transforming hCD33M into hCD33m through Cre recombination</i>	<i>108</i>
<i>Figure 4.10 A reporter system for discovering drug candidates that induce exon-2 skipping.</i>	<i>110</i>
<i>Figure 4.11 Validation of splicing reporter Construct I.....</i>	<i>111</i>
<i>Figure 4.12 Validation of reporter splicing Construct II.</i>	<i>112</i>

<i>Figure 4.13 Validation of splicing reporter Construct III.</i>	<i>114</i>
<i>Figure 4.14 An inducible system for turning hCD33M to hCD33m.</i>	<i>118</i>

List of Abbreviations

Abbreviations

2xSTOP	Two Stop Codons
3SNL	α 2-3 Sialyl LacNAc
3xFLAG-tag	Triple FLAG tag
6SNL	α 2-6 sialyl LacNAc
AMLs	Acute-Myeloid-Leukemias
BCR	B-Cell Receptor
BRCA	Breast Invasive Carcinoma
CHO	Chinese Ovary Hamster
CHST	Carbohydrate Sulfotransferase
CID	Collisional Induced Dissociation
CV	Column Volume
DAMPs	Damage-Associated Molecular Patterns
Endo H	Endoglycosidase H
ESI	Electrospray Ionization
Fluc	Firefly Luciferase
Gal	Galactose
GalNAc	N-Acetyl galactosamine
GBM	Glioblastoma Multiforme
GBPs	Glycan Binding Proteins
Glc	Glucose
GlcNAc	N-Acetyl Glucosamine
GWAS	Genome Wide Association Studies

hCD33M	Human Siglec-3 Long Isoform
hCD33m	Human Siglec-3 Short Isoform
HTS	High-throughput Screening
ITAMs	Immunoreceptor Tyrosine-based Activatory Motifs
LB	Lysogeny Broth
LIHC	Liver Hepatocellular Carcinoma
LOAD	Late-Onset Alzheimer's Disease
LUAD	Lung Adenocarcinoma
<i>m/z</i>	Mass to Charge Ratio
MAA	<i>Maackia amurensis</i> Lectin
Man	Mannose
mCD33	Mouse Siglec-3
Meso	Mesothelioma
MOE-ASO	2'-O-methoxyethylribose Antisense Nucleotides
Neu5Ac	5-N-acetyl-neuraminic acid
NeuA	Neuraminidase A
NeuS	Neuraminidase S
Nluc	NanoLuc Luciferase
OV	Ovarian Serous Cystadenocarcinoma
PAAD	Pancreatic Adenocarcinoma
PAMPs	Pathogen-Associated Molecular Patterns
PAPS	3'-Phosphoadenosine-5'-phosphosulfate
PGMs	Post-Glycosylational Modifications
Siglec	Sialic acid-binding immunoglobulin-like lectins
SLe ^x	Sialyl Lewis X

SNA	<i>Sambucus Nigra Agglutinin</i>
SNP	Single Nucleotide Polymorphism
STAD	Stomach Adenocarcinoma
Streptactin-AF647	Streptactin labeled with Alexa647
STs	Sialyl Transferases
TCGA	The Cancer Genome Atlas
TEV	<i>Tobacco Etch Virus</i>
UVM	Uveal Melanoma

Glossary of Terms

Alternative splicing	an RNA splicing process where introns are removed to give rise to the mature mRNA transcript.
Bioluminescence	a chemical reaction between luciferase enzyme and its substrate to produce light.
Carbohydrate sulfotransferases (CHSTs)	sulfotransferases transfer sulfates to glycoconjugates. It can enhance Siglec-ligand interactions.
<i>Cis</i> interactions	Siglec-sialoglycan interactions when Siglecs recognizes ligands on the same cell.
Collisional induced dissociation (CID)	A mass spectrometry technique to induce fragmentation of selected ions in the gas phase. Low-energy CID can dissociate non-covalent protein-ligand interactions.
Electrospray Ionization	a technique to generate ionic species in solution using an electrospray in which a high voltage is applied to a liquid to create an aerosol.
Endoglycosidase H (Endo H)	a glycosidase that cleaves asparagine-linked high-mannose type <i>N</i> -glycans.
Genome wide association study (GWAS)	A statistical approach to identify generic variants associated with a risk for a disease or a particular trait
Immunoreceptor tyrosine-based activatory motifs (ITAMs)	A conserved peptide sequence located in the cytoplasmic tails of signaling molecules and adaptors such as TYROBP. Signal activation can be initiated by the initial phosphorylation of tyrosine residues on these motifs.
Immunoreceptor tyrosine-based inhibitory motifs (ITIMs)	A conserved peptide sequence located in the cytoplasmic tails of many inhibitory receptors. when inhibitory receptors get activated, tyrosine residues on these motifs can get phosphorylated by Src family kinases leading to recruitment of phosphatases to inhibit signaling.
Late-onset Alzheimer's Disease (LOAD)	A common form of Alzheimer's disease where symptoms become apparent in the mid-60s or later. Two AD pathology hallmarks are accumulations of amyloid plaques and neurofibrillary tangles.
N-glycosidase F (PNGase F)	an amidase, which cleaves between the innermost GlcNAc and asparagine residues of all types of <i>N</i> -glycans.
Sialic acid	nine-carbon keto sugars terminate <i>N</i> -, <i>O</i> -glycans and gangliosides, which is essential for being Siglec ligands.
<i>Trans</i> interactions	Siglec-sialoglycan interactions when Siglecs recognizes ligands on the other cells.

Chapter 1

Introduction

1.1 Introduction of Siglecs

The response of immune cells is fine-tuned by many activatory and inhibitory receptors that have opposing effects on cell signaling. Improper control of these signals can be a driver of autoimmune disease, which occurs when an undesired immune response towards the tissues and cells of our own bodies takes place.¹ Among the inhibitory receptors that help maintain immunological tolerance towards our own bodies, Sialic acid-binding immunoglobulin-like lectins (Siglecs) are a 15-membered family, which recognize sialic acid containing glycan conjugates (sialosides). Siglecs are expressed primarily on a variety of white blood immune cells.² (**Figure 1.1a**) Each type of white blood cell has a unique expression pattern for Siglecs.³ (**Figure 1.1b**) A leading hypothesis for the role of Siglecs is that they help to prevent immune response toward “*self*” antigens. For example, a study on aging C57BL/6 mice showed that Siglec-G deficiency led to the development of an autoimmune phenotype characterized by increased autoantibody levels and resulting mild glomerulonephritis.⁴ Moreover, Siglecs can antagonize diverse immune responses mediated by pathogen-associated molecular patterns (PAMPs) or damage-associated molecular patterns (DAMPs).⁵ For instance, microglia – the major immune cell in the brain – lacking Siglec-E secreted higher levels of proinflammatory cytokines in response to LPS-mediated stimulation.⁶ These examples and many others support an immunosuppressive role for Siglecs in a variety of immune responses.⁷

1.1.1 The functional parts that make up a Siglec

Siglecs are type 1 membrane bound glycoproteins and consist of four major parts: a ligand binding domain (V-set domain), a variable number of immunoglobulin-like domains (C2 domains),

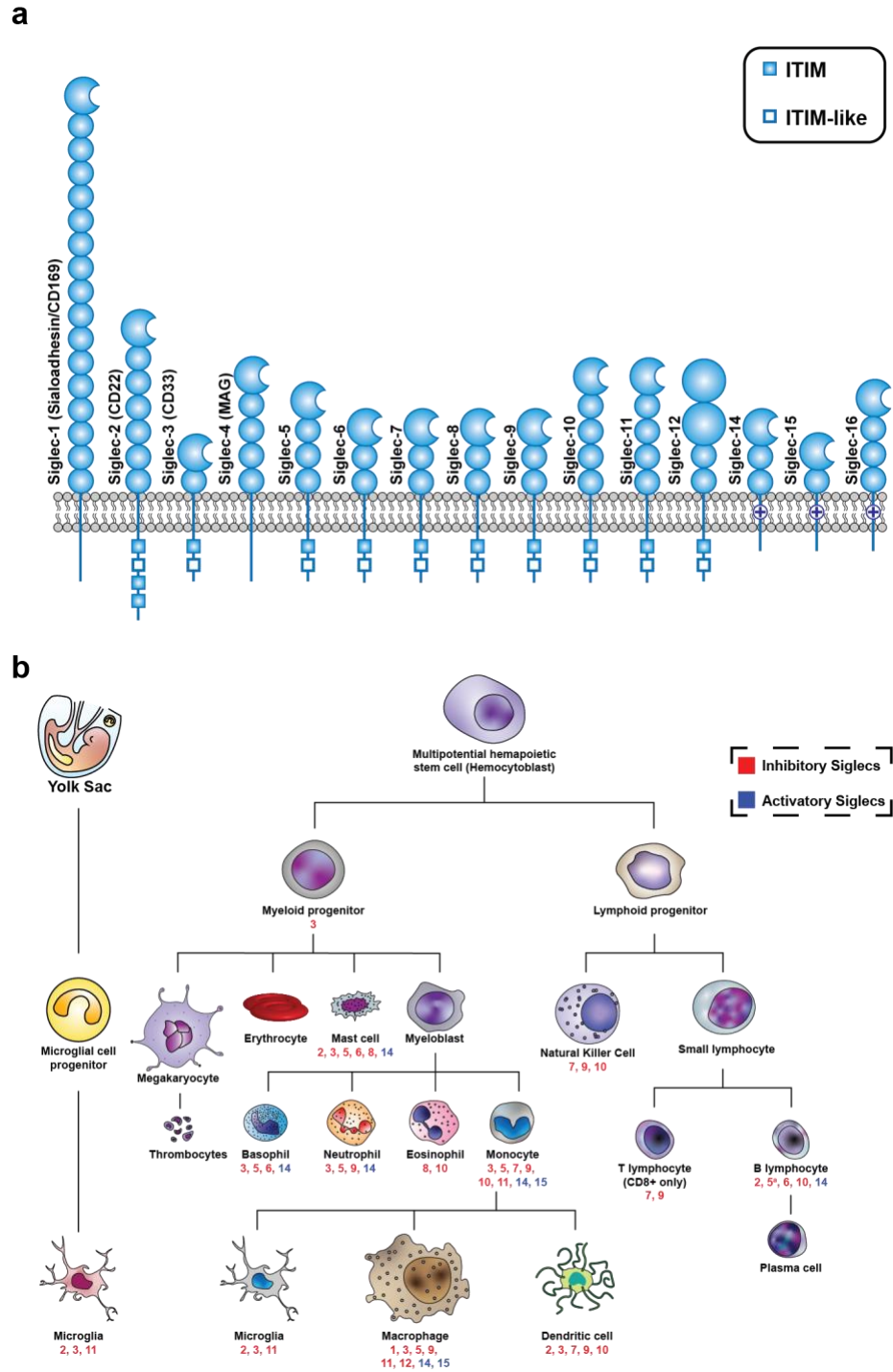


Figure 1.1: The Siglec family and their expression pattern across immune cells. (a) The Siglec family consists of 15 members of cell surface protein receptors belonging to immunoglobulin superfamily. (b) Siglecs are differentially expressed on numerous immune cells. For each type of white blood cell, the number indicates each Siglec that is expressed. Red color represents inhibitory Siglecs and blue color represents activatory Siglecs.

a transmembrane segment, and a cytosolic tail containing signaling motifs. (**Figure 1.2**) The V-set domain is composed of anti-parallel β sheets and loops forming a shallow groove where sialic acid-containing glycans can bind. Within the binding groove, a key ionic interaction between the negatively charged carboxyl group on sialic acid and the positively charged guanidinium group on an essential and conserved arginine residue is arguably the most important for ligand binding.⁸ Evidence for the importance of this interaction is that Siglec-glycan binding is typically completely lost upon mutation of this arginine. Siglec-sialoside interactions help cell adhesion, engagement of Siglecs near immunological synapse, and cargo internalization.^{2,9} Siglecs have a varied number of the conserved C2 domains from one domain for Siglec-3 (CD33) and Siglec-15, to sixteen domains for Siglec-1. These extracellular domains can be glycosylated,¹⁰ and the number of C2 domains can influence how far the Siglec protrudes from the cell surface, which in turn affects its ability to interact with glycan ligands on the same cell versus another cells.¹¹

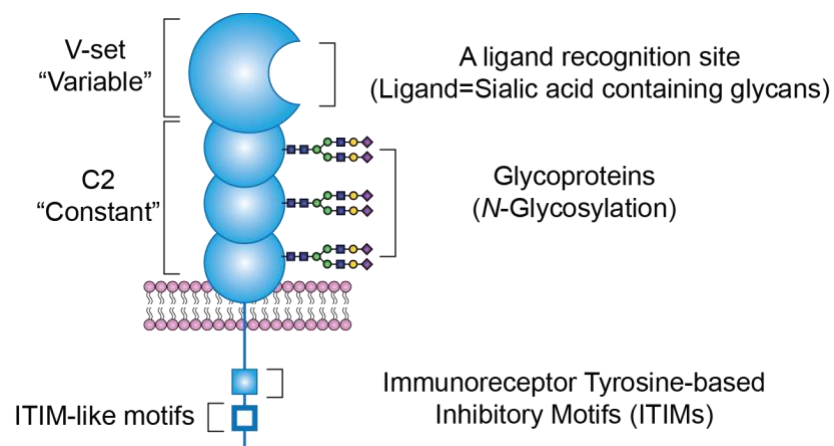


Figure 1.2: Schematic of a Siglec and its different parts. The V-set domain of Siglecs mediates ligand binding, which is crucial for cell surface localization. Each Siglec has varied number of C2 constant domains. After a transmembrane segment, a cytosolic tail of Siglec contains ITIM and/or ITIM-like motifs, which can recruit phosphatases to break the signal cascades under certain circumstances.

In humans, cytosolic tails of most Siglecs contain immunoreceptor tyrosine-based inhibitory motifs (ITIMs) and ITIM-like motifs, which characterizes Siglecs as inhibitory

checkpoint receptors.¹² (**Figure 1.2**) Numerous antigenic molecules can stimulate activatory receptors on various immune cells leading to the initial phosphorylation of immunoreceptor tyrosine-based activatory motifs (ITAMs) by Src-family kinases, such as Lyn, Syk, and Fyn.¹³ This early phosphorylation further recruits a series of kinases required for signal activation. (**Figure 1.3a**) Siglecs can be engaged in this activatory signaling through interactions with sialosides. When Siglecs are colocalized near activatory receptors in an immunological synapse, ITIMs at their cytosolic tails can also get phosphorylated by Src-family kinases in company with ITAM. The following phosphorylated ITIMs create a docking site for SHP-1 or -2 phosphatases, which can bring about the recruitment of SHP-1 or -2 phosphatases and signal inhibition.¹² These two phosphatases sequentially deactivate signal cascades by hydrolyzing phosphate groups on phosphotyrosine residues on the ITAMs.⁷ (**Figure 1.3b**) In contrast, Siglec-14, -15, and -16 have a distinct transmembrane segment containing either a lysine or arginine residue, which requires them to pair with adaptor proteins such as DAP-12 (also called TYROBP). In mice, CD33, Siglec-H, and Siglec-15 have this feature. DAP-12 has a transmembrane peptide sequence containing aspartic acid residue that stabilizes the cationic amino acid residues within the transmembrane segment of these Siglecs. DAP-12 carries ITAM on its cytosolic tail, so the Siglec-DAP-12 complex have a potential to transduce activatory signals.¹⁴

1.1.2 Ligands for Siglecs

Glycosylation is a major biosynthetic process to generate oligosaccharides containing sialic acids (Neu5Ac), such as *N*- or *O*-glycans and gangliosides, which can be ligands for Siglecs. Monosaccharides such as glucose (Glc), galactose (Gal), mannose (Man), *N*-acetyl-glucosamine (GlcNAc), *N*-acetyl-galactosamine (GalNAc) can make up a variety of glycan structures and Neu5Ac is typically the last monosaccharide added to the non-reducing end. Notably, twenty

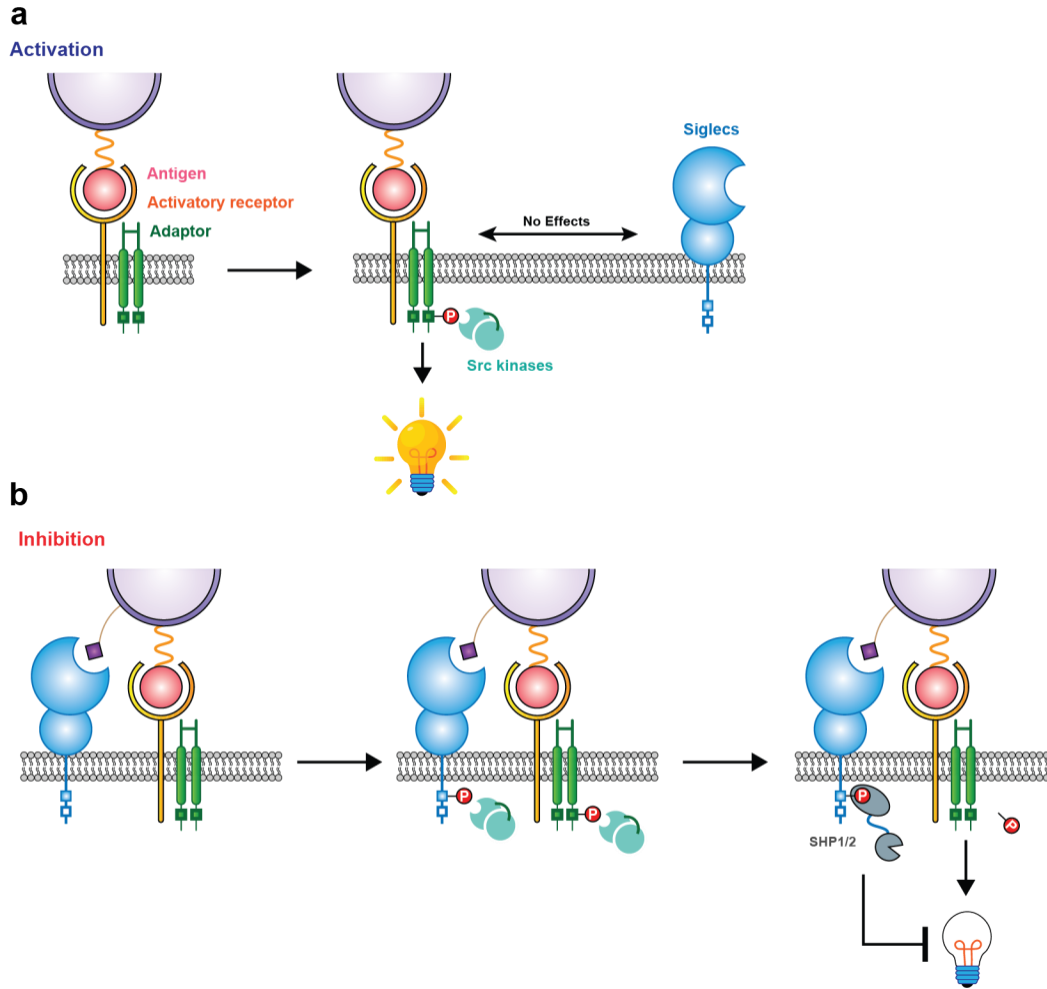


Figure 1.3: The localization of Siglecs is the key for Siglecs to antagonize immune responses. (a) When activatory receptors become stimulated by antigens, the phosphorylation of ITAMs by Src kinases takes place. This initial phosphorylation on ITAMs stimulates successive kinase activation mediating a signal cascade. (b) The presentation of sialic acid-containing glycans on the other cells can recruit Siglecs near activatory receptors. The Src kinases can also phosphorylate the ITIMs at Siglecs' cytoplasmic tails, which lead to the recruitment of SHP1/2. Consequently, these phosphatases hydrolyze phosphorylated tyrosine residues on ITAMs and the ITAM-mediated signal cascades become suppressed.

different sialyltransferases (STs) in human use CMP-Neu5Ac as their donor to transfer Neu5Ac to glycan acceptors terminated with Gal, GalNAc, or Neu5Ac residues.^{15,16} Sialic acids can be linked to 3- or 6- hydroxy group on Gal, 6-hydroxy group on GalNAc and 8-hydroxy group on Neu5Ac in mammals. The position of hydroxyl group on Gal, GalNAc, or Neu5Ac residues where sialic acid is linked defines three different linkages: α 2-3, α 2-6, and α 2-8, which is critical for

determining specificity of each Siglec. (**Figure 1.4**). For example, Siglec-2 and Siglec-10 have strong preference to α 2-6 linked sialoside while Siglec-1 prefer α 2-3 linkage and Siglec-7 binds α 2-8 linkage more strongly than the other linkages.⁷ (**Table 1.1**) Subsidiarily, fucosylation and post-glycosylational modifications (PGMs) - such as *O*-sulfation, and *O*-acetylation - also influence affinity.^{17,18}

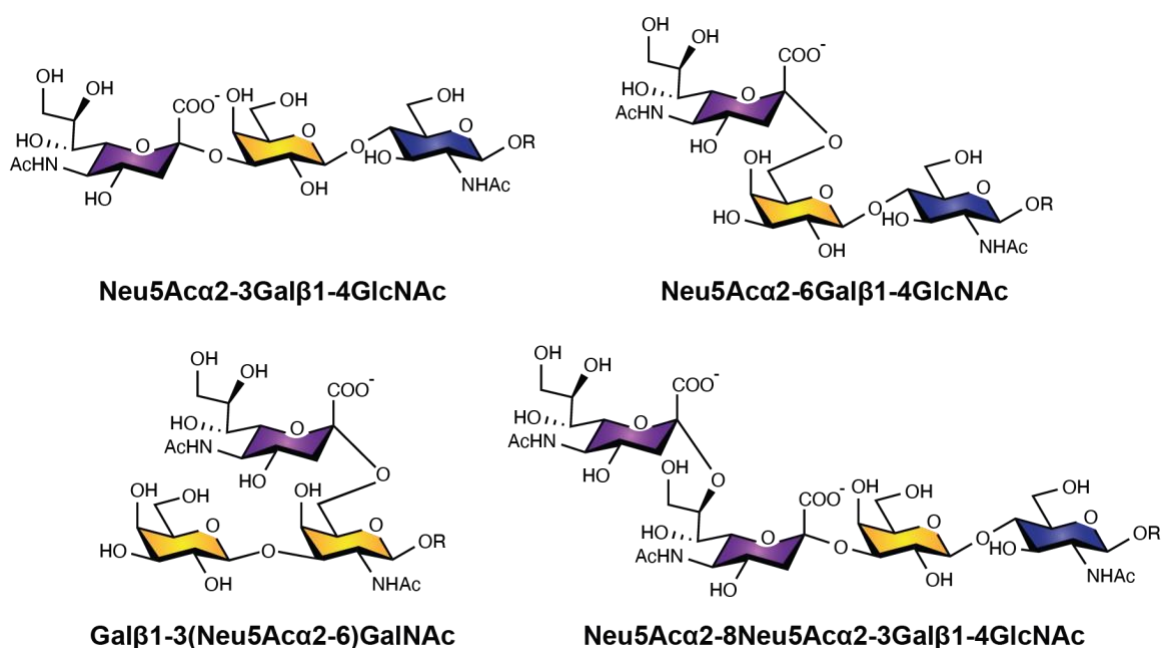











Figure 1.4: General types of sialosides in mammals. There are three major linkages of sialosides: α 2-3, α 2-6, and α 2-8, which is a determinant factor for ligand specificity of Siglecs.

1.1.3 Examples of the effects of Siglec-glycan interactions on immune responses

Sialosides can modulate the function of Siglecs to maintain immune balance. For example, self-associated molecular patterns (SAMPs) mediated by endogenous sialosides are one mechanism preventing immune responses toward ‘self’ antigens.⁵ On the other hand, the roles of Siglecs as a “brake” can be exploited by sialylated pathogens or cancer cells to evade immune clearance.¹⁹⁻²¹ A primary mechanism for escape of immune destruction of cancer is the ability of

cancer cells to exploit inhibitory receptors (also called immune checkpoint inhibitors), to shut down immune cell function.²² Despite the crucial role of Siglecs in maintaining immune homeostasis in disease and health, glycan ligands for Siglecs are not completely understood. Studying the terminal epitopes of sialosides was mainly focused due to the complexity and diversity of the advanced glycan structures. Nevertheless, there have been efforts for discovering ligands for Siglecs with simple *N*-, *O*-mucin-type, *O*-mannose glycans, and gangliosides.²³⁻²⁶ To advance a deeper understanding of Siglec ligands, novel strategies and tools are required for probing Siglec-ligand interactions.

Table 1.1 Discovered sialoside specificity for Siglecs through glycan microarrays.

Name	Ligands
Siglec-1	
Siglec-2	
Siglec-3	
Siglec-4	
Siglec-5	
Siglec-6	
Siglec-7	
Siglec-8	
Siglec-9	

1.1.4 The two key types of Siglec-glycan interactions

A primary means by which the glycan ligands of Siglecs impact the function of Siglecs is by mediating the spatial proximity of Siglecs to other activatory receptors.² Siglec-glycan interactions can take place in *cis* or *trans*. *Cis* interactions occurs when Siglecs recognize their endogenous ligands on the same cell surface. **(Figure 1.5a)** An example for the biological importance of *cis* interactions is CD22 (Siglec-2), which is highly expressed on B-cells. *Cis* ligand interactions keep CD22 within clusters,²⁷ which prevents CD22 from repressing B-cell receptor signaling. In contrast, *trans* interactions are those in which Siglecs recognize glycan ligands on another cell. **(Figure 1.5b)** *Cis* interactions can ‘mask’ Siglecs, which means when binding sites of Siglecs are preoccupied with *cis* ligands. *Masked* Siglecs have limited capacity for interacting with *trans* ligands. For example, depleting *cis* ligands on COS cells expressing CD33 with sialidase led to increased adhesion to HL-60 cells.^{28,29} *Trans* interactions have potential to locate Siglecs near activatory receptors, which can lead to immune suppression. For example, CD22 recruitment near B-cell receptors (BCR) through a multivalent scaffold displaying glycan ligands for CD22 and antigens for inducing BCR activation was shown to suppress the BCR activation.³⁰ Accordingly, Siglecs can modulate immune system through *cis* or *trans* interactions, which can help relocate Siglecs around clustered activatory receptors to suppress activatory signal pathway.²

1.1.5 Therapeutic aspects of Siglecs in cancer

Hypersialylation of cancer cells is a key mechanism of exploiting Siglecs to evade immune clearance.²¹ Hence, blocking the engagement of Siglecs near the signaling synapse has become one of attractive therapeutic applications to improve antibody-mediated cancer therapies. For example, sialidase-conjugated anti-HER2 antibody has been shown to improve anticancer therapies, whose mechanism is to block the engagement of Siglecs by depleting sialic acids in the

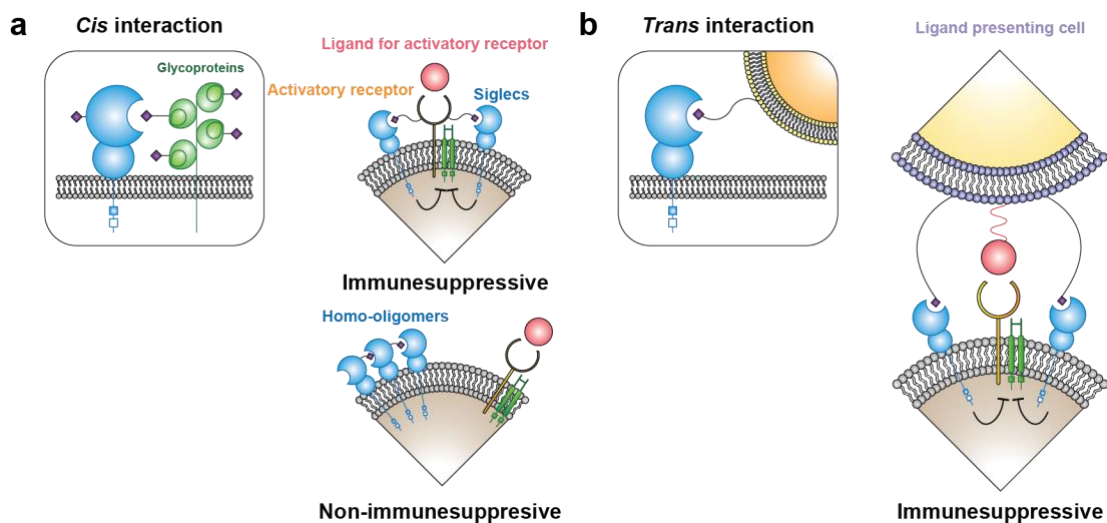


Figure 1.5: The two types of interactions between Siglecs and their glycan ligands. (a) *Cis* interactions have the potential to recruit Siglecs near activatory receptors to inhibit signaling at normal state. (b) *Trans* interactions can also recruit Siglecs to activatory receptors in the context of an immunological synapse to inhibit immune signaling.

signaling synapse.³¹ Alternatively, targeted protein degradation by attaching cancer specific nanobody to mucin-selective protease has shown to strip *O*-glycans on tumour, which led to promoted cell death in the cell culture model and reduced tumour size in mice model.³²

1.2 Soluble versions of Siglec to probe their glycan binding

Transmembrane proteins account for more than 40% of druggable targets but studying their interaction with ligands presents challenges due to their hydrophobic transmembrane segment. Indeed, transmembrane proteins are generally insoluble in water and hard to crystalize for x-ray crystallographic studies.³³ Detergent can help dissolve these hydrophobic proteins in water, but it can also denature the tertiary structures of proteins, which can lead to functional loss.³⁴ To resolve these intrinsic problems, combining the functionally soluble ligand-binding domains of the transmembrane proteins with the constant Fc region of antibody has been widely-used. This

approach has applied to Siglecs to investigate their structures, post-translational modifications, functions.³⁵

1.2.1 Siglec-Fc chimeric fusion proteins

Fc chimeric proteins are composed of three major parts. At the DNA level, the 5'-terminal sequence generally encodes an optimized signal peptide and a V-set domain and one or two C2 domains of Siglecs devoid of transmembrane region, followed by nucleotide sequence encoding a linker peptide to stabilize the connection between the extracellular domains of Siglecs and Fc constant domains of antibody at C-terminal.³⁶ (**Figure 1.6**) The last part is the DNA sequence encoding the IgG Fc domains, which notably have a *N*-glycosylation site at Asn297.

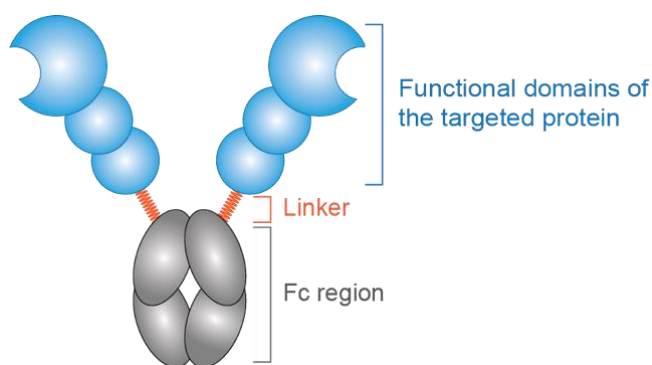


Figure 1.6 Schematic of a Siglec-Fc. The Fc recombinant chimeric proteins consist of extracellular domains of Siglecs lacking transmembrane sequence, a linker peptide, and the Fc region of IgG. The native or optimized signal peptides can be located at N-terminus of the fusion proteins.

1.2.2 The advantages of the Fc chimera

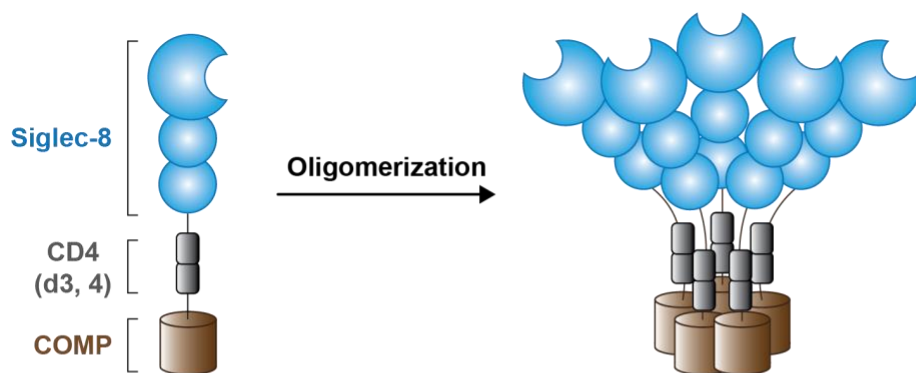
The Fc domains confer several benefits to Fc chimeric proteins. First, the Fc domains improve the stability of the chimeric proteins in solution. The high-affinity interaction between the

IgG Fc and protein A or G enables protein purification from the cultured supernatants.³⁷ Notably, the dimeric conformation of the IgG Fc domains enhances the avidity of the targeted proteins, which is feasible for investigating the physical interactions between Siglecs and glycan ligands. By taking advantages of IgG-Fc γ receptor interactions, Fc chimeric proteins can be designed to mark specific pathogenic cells to induce immunogenicity on innate immune cells, such as natural killer cells.³⁶ However, this IgG-Fc γ receptor interactions can be double-edged sword since these interactions can disturb immunostaining applications for cells or tissues.³⁸ The choice of IgM as the Fc portion instead of IgG can enable multivalent presentation of the targeted proteins. Alternatively, high multivalency with Siglec-Fc can be achieved by interaction between Siglec-Fc and protein-A or G-coated beads.³⁵ These multivalent presentations have a potential to study Siglecs, most of which have generally very weak binding affinity.¹² This multivalent presentation has been helpful for many applications such as microarray, immunohistochemistry, western blot, and flow cytometry.³⁶ Therefore, this strategy combining Siglecs with the IgG Fc domains has been and continues to be useful in a wide variety of applications.

1.2.3 Alternatives to Siglec-Fc

Engaging Fc γ receptor by Siglec-Fc is a key drawback in its applications. To avoid undesired interactions with Fc γ receptors and take advantage of multivalency, two alternative strategies have been reported. First, the first three domains of Siglec-8 were fused to two IgG domains from CD4 and a spontaneously pentamerizing domain from cartilage oligomeric matrix

a



b

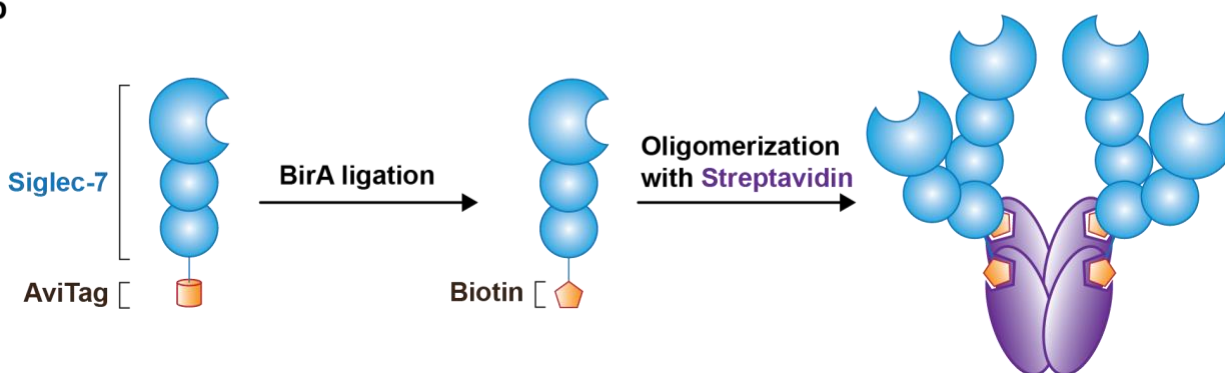


Figure 1.7 Alternatives to Siglec-Fc recombinant soluble proteins for probing Siglec-glycan interactions. (a) C-terminal COMP enables Siglec-8 to be pentamerized for enhancing avidity. (b) Biotinylation on AviTag by BirA ligase enables Siglec-7 to be precomplexed with Streptavidin in a tetramer.

protein with C-terminal His₆ tag, which is named Siglec-8-COMP. Siglec-8-COMP increases avidity through pentameric presentation and it was validated by glycan microarray.³⁹ (**Figure. 1.7a**) In another example, the first three domains of Siglec-7 containing biotinylation peptide sequence (AviTag) and His₆ tag at C-terminal was expressed in HEK293T cells. Subsequently, this Siglec-7 construct was biotinylated by BirA ligase, which helps to form tetrameric presentation with streptavidin. (**Figure. 1.7b**) This Siglec-7 and streptavidin tetramer was validated through glycan microarray and applied to flow cytometry with primary B cells, T cells, and NK cells to evaluate expression level of Siglec-7 ligands.⁴⁰

1.3 Probing the glycan ligands of Siglecs with Siglec-Fc proteins

Siglec-1 (CD169) is one of the first discovered member of Siglec family.^{41,42} Siglec-2 (CD22) and Siglec-3 (CD33) are two more of the oldest and most studied Siglec. While the strict specificity of Siglec-2 for α 2-6 linked sialosides and biological roles for CD22 have been advanced,⁴³ there is still much to be learned about CD33. Siglec-Fc proteins have been instrumental in elucidating the specificity of Siglecs. Here, the methods for studying Siglec ligands are reviewed, with focus on microarrays and flow cytometry.⁴²

1.3.1 Glycan microarrays

Glycan microarrays are a tool for investigating interactions between glycan binding proteins (GBPs) and multiple glycans or glycoconjugates. A glycan microarray is composed printing glycans or glycoconjugates onto a solid support that can be probed with a GBP. The contents of glycan libraries are critical for the high odds of discovering ligands for GBPs. Glycan libraries can be synthetically prepared glycans, consisting of *N*- or *O*-glycans, gangliosides, glycosaminoglycans, and synthetic glycans including PGMs like acetylation, sulfation and phosphorylation.⁴⁴ Alternatively, glycans for libraries can be isolated and fractionated from natural sources.⁴⁵

Using glycan microarrays with Siglec-Fc proteins have led to many discoveries about the glycan ligands recognized by each Siglec. Overall, glycan microarrays have been widely used to discover glycan ligands for Siglecs. (**Table 1.1**) In 2000, Varki and co-workers reported the binding of Siglec-1, -2, -3, -4, -5, and -6 towards ten sialosides on a glycan microarray.¹⁸ The sialosides were composed of basic di- or tri-saccharides, such as α 2-3 or α 2-6 sialyl Lac or sialyl LacNAc. CD22 showed a strong α 2-6 preference and Siglec-1 showed a preference for α 2-3 over α 2-6 sialosides. Siglec-3 bound both linkages, but α 2-6 sialosides displayed moderately higher

signal. Siglec-5 bound both α 2-3 or α 2-6 linkages to a similar extent. Sialyl Tn antigen was the only ligand for Siglec-6 in this study.¹⁸ In 2004, a study led by Paulson and co-workers, the effect of α 2-6 sialic LacNAc with 6-*O*-sulfation at GlcNAc on Siglec-2 affinity was firstly discovered through glycan microarray with a library of 200 defined glycoconjugates.²⁵ In 2005, Schnaar and coworkers discovered that Siglec-8 has strong specificity to sialyl lewis X (SLe^X) structure with 6-*O*-sulfation at Gal.⁴⁶ In the following year, sulfation on SLe^X as a factor for improving affinity was studied with hSiglec-7, -8, -9, and mSiglec-2 and -F.⁴⁷ Sulfation on 6-*O*-Gal in α 2-3 sialyl LacNAc moiety enhanced Siglec-7, and -8 and mSiglec-F binding greatly while Sulfation on 6-*O*-GlcNAc improved Siglec-9 and mSiglec-2 binding.⁴⁷ In 2014, Varki and co-workers studied Siglec-3, -5, and -9 in three different species (human, chimpanzee, and baboon). Specificity for each Siglec was evaluated on α 2-3 and α 2-6 linkages, effects of Neu5Gc, 9-*O*-acetylation and 6-*O*-sulfation on GlcNAc were analyzed. Siglec-3, and -5 bound to both linkages at similar extent while Siglec-9 slightly prefer to α 2-6 linkage. it was turned out that 9-*O*-acetylation can be disadvantageous for Siglecs. Siglec-9 showed sulfation dependent binding enhancement consistent with the 2006 study.⁴⁷ Sialosides containing Neu5Gc had higher signal rank than the one with Neu5Ac in Siglec-3, -5, and -9.⁴⁸ In 2017, Schnaar and co-workers investigated different ligand binding patterns in two pairs of functional paralog between hSiglec-8 and mSiglec-F and functional ortholog between hSiglec-9 and mSiglec-E through glycan microarray. The difference emerged where hSiglec-8 and mSiglec-F preferred SLe^X with 6-*O*-sulfation at Gal while hSiglec-9 and mSiglec-E favored SLe^X with 6-*O*-sulfation at GlcNAc.^{26,46} Identifying the difference between paralogs or orthologs in their ligand recognition is critical for studying mouse model as Siglec-8 binding was abrogated in the absence of 6-*O*-sulfation at Gal but Siglec-F binding was still retained, which might be because of its broader range of glycan ligands. In 2019, Cummings

and co-workers screened twelve Siglecs against the glycan microarray printed with multiantennary *N*-glycans and only Siglec-1, -2, -9, and -10 generated signals. In this study, Siglec-1, and -9 preferred α 2-3 linked *N*-glycans while Siglec-2, and -10 bound to α 2-6 linked *N*-glycans higher.⁴⁹

1.3.2 Elucidation of Siglec ligands using multivalent ligand presentations

A challenge in discovering ligands for Siglecs is their intrinsically weak binding affinity. One way to circumvent this is using multivalency. To improve avidity in Siglec-glycan ligand interactions, streptavidin-alkaline phosphatases (SAAP) was developed. In principle, sialosides containing biotin at reducing ends formed a tetrameric presentation with SAAP, which was called sialoside-SAAP probe. In this study, a panel of 25 synthetic sialosides were designed to contain biotin with a different combination of aglycons and biotin spacers at reducing ends of each sugar.⁵⁰ Subsequently, sialosides with biotin were complexed with SAAP to form tetrameric displays. Next, a panel of Siglec-Fc chimeras (human Siglec-2, 3, 5, 7, 8, 9, and 10 and mouse Siglec-1, 2, and 4) were fixed onto microtiter wells and sialoside-SAAP complexes were incubated on each well. Subsequently, *p*-nitrophenyl phosphate was used for detecting the bound sialoside-SAAP. Both human Siglec-2 (hSiglec-2) and mouse Siglec-2 (mSiglec-2) exclusively preferred α 2-6 linked sialosides. hSiglec-3 bound both α 2-3 and α 2-6 sialosides at a similar degree. Siglec-5 presented α 2-3 sialyl LacNAc (3'SLN) specificity. hSiglec-8 recognized 3'SLN more than α 2-6 sialyl LacNAc (6'SLN). Under lower amount of Siglec-Fc, hSiglec-7 presented outstanding binding to Neu5Ac α 2-8Neu5Ac α 2-3Gal β 1-4GlcNAc. Siglec-10 bound both Neu5Gc α 2-6Gal β 1-4GlcNAc and Neu5Gc α 2-6Gal β 1-4Glc. (**Table 1.1**)

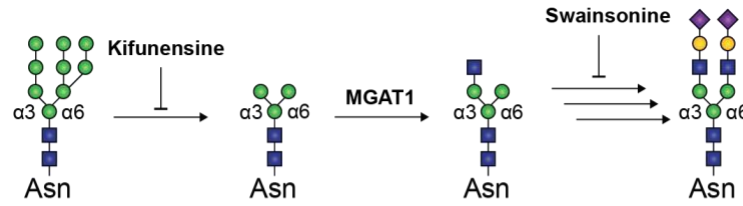
1.3.3 Cell- and tissue-based assay

Chemical synthesis of all possible glycan structures represented on mammalian cells is very challenging. Instead, introducing or deleting genes into cell lines provides an alternative way

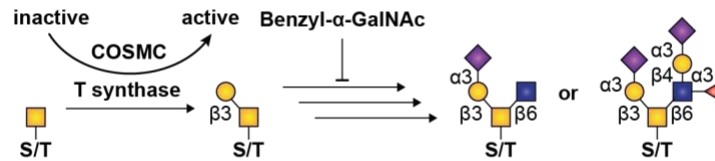
to probe a diversity of glycan structures. For example, Miyazaki *et al.* transfected the gene encoding fucosyltransferase 7 and carbohydrate sulfotransferase (CHST) 2 to ECV304 cells to enrich SLe^x with 6-*O*-sulfation at GlcNAc on the surface and binding of Siglec-7-Fc and Siglec-9-Fc to those cells was demonstrated by flow cytometry.⁵¹ In this study, consistent with the previous finding,^{26,47} Siglec-7 and -9 showed enhanced binding to the transfected cells where 6-*O*-sulfation at GlcNAc was abundant.⁴⁷ The earlier study used two antibodies recognizing Neu5Ac α 2-6Gal β 1-4(6S)GlcNAc (clone KN343) and Neu5Ac α 2-3Gal β 1-4(6S)GlcNAc (clone G152) to assess the importance of a 6'SLN moiety with 6-*O*-sulfation at GlcNAc as CD22 ligands. First, KN343 was shown to block the epitope of Siglec-2-Fc in a competitive binding assay. Furthermore, sodium chlorate was treated on KN343 reactive cells, which is an analogue inhibitor of a sulfate required for biosynthesis of 3'-Phosphoadenosine-5'-phosphosulfate (PAPS). As sodium chlorate reduced cellular carbohydrate sulfation, CD22 binding to the KN343 reactive cells was decreased. These results supported sulfation as an important modification for producing high-affinity ligands for CD22.^{25,52}

To assess which types of glycosylation matter in Siglec binding, several inhibitors perturbing glycosylation pathway have been developed and applied. In general, glycosylation is stepwise, so blocking early stages of biosynthesis can block maturation of glycans. Kifunensine is an alkaloid compound that acts as inhibitors of mannosidase, which blocks *N*-glycan maturation, thereby preventing the appears of hybrid- and complex *N*-glycans, keeping *N*-glycans in their high mannose state.⁵³ To block mucin type *O*-glycans, benzyl- α -GalNAc is widely used.⁵⁴ Inhibition of the biosynthesis for ganglioside can be made by Genz-123346, which deactivates UGCG enzymatic activity.⁵⁵ **(Figure 1.8)**

a N-Glycosylation inhibition



b O-Glycosylation inhibition



c Glycolipid inhibition

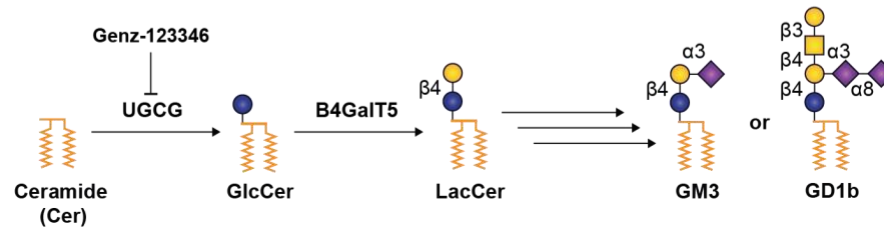


Figure. 1.8: The key target gene or inhibitors for blocking the three major glycosylation pathways. (a) Kifunensine blocks mannosidases essential for generating complex or hybrid types of *N*-glycans. Knocking out *MGAT1* maintains *N*-glycans in their high mannose state. (b) *COSMC* is the molecular chaperone essential for functional T synthase, so knocking out *COSMC* leads to inhibition of mucin type *O*-glycosylation. Alternatively, Benzyl-α-GalNAc can block the biosynthesis of mucin-type *O*-glycans. (c) Knocking out either *UGCG* or *B3GalT5* can lead to inhibition of ganglioside biosynthesis. Alternatively, a pharmacological inhibitor, Genz-123346 can suppress UGCG activity.

Instead of using glycosylation inhibitors, CRISPR/Cas9 provides a convenient way for knocking out targeted genes in cells. In general, CRISPR/Cas9 forms a complex with a guide RNA (gRNA), which helps CRISPR/Cas9 to be guided to the gene of interest in chromosome. This leads to disability of intact transcription by breaking double stranded DNA near gRNA hybridization area and inducing non homologous end joining.⁵⁶ The consequence is the formation of indel in the targeted gene causing frameshift, deletion, or mutations on nucleotide sequence. CRISPR has been successfully applied to blocking the expressions of numerous glycosyl transferases. To block

complex or hybrid types of *N*-glycosylation, *MGAT1* is widely knocked out.⁵⁷ *MGAT1* encodes alpha-1,3-mannosyl-glycoprotein 2-β-*N*-acetylglucosaminyltransferase transferring the first GlcNAc residue to *N*-glycan backbone where high mannose structures get trimmed down by a series of mannosidases. Knocking out *COSMC* can block *O*-linked glycosylation initiated by T synthase.⁵⁸ *COSMC* encodes a molecular chaperone helping T synthase fold properly, so *COSMC* KO leads to abundant Tn antigen expression on the surface. Lastly, to inhibit the biosynthesis of glycolipids, *UGCG* or *B4GALT5* can be targeted.⁵⁹ **(Figure 1.8)** These are the initial two key enzymes required for more advanced structures for gangliosides.

1.4 CD33 biology in Late-Onset Alzheimer's Disease

In 1990s, CD33 gained attention as a therapeutic target for acute-myeloid-leukemias (AMLs) since a wide range of AMLs derived from malignant stem cells are CD33 positive. Consequently, Lintuzumab and Gemtuzumab were developed as therapeutic antibodies for targeting CD33 as antibody-drug conjugate.⁶⁰⁻⁶² More recently, GWAS (genome wide association studies) have revealed the correlation between CD33 and late-onset Alzheimer's disease (LOAD) susceptibility. In the brain, CD33 is expressed in the major immune cells of the brain, which are called microglia, and regulates cell processes such as phagocytosis.^{63,64} Hence, its function in Aβ-associated pathology becomes great interest.⁶⁵

1.4.1 Genetics of the rs12459419 SNP in CD33

GWAS are meta-analyses of genetic variants that aim to discover variants of genes associated with disease susceptibility. Several different SNPs in *CD33* have been investigated for AD susceptibility **(Figure 1.9a)**. The link between CD33 and AD was first discovered through GWAS in 2008, wherein the rs3826656 SNP was identified as a susceptibility factor for LOAD.⁶⁶

Several follow up studies in larger cohorts led to the identification of another nearby SNP, rs3865444, that also correlates with AD susceptibility.⁶⁷⁻⁶⁹ In the brain, CD33 is predominantly

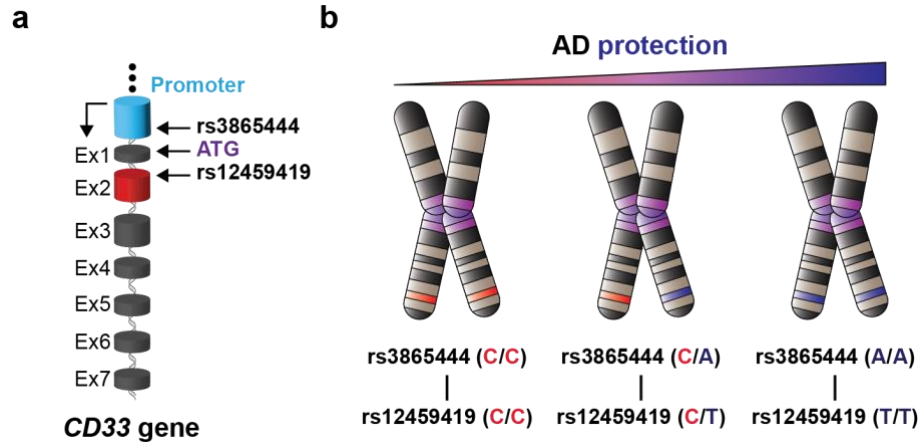


Figure 1.9: CD33 polymorphisms implicated in AD susceptibility. (a) rs3865444 (C or A) and rs12459419 (C or T) are coinherited SNPs and C is the common risk allele for both SNPs, while A or T is the protective allele in late-onset AD. (b) Populations with heterozygous (C/A or C/T) or homozygous (A/A or T/T) in rs3865444 and rs12459419 have lower chance to develop AD compared to the population with the risk homozygosity.

expressed by microglia.^{63,70} This suggests that the strong genetic correlation between CD33 and AD susceptibility is due to CD33 regulating microglial cell function, although a contributing role in myeloid cells outside the brain is difficult to formally rule out. Initial findings showed that AD brains have increased expression of CD33 on microglia.⁶⁹ Additionally, the number of CD33-expressing microglia were reported to positively correlate with plaque burden and cognitive decline.^{63,69} But higher CD33 expression in AD could be a response to disease pathology; indeed, inflammation and certain cytokines (IL-15) have been shown to induce upregulation of CD33.^{71,72}

Large scale GWAS in European, North American, and Chinese populations further supported the polymorphisms in *CD33* associating with AD susceptibility.⁷³ As the rs3865444 SNP is in the *CD33* promoter, 372 bp upstream of the transcription start site, this identified SNP

was originally speculated to deliver its effect through modulating *CD33* gene expression (**Figure 1.9b**).⁷⁰ However, further analysis led to the discovery of a nearby co-inherited SNP, rs12459419, located fourth nucleotides on exon 2 of *CD33* that impacts mRNA splicing. The rs12459419 SNP enhances the expression of a short protein isoform, called hCD33m (h= human, m=minor), and consequently decreases the expression of the long protein isoform, called hCD33M (M=Major), through alternative mRNA splicing.^{70,74} (**Figure 1.10**) Accordingly, rs12459419 is now considered a functional proxy of the AD-associated rs3865444 SNP. Specifically, the AD-susceptible

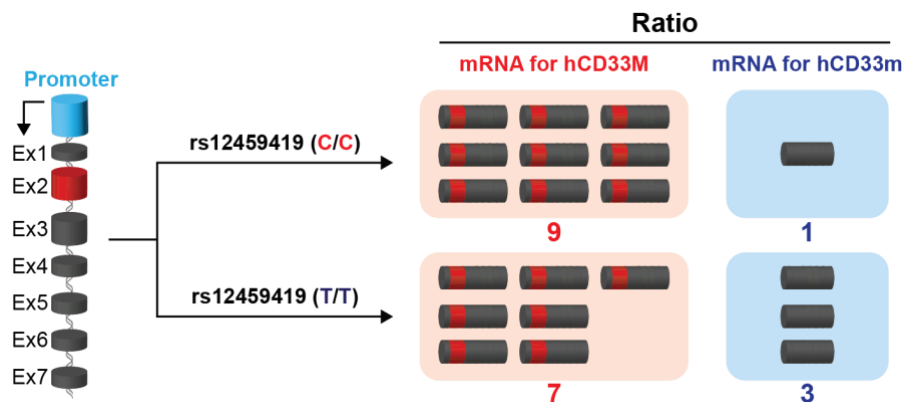


Figure 1.10: rs12459419 modulates the ratio of the two isoforms of CD33. The T allele has a higher chance to mediate alternative splicing leading to exon-2 exclusion. Depending on the presence of exon-2, two mRNA transcripts can be transcribed leading to the expression of the two isoforms of CD33. hCD33M has sialic-acid recognition domain while hCD33m lacks.

rs12459419C allele (which is co-inherited with the rs3865444C allele) results in a hCD33M:hCD33m transcript ratio of 9:1, while the AD-protective rs12459419T allele (i.e. rs3865444A) shifts this ratio to 7:3.⁷⁰ Consistent with these findings at transcript levels, cells expressing the rs12459419T allele showed decreased levels of hCD33M at the protein level.^{70,75} A corresponding increase in hCD33m protein levels has been more challenging to demonstrate. A recently identified *CD33* SNP, rs2455069, was proposed to be associated with risk of developing AD in a small cohort of Italian patients.⁷⁶ *In silico* analysis suggested that an amino acid switch at

position 69 of hCD33 from an arginine to glycine in the rare rs2455069 SNP, may enhance the affinity for sialic acid containing ligands. More testing is required to establish the association of this SNP with AD susceptibility within a larger cohort.

1.4.2 The two isoforms of CD33 and their function

The most obvious difference between hCD33M and hCD33m is the presence and absence of the V-set domain, respectively. However, there are at least three additional more subtle differences between the two isoforms (**Figure 1.11**). The first subtle difference is the number of glycans; hCD33m is less glycosylated compared to its counterpart isoform because two sites of *N*-glycosylation are located on the V-set domain.^{77,78} But the repertoire of glycans on the C2-set domain of hCD33M and hCD33m have never been investigated. The second subtle difference is that hCD33m contains an unpaired cysteine residue that is normally engaged in an interdomain disulfide bond with a cysteine residue in the V-set domain of hCD33M.^{79,80} The third subtle difference between hCD33M and hCD33m is in their signal peptide. The signal peptide cleavage site of hCD33M is encoded without exon 2, therefore, hCD33m contains a different signal peptide cleavage site. What is not so subtle is that hCD33m appears to prefer an intracellular location. While hCD33M is predominantly located on the cell surface, the hCD33m isoform is not.⁸⁰ Although numerous studies have observed partial cell surface localization of hCD33m, these are all under overexpressed conditions.^{79,81-83} Upon investigating the intracellular pool of CD33 in monocyte-derived macrophages and a human microglia cell line (CHME-5), it was found that hCD33m co-localized with catalase and PMP70, suggesting it resides in the peroxisome.⁸⁰ It is speculated that the unpaired Cys residue in hCD33m may contribute to the different intracellular trafficking of this isoform, which is in agreement with an earlier report suggesting that a C36A mutation in the V-set domain of hCD33M results in its accumulation inside the cell.^{79,80} Overall,

the notable differences in subcellular localization and cellular trafficking of the two hCD33 isoforms could be major determinant of their distinctive functions, but this has yet to be rigorously tested.

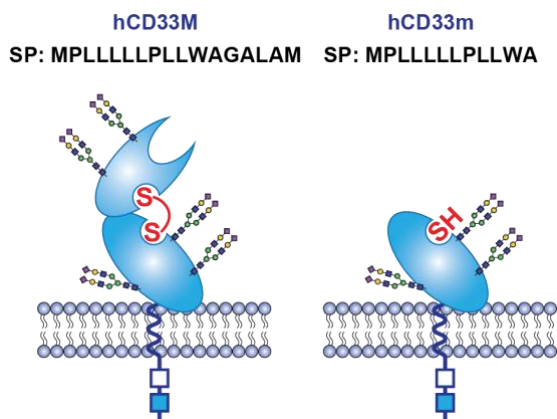


Figure 1.11: Schematic of the two human isoforms of CD33 implicated in AD susceptibility. The differences between the two isoforms are: 1) an altered signal peptide, 2) presence of the glycan binding domain, 3) an interdomain disulfide bond, and 4) the number of *N*-glycosylation sites. Both isoforms contain ITIM and ITIM-like signaling motifs.

1.4.3 Ligands for CD33

Despite of its crucial role in LOAD susceptibility, the mechanisms how hCD33M correlates with A β accumulation in the brain remains unknown. To understand the role of hCD33M in microglia, clarity is needed on its ligands in the brain. There has been over 25 years of tracking down glycan ligands of CD33. (**Figure 1.12**) In 1995, the sialic acid linkage specificity for CD33 was first reported. Desialylation and reintroduction of linkage specific sialic acids on *N*- or *O*-glycans on human red blood cells, followed by probing with CD33-Fc revealed that CD33 prefers α 2-3 to α 2-6 linkage.²⁸ In 2000, it was reported that CD33 recognizes both α 2-3 and α 2-6 linked sialosides such as sialyl Lac and LacNAc and α 2-6 linked sialyl GalNAc.¹⁸ In 2003, IC_{50} values of CD33 on both α 2-3 and α 2-6 linked sialosides were reported, where both were 4.72 mM.⁵⁰ In 2014, it was found that 9-*O*-acetylation at Neu5Ac has a negative effect on CD33 ligand binding.

Neu5Gc at non-reducing end compared to Neu5Ac was shown to slightly enhanced CD33 binding.⁴⁸ The gene encoding cytidine monophospho-N-acetylneuraminic acid hydroxylase

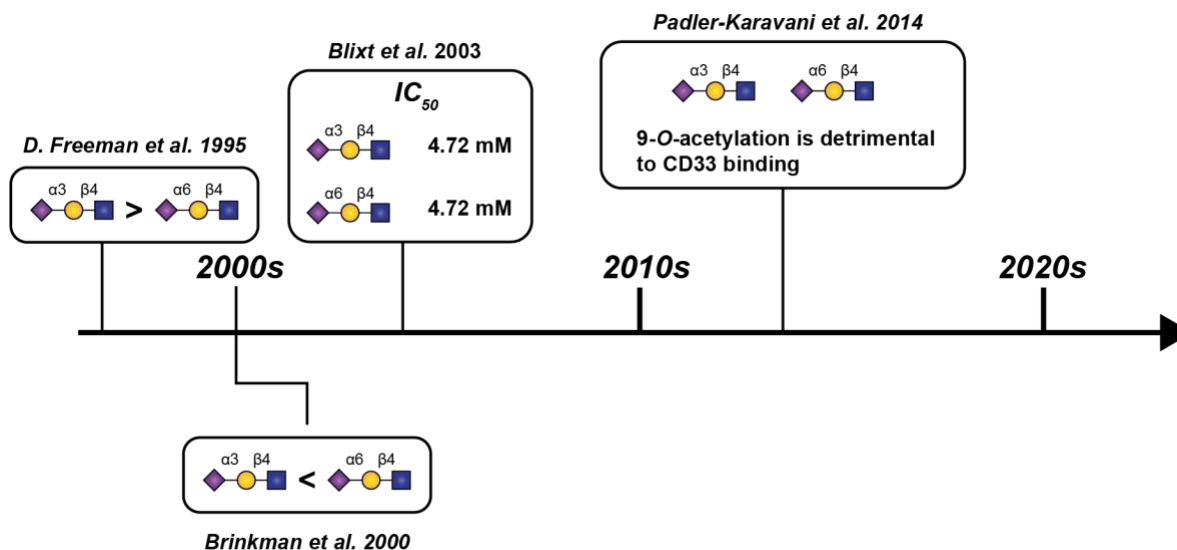


Figure 1.12: The history of the ligand discovery for CD33. In the early 2000s, the terminal epitopes of sialoglycans were mainly investigated with CD33. In 2014, it was shown that 9-*O*-acetylation on sialic acid is disadvantageous for CD33 binding.

(CMAH) required for the biosynthesis of Neu5Gc is inactivated in human,⁸⁴ so the interactions between CD33 and glycan containing Neu5Gc might be less likely to have any biological importance. Of note, there has been a controversy on which linkage of sialic acid is better than another. Crocker and coworkers in 1995 showed CD33 preferred $\alpha 2$ -3 linkage to $\alpha 2$ -6 linkage, which is opposite to the work from Varki and coworkers in 2000. As Paulson and coworkers in 2003 reported CD33 had similar IC_{50} values on both $\alpha 2$ -3 and $\alpha 2$ -6, clarity on this controversy needs to be addressed.

1.4.4 Differences between hCD33 and mCD33

A comparison between hCD33 and mouse CD33 (mCD33) is important to consider, particularly in light of the utility of mouse models in studying AD.⁸⁵⁻⁸⁷ The *CD33* gene in humans

and *Cd33* gene in mouse have a similar structure and chromosomal position, and the extracellular domains of the two expressed proteins show a cumulative sequence identity of 62%, albeit the sequence identity in their V-set domain is significantly less, at 50%.^{2,88} Nevertheless, three significant differences exist between CD33 from humans and mice. The first difference is in the transmembrane segment, where mCD33 has a lysine residue that is absent in hCD33 (**Figure 1.13a**). This lysine residue enables to pair with another transmembrane protein that contains a complementary acidic residue in their transmembrane segment, such as DAP10 and DAP12 (**Figure 1.13b**).¹⁴ The second significant difference between hCD33 and mCD33 is found within

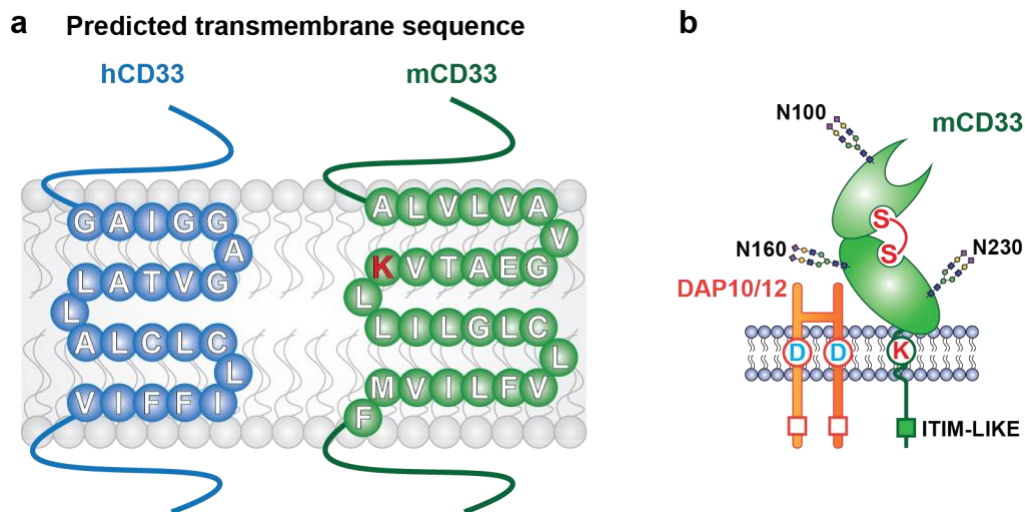


Figure 1.13: Schematic showing the major differences between hCD33 and mCD33. (a) Transmembrane sequence comparison between hCD33 and mCD33. **(b)** mCD33 has a transmembrane containing a lysine residue, which pairs with adaptor proteins to be stably on the surface. Notably, the ITIM is absent in mCD33.

their cytoplasmic tail. While hCD33 contains a *bona fide* ITIM, mCD33 lacks this signaling motif (**Figure 1.13b**).⁸⁹ The third significant difference between hCD33 and mCD33 is in their ligand binding properties. Remarkably, mCD33 has not shown any binding to glycan ligands.⁸⁹ The more profound divergence in identity within the V-set domain of hCD33 and mCD33, in comparison to

the C2-set domain (52% sequence identity in the V-set; 72% in the C2-set domain), is likely a key contributing factor for this striking contrast in ligand recognition.⁸⁹

1.5 Aims of this Thesis

The link between CD33 and LOAD susceptibility is strong and a potential role for glycan ligands of CD33 in modulating microglial function motivates the development of biochemical tools and approach to better understand CD33 and its glycan ligands. Regarding its glycan ligands, it has been discovered that CD33 recognizes 3'SLN and 6'SLN, but there has been a controversy which linkages are better.^{18,50} For example, a microarray with a defined *N*-glycan library or oligosaccharide library has suggested that CD33 prefers α 2-6 linked sialic acid on biantennary structures.²⁴ Also, α 2-6 disialylated biantennary *N*-glycan on hHBV further supports CD33's preference to α 2-6 linkage.⁹⁰ On the other hand, there are other studies insisting CD33 prefers α 2-3 linkage.^{28,50} In general, cell or tissue-based study has pointed that a α 2-3 linkage is preferred in CD33 binding. This may indicate that α 2-3 linkage is more in certain tissues, so it appears more dominant than α 2-6 linkage. Of note, *O*-glycans do not contain 6'SLN moiety, so the significance of α 2-3 linked sialosides should not be overlooked.²³ The first aim of this thesis is to develop screening tools to discover natural ligands for CD33 from a biologically relevant cell line. Second, as carbohydrate sulfation has been shown to enhance Siglec binding,^{25,26,39,47,91} carbohydrate sulfation as an important PGM in the brain will be investigated regarding its effect on CD33 affinity. Lastly, the effects of the two isoforms of CD33 associated with AD susceptibility is studied as rs12459419 rare allele have shown AD protection among population.

In Chapter 2, a new platform of Siglec-Fc was developed. Creating versatile Siglec-Fc proteins was carried out to overcome several intrinsic challenges in working with Siglec-Fc chimeras. This new version of Siglec-Fc was devised to minimize the undesired engagement of

Fc γ receptors, expand multivalency, and be transformed to the homogeneous Siglec fragments for mass spectrometry (MS) applications. With the intact Siglec-Fc and MS-based approach, high-throughput library screening was optimized and the ability to generate the homogeneous Siglec fragment was applied to determining the dissociation constants in collaboration with the Klassen group. This application was applied to studying glycan ligands for CD33. It is shown that determined dissociation constants for CD33 on both α 2-3 and α 2-6 sialosides are similar in solution. Also, natural *N*-glycan compositions of CD33 ligands from U937 cells are elucidated. Furthermore, both the effects of α 2-3 and α 2-6 sialosides in U937 cells were assessed by measuring *cis* and *trans* interactions.

In Chapter 3, I tested a hypothesis that carbohydrate sulfation plays larger than previously anticipated role in enhancing the affinity of glycans for Siglecs. Previously, CD22, Siglec-7, -8, and -9 were known to bind better to sulfated sialosides.⁷ A panel of cells overexpressing CHSTs were created and it was discovered that CHST1, which transfers a sulfate group to 6-*O*-Gal, was shown the major enzyme improving CD33 binding affinity. Furthermore, it was discovered that CHST1 can also enhance the binding of Siglec-5 and Siglec-15, which is novel. In an ESI-MS-based binding assay, dissociation constants with 3'SLN containing different degrees of 6-*O*-sulfation at either Gal or GlcNAc, or both were determined against engineered CD33 fragments. In addition, it was discovered that disulfation, on both 6-*O*-Gal and 6-*O*-GlcNAc, has synergetic effects on CD33 affinity. Last, it was shown that MDA-MB-231 and U251 cancer cells naturally upregulate CHST1 as a mechanism that enhances Siglec ligands on their surface.

In Chapter 3, the two CD33 protein isoforms were investigated through developing new biochemical tools and assays to study the short isoform expression mediated by alternative mRNA splicing. Two hCD33m-specific monoclonal antibodies were developed and were shown, for the

first time, that the rs1245419T allele gives rise to enhanced expression of hCD33m at the protein level. The propensity of hCD33m to maintain an intracellular localization was studied and it was discovered that the unpaired Cys42 on hCD33m is required for preventing it from reaching the cell surface. Given all the evidence pointing to a gain-of-function role for hCD33m, new cell reporters were developed that provide a readout on the amount of short isoform transcript produced. These new reporters can be used in the future for discovering small molecules that enhance the expression of hCD33m as a potential AD-protective therapeutic strategy.

Chapter 2

Development of a New Generation of Siglec-Fc for Characterizing Ligands for CD33

2.1 Acknowledgements

All mass spectrometry analysis was done by the Klassen group. Specifically, I would like to thank Duong Bui, Dr. Heajin Park, and Dr. Elena Kitova for determining the dissociation constants of CD22 and CD33, and probing CD33 against a defined oligosaccharide library, a synthetic *N*-glycan library and a U937 cell-derived *N*-glycan library. We thank Dr. Gour Daskhan for preparing liposomes displaying high affinity CD33 ligands.

Section 2.3.3 and 2.3.4 were published in Rodrigues *et al.*, *Nature Communications*, 2020.

Section 2.3.5 were published in Park *et al.*, *ACS Analytical Chemistry*, 2020.

2.2 Introduction

The presence and absence of the glycan-binding on CD33 isoforms that correlate with AD susceptibility suggest that the glycan ligands of CD33 may play a role in AD pathogenesis. Despite this connection, the natural physiological glycan ligands of CD33, let alone in the brain, have not been elucidated.^{63,92} Numerous methods for studying the glycan ligands of Siglecs have been developed.⁴² Terminal epitopes consisting of a trisaccharide have been the focus, but what type of glycan scaffold they are presented from - *N*-glycosylation, mucin-type *O*-glycosylation, *O*-mannose glycosylation, or glycolipids - is less established.^{18,46,47,50} More recently, complex glycans containing core structures have been synthesized and studied as ligands of Siglecs. For example, 98 synthesized *N*-glycans were probed in glycan microarray, which characterized the preference of human CD33.²⁴ In another study, 73 *O*-mannose glycans were used to identify ligands for CD33.²³ Although these synthesized glycans have helped to tease out what glycan ligands for Siglecs are, there are challenges associated with this approach: 1) synthesizing the entire repertoire of sialosides present on the cell surface is not possible; 2) there can be undesired effects from aglycon;⁹³ and 3) printed glycans have less flexibility to interact with Siglecs.⁹⁴ Therefore, other systems should be considered for the discovery of Siglecs ligands.

Methods for isolating glycan components from cells or tissues have been developed. To isolate an *N*-glycan library, an enzyme called PNGase F is widely used, which can remove *N*-glycans from glycopeptides or glycoproteins.⁹⁵ As glycosidases for isolating complex *O*-glycans are not available, reductive β -elimination is widely used. In this reaction, sodium hydroxide (NaOH) liberates *O*-linked glycans from glycoproteins. This reaction is susceptible to an additional peeling mechanism, which can hydrolyze the linkage of Gal β 1-3GalNAc.⁹⁶ To prevent undesired outcome, sodium borohydride (NaBH₄) is used to convert the reducing end of *O*-glycans to an

alditol. To derivatize reducing end of released *O*-glycan, non-reductive β -elimination has been also developed.⁹⁷ To release glycans from glycosphingolipids, enzymatic methods using endoglycoceramidas or chemical methods such as ozonolysis have been developed.^{98,99} Lastly, oxidative release of natural glycans has been shown to isolate *N*-, *O*-glycans, and glycan moieties from gangliosides.¹⁰⁰

A need for discovering Siglec ligands directly from intact and natural glycan library motivated the development of native MS-based approach to study protein-ligand complexes. In general, the first component of MS is an ion source, which generates gas-phase ions. Subsequently, these gas phase ions are electrically separated based on their mass-to-charge (m/z) ratios in the second component, a mass analyzer. The last step is that the gas-phase ions reach a mass detector, transforming the collected data to a MS spectrum where x-axis represents “ m/z ” and “ion counts” is on the y-axis.¹⁰¹ Notably, electrospray ionization (ESI) was chosen in this study due to its biocompatible advantages over the other ion sources, such as matrix-assisted laser desorption/ionization. For example, ESI-MS requires low sample amount for analysis. Also, its mild ionization helps to analyze non-covalent ligand-protein interactions.¹⁰² Notably, there are two feasible MS applications. First, in an ESI-MS-based binding assay, ligand titration with a protein enables a dissociation constant to be determined by analyzing the abundance of m/z values representing protein-ligand complexes at each concentration. On the other hand, indirect measurement is ideal for library screening. The key process is to ionize a series of protein-ligand complexes mildly, select the right range of m/z , and break the complexes with collisional induced dissociation (CID). The final step is to analyze the released ligands.¹⁰³ This specific application is called Catch-and-Release ESI-MS (CaR-ESI-MS), which was pioneered by Dr. Klassen. **(Figure 2.1)**

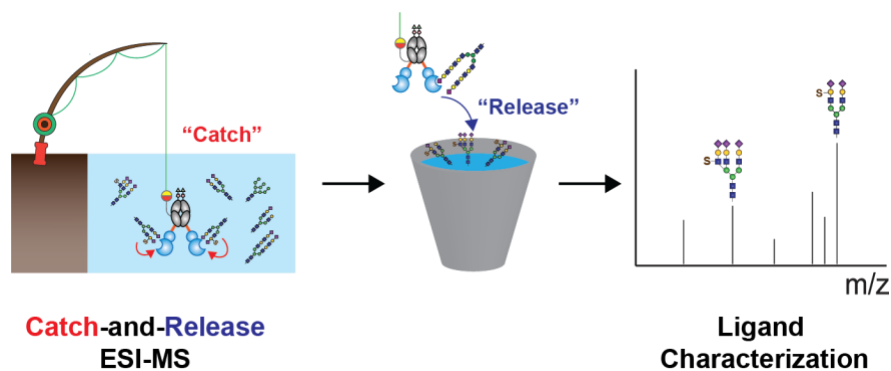


Figure 2.1 CaR-ESI-MS for discovering glycan libraries against Siglec-Fc. During the “Catch” step, the Siglec-Fc complexes with glycan ligands are formed. Analytes become ionized and reach to MS detector. After detection of the m/z range of the Siglec-ligand complexes, the analytes with right m/z ranges are selected, and CID is applied to dissociate the complexes, leading to the release of the ligands. The released ligands are detected by the MS detector for characterization.

To discover ligands for CD33 through ESI/MS-based approaches, a new version of versatile recombinant Siglec-Fc was developed. This is because a highly pure and homogenous protein is required for native MS-based applications. In this chapter, a new soluble version of CD33 (CD33-Fc) was expressed and optimized for ESI-MS and flow cytometry applications, providing a great analytical tool to elucidate glycan ligand specificity of CD33.

2.3 Results

2.3.1 A new generation of CD33-Fc

Siglecs are transmembrane proteins, making it challenging to study their native full-length protein in solution due to solubility issues. A solution has been to use extracellular domains of Siglecs combined with Fc domains of IgG to investigate their glycan ligands. These Siglec-Fc chimeras have been invaluable in studying Siglec-ligand interactions. However, these chimeric proteins have several undesirable properties. First, they have the potential of engaging Fcγ receptors through their Fc regions of the Fc chimeric Siglecs. Second, although the dimeric presentation enhances avidity of Siglecs, most Siglec-Fc constructs - except a few Siglec-Fc

constructs such as Siglec-2-Fc - still have very weak binding affinities, which necessitates higher multivalency to enhance avidity. Third, this dimeric presentation of the Siglec-Fc constructs may not be ideal for determining a purely one-to-one binding affinity of Siglec for its ligand. Lastly,

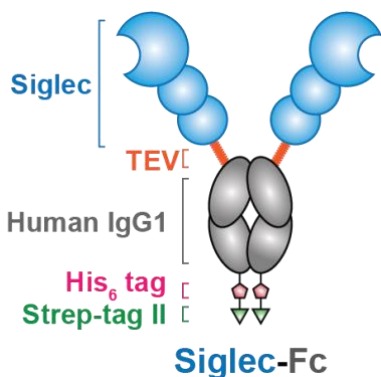


Figure 2.2 Components of a new versatile Siglec-Fc chimera. This new generation of Siglec-Fc has four key new features: 1) a TEV recognition site as a linker; 2) an engineered human IgG1 to minimize non-specific binding to Fc γ receptors; 3) a His₆ tag for purification purposes; and 4) Strep-tag II for both purification and pre-complexing.

commercial Siglec-Fc constructs from R&D systems are prohibitively expensive, and do not come with a mutated version of the protein that is unable to bind its ligands as an ideal negative control.

Desiring to create more versatile Siglec-Fc proteins, multifunctional features were embedded into the new generation of Siglec-Fc. (**Figure 2.1**) This next-generation of Siglec-Fc chimeras contains an engineered human IgG1 Fc region (L234A, L235A, G237A, H268A, P238S, A330S, P331S) designed for avoiding non-specific interactions with Fc γ receptors.¹⁰⁴ Also, a *Tobacco Etch Virus* (TEV) recognition site (Glu-Asn-Leu-Tyr-Phe-Gln/Ser) was inserted between extracellular domains of Siglec and Fc domains, making it possible to generate monomeric Siglec fragments by TEV digestion, which is useful for determining dissociation constants.^{105,106}

The genes encoding these new Siglec-Fc proteins were stably transfected into Chinese Ovary Hamster (CHO) cells through a Flp-in system.¹⁰⁷ The Flp-in system integrates genes of interest into the host genome through the use of the Flp recombinase and a target site previously incorporated into a transcriptionally active site within the genome. Aiming to better understand

glycan ligands for Siglecs and link this to the role of CD33 in AD susceptibility, CD33-Fc was cloned, expressed, and purified. Specifically, CHO cells stably expressing CD33-Fc were cultured and supernatant was collected a week after confluency. A double purification was carried out with a nickel column and a protein-G column to achieve high purity. **(Figure 2.3a)** The second purification step, protein-G, can be replaced by a Streptactin column to avoid dramatic pH change required for elution.¹⁰⁶ A quadruplet on mass spectrum ranging from 6,000 to 7,000 m/z indicates the presence of CD33-Fc. **(Figure 2.3b)** The broad range of peaks were caused by the heterogeneity of glycoforms coming from five predicted *N*-glycosylation sites on the CD33 portion of the protein along with one site on the Fc, meaning the entire CD33-Fc may have twelve *N*-glycans.

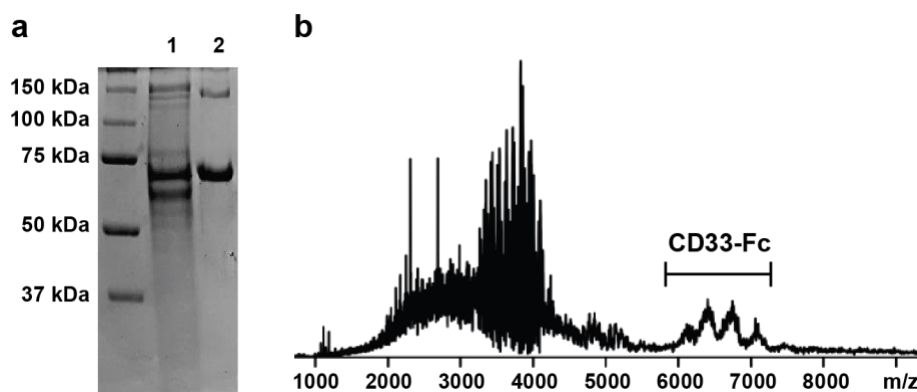


Figure 2.3 Preparation of highly pure CD33-Fc. (a) CD33-Fc was prepared by two successive purifications and run on SDS-PAGE. Eluents from a single purification with a nickel column contain impurities (Lane 1). Double purification with a combination of nickel and protein-G columns yields highly pure CD33-Fc (Lane 2). (b) A MS spectrum on the doubly purified CD33-Fc sample.

2.3.2 Validation of Catch-and-Release electrospray ionization mass spectrometry

CaR-ESI-MS is an indirect MS-based technique pioneered by the Klassen lab for high-throughput screening (HTS) of glycan libraries against GBPs. ESI-MS has outstanding benefits due to its feasibility for ionizing intact ligand-protein complexes, which can be also applied to

determining dissociation constants.¹⁰² CaR-ESI-MS has two major steps: the “Catch” step and the “release” step. During the “catch” step, GBPs and a glycan library are mixed to form the protein-ligand complexes, which are ionized and sprayed into a quadrupole MS detector.¹⁰³ Subsequently, the Siglec-ligand complexes are dissociated by CID, and the ligands are released from the complexes and analyzed on a mass spectrum, which is called the “Release” step. (**Figure. 2.4**)

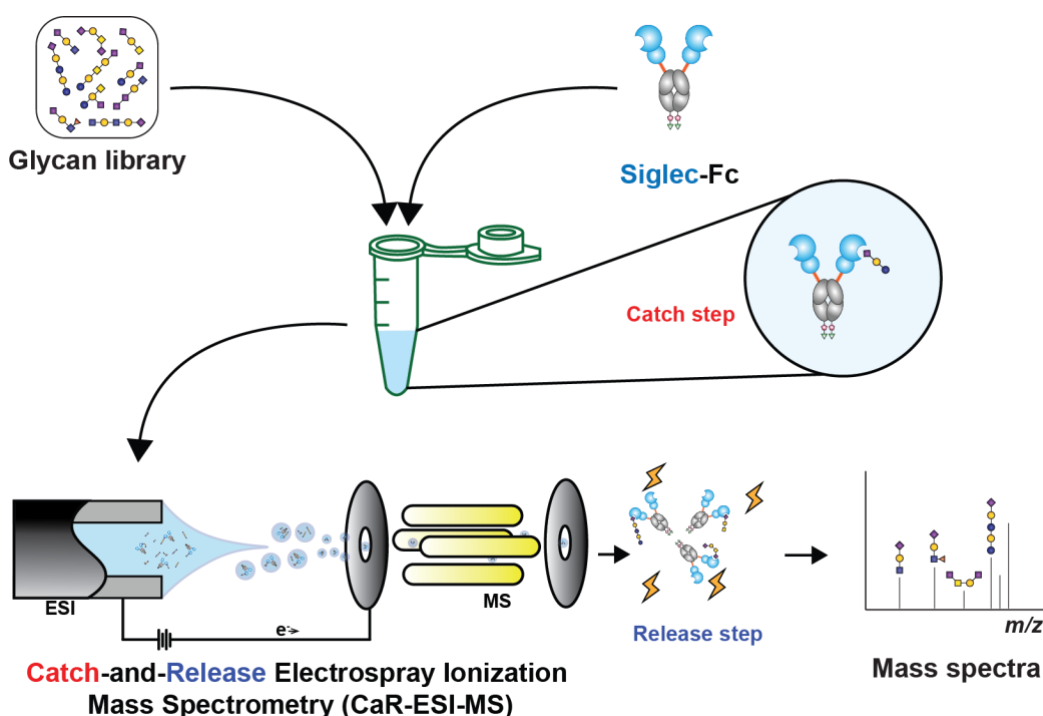


Figure 2.4 Scheme for CaR-ESI-MS In CaR-ESI-MS, Siglec-Fc is incubated with a library containing glycan ligand candidates to encourage the interactions before ESI ionization of the complexes. After the ionized complexes pass through quadrupole MS detector, the range of m/z arising from the complexes are selected and dissociated by CID. The released glycans are analyzed.

To validate the feasibility of CaR-ESI-MS with a Siglec-Fc probe, CD22-Fc (or hSiglec-2-Fc) was first validated against a defined glycan library consisting of 167 oligosaccharides. CD22-Fc has a well-known preference to α 2-6 linked sialosides, with a binding affinity of approximately 150 μ M.⁵⁰ As expected, the top three ranked glycans were α 2-6 linked sialosides. (**Figure 2.5**) Some marginal intensity from α 2-3 linked sialosides is considered noise in this assay,

since CD22 does not recognize α 2-3 linked sialosides.^{18,50} Consequently, CaR-ESI-MS with CD22-Fc and a defined oligosaccharide library reproduced the previous finding on specificity for CD22, whose binding intensively prefers α 2-6 sialosides.^{18,50}

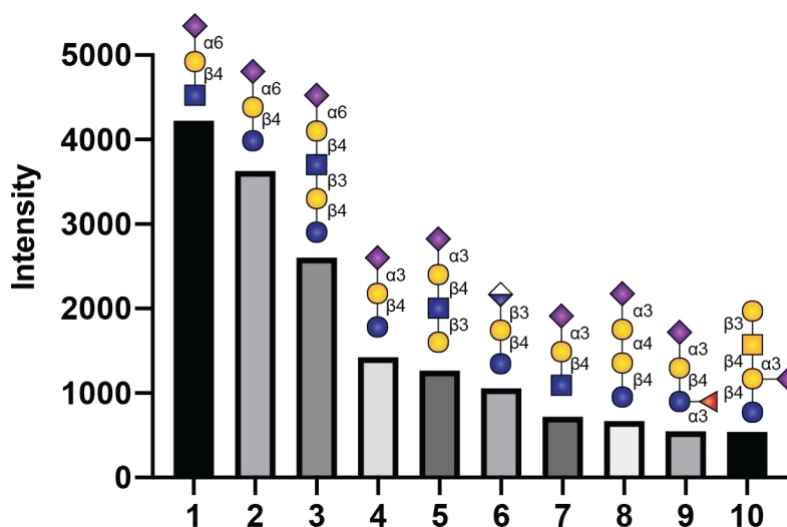


Figure 2.5 CaR-ESI-MS with CD22-Fc and a defined glycan library. CD22-Fc was screened against a defined oligosaccharide library consisting of 167 unique glycan structures. CD22-Fc recognized α 2-6 linked sialosides, which validates the feasibility of CaR-ESI-MS. This graph shows the top ten oligosaccharide structures.

After validating the assay with CD22-Fc, CD33-Fc was screened against the defined glycan library. Generally, low signal intensity was observed compared to the assay with CD22, which is presumably due to the weak affinity that CD33 exhibits towards its glycan ligands. Several studies previously found that CD33 has a small preference for α 2-6 linked sialosides,^{18,24,90} and although α 2-6 sialyl lactose was the top hit, α 2-3 sialyl lactose was not far behind and not considered significantly different. (**Figure 2.6a**) Next, a defined *N*-glycan library was screened with CD33-Fc. (**Figure. 2.6b**) In this case, α 2-3 linked sialyl lactose moiety had higher signal than α 2-6 linked sialyl lactose, so this suggests that CD33 recognizes both linkages, but its preference can be slightly different depending on analytical methods. On the biantennary *N*-glycans, both mono- and disialylated *N*-glycans with either α 2-3 or α 2-6 linkage were similarly recognized by CD33-Fc.

2.3.3 Development of monomeric and homogenous CD33 for an ESI-MS-based binding assay

ESI-MS is a great method for determining binding affinity through direct measurements of ligand-protein complexes at different ligand concentrations.¹⁰⁸ For this approach, a homogeneous protein is required because multiple protein glycoforms result in significant spectral overlap between protein and protein-ligand complexes. CD33 has five predicted *N*-glycosylation sites based on the *N*-glycosylation sequon (Arg-Xaa-Ser/Thr, Xaa indicates any amino acid except Pro), and glycan structures at *N*-glycan site are necessarily heterogenous, leading to many glycoforms.¹⁰⁹ **(Figure 2.3b)** To produce homogeneous CD33, CD33-Fc was expressed from Lec-1 CHO cells, which genetically lacks a gene called *MGAT1*.¹¹⁰ This gene encodes the first GlcNAc transferase initiating the biosynthesis of hybrid and complex types of *N*-glycans. Hence, Lec-1 CHO cells predominantly express high-mannose type (Man₅) *N*-glycans.¹¹⁰ Notably, the high-mannose types of *N*-glycans can be enzymatically trimmed down by Endoglycosidase H (Endo H) to greatly reduce heterogeneity of CD33. Recombinant Endo H from *Streptomyces picatus*, is commonly employed to hydrolyze between a chitobiose core of hybrid and high-mannose *N*-glycans, leaving a single GlcNAc residue at each *N*-glycan site on glycopeptides or glycoproteins.¹¹¹ The residual GlcNAc at each site is important for maintaining stabilizing effects of glycosylation on protein folding.¹¹² In contrast, PNGase F, which clips off the entire *N*-glycan including the first GlcNAc residue, can impact protein stability. To demonstrate successful removal of glycans, CD33-Fc was expressed in both WT and Lec-1 CHO cells and treated with either PNGase F or Endo H. **(Figure 2.7)** As CD33-Fc from Lec-1 CHO cells lacks sialylated *N*-glycans, it migrated faster than the one from WT CHO cells. PNGase F deglycosylated CD33-Fc recombinant proteins from the two CHO cell lines, but Endo H only digested CD33-Fc from the Lec-1 CHO cells.

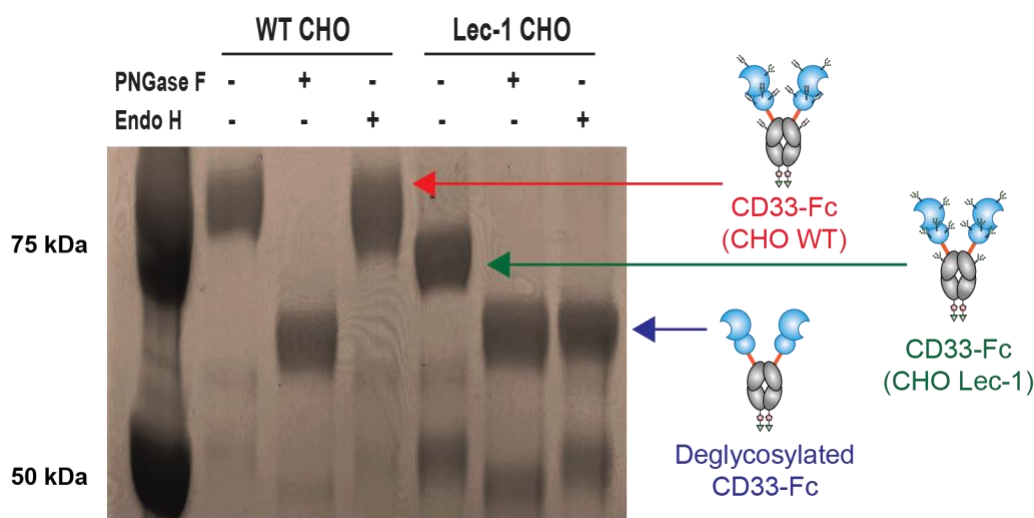


Figure 2.7 The effect of two glycosidases on two CD33-Fc constructs expressed in WT and Lec-1 CHO cells. In SDS-PAGE, CD33-Fc expressed from WT CHO cells was deglycosylated by PNGase F, while CD33-Fc expressed from Lec-1 CHO cells was sensitive to both PNGase F and Endo H digestion.

In addition to producing deglycosylated CD33-Fc, we also aimed to reduce its size for ESI-MS-based binding assay by removing the Fc. This was accomplished by TEV protease digestion. Accordingly, a protocol was developed whereby TEV digestion was followed by Endo H digestion. **(Figure 2.8a)** When TEV liberated CD33 fragments from Siglec-Fc, the band on SDS-PAGE was shifted from about 75 kDa to 37 kDa. Additional Endo H digestion further reduced the size of the CD33 fragments, which was caused by a loss of glycosylation. **(Figure 2.8b)**

Unexpectedly, the digested CD33 fragments ran as a doublet on SDS-PAGE, and MS analysis helped to identify the three isoforms of the CD33 fragments. **(Figure 2.8c)** The mass difference between isoforms corresponds to the size of Neu5Ac₁Hex₁HexNAc₁ (P2) and Neu5Ac₂Hex₁HexNAc₁ (P3), which represents *O*-glycans corresponding to sialyl and disialyl T antigens. An *O*-glycosylation site on CD33 has never been reported. Consistent with this prediction, treatment of the CD33 fragment with neuraminidase A (NeuA) reduced the mixture to two glycoforms with the absence and presence of the T antigen. **(Figure 2.8c)**

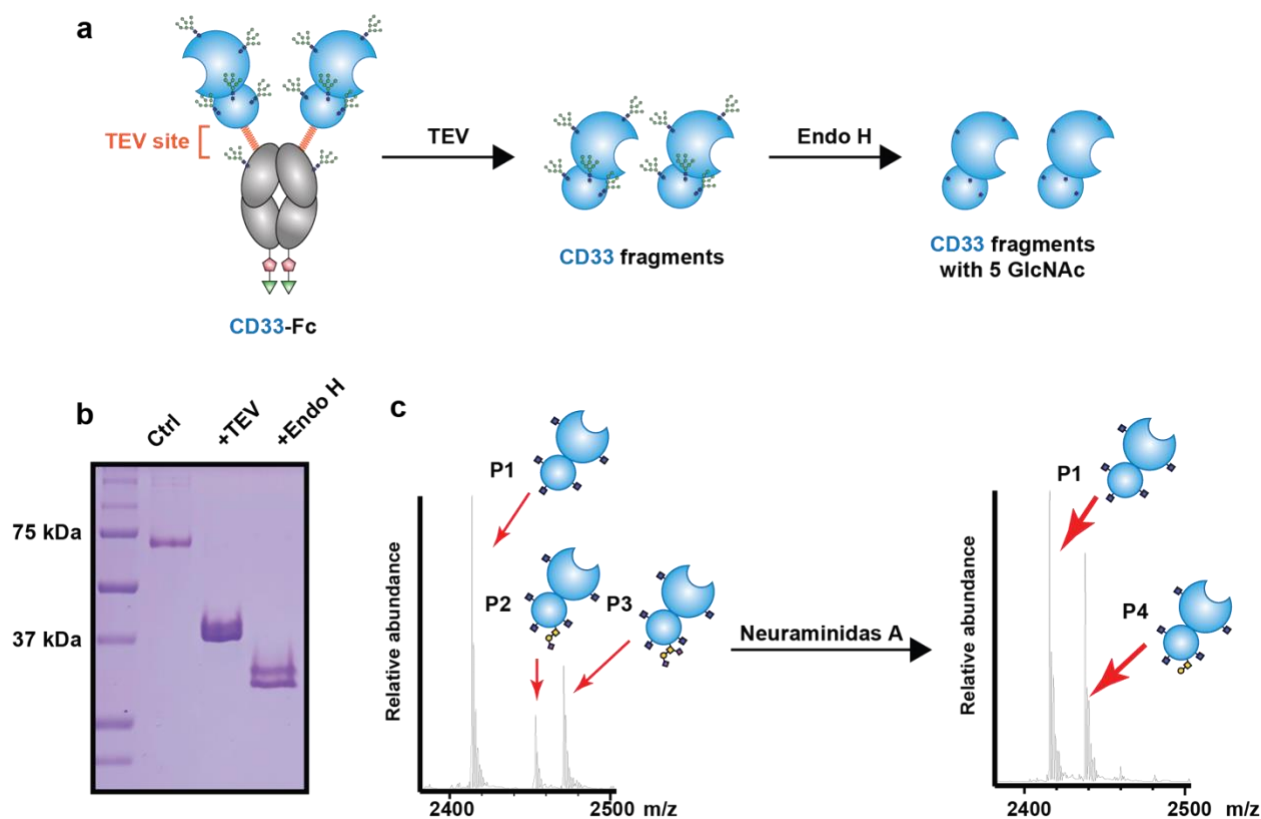


Figure 2.8: Generation of a homogeneous CD33 fragment. (a) Sequential enzymatic reactions help to produce the homogeneous CD33 fragment. TEV protease was used to cleave the extracellular domains of CD33 from the CD33-Fc fusion chimera, and Endo H cleaved the chitobiose core of high-mannose *N*-glycans on CD33, leaving a single GlcNAc residue per *N*-glycosylation site. (b) The intact CD33-Fc expressed from Lec-1 CHO cells, CD33 fragment after TEV digestion, and CD33 fragment after Endo H were run on SDS-PAGE. (c) The fully deglycosylated CD33 fragment was analyzed on an ESI mass spectrum, which had three isoforms. After NeuA digestion, the number of CD33 isoforms was reduced to two. Notably, the presence of CD33 comes from a novel *O*-glycosylation site.

2.3.3.1 Validation of MS-based determination of dissociation constant

To optimize ESI-MS-based K_d determination, CD33-ligand complex on a mass spectrum was first analyzed. (Figure 2.9a) Since the CD33-ligand complex, corresponding to the P1 glycoform, overlaps with the P2 glycoisoform, the P3 glycoisoform containing the disialyl T antigen was selected to determine K_d value by a titration. (Figure 2.9b) The measured K_d value

was 87 μM , which is 8-fold higher than the reported IC_{50} value (11 μM).¹¹³ Subsequently, α 2-3 or α 2-6 linked sialyl LacNAc were titrated against the CD33 fragment to determine dissociation constants. (**Figure 2.10a-c**) The CD33 fragments have similar K_d values against 3'SLN (2.7 ± 0.1 mM) and 6'SLN (2.7 ± 0.4 mM), which is consistent with CaR-ESI-MS screening with CD33-Fc and a defined *N*-glycan library. Again, this indicates CD33 recognize both types of sialic acid linkages with similar affinity, which is in line with the catch-and-release findings presented above. (**Figure 2.6**)

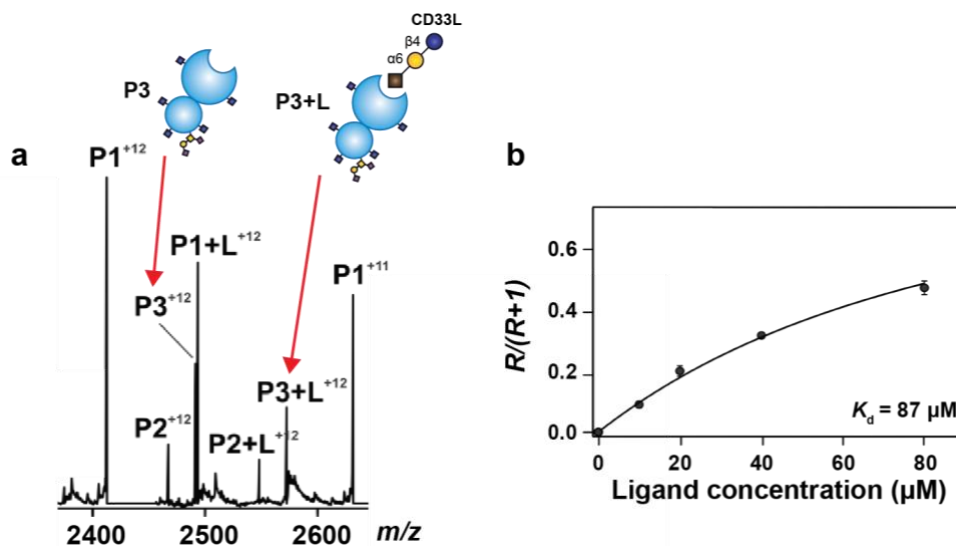


Figure 2.9 Affinity measurement of CD33 with its high affinity ligand. (a) A MS spectrum of the complexes of CD33 and the CD33 high affinity ligands. P1-P3 indicates the three different CD33 isoforms and L stands for the high affinity ligand. (b) Dissociation constant between CD33 fragment and CD33L was determined by an ESI-MS titration method.

2.3.4 Both α 2-3 or α 2-6 linked sialosides in U937 cells play a role in masking CD33

The glycan ligands of CD33 on cells have not been thoroughly studied. Yet, the ability of glycan ligands of Siglecs to modulate their function and the potential connection between CD33 ligands and AD motivated a development of a cell-based assay. U937 cells, a monocytic cell line, were used for this study because they naturally express high level of endogenous CD33 and have

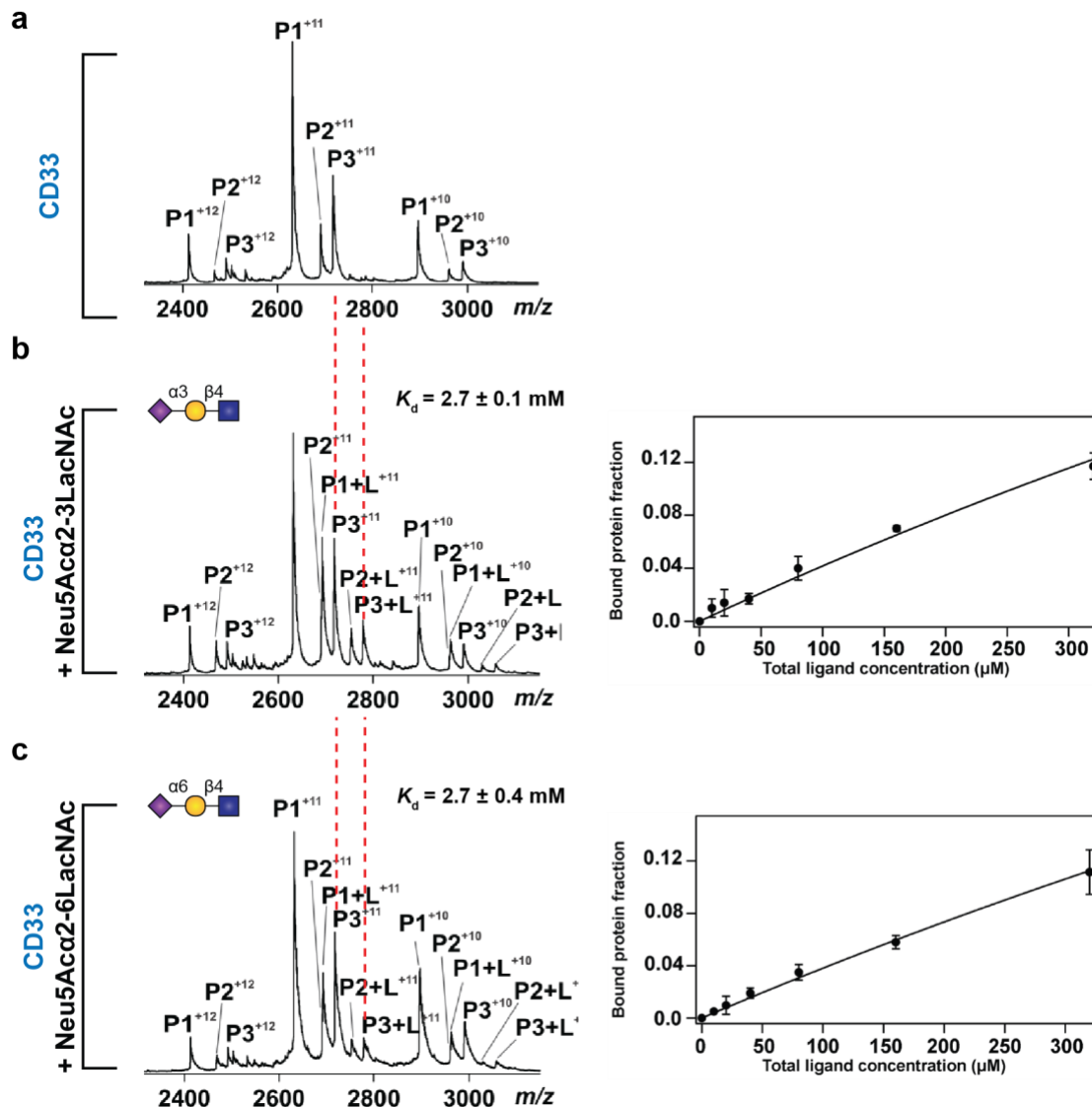


Figure 2.10: Affinity measurement of CD33 with its K_d determination of $\alpha 2$ -3 or $\alpha 2$ -6 linked sialyl LacNAc with the CD33 fragment. Ligand titrations enable the determination of dissociation constants for $\alpha 2$ -3 or $\alpha 2$ -6 sialyl LacNAc. The different charge states, CD33 isoforms and their CD33-ligand complexes were plotted on mass spectra with titration curve graphs for (a) CD33 fragment, (b) CD33 fragment and 3'SLN, and (c) CD33 fragment and 6'SLN. The dissociation constants between CD33 fragment and the ligands were calculated by an ESI-MS titration method.

been used extensively to study the function of CD33. To assess how both $\alpha 2$ -3 or $\alpha 2$ -6 linked sialosides engage CD33 through *cis* or *trans* interactions, two types of neuraminidases were used. One is recombinant neuraminidase A (NeuA) from *Arthrobacter ureafaciens*, which cleaves all

sialic acid linkages.¹¹⁴ Another is recombinant neuraminidase S (NeuS) from *Streptococcus pneumoniae*, which has high specificity for α 2-3 linked sialic acids.¹¹⁵ For this study, an assay was needed where engagement of CD33 on cells by glycan ligands can be monitored and quantified. Accordingly, fluorescent liposomes carrying CD33L were used as a *trans* ligand, enabling an investigation into the type(s) of sialic acid linkages that “*mask*” CD33. **(Figure 2.11a)** The concept of depleting cell-surface sialic acids by neuraminidase can “*unmask*” CD33, leading to increased interactions with the CD33L liposomes. Compared to a WT control, NeuS digestion led to increased liposome binding, which indicates that the elimination of α 2-3 linked sialic acids on the surface *unmasks* CD33. NeuA digestion further enhanced liposome binding, suggesting that CD33 is *masked* by both α 2-3 and α 2-6 linked sialic acids. **(Figure 2.11b)** Lack of liposome binding to CD33^{-/-} U937 cells demonstrates that the CD33L is highly specific for engaging CD33 and not any other Siglecs expressed on U937 cells.

Precomplexing the new version of CD33-Fc with Streptactin conjugated to Alexa647 (Streptactin-AF647), which is a mutated form of Streptavidin that binds Strep-tag II on CD33-Fc, enables the probing of natural *trans* ligands on U937 WT cells. The leveraged avidity in the octameric presentation of CD33 performed better in staining the WT U937 cells than general primary and secondary staining. **(Figure 2.12a)** Moreover, avoiding a washing step for secondary staining minimizes the loss of primary CD33-Fc binding to the cells. To assess the effect of both NeuS and NeuA on CD33 *trans* interactions, U937 cells were probed with the octameric complex of CD33 and Streptactin after sialidase treatment. **(Figure 2.12b)** The removal of α 2-3 linkages led to the decreased octamer binding. The lack of both α 2-3 and α 2-6 linkages significantly abrogated the binding of the CD33-Streptactin complexes compared to Streptactin staining with

absence or presence of CD33R where an essential Arg119 was mutated to Ala. This result provides further evidence that cellular ligands of CD33 are both α 2-3 and α 2-6 linked sialosides.

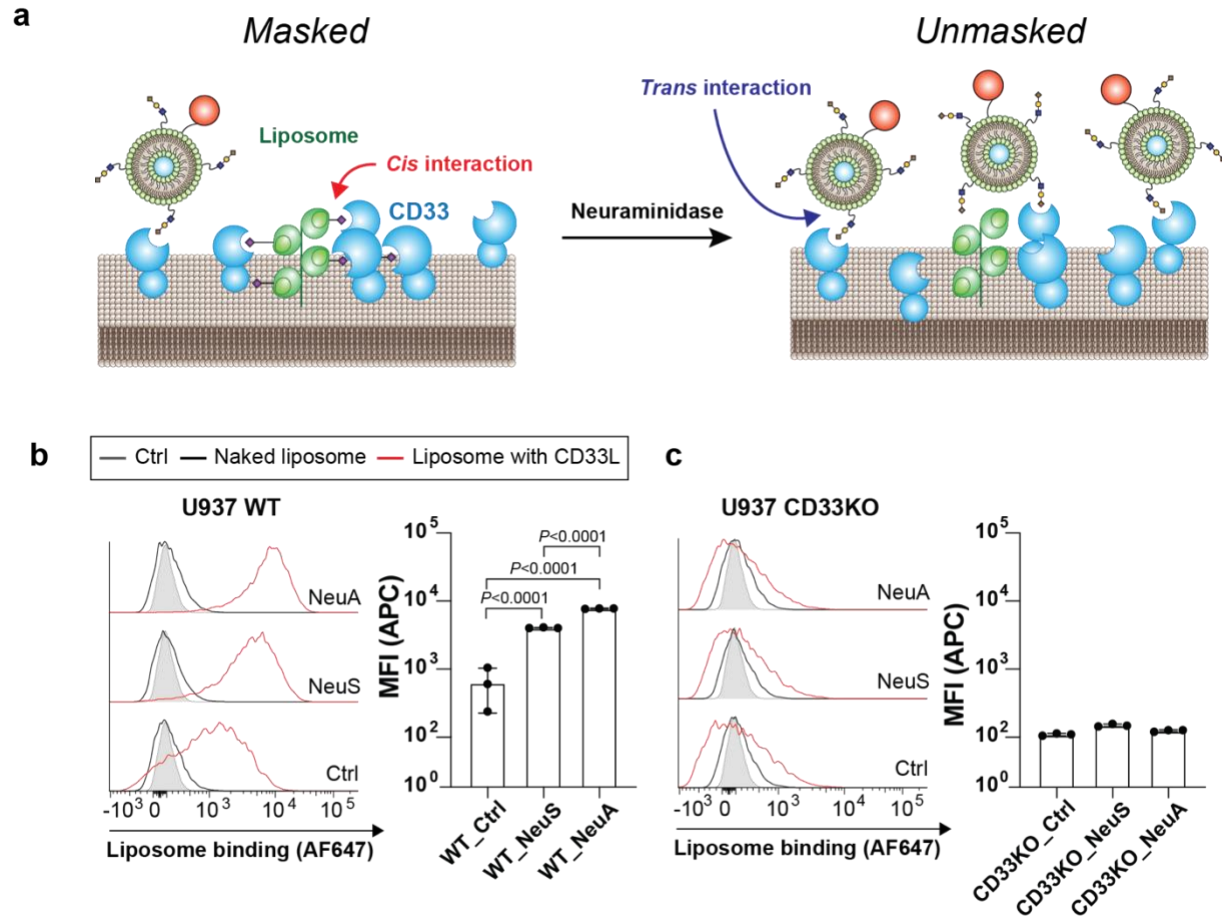


Figure 2.11: CD33 is masked by both α 2-3 or α 2-6 linked sialic acids. (a) CD33 is *masked* by *cis* interactions with endogenous sialosides. Neuraminidase treatment enables CD33 to interact with *trans* ligands by releasing the *masked* CD33. (b) NeuS and NeuA treatments increased CD33L liposome binding compared to a naked liposome, which lacks the presentation of CD33L. (c) CD33L liposome binding observed in U937 WT was not reproduced in CD33 KO U937 cells, which indicates the whole process was mediated by the cell surface CD33. Statistical significance calculated based on a one-way ANOVA with Tukey's multiple comparisons test.

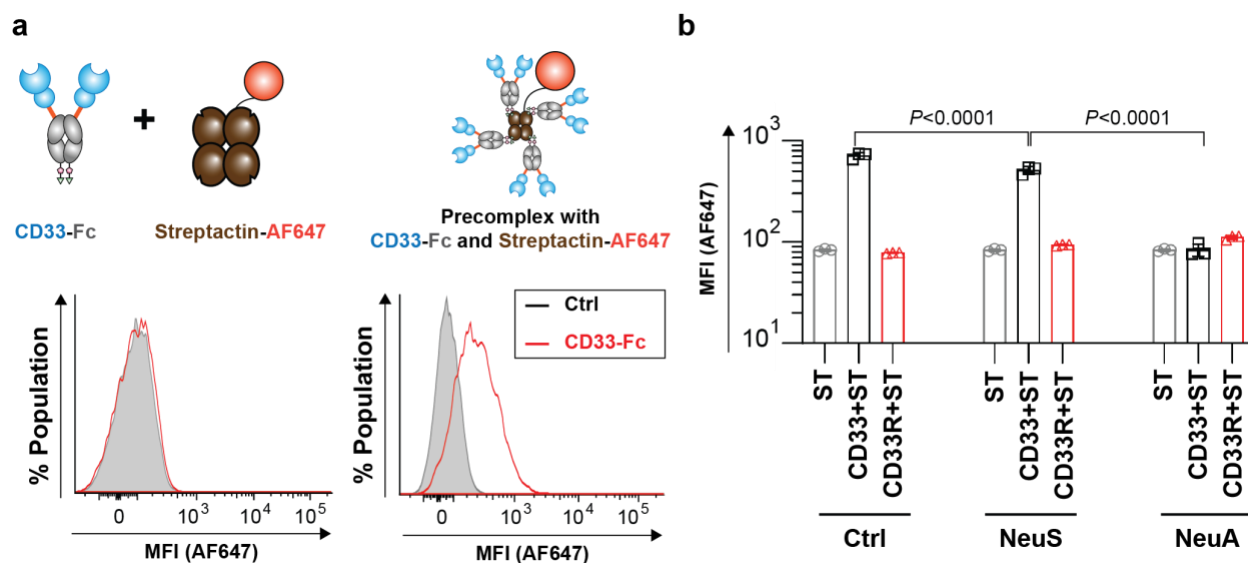


Figure 2.12: CD33-Fc binds to U937 cells in NeuA- and NeuS-dependent manners. (a) Precomplexing CD33-Fc with Streptactin-AF647 enables better binding to U937 cells than the two steps of the primary and secondary staining. (b) The elimination of both α 2-3 and α 2-6 linked sialic acids led to the binding loss of CD33-Streptactin precomplexes, which indicates both linkages of sialosides are important for *trans* interactions with CD33. CD33R indicates CD33 with R119A, which lacks its sialic acid dependent binding. Statistical significance calculated based on a one-way ANOVA with Tukey's multiple comparisons test.

2.3.5 CaR-ESI-MS with CD33-Fc and a natural N-glycan library from U937 cells

An advantage of CaR-ESI-MS is that it can be used in a 'shotgun' style approach for studying the binding to natural libraries of glycans derived from a biological source. As both NeuA and NeuS can abrogate *cis* and *trans* interactions of CD33 on the surface of U937 cells, it is hypothesized that N-glycans terminating with both α 2-3 and α 2-6 linked sialic acids are potentially ligands for CD33. First, a natural N-glycan library was prepared from U937 cells by enzymatic digestion with trypsin and PNGase F, followed by purification with a Seppak-C18 column. (**Figure 2.13a**) The mass spectrum for the N-glycan library show various N-glycan structures with different degrees of fucosylation, sialylation, and antenna. (**Figure 2.13b**) In a separate study, it was analyzed that 97 unique N-glycan structures were resolved, which were composed of 40% of neutral glycans and 60% of acidic glycans by combining 2-aminobenzamide-based fluorescent

labeling of the *N*-glycan library with a hydrophilic interaction–ultrahigh-performance liquid chromatography technique.⁴⁵

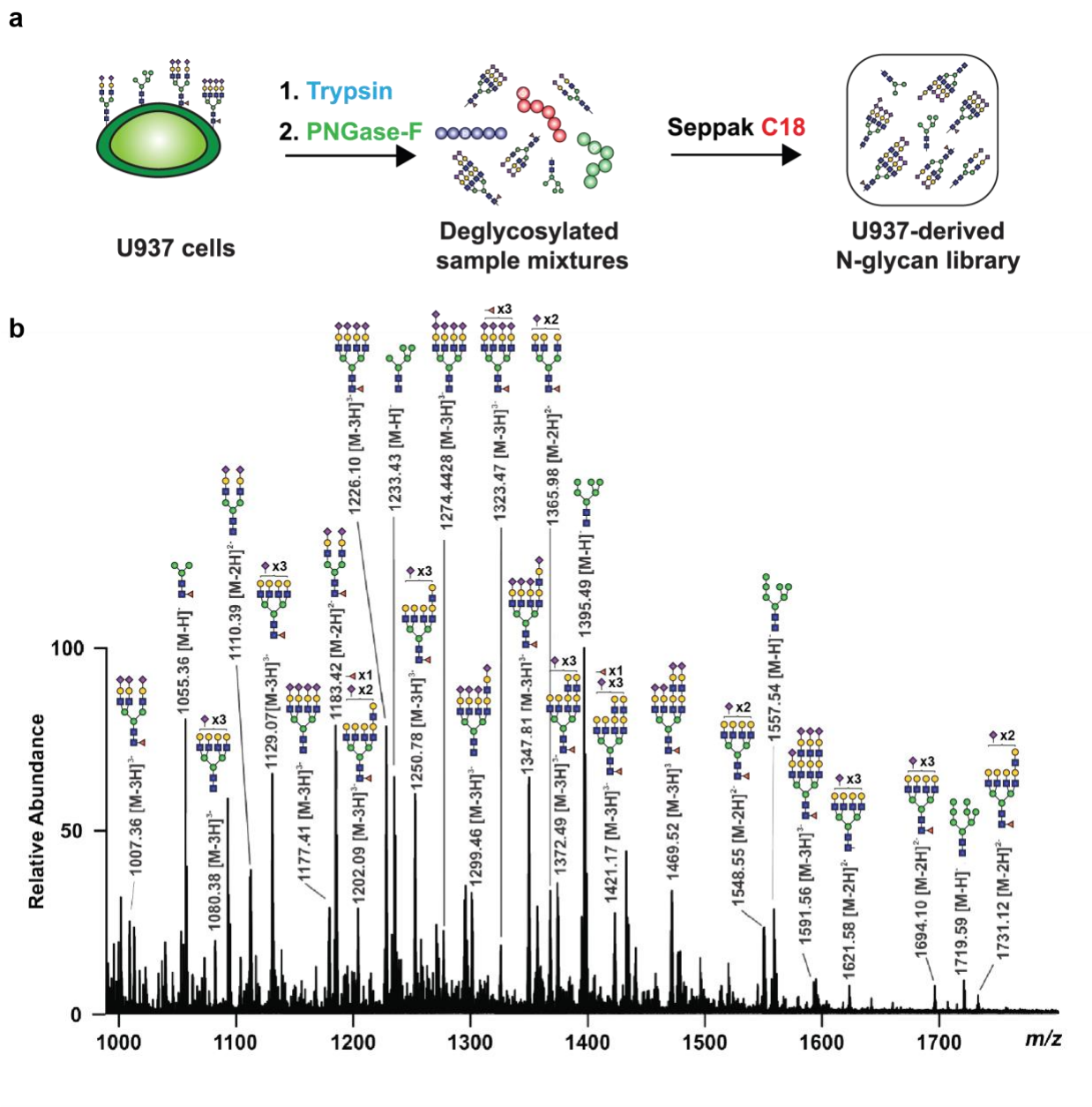


Figure 2.13: Mass spectrum *N*-glycans derived from U937 cells. (a) Trypsin and PNGase F help to isolate glycan structures from the U937 cell lysate, which is further purified by a Seppak C18 reverse-phase column. (b) The mass spectrum was gained from the natural *N*-glycan library from the U937 cells in negative ion mode.

CD22-Fc was used to optimize the CaR-ESI-MS with the U937 derived *N*-glycan library. (Figure 2.14) As expected, neutral glycans were not released. CD22 strongly prefers α 2-6 linkage, so these *N*-glycans with mono-, bi-, tri-, and tetra sialylation could contain more than one sialic acid with a α 2-6 linkage. Of note, CD22-Fc bound to mono α 2-6 linked sialic acid on biantennary structures, and it preferred α 2-6 linked sialic acid present on α 1-3 branch to α 1-6 branch on the *N*-glycan core structure.⁴⁵ Also, the combination of α 2-6 linked sialic acid on α 1-3 branch and α 2-3 linked sialic acid on α 1-6 branch was ranked higher than the disialylated counterpart with α 2-6 linkages on both branch. (Figure 2.14)

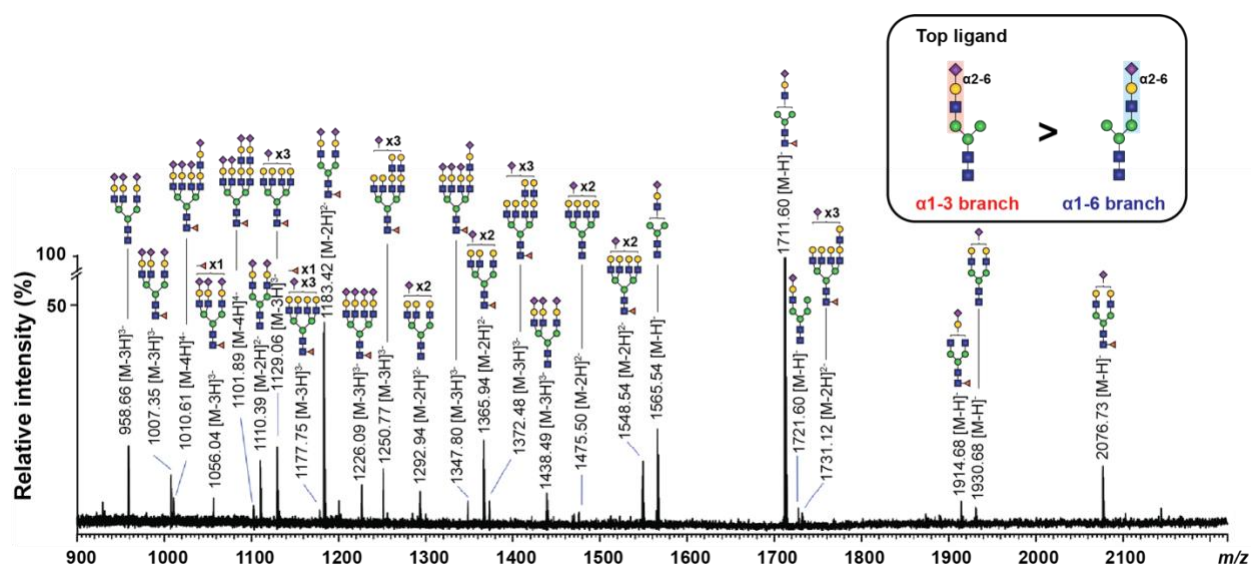


Figure 2.14: CaR-ESI-MS spectra for CD22-Fc against *N*-glycans from U937 cells. The *N*-glycan library isolated from the U937 cell lysates are screened with CD22-Fc and various *N*-glycans were released after CID is applied to the complexes of CD22-Fc and ligands. HCD mass spectrum acquired for ions with *m/z* of 6,500–9,000 produced from a solution (0 °C) of hCD22-Fc (6 μ M) *N*-glycan library (500 μ g/mL) using a collision energy of 110 eV.

Having validated the CaR-ESI-MS approach with CD22, the natural *N*-glycan library was used to probe *N*-glycan ligands for CD33-Fc. (Figure 2.15) Relatively low numbers of *N*-glycan compositions compared to CD22 were discovered, which is due to the weaker binding of CD33

for glycan ligands.¹⁰⁶ In this assay, a composition representing disialylated biantennary *N*-glycans without core fucosylation was the most abundant. Since CD33 binds both α 2-3 and α 2-6 linkages, both linkages of sialic acid on biantennary *N*-glycans may be enriched in this *N*-glycan composition. CaR-ESI-MS with a defined *N*-glycan library also showed that the disialylated species have slightly higher signals than monosialylated counterparts. (**Figure 2.6**) The low intensities from the incomplete LacNAc formation on the α 1-6 may be caused by either relatively lower binding affinity or insufficient amount of the structure in the library. For CD22, the composition of sialyl LacNAc present only on the α 1-3 branch was top ranked, but CD33 seems to prefer intact LacNAc on both branches with sialylation.

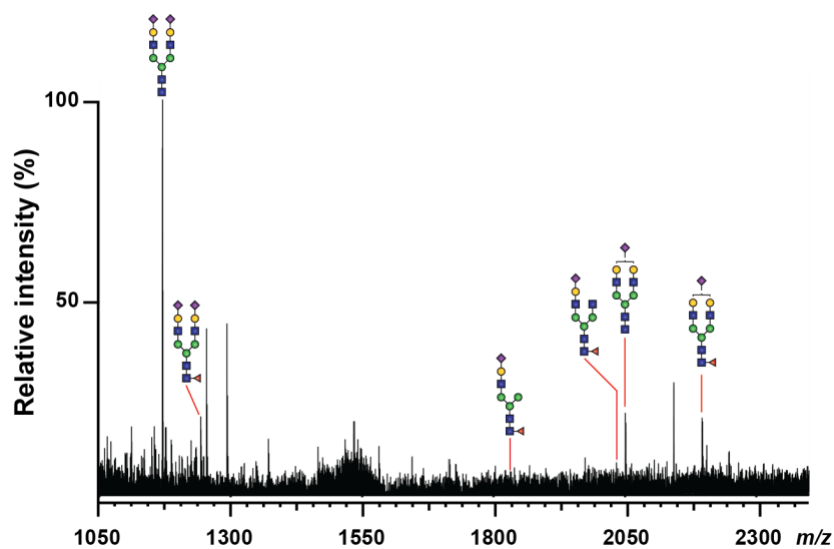


Figure 2.15: CaR-ESI-MS spectra for CD33-Fc against *N*-glycans from U937 cells. CaR-ESI-MS with CD33-Fc and the natural *N*-glycan library discovered 6 *N*-glycan compositions as potential ligands. Sialylated biantennary *N*-glycans carrying LacNAc on each branch had higher intensities than the one carrying a single LacNAc on either branch.

2.4 Discussion

A new generation of Siglec-Fc was developed and applied to multiple applications successfully. The focus was on showing the utility of CD33-Fc in applications aimed at

understanding its glycan ligands because of the association between CD33 and LOAD susceptibility. The intact CD33-Fc was applied to the CaR-ESI-MS approach to elucidate its natural *N*-glycan ligands. Moreover, the internal TEV recognition site within the linker between CD33 and IgG constant domains provided an additional benefit, enabling the generation of an optimized molecular weight for ESI-MS applications. Endo H digestion enabled the CD33 fragment to be homogenous and suitable for an ESI-MS-based binding assay. The C-terminal Strep-tag II on CD33-Fc enabled the formation of a highly multivalent probe to investigate ligands on the U937 cells after precomplexing with AF647-labelled Streptactin. Most studies use an anti-human IgG secondary, to effectively generate tetrameric presentation of the Siglec. However, several drawbacks in a recombinant Siglec-Fc with human IgG isotype are present such as non-specific interactions with Fc γ receptors, and insufficient avidity. To circumvent these limitations, Schnaar and colleagues developed a pentameric COMB construct with added benefits of multivalency and avoidance of the Fc domains.³⁹ Our octameric presentation is the most multivalent system developed to date with the added benefit of mutations in the Fc that avoid concerns about binding to Fc γ receptors on primary cells or tissues.

One key application with this new recombinant CD33 was to assess the affinity towards sialoside ligands. There have been numerous approaches used for determining the affinity of CD33 for its glycan ligands. For example, an ELISA based method has been generally used, which monitors the binding decrease of Siglecs upon titration of target compounds. To be specific, glycan ligands for testing are immobilized on a surface of a microwell plate, and Siglec-Fc is incubated and a maximum signal from the Siglec-ligand interaction per well is recorded. Subsequently, molecules competing with the standard molecule are titrated for calculating IC_{50} values.⁵⁰ Similarly, a glycan ligands anchored to beads was applied to the IC_{50} calculation by flow

cytometry.¹¹³ Dissociation constants for Siglecs have been also determined by NMR titration, isothermal titration calorimetry, and surface plasmon resonance (SPR).^{91,116} To study the dissociation constant of glycoproteins by nature MS, heterogeneity CD33 had to be first addressed, because the presence of many glycoisoforms made it impossible to work with the fully glycosylated protein. To address this, CD33-Fc was expressed from CHO Lec-1 cells and digested by both TEV proteases and Endo H to achieve the homogenous CD33 fragments with a suitable molecular weight for ESI-MS applications. This approach maintained a single GlcNAc residue at each *N*-glycosylation site, which retains beneficial effects on protein folding compared to eliminating the entire *N*-glycan.¹¹² Compared to the reported IC_{50} of CD33 on the same panel of trisaccharides,¹¹³ the ESI-MS based K_d measurements yielded a eight-fold higher value. As IC_{50} value is calculated based on how effectively a target molecule interferes with an interaction between protein and its ligand, the difference could stem from different analytical ways in which the K_d or IC_{50} values were measured. In 2019, a co-crystal structure of the CD33-CD33L complex was determined and the dissociation constant was determined by SPR as $118 \pm 41 \mu\text{M}$, which is comparable to the value obtained by our MS approach. Consequently, MS-based determination of dissociation constants provides feasible and practical way as it can generate comparable results to the value calculated by SPR approach.¹¹⁶ After method validation between the CD33 fragment and CD33L, $\alpha 2-3$ and $\alpha 2-6$ sialyl LacNAc were titrated against the CD33 fragment. Notably, the calculated K_d values for the two compounds were similar to each other. Likewise, an identical IC_{50} values, determined from an ELISA, between $\alpha 2-3$ and $\alpha 2-6$ sialyl LacNAc on CD33 were reported, which was 4.72 mM.⁵⁰ Even though a direct comparison between K_d and IC_{50} is not ideal, the two calculated values on 3'SLN and 6'SLN were only 2-fold different. The important finding is that both $\alpha 2-3$ and $\alpha 2-6$ linked sialosides may play a role in modulating CD33 in immune cells.

From the defined glycan libraries where all glycan members are characterized, CD33 was shown to recognize both α 2-3 and α 2-6 linked sialosides. To expand this application, a *N*-glycan library from U937 cells was prepared by a series of enzymatic reactions and purifications. The CaR-ESI-MS with CD33-Fc and the natural *N*-glycan library elucidated 6 *N*-glycan ligand compositions. Among these compositions, disialylated biantennary *N*-glycan was the top hit, but which combination of sialic acid linkages is the best remains undefined. Inconsistent with this result, it was reported α 2-6 linked sialic acid on α 1-3 branch of biantennary structures was the most strong ligand part while α 2-3 linked sialic acid on α 1-6 branch was a weaker ligand for CD33.²⁴ Because CaR-ESI-MS with CD33-Fc and a defined *N*-glycan library did not show a great difference between both α 2-3 and α 2-6 linkages, these variations might be derived from the nature of the two different methods between CaR-ESI-MS and microarray. To assess effects of α 2-3 and α 2-6 linked sialic acids on the interactions with CD33, generating *N*-glycan library from U937 cells with different genetic backgrounds can be considered in the future. For example, knocking out *ST6Gal1* to deplete α 2-6 linked sialic acids on most *N*-glycans, or *ST3Gal4* to deplete α 2-3 linked sialic acids would be logical ways to complete these studies. Alternatively, NeuS can help to deplete α 2-3 linked sialosides on *N*-glycans, but the lack of neuraminidase specific for α 2-6 sialosides can be an obstacle. Hence, the combination of genetic knock-out and neuraminidase may need to be pursued.

Another key application with our recombinant CD33-Fc was assessing its glycan ligands on cells. A liposome-based assay helped to extend understanding on both α 2-3 and α 2-6 linkages of sialic acids play a role in regulating CD33 interactions on U937 cells. In a separate experiment performed by a colleague in the lab, primary human monocytes were tested and the same results were found.¹⁰⁶ In the future, genetic or pharmaceutical approaches to inhibit each glycosylation

pathway would be a valuable way to assess which glycosylation can affect CD33 interactions. Nevertheless, CaR-ESI-MS with CD33-Fc and the U937 cell-derived *N*-glycan library helps to understand that *N*-glycans can be one of players modulating CD33 interactions. Furthermore, an investigation into different types of glycosylation present on more physiologically relevant cells and tissues such as *O*-glycans or gangliosides needs to be investigated.

2.5 Conclusions

A new generation of CD33-Fc was applied to ESI-MS and flow cytometry-based applications. This new CD33-Fc was shown to be versatile in the applications shown here. Affinity tags afforded the generation of CD33-Fc with high purity for use in ESI-MS based applications. The screening with a defined *N*-glycan library revealed CD33 can interact with both α 2-3 and α 2-6 linked sialic acids on biantennary *N*-glycans. To expand this finding, *N*-glycan library was prepared from more biologically relevant U937 cells and screened with CD33-Fc. Consistently, 6 compositions of *N*-glycans are discovered and these are all biantennary *N*-glycans. To evaluate the biological significance of both linkages, U937 cells were desialylated by NeuA and NeuS, and *masking* effects on the endogenous CD33 was evaluated by a liposome-based assay. It was shown that both linkages played a role in modulating *cis* and *trans* interactions of CD33 on U937 cells. Lastly, K_d values for between CD33 fragment and both α 2-3 and α 2-6 sialyl LacNAc were determined, and we discovered that CD33 recognizes both linkages without a huge difference in K_d . Consequently, the new generation of CD33-Fc with novel features was shown to be successfully applied to both ESI-MS and flow cytometry applications.

2.6 Methods

2.6.1 Neuraminidase-A and -S preparation

The coding sequence of Neuraminidase-A and Neuraminidase-S (GenBank accession number: AY934539.2 for NeuA and ABJ55283 for NeuS) were commercially synthesized into pET100/D-TOPO vectors. (GeneArt Gene Synthesis, CA) The pET100/D-TOPO vectors containing each sialidase gene were chemically transformed into BL-21 competent cells (New England Biolabs, CA). The transformed cells were cultured in 50 mL of lysogeny broth (LB) containing 100 µg/mL of ampicillin at 37 °C for 18 h on the shaking incubator with 20 rcf. The cultured cells were scaled up to 1 L of LB-amp for 4 h. After OD₆₀₀ reached 0.6, 1 mM of IPTG was added and the culture was shaken at 25 °C for 18 h. The culture was centrifuged at 13,000 rcf for 30 min. The cell pellet was resuspended into 40 mL of cell lysis buffer (50 mM of Na₂PO₄, 400 mM NaCl, 0.1% w/v of triton-100, 1 mg/mL of lysozyme, pH 7.7) and incubated at 37 °C for an hour. Then, the cells were sonicated with half intensity of 50% pulses with three intervals of 10 s sonication and 50 s rest. The cell lysates were centrifuged at 18,000 rcf for 30 min. The supernatant was filtered through a 0.22 µm filter and loaded onto a 1 mL HisTrap HP (GE LifeSciences) that was equilibrated with 10 CV of lysis buffer. The column was then washed with washing buffer (50 mM sodium phosphate, 400 mM NaCl, pH 7.7), and then eluted with elution buffer (50 mM sodium phosphate, 400 mM NaCl, 500 mM of imidazole, pH 7.7). The elution fractions containing proteins were pooled together and dialyzed three times with a storage buffer (50 mM NaCl, 20 mM Tris-HCl, 1 mM EDTA, pH 7.5). After dialysis, each aliquot was stored at –80 °C until use.

2.6.2 TEV protease preparation

The gene encoding TEV with a N-terminal His₆-tag was transformed into BL21 pLysS competent cells was grown in 30 mL of LB media with 100 mg/L ampicillin and 25 mg/L chloramphenicol overnight at 37 °C. The following morning, 10 mL of the starter culture was added to 1 L of LB media with the same ratio of antibiotic and allowed to grow at 37 °C until OD₆₀₀ = 1.5. After, 1 mM isopropylthio-β-galactoside (IPTG) was added and the culture was shaken at 18 °C for 18 h. The culture was spun down at 14,000 rcf for 30 min at 4 °C, after which the pellet was resuspended in lysis buffer (20 mM Tris, 500 mM NaCl, 5 mM Imidazole, 5% Glycerol, 1 mM DTT, pH 8.0) and cells were lysed using a tissue homogenizer. The eluted solution was filtered using a 0.45 μm PES filter unit (Fisher Scientific) and loaded onto a HisTrap HP (GE LifeSciences) that was equilibrated with 10 column volume (CV) of lysis buffer. The protein was eluted off the column using 10 CV of elution buffer (20 mM Tris, 500 mM NaCl, 500 mM Imidazole, 5% glycerol, 1 mM DTT, pH 8.0) and then buffer exchanged into storage buffer (20 mM Tris, 20% glycerol, 1 mM DTT, pH 8.0).

2.6.3 Endo H preparation

The coding sequence of endo-beta-N-acetylglucosaminidase H (GenBank accession number: K02182.1) was commercially synthesized into a pET-DEST40 vector. After transformation into BL21(DE3) competent cells, colonies were selected on LB-Agar containing 100mg/L Ampicillin. The rest of steps are the same with the preparation of neuraminidase above.

2.6.4 CD33-Fc preparation

CD33 gene containing 2 domains with pairs of forward and reverse primers were cloned into pcDNA5/FRT/V5-His-TOPO vector (Invitrogen), which contained a C-terminal hIgG1-Fc, TEV cleavage site on the N-terminal side of the Fc, as well as a C-terminal His₆ tag and a Strep-

tag II. (Appendix A) The vector containing CD33-Fc was co-transfected with pOG44 vectors into Chinese Ovary Hamster cell lines after incubating with Lipofectamine LTX (Thermo Fisher) at RT for 30 mins as previously described.¹⁷ Subsequently, transfected cells were selected under DMEM-F12 (Gibco) media containing 10% FBS, 50 U/mL Penicillin/Streptomycin, 100 mM HEPES, and 1 mg/mL of Hygromycin (Thermo Fisher). Selected cells were cultured in DMEM-F12 media containing 5% FBS, 50 U/mL Penicillin/Streptomycin, and 100 mM HEPES for one week after cells reached 90% confluency. Finally, the supernatants were filtered through a sterile filter unit (0.22 μ m pore size, Thermo Fisher) and stored at 4 °C.

2.6.5 CD33 fragment preparation

Siglec-Fc proteins were purified by Histrap™ Excel (Cytiva) according to manufacturer guidelines and dialyzed into a PBS buffer overnight. The dialyzed protein was subsequently purified by Hitrap™ Protein G (Cytiva) as manufacturer guidelines and 2 mL of the elution fractions were pooled together and adjusted at pH 7 for TEV and Endo H digestion for 3 hr at 37 °C and overnight at 4 °C. The reaction mixture was diluted with 10 mL of PBS buffer, passed through a Histrap™ Excel column. The flow through was collected for further dialysis three times with 200 mM of ammonium acetate at 4 °C. The proteins were concentrated with Amicon (cutoff 10 kDa, Sigma) to achieve around 0.200 mg/mL of protein solution for mass spectrometry.

2.6.6 N-glycan library preparation

N-glycans from U937 cells were prepared based on a reported procedure. Briefly, the cells (1.5×10^8) were homogenized in 15 mL of lysis buffer (25 mM TRIS, 150 mM NaCl, 5 mM ethylenediaminetetraacetic acid and 0.5% (w/v) 3-((3S4cholamidopropyl)dimethylammonio)-1-propanesulfonate, pH 7.4). The cell lysate was centrifuged at 3,400 rcf for 30 mins and the supernatant was dialyzed against 50 mM ammonium bicarbonate three times and lyophilized.

Samples were re-suspended with 2 mg mL⁻¹ of 1,4-dithiothreitol (Sigma-Aldrich Canada) in 2 mL of 0.6 M TRIS buffer (pH 8.5) and kept at 50 °C for 1.5 h. After cooling, 12 mg/mL of iodoacetamide (Sigma) in additional 2 mL of 0.6 M TRIS buffer (pH 8.5) was added and placed in the dark at room temperature for 1.5 h. The sample was dialyzed against 50 mM ammonium bicarbonate at 4 °C three times and lyophilized. The reduced and alkylated sample was re-suspended in 2 mL of 50 mM ammonium bicarbonate and then digested with 1 mg of trypsin (Sigma-Aldrich) at 37 °C for 18 h followed by purification using a C18 Sep-Pak (200 mg) column (Waters). The column was conditioned with 3 mL of methanol, 3 mL of 5% of acetic acid, 3 mL of 1-propanol, and 3 mL of 5% of acetic acid. After loading the sample, the column was washed with 6 mL of 5% acetic acid. The peptides were eluted sequentially with 2 mL of 20% 1-propanol, 2 mL of 40% 1-propanol and then 2 mL of 100% 1-propanol. The eluted fractions were pooled together and lyophilized. The dried sample was resuspended in 1 mL of 50 mM ammonium bicarbonate and incubated with 3 µL of PNGase F (500,000 units/mL, New England Biolabs) at 37 °C for 4 h, and further incubated with additional 5 µL of PNGase F for 18 h. The released *N*-glycans were purified using C18 Sep-Pak (200 mg) column conditioned with 3 mL of methanol, 3 mL of 5% of acetic acid, 3 mL of 1-propanol, and 3 mL of 5% of acetic acid. The sample was added, followed by washing with 6 mL of 5% of acetic acid. The flow-through from the sample load and the washing steps which contains *N*-glycans were pooled and lyophilized. The lyophilized sample was stored at -20 °C until needed.

2.6.7 Flow cytometry with two steps or precomplexing CD33-Fc with Streptactin

Dimeric CD33-Fc (200 nM) and monomeric Streptactin (80 nM) were pooled together in a single volume. For two steps, 50 µL of CD33-Fc solution was added to the cells and incubated on ice for 15-20 min. Subsequently, the secondary Streptactin was added after washing the cells

once with PBS. For pre-complexing, the mixture was incubated on ice for 15-20 mins. Next, 50 μ L of the mixtures was added to each well containing 100,000 cells on a 96-well U-bottom plate. After incubation on ice for 15 mins, the plate was centrifuged (300 rcf, 5 min). After washing once with 200 μ L of flow buffer (HBSS, 1% BSA, 0.1% EDTA) with centrifugation, cells per well are resuspended in 200 μ L of flow buffer and transferred to flow tubes. After gating cell population by ranging forward scatter and side scatter, intensities in an APC channel were measured to access CD33 binding.

2.6.8 Formulation of CD33 ligand-targeted liposomes

Commercially available lipids—DSPC and cholesterol—were suspended in chloroform and an appropriate volume of each lipid solution in chloroform was transferred into a glass test tube to reach the desired mol% (60.9% DSPC and 38% cholesterol) of each lipid. The solvent was removed under nitrogen gas to form the lipid mixtures. Once all visible chloroform was removed, 100 μ L of dimethyl sulfoxide (DMSO) was added to the test tube. CD33L-DSPE (compound 6, 1 mol%) and DSPE-PEG-A647 (0.1 mol%) in DMSO were then added to the lipid mixture. The samples were then placed at -80°C until completely frozen and excess DMSO was removed via lyophilization overnight and then they dried liposomes were stored at -80°C until they were extruded. Dried lipids were then allowed to warm to RT and were then hydrated with 1.0 mL of phosphate-buffered saline pH 7.4. The hydrated lipids were then sonicated in a cycle of 1 min on, 4–5 min off until all lipids were uniformly suspended. The lipids were then extruded with an 800 nm filter and then 100 nm filters. The average diameter of the liposomes was verified by dynamic light scattering (Malvern Panalytical Zetasizer Nano S) to be 110 ± 20 nm. Liposomes were stored at 4°C .

2.6.8 Liposome assay

A million of cells were digested by 50 μ L of NeuA (1 mg/mL) or NeuS (1 mg/mL) in 500 μ L PBS for 2h, cells were washed with flow buffer twice. Next, 100,000 cells per well were added to a 96-well U-bottom plate in 200 μ L RPMI containing 10% FBS. Cells were centrifuged (300 rcf, 5 min) and the supernatant was discarded. The cell pellet was re-suspended in 50 μ L of fresh media and 50 μ L of media containing naked (without CD33 ligand) or CD33 ligand (CD33L) tagged liposomes was added to it. The final concentration of liposomes was 100 μ M (total concentration of lipid). These suspensions were incubated for 30 min at 37 °C. Following this incubation, 100 μ L of media was added to each sample and they were centrifuged (300 rcf, 5 min). The supernatant was discarded, and the cell pellet was suspended in a flow buffer and further analyzed by flow cytometry.

2.6.9 Flow cytometry

Flow cytometry was performed on a BD LSRFortessa™ X-20 cell analyzer using BD FACSDiva Software Version 8.0.1 and data was processed using FlowJo LLC. Version 9.9.6 and Graphpad Prism 8. The live, single cell populations were gated out with two gates to obtain the median fluorescence intensity (MFI), which indicates binding of the Siglec-Fc-secondary complex to the cell surface. Control cells that were incubated with Streptactin-AF647 (50 μ L of 80 nM) was used to determine the baseline fluorescence. Number of cells (# cells) plotted indicates percentage of max (% max).

2.6.10 ESI mass spectrometry

The affinities of α 2-3 and α 2-6 sialosides for the three Siglec fragments were monitored using the ESI-MS-based binding assay. In all cases, a reference protein (P_{ref}), cytochrome C, was added to the solution in order to correct mass spectra for any nonspecific binding that occurred

during the ESI process. For a 1:1 protein–ligand complex, the dissociation constant (K_D) can be calculated from the abundance (Ab) ratio (R) of the ligand bound (PL) to free protein (P) ions (Eq. (1)) measured by ESI-MS for solutions of known initial concentration of protein ($[P]_0$) and ligand ($[L]_0$), Eq. (2):

$$R = \frac{AB(PL)}{AB(P)} = \frac{[PL]_{eq}}{[P]_{eq}} \quad (1)$$

$$K_d = \frac{[L]_0}{R} - \frac{[P]_0}{R+1} \quad (2)$$

In principle, K_d can be determined from measurements performed using a single concentration of P and L. However, to establish a reliable K_D for a low affinity interaction it is generally preferable to utilize a titration approach, wherein $[P]_0$ is maintained constant and $[L]_0$ is varied. The affinity is determined by fitting Eq. (2) to a plot of fraction bound protein ($R/(R + 1)$) versus $[L]_0$:

$$\frac{R}{R + 1} = \frac{[P]_0 + [L]_0 + K_D - \sqrt{([P]_0 + [L]_0 + K_D)^2 - 4[P]_0[L]_0}}{2[P]_0} \quad (3)$$

The reported errors correspond to 2 standard deviations (95% confidence interval). Direct ESI-MS measurements were performed in positive ion mode on a Synapt G2S quadrupole-ion mobility separation-time of flight (Q-IMS-TOF) mass spectrometer (Waters UK Ltd., Manchester, UK) and Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific). In both cases, nanoflow ESI (nanoESI) was used. The nanoESI tips, with $\sim 5 \mu\text{m}$ outer diameters (o.d.), were produced in-house from borosilicate capillaries (1.0 mm o.d., 0.68 mm i.d.) using a P-1000 micropipette puller (Sutter Instruments, Novato, CA). A voltage of $\sim 1 \text{ kV}$ was applied to a platinum wire inserted into the back end of the nanoESI tip and in contact with the sample solution. For the measurements performed on the Synapt G2S instrument, the Source temperature was 60°C

and the Cone, Trap and Transfer voltages were 20 V, 3 V, and 1 V, respectively. All other instrumental conditions were set to the default parameters. Data acquisition and processing were carried out using MassLynx (Waters, version 4.1). For the Q-Exactive Orbitrap instrument, the automatic gain control target, the maximum injection time, capillary temperature, and S-lens RF level were set to 1×10^6 , 100 ms, 150 °C and 100, respectively. The resolution was 17,500 at m/z 200. Data acquisition and processing were carried out using Xcalibur (Thermo Fisher, version 4.1).

2.6.11 CaR-ESI-MS Screening

Measurements were performed in negative mode using a Q Exactive Orbitrap mass spectrometer with Ultra High Mass Range (Q Exactive UHMR, Thermo Fisher Scientific) equipped with a modified nanoflow ESI (nanoESI) source. NanoESI tips with an outer diameter (o.d.) of $\sim 5 \mu\text{m}$ were pulled from borosilicate glass (1.0 mm o.d., 0.78 mm inner diameter) with a P-1000 micropipette puller (Sutter Instruments, Novato, CA). Mass spectra were acquired using a spray voltage of 0.9–1.0 kV. The capillary temperature was 200 °C, and the S-lens RF level was 100; an automatic gain control target of 5×10^6 and maximum injection time of 200 ms were used. The resolving power for full MS and CID [referred to in the present work as HCD (higher energy CID)] spectra was set to 3125 and 25,000, respectively. For the CaRESI-MS measurements, HCD spectra were acquired using a range of collision energies. The temperature of the solution in the nanoESI tip was controlled using a custom temperaturecontrolled device, as described previously.²³ Data acquisition and analysis were performed using Xcalibur version 4.1.

Chapter 3

Carbohydrate sulfation as a mechanism for fine-tuning Siglec ligands

3.1 Acknowledgements

I thank Nitin, and Dr. Wesley F. Zandberg for analyzing the compositions of *N*-glycans from each CHST overexpressing U937 cells. I thank Parisa Raeisimakiani, and Dr. Lala K. Mahal for carrying out Lectin array. I thank Chu-Wei Kuo, and Dr. Kay-Hooi Khoo for identifying *O*-glycosylation sites on the CD33 fragment. I thank Po-Han Lin, and Dr. Xuefei Huang for synthesizing a series of sulfated LacNAc disaccharides. I thank Fahima Mozaneh for installing α 2-3 sialylation on the sulfated LacNAc disaccharides. I thank Duong T. Bui and Dr. John S. Klassen for determining dissociation constants between the CD33 fragment and the mono- or di- sulfated 3'SLN trisaccharides. I thank Jhon R. Enterina for performing statistical analyses on the correlation between CHST1 and 14 tumour types from TCGA. I thank the GlycoNet Synthetic core for assistance in preparing the high affinity CD33 ligand. I thank Dr. Gour C. Daskhan for conjugating the high affinity CD33 ligand to a liposome. I thank Chris D. St. Laurent for the advice on the viral transduction.

Chapter 3 was published in Jung *et al.*, *ACS Chemical Biology*, 2021.

3.2 Introduction

A growing body of evidence points to sulfation of the underlying glycan as being an important modification to the glycan structure that can increase the affinity of glycans for Siglecs.²⁶ For example, human B-cells contain high levels of 6-*O*-GlcNAc sulfation, which maintains CD22 in a *masked* state on naive B-cells.^{43,117} By glycan microarray, SLe^X containing 6-*O*-sulfated Gal (6'-S-SLe^X) was the top hit as a ligand for Siglec-8-Fc.⁴⁶ Elegant affinity measurements and structural studies on Siglec-8 with sulfated SLe^X demonstrated that Siglec-8 has a 40-fold enhanced affinity for 6'-S-SLe^X and a 4-fold enhanced affinity for the isomer bearing the sulfate appended to GlcNAc (6-S-SLe^X).⁹¹ Interestingly, the disulfated compound (6,6'-S,S-SLe^X) displayed only modest additivity of these affinity gains, with a further 1.6-fold enhancement in affinity relative to 6'-S-SLe^X. The structural basis for the affinity gain provided by 6-*O*-Gal sulfation was revealed to be an ionic interaction between the sulfate and an arginine contained within a loop adjacent to the glycan binding site. It is unknown if this mode of recognition of the sulfate is shared with other Siglecs but the primary sequence of this loop is not well conserved (in both size and composition) between Siglec family members. More recently, sialylated keratan sulfate was discovered as a ligand for Siglec-8 in human airways, which is significant as keratan sulfate often contains sulfation on both the underlying Gal and GlcNAc.³⁹ Glycan microarrays also revealed that Siglec-7 has enhanced binding to 6-S-SLe^X and 6'-S-SLe^X over SLe^X, whereas Siglec-9 displayed a preference for 6-S-SLe^X,²⁶ which is consistent with enhanced binding of Siglec-7 and Siglec-9 to ECV304 cells transfected with *carbohydrate sulfotransferase 2* (*CHST2*), which installs a sulfate onto the 6-position of GlcNAc.⁵¹ Very recently, probing Siglec ligands on HEK293 cells transfected with a series of *CHST*s genes demonstrated that overexpression of *CHST1*, which catalyzes 6-*O*-Gal sulfation,¹¹⁸ enhanced the binding of Siglec-3, -7, -8, and -15.¹¹⁹

Although it is increasingly clear that sulfation plays a role in Siglec biology, numerous unanswered questions remain about the ability of CHST1 to enhance Siglec ligands, including context dependence, which was not studied in earlier work. In this chapter, it is demonstrated that overexpressed eight CHSTs in a variety of cell lines and discovered both enhanced and diminished binding of human and mouse Siglec-Fc proteins. Several of these enhanced interactions were cell type dependent. Through determining the affinity of CD33 for several synthetic glycan ligands, a disulfated ligand, Neu5Ac α 2-3(6-sulfo)Gal β 1-4(6-sulfo)GlcNAc (6,6'-S,S-3'SLN), was found to bind 36-fold tighter than its corresponding non-sulfated counterpart. Genetic analyses reveal that many cancer types upregulate *CHST1*, and several show poorer survival rates in patients with higher levels of *CHST1*. Moreover, inhibiting carbohydrate sulfation with sodium chlorate lowers Siglec binding to cancer cell lines that express high levels of *CHST1* or *CHST2*. Taken together, our study reveals new insights into carbohydrate sulfation as a mechanism for upregulating cellular Siglec ligands, which is prevalent in cancer.

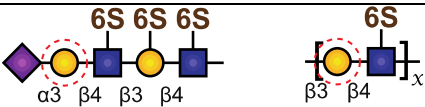
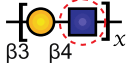

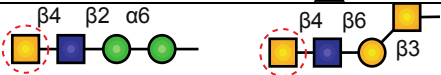
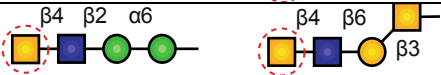
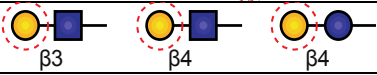
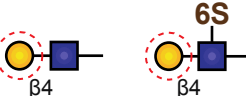
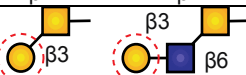
3.3 Results

3.3.1 Enhanced glycan sulfation by CHST overexpression

CHSTs are Golgi-localized enzymes that install sulfate at defined locations on glycans.¹²⁰ Of the large family of human CHSTs, at least eight (CHST1, CHST2, CHST4, CHST8, CHST9, Gal3ST2, Gal3ST3, and Gal3ST4) are known to install sulfate onto glycan subclasses that contain sialic acid (e.g. *N*-glycans and mucin-type *O*-glycans),¹²¹ which could potentially serve as Siglec ligands. (**Table 3.1**) Therefore, the genes encoding these eight CHSTs were overexpressed in U937 cells through lentiviral transduction (**Figure 3.1a**). U937 cells are a monocytic cell line and were chosen because they are known to express Siglec ligands and interactions between Siglecs with their glycan ligands on white blood cells is physiologically relevant.¹⁰⁶ Three different methods

were used to characterize the glycosylation on the CHST-overexpressing cells: mass spectrometry (MS) of released *N*-glycans, lectin microarrays, and lectin staining in conjunction with flow cytometry.

Table 3.1 The known substrates for each CHST.

Gene	HGNC ID	Substrate	Reference
<i>CHST1</i>	HGNC:1969		122
<i>CHST2</i>	HGNC:1970		123
<i>CHST4</i>	HGNC:1972		124
<i>CHST8</i>	HGNC:15993		125
<i>CHST9</i>	HGNC:19898		126
<i>Gal3ST2</i>	HGNC:24869		127
<i>Gal3ST3</i>	HGNC:24144		128
<i>Gal3ST4</i>	HGNC:24145		129

(Red dashed circles indicate substrates where each CHST acts)

By MS, enhanced sulfation was evident in all eight CHST-overexpressing lines (**Figure 3.1b**) Indeed, 20-40% of unique, complex-type *N*-glycans contained a sulfate (accounting for an average of $4.2 \pm 1.4\%$ of the MS peak area) in the CHST-overexpressing lines, while no sulfated complex *N*-glycans were detected from the U937 cells transduced with control lentivirus. By lectin microarray, *Sambucus Nigra* lectin (SNA), which is specific for α 2-6 linked sialosides independent of sulfation, showed enhanced binding to the *CHST1*-overexpressing cell lysate. Binding of *Maackia amurensis* lectin (MAA), which recognizes α 2-3 linked sialosides or LacNAc-terminated glycans with 3-*O*-sulfated Gal, was highly enriched in the *Gal3ST2*- and *Gal3ST3*-overexpressing

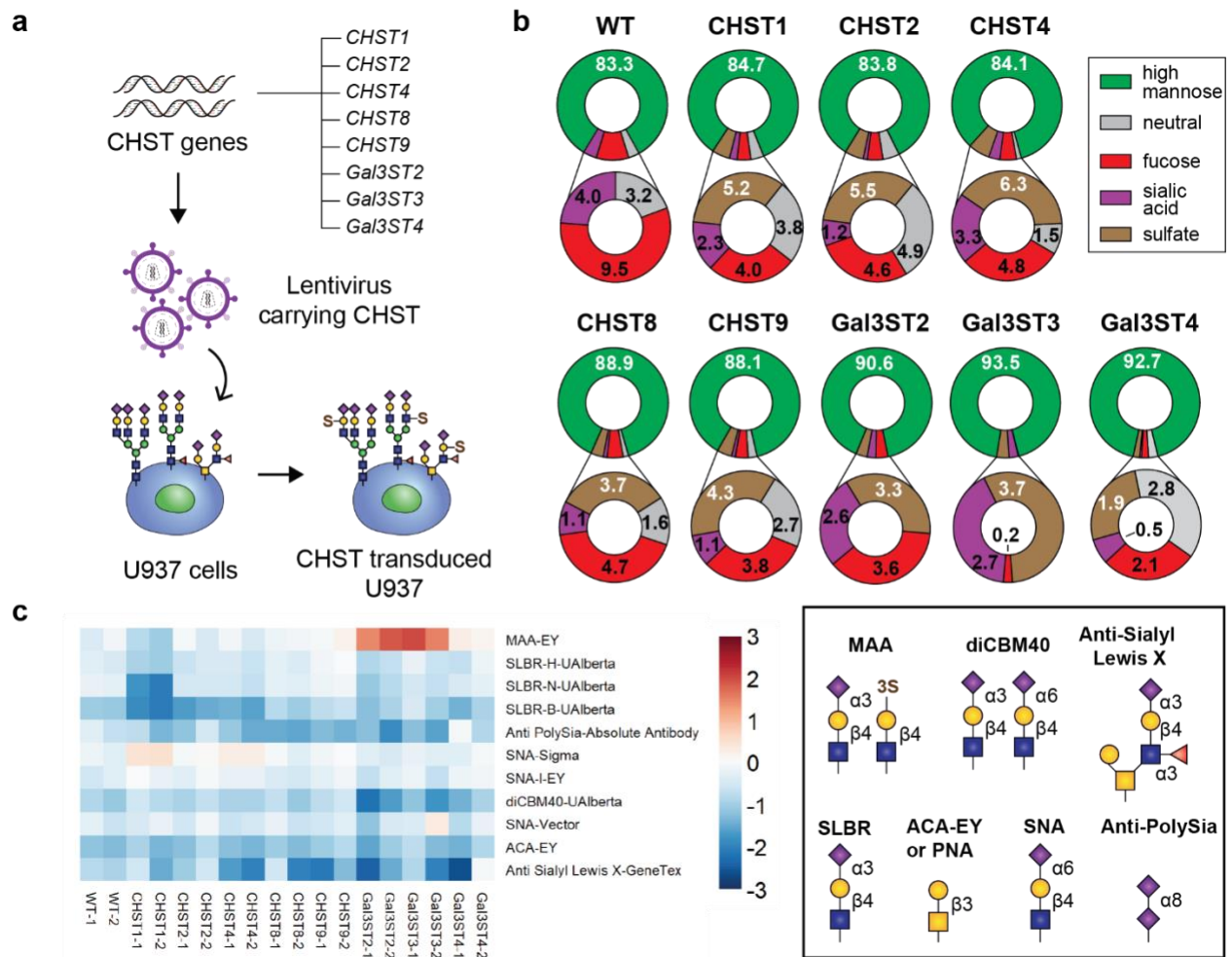


Figure 3.1 Characterization of CHST-overexpressing U937 cells. (a) Eight genes encoding CHSTs were transduced into U937 cells with lentivirus. (b) Mass spectrometry analysis of PNGase F released N-glycans from the CHST-transduced cells. The percentage of each glycan class relative to the total is noted. (c) Select data from lectin microarray analysis on lysates from the CHST-transduced U937 cells. Lysates from each cell line were run in duplicate against a pooled reference. Median-normalized log₂ ratio (Sample (S)/Reference (R)) is shown. Glycan structures recognized by each lectin are shown to the right.

cell lysates (**Figure 3.1c**).^{45,130} SLBR-H and -N, which are specific for α 2-3 linked sialosides on *O*-linked glycans,¹³¹ and diCBM40, which recognizes both α 2-3 and α 2-6 linked sialosides,¹³² showed a decrease in signal towards the *Gal3ST2*- and *Gal3ST4*-overexpressing cell lysates. Likewise, a decrease in SLe^x structures were observed in the *CHST2*- and *Gal3ST4*-overexpressing cell lysates. The enhanced binding of MAA and decreased levels of α 2-3 linked

sialosides in *Gal3ST2*- and *Gal3ST3*-overexpressing cells strongly suggest a direct competition between sialylation and sulfation at the 3-position of terminal Gal.

3.3.2 *CHST*-dependent binding of Siglecs to U937 cells

Recently, a new panel of soluble human Siglec-Fc proteins for studying Siglec ligands on cells and tissues were produced with features that maximize recognition of their cellular glycan ligands in a sialic acid-dependent manner (**Figure 3.2a**).¹⁰⁶ Here, these Siglec-Fc probes were applied to study the effects of increased sulfation on human (**Figure 3.2b-f**) and mouse (**Figure 3.3a-d**) Siglec ligands. Significantly enhanced binding of CD33, Siglec-5, -7, -8, -14, and -15 were observed towards the *CHST1*-overexpressing U937 cells. Results for Siglec-8 were anticipated based on previous work,^{39,47,91} while results with CD33, Siglec-7, and Siglec-15 are consistent with a recent study that overexpressed *CHST1* in HEK293 cells.¹¹⁹ However, the ability of *CHST1* overexpression to upregulate ligands for Siglec-5 and -14, which share an identical amino acid sequence in their first three domains¹³³ (herein described as Siglec-5/14), have not been described previously. A modest enhancement in binding of CD22 to the *CHST1*-overexpressing U937 cells was also not anticipated, but this may be due to enhanced expression of α 2-6 linked sialosides, as evidenced by increased staining of SNA, CD22 and Siglec-9 showed greatly enhanced binding to the *CHST2*- and *CHST4*-overexpressing U937 cells, which were anticipated from previous studies.^{26,47,51} Close inspection of the data indicates that binding of CD33, Siglec-5/14, Siglec-7, Siglec-8, and Siglec-15 also showed modestly enhanced binding to the *CHST2*- and *CHST4*-overexpressing U937 cells. It was also observed several interesting CHST-dependent decreases in binding. These include: Siglec-1 to *CHST1*-overexpressing U937 cells; Siglec-1, CD22, and CD33 to *Gal3ST3*-overexpressing U937 cells; and Siglec-8 to *Gal3ST2*- and *Gal3ST3*-overexpressing

U937 cells. These latter results provide an intriguing possibility that sulfation can not only enhance Siglec binding, but also downregulate Siglec ligands.

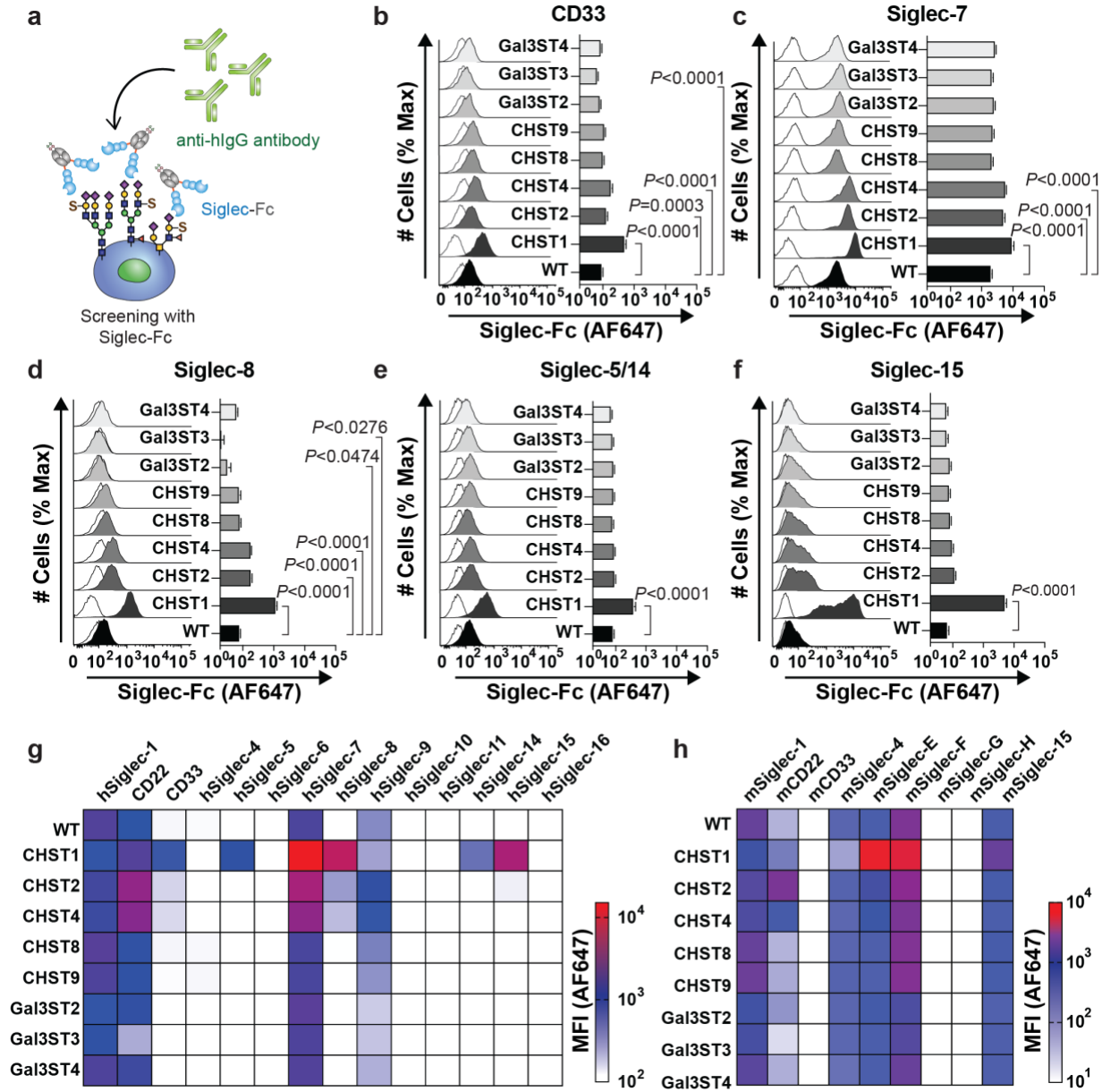


Figure 3.2 Probing CHST-overexpressing U937 cells for Siglec ligands. (a) CHST-overexpressing cells were screened with Siglec-Fc proteins complexed with anti-human IgG-AF647. (b-f) Flow cytometry results (*left panels*) and summary of MFIs (*right panel*) for each Siglec-Fc screened against each CHST-overexpressing U937 cells. (g,h) Heatmap for both (g) human and (h) mouse Siglec-Fc binding to CHST-overexpressing U937 cell. Data plots in b-f are presented as mean \pm SD. Statistical significance was calculated using a one-way ANOVA with Tukey's multiple comparisons test.

Our previous study developing the new Siglec-Fc scaffold only focused on human Siglecs.¹⁰⁶ Given the striking evolutionary differences between Siglecs in mouse and human,¹² and widespread use of mouse models to study concepts related to human health and disease,¹³⁴ a complementary set of mouse Siglec-Fc soluble proteins for this study were made and used to examine binding to the CHST overexpressing U937 cells (**Figure 3.2h, Figure 3.3**). it was found

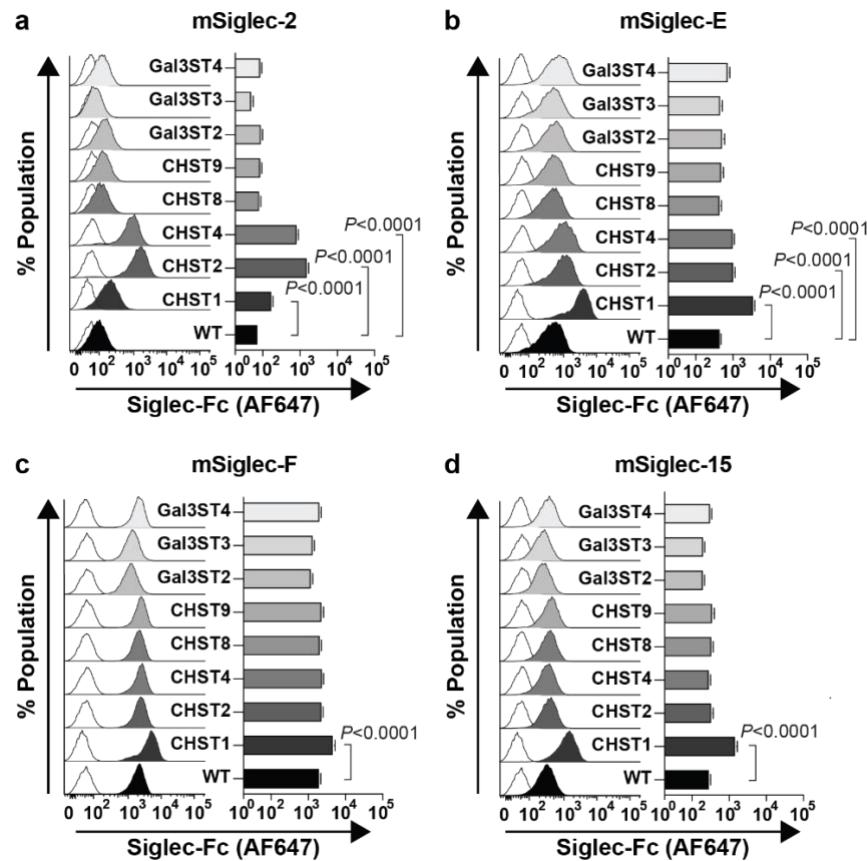


Figure 3.3 mSiglec binding to the CHST transduced cells. (a) mCD22, (b) mSiglec-E, (c) mSiglec-F, and (d) mSiglec-15. Data plots are presented as mean \pm SD. Statistical significance was calculated using a one-way ANOVA with Tukey's multiple comparisons test.

that mouse Siglecs conserved between mouse and human (Siglec-1, CD22, and Siglec-15)⁴² showed similar patterns as their human counterparts. Specifically, mouse CD22 (mCD22) showed greatly enhanced binding to *CHST2* and *CHST4*-overexpressing U937 cells, while mSiglec-15

showed enhanced binding to *CHST1*-overexpressing U937 cells. The significant binding enhancement for mCD22 was anticipated based on a 9-fold increase in affinity observed previously for Neu5Aca2-3Gal β 1-4(6-*O*-sulfo)GlcNAc relative to its non-sulfated counterpart.⁴³ For the more divergent CD33-related Siglecs, less predictable patterns were observed. Binding of mouse CD33, which has divergent properties compared to its human counterpart,⁹² was not enhanced in the *CHST1*-overexpressing U937 cells. These results have important implications for studying the role of CD33 on brain microglia and its relationship to Alzheimer's disease susceptibility and the use of mouse models to study the relationship of CD33 to neurodegeneration.^{92,135} Siglec-E, which is described as the mouse ortholog of Siglec-9,²⁶ showed a similar binding pattern to Siglec-7, where binding to cells was enhanced by either *CHST1* or *CHST2/4* overexpression, which is consistent with findings from a previous study on Siglec-E.¹³⁶ Lastly, Siglec-F binding was increased to *CHST1*-overexpressing U937 cells, which was anticipated based on previous work.²⁶

3.3.3 Sialic acid-dependent binding and generality of CHST-dependent Siglec binding

To examine whether the enhanced binding of Siglecs to CHST-overexpressing cells was sialic acid dependent, a complementary set of the CHST-overexpressing U937 cells were generated on a *CMAS*^{-/-} background, which was generated previously.¹⁰⁶ In all cases, no binding was observed compared to WT U937 cells (**Figure 3.4a**). Moreover, *Arthrobacter Ureafaciens* neuraminidase was used to hydrolyze sialic acid residues, which likewise produced a near complete disappearance of ligands for CD33, Siglec-8, and Siglec-15 in *CHST1*-overexpressing U937 cells as well as CD22 and Siglec-9 in *CHST2*-overexpressing U937 cells (**Figure 3.4b,c**). These results establish that enhanced binding to sulfated cellular glycans still requires sialic acid. To investigate to what extent our findings of enhanced Siglec binding to CHST-overexpressing U937 cells extend to other cells, *CHST1* and *CHST2* were transduced into four additional cell lines:

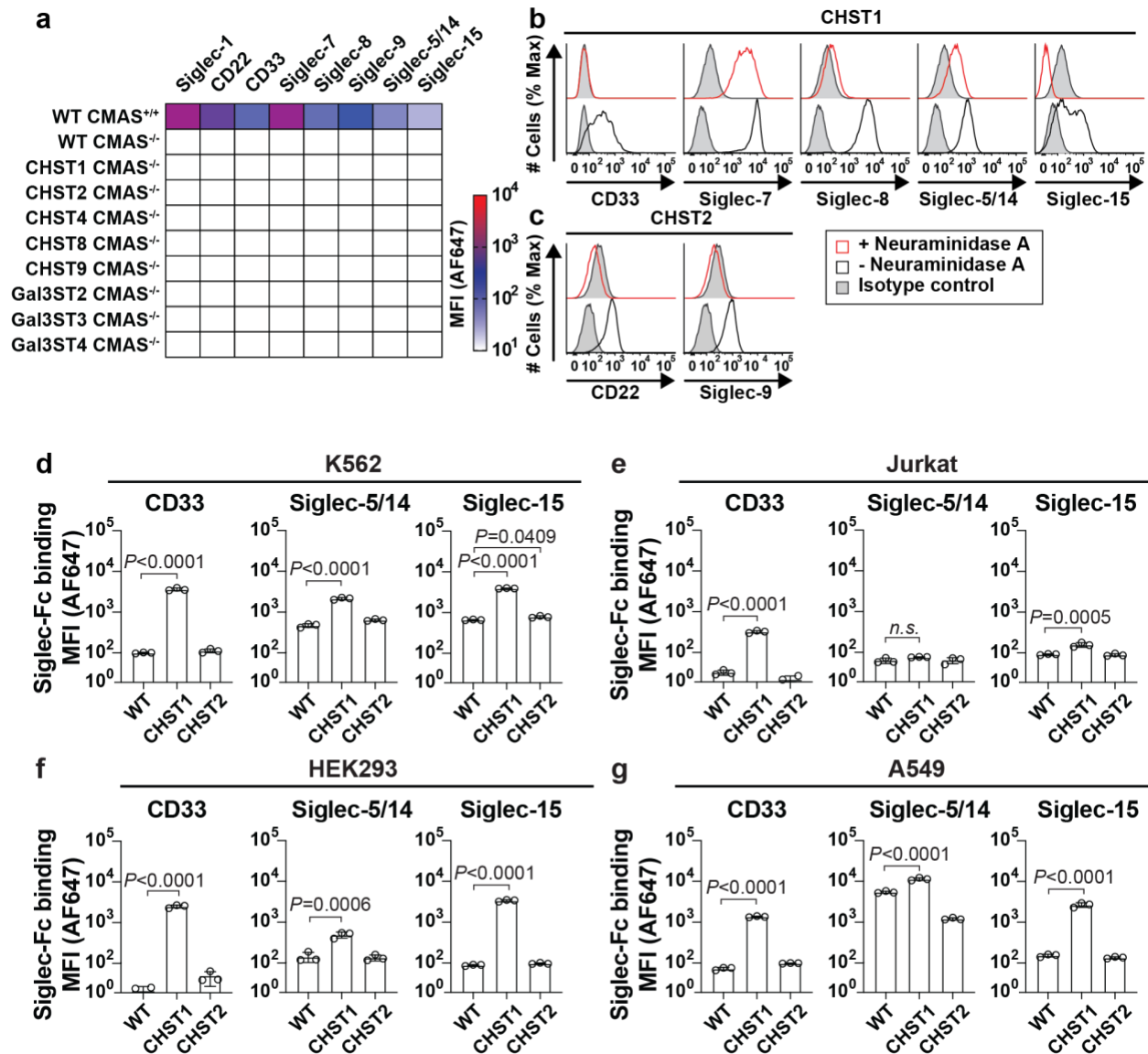


Figure 3.4 Sialic acid-dependent Siglec binding to CHST overexpressing cells. (a) Siglec-Fc binding to the CHST overexpressing *CMAS*^{-/-} U937 cells. (b,c) Probing the CHST1 and CHST2 overexpressing U937 cells with Siglec-Fc after neuraminidase digestion. (d-e) the four additional cell lines (K562, Jurkat, HEK293, and A549) transduced with either CHST1 or CHST2 were probed with CD33, Siglec-5/14, and Siglec-15. Data plots in d,e,f, and g are presented as mean \pm SD. Statistical significance was calculated based on a one-way ANOVA with Tukey's multiple comparisons test.

K562, Jurkat, HEK293, and A549 cells. Similar trends were observed for many of the key CHST-upregulated Siglec binding interactions in U937 cells (**Figure 3.4d-g**). Interestingly, enhanced binding of Siglec-5/14 to the *CHST1*-overexpressing HEK293 cells was not observed

previously,¹¹⁹ but was readily observed in our studies, which may be attributed to non-functional commercial Siglec-5/14-Fc. In contrast with the other cells, CD33, Siglec-5/14, and Siglec-15 showed either minimal or no increase of binding to the *CHST1*-overexpressing Jurkat cells. Moreover, the modest binding increase of CD33 in the *CHST2*-overexpressing U937 was not observed in the other cells. These results suggest that the correct ensemble and/or the unique repertoire of proteins expressed in each cell type impact the ability of CHST1 and CHST2 to create Siglec ligands, but overall point to carbohydrate sulfation as being an important mechanism for upregulating Siglec ligands.

3.3.4 Pharmacological perturbation of key cellular glycans

Kifunensine, benzyl- α -GalNAc, and Genz-123346 are inhibitors for the three major classes of glycans that contain sialic acid (**Figure 3.5a**). These three inhibitors were used in the CHST-transduced U937 cells to examine what sub-class of glycans are elaborated with sulfation and enhance binding of Siglecs (**Figure 3.5b-f**). As anticipated based on previous observations for CD22 preferring sialoside presented on *N*-glycans,¹³⁷ CD22-Fc binding was completely abolished in the kifunensine-treated *CHST1*- and *CHST2*-overexpressing U937 cells (**Figure 3.5b**). CD33, Siglec-5/14, Siglec-15 were affected by inhibition of either *N*- or *O*-glycosylation (**Figure 3.5c,e f**), suggesting that their sulfated ligands are presented on both classes of glycans. Benzyl- α -GalNAc significantly reduced Siglec-7 binding to the *CHST1*- and *CHST2*-overexpressing U937 cells (**Figure 3.5d**), which is in line with a recent study that found Siglec-7 ligands are sensitive to treatment with the mucin-selective protease, StcE.¹³⁸ Interestingly, Siglec-8 binding was not perturbed by either kifunensine or benzyl- α -GalNAc treatment. Pharmacological blockade of glycolipid biosynthesis did not perturb Siglec ligands in any instance.

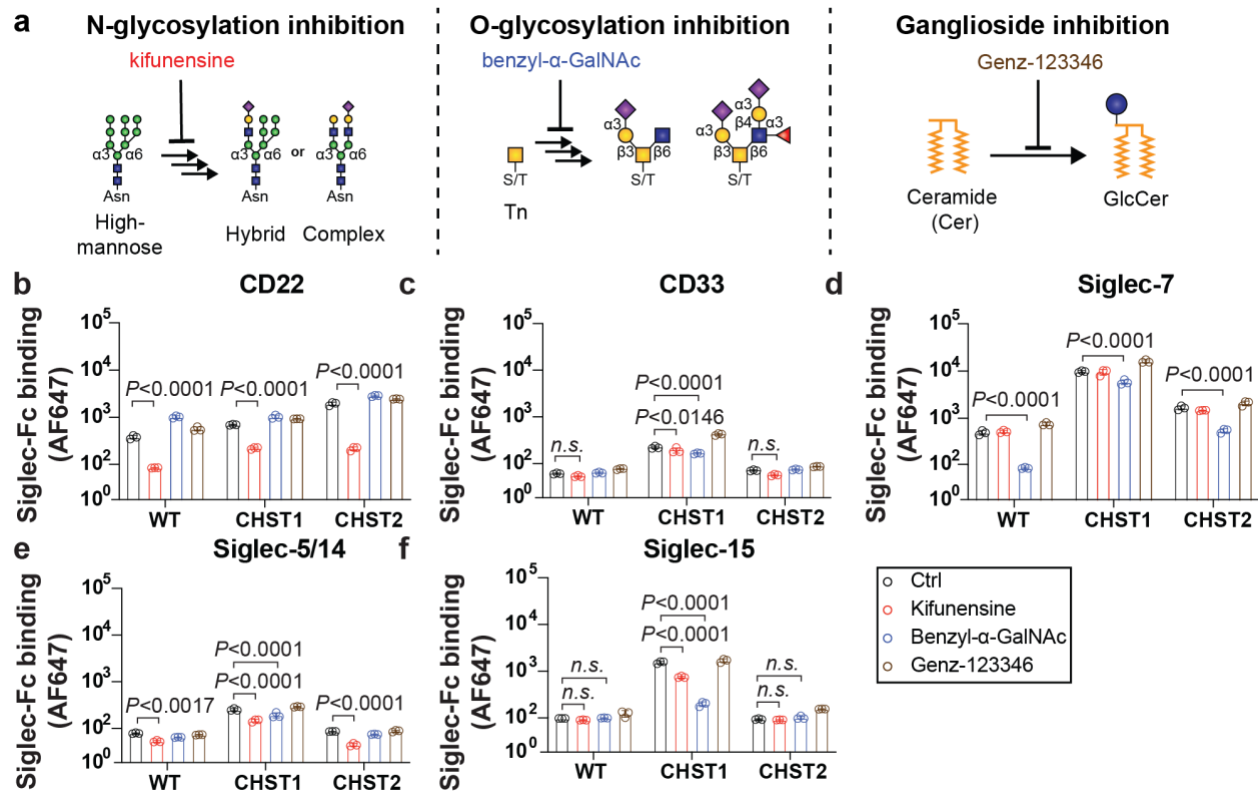


Figure 3.5. Inhibitors of cellular glycosylation reveal a cell type-specific pattern of Siglec ligands. (a) Inhibitors for *N*-glycosylation, mucin-type *O*-glycosylation, and ganglioside biosynthesis. (b-h) binding of CD22, CD33, Siglec-7, 8, 9, 14, and 15 to the WT, *CHST1*-, and *CHST2*-overexpressing U937 cells with inhibition of each glycosylation pathway. Data are presented as mean \pm SD. Statistical significance was calculated using a two-way ANOVA with Tukey's multiple comparisons test.

To validate whether the glycosylation inhibition patterns are conserved in other cells, K562, Jurkat, HEK293, and A549 cells transduced with empty vector (EV), *CHST1*, or *CHST2* for ligands of CD33, Siglec-5/14, and 15 were tested. (Figure 3.6a-d) Cell-specific patterns emerged; benzyl- α -GalNAc was dominant in decreasing the binding of all three Siglecs in all types of K562 and A549 cells (Figure 3.6a,d), while kifunensine effectively decreased binding of all Siglecs in the *CHST1*-overexpressing Jurkat cells (Figure 3.6b). This later observation may be partially explained by the fact that mucin-type *O*-glycans in Jurkat cells are highly truncated, being composed primarily of the T antigen (Gal α 1-3GalNAc).¹³⁹ Kifunensine and benzyl- α -GalNAc had

only minor effects in blocking Siglec binding in HEK293 cells (**Figure 3.6c**). Previous studies overexpressing CHSTs in HEK293 cells used genetic knockouts of subclasses of glycans to make conclusions about classes of glycan present sulfated glycans as Siglec ligands,¹¹⁹ which differ to some degree with the results presented here using these pharmacological approaches. Overall, these results strongly suggest that binding enhancement of Siglecs in *CHST1* or *CHST2*-overexpressing cells is cell type-specific and may be presented on different sub-classes of glycans on different cell types.

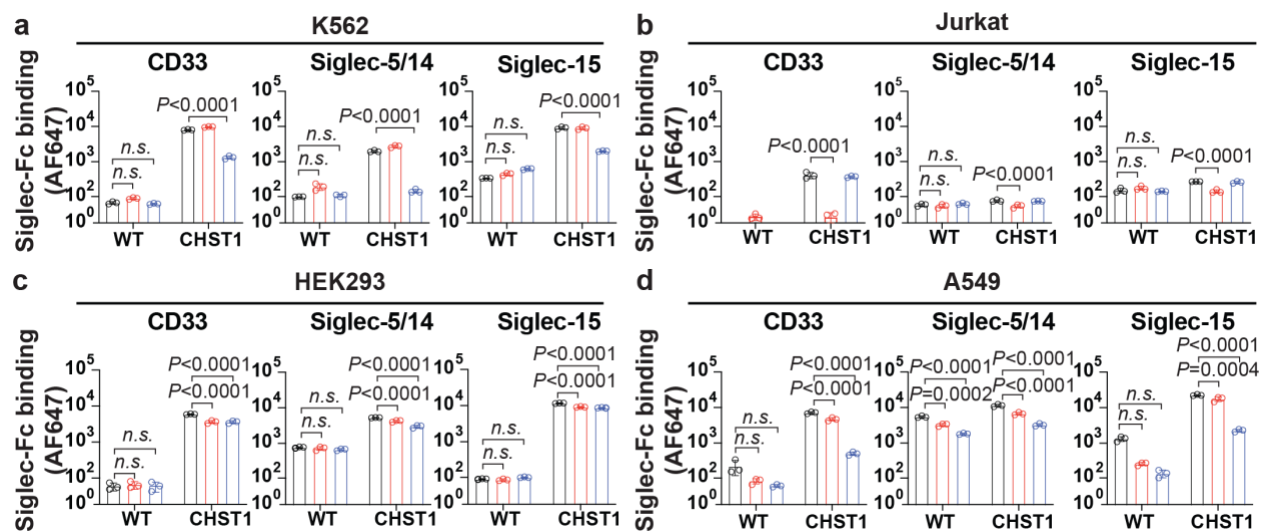


Figure 3.6 The effect of Kifunensine or benzyl- α -GalNAc on CD33, Siglec5/14, and Siglec-7 binding to four different cell lines expressing CHST1. (a-d) K562, Jurkat, HEK293, and A549 overexpressing *CHST1* or EV were probed for ligands of CD33, Siglec-5/14 and Siglec-15 in cells treated with kifunensine or benzyl- α -GalNAc. Data are presented as mean \pm SD. Statistical significance was calculated using a two-way ANOVA with Tukey's multiple comparisons test.

3.3.5 Quantifying Siglec binding to sulfated glycans

Previously, ESI-MS assay was validated for quantifying the affinity of Siglec-ligand interactions.^{45,106} Advantages of this approach are that it is label free, fast, has low sample consumption, can quantify very weak interactions, and the results are consistent with biochemical approaches such as isothermal titration calorimetry.¹⁰² The enhanced binding of human CD33 to

several of the CHST-overexpressing cell lines and the potential relevance of these findings to glycan ligands of CD33 in Alzheimer's disease susceptibility motivated us to apply this MS-based approach to quantify the affinity enhancement afforded by sulfation.^{92,135} From the previous study, it was unexpected that soluble CD33, expressed as an Fc chimera, was *O*-glycosylated.¹⁰⁶ Heterogeneity associated with approximately half of the protein containing an *O*-glycan makes analyzing ligand binding more challenging because of spectral overlap of peaks corresponding to the *O*-glycosylated glycoforms and peaks corresponding to protein without the *O*-glycan bound to ligands. Attempting to remove the *O*-glycans enzymatically, with *O*-glycosidase, was not successful even after first removing the sialic acid residues with a neuraminidase. Accordingly, sites of *O*-glycosylation were mapped by MS and evidence of *O*-glycosylation was observed at Thr239 and Thr260 (**Figure 3.7a,b**). The latter site is located within the linker between CD33 and the Fc encoded by the AgeI restriction site used for the molecular cloning. Accordingly, a double mutant (T239A/T260A) of CD33-Fc was prepared in LEC1 CHO cells. Following removal of the Fc and high mannose *N*-glycans by TEV protease and Endo H, respectively. SDS-PAGE showed a single band for T239A/T260A CD33 in contrast to WT CD33, which runs as doublet by SDS-PAGE (**Figure 3.8a,b**). The mass spectrum of the T239A/T260A CD33 confirmed that it lacks *O*-glycosylation.

Four differentially sulfated compounds were chemo-enzymatically prepared based on a Neu5Ac α 2-3LacNAc-propylamine scaffold. Specifically, versions of this scaffold without sulfate **7** (3'SLN), with 6-*O*-sulfate at the Gal **11** (6-S'-3'SLN), with 6-*O*-sulfate at the GlcNAc **14** (6-S-3'SLN) was prepared, as well as a disulfated derivative with sulfate at both positions **18** (6,6'-S,S-3'SLN). The T239A/T260A CD33 protein was titrated with these ligands to determine the

dissociation constant (K_d) by ESI-MS (**Figure 3.8c**). The affinity measured against **7** was surprisingly weak, with a K_d value of 25 ± 3 mM (**Figure 3.8d**). Such a weak affinity was not

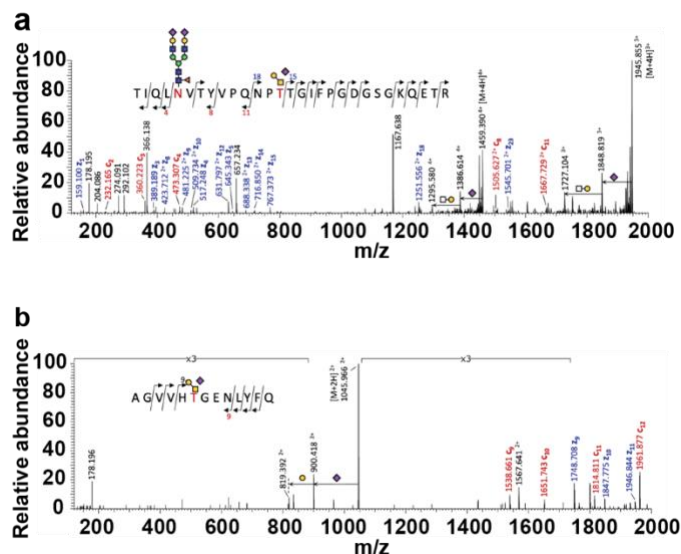


Figure 3.7 Mapping out the novel O-glycosylation sites on the CD33 fragment. (a,b) EThcD MS² analyses of glycopeptides derived from CD33 unambiguously identified O-glycosylation at (a) T239 and (b) T260. In addition to HexNAc₁Hex₁NeuAc₁ shown here, two other O-glycans were detected for T239, corresponding to HexNAc₁Hex₁, and HexNAc₁Hex₁NeuAc₂.

anticipated given that our previous measurements of CD33 against 3'SLN, lacking the propylamine aglycone, yielded a K_d value of 2.7 ± 0.1 mM.¹⁰⁶ To test whether a lack of the O-glycosylation on CD33 decreased affinity, the WT CD33 fragment was tested and it shows a similarly weak affinity for 3'SLN-propylamine (**Figure 3.9a**). Moreover, the T239A/T260A CD33 fragment yielded a similar K_d for 3'SLN without the propylamine linker, in line with the WT CD33 fragment (**Figure 3.9b**). Thus, the propylamine aglycone has an unexpected deleterious effect on the binding affinity of CD33. The crystal structure of CD33 presents no obvious rationale for this loss of affinity.¹¹⁶ This surprising finding has important implications for the future design of high affinity glycans ligands for targeting CD33. Aware of this linker effect, the T239A/T260A CD33 against **11**, **14**, and **18** (**Figure 3.8d-g**) was tested. It was shown that the 6-O-sulfate

modification at the Gal enhances the affinity by 10-fold ($K_d = 2.5 \pm 0.1$ mM), (**Figure 3.8e**) while the 6-*O*-sulfate modification at the GlcNAc enhances the affinity by 3-fold ($K_d = 8.5 \pm 0.5$ mM) (**Figure 3.8f**). The disulfated compound **18** displayed the highest affinity ($K_d = 0.70 \pm 0.1$ mM) for CD33, which is a 28-fold affinity enhancement relative to the non-sulfated **7**. (**Figure 3.8d,g**).

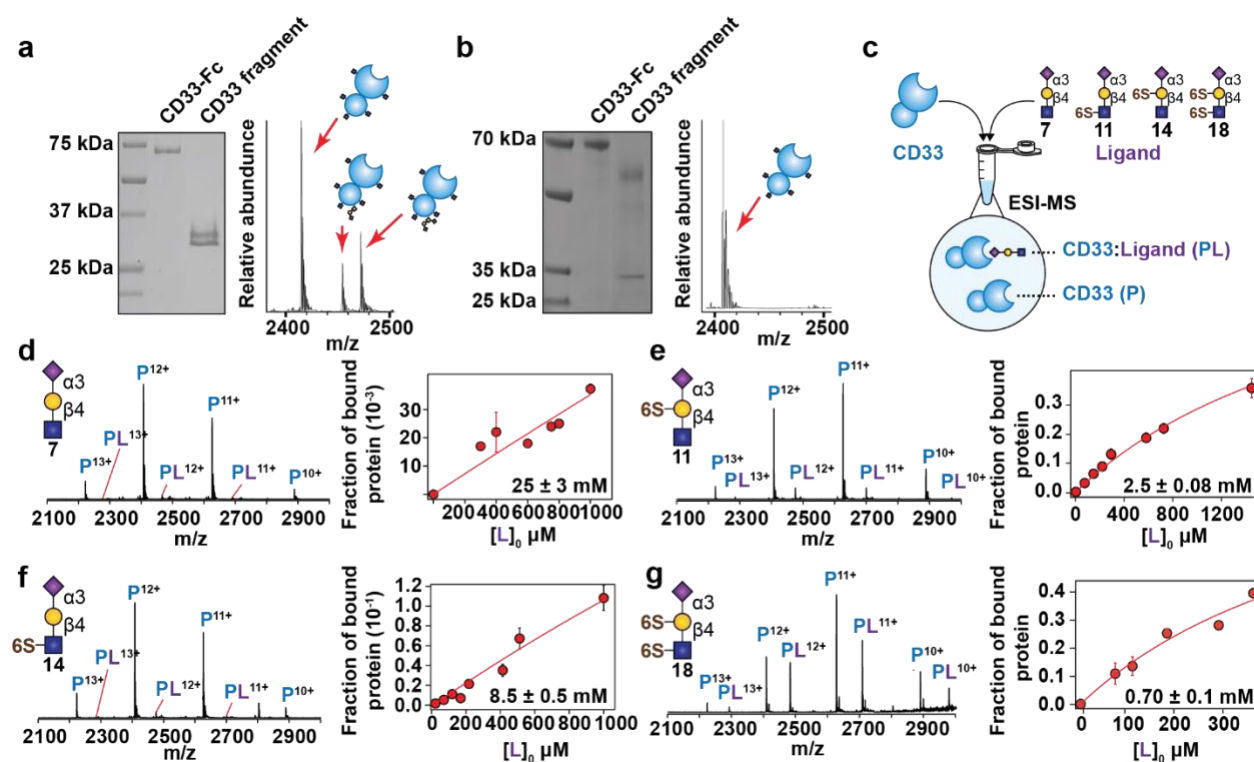


Figure 3.8 Generation of homogeneous CD33 fragment and determination of dissociation constant (K_d). (a-b) SDS-PAGE of the (a) WT CD33 fragment and (b) T239A/T260A CD33 fragment. (c) Scheme for direct binding measurement by ESI-MS to determine the affinity of the CD33 against compounds **7**, **11**, **14**, and **18**. (d-g) Results of ESI-MS binding studies of compounds **7**, **11**, **14**, and **18** with the T239A/T260A CD33 fragment. For each compound a representative spectrum is shown at a single concentration (**7**, 400 μM ; **11**, 280 μM ; **14**, 400 μM ; and **18**, 100 μM) in the left panel and a summary of the titration is shown in the right panel.

The only other study to perform quantitative binding measurements on a similar series of sulfated compounds was an investigation of Siglec-8 with SLe^x structures.⁹¹ This previous study found that a 6-*O*-sulfation at the Gal dominated the affinity gains, with a disulfated ligand showing only modestly increased affinity 1.6-fold relative to the best monosulfated ligand. Our binding

measurements with CD33 demonstrate a 3.6-fold increase in affinity for the disulfated ligand relative to the best monosulfated ligand. To our knowledge, our results with CD33 represents the largest enhancement in affinity provided by disulfation. Glycan microarray results, which represent semi-quantitative binding, suggest that additivity between the two sulfate modifications may also be at play for Siglec-7.⁴⁷

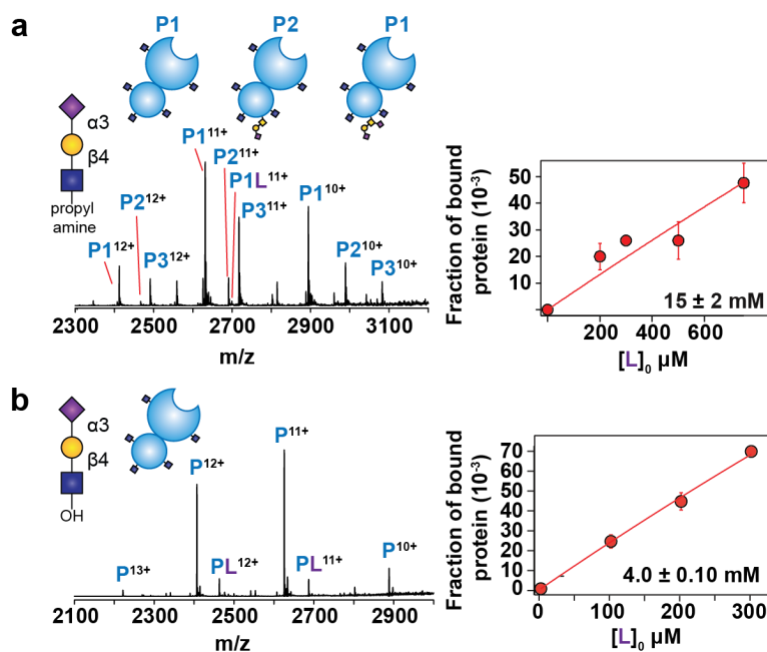


Figure 3.9 Comparison of the T239A/T260A CD33 fragment with the WT CD33 fragment. (a) K_d determination of the WT fragment with a disialyl T antigen on 3'SLN with a propylamine linker. (b) K_d determination of the T239A/T260A CD33 fragment on 3'SLN without any linker.

3.3.6 Overexpression of both *CHST1* and *CHST2* in U937 cells greatly enhances Siglec ligands

The quantitative binding measurements above for the two monosulfated trisaccharides **11** and **14** are in line with results for CD33-Fc binding to the *CHST1* and *CHST2* or *CHST4* overexpressing cells, respectively (**Figure 3.2b**). The additive effect in the disulfated species **18** suggests that dual expression of CHST1 and CHST2 will significantly enhance cellular ligands of

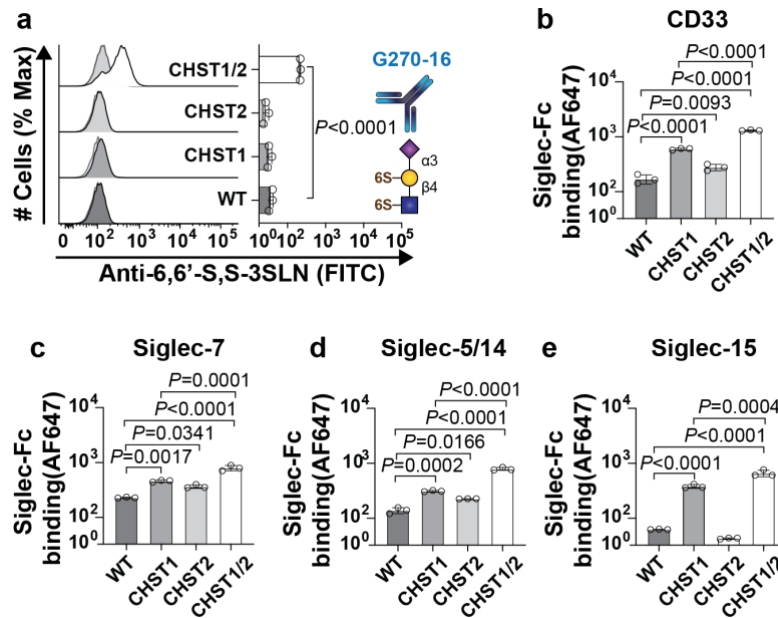


Figure 3.10. Evaluation of the level of disulfated sialosides by G270-16 and its effects on Siglec binding. (a) Staining of U937 cells transduced with both *CHST1* and *CHST2* with an antibody that detects the 6,6'-S,S-3'SLN epitope (clone G270-16). (b-e) Probing *CHST1/2*-overexpressing cells with (b) CD33, (c) Siglec-7, (d) Siglec-5/14, and (e) Siglec-15. Data are presented as mean \pm SD. Statistical significance was calculated using a one-way ANOVA with Tukey's multiple comparisons test.

CD33. To test this, *CHST1*-transduced U937 cells were further transduced with lentivirus carrying *CHST2* or an EV control. To accomplish this, our lentiviral vector was re-engineered to express an orthogonal fluorescent protein (EGFP) to mAmertine, enabling sorting of doubly transduced cells. The U937 cells overexpressing both *CHST1* and *CHST2* (herein described as *CHST1/2*) showed enhanced staining with an antibody that recognized a 6,6'-S,S-3'SLN (clone G270-16), providing evidence that this strategy had been successful (**Figure 3.10a**). Comparing the binding of CD33 to the *CHST1/2*-overexpressing U937 cells to the WT, *CHST1*-, and *CHST2*-overexpressing cells, enhanced binding to the former (**Figure 3.10b**) was observed. Similar results were also observed for Siglec-7, Siglec-5/14, and Siglec-15. (**Figure 3.10c-e**).

Staining of cells with CD33-Fc represents *trans* ligands. To examine *cis* ligands of CD33, binding studies with fluorescent liposomes displaying a high affinity and selective synthetically-modified ligand of CD33 was performed, described previously.¹¹³ (**Figure 3.11a**) The ability of CD33 on U937 cells to engage the liposomes is a measure for *cis* ligands, with strong *cis* ligands expecting to mask and hinder CD33-liposome interactions.¹⁴⁰ CD33 levels on U937 cells were

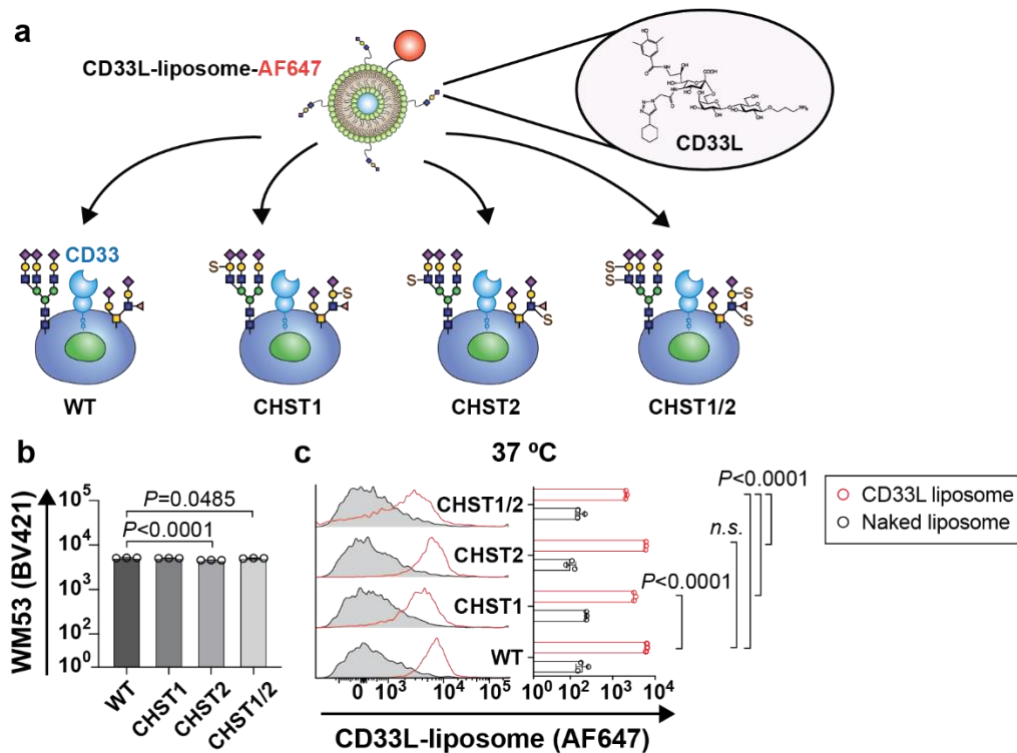


Figure 3.11 CD33L-liposome binding assay on the *CHST1*, *CHST2*, and *CHST1/2* overexpressing U937 cells. (a) Scheme for the masking assay to assess *cis* ligands of CD33. Liposome binding is measured using flow cytometry. The structure of the CD33 high affinity ligand is shown in the inset.¹⁴⁰ (b) Anti-CD33 (clone WM53) staining on CHST-transduced U937 cells to evaluate CD33 expression. (c) Results of the liposome assay performed at 37 °C. Data plots are presented as mean \pm SD. Statistical significance was calculated using one-way ANOVA with Tukey's multiple comparisons test.

first verified to be unchanged in the WT, *CHST1*-, *CHST2*-, and *CHST1/2*-overexpressing U937 cells (**Figure 3.11b**). Examining liposome binding by flow cytometry, it was shown that *CHST1*-overexpressing U937 cells showed significantly reduced engagement by CD33 ligand liposomes.

CHST2-overexpressing U937 showed a small, but not statistically significant, decrease in liposome binding relative to the WT U937 cells (**Figure 3.11c**). *CHST1/2*-overexpressing U937 showed the least amount of binding to the liposomes, demonstrating that CD33 is strongly masked on these cells. These results for *cis* ligands on U937 cells overexpressing both *CHST1* and *CHST2* provide further support that a disulfated ligand is a strong ligand for CD33.

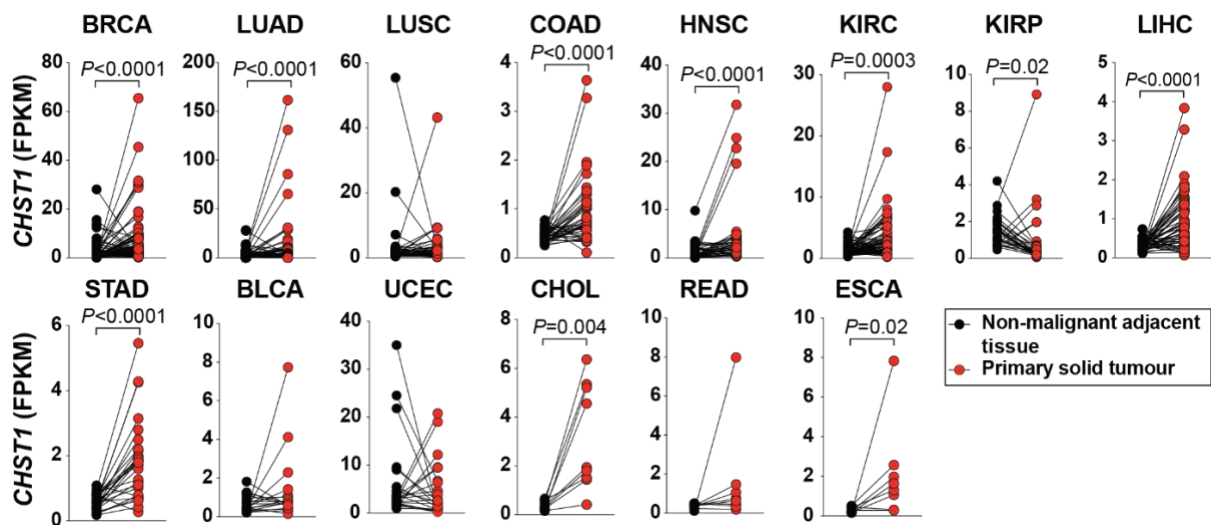


Figure 3.12 Upregulated *CHST1* expression in solid tumours correlates with poor patient outcome in several cancer types. Transcript profiling of *CHST1* in patient-matched non-malignant adjacent and cancerous tissues in 14 cancer types from The Cancer Genome Atlas (TCGA) Consortium. Normalized *CHST1* expression (FPKM) data from cancer types with >5 matched samples were downloaded through the University of California Santa Cruz (UCSC) Xena tool (<https://xenabrowser.net>). TCGA cancer types included in the analysis are as follow: BRCA (breast cancer), LUAD (lung adenocarcinoma), LUSC (lung squamous cell carcinoma), COAD (colon adenocarcinoma), HNSC (head and neck squamous cell carcinoma), KIRC (kidney renal clear cell carcinoma), KIRP (kidney renal papillary cell carcinoma), LIHC (liver hepatocellular carcinoma), STAD (stomach adenocarcinoma), BLCA (bladder urothelial carcinoma), UCEC (uterine corpus endometrial carcinoma), CHOL (choleangiocarcinoma), READ (rectum adenocarcinoma), and ESCA (esophageal carcinoma). Statistical analysis was performed using Wilcoxon matched-pairs signed-rank test.

3.3.7 Upregulated expression of *CHST1* in cancer

Elevated levels of sialic acid on cancer cells have emerged as a key player in controlling anti-tumour immune response, by engaging with Siglecs on immune cells.⁷ Given the enhanced

binding of Siglecs to their sulfated glycan ligands generated by *CHST1* overexpression, it was hypothesized that cancer cells utilize carbohydrate sulfation to further increase Siglec engagement and potentially enhance immune-evasion. To evaluate expression levels of *CHST1* in cancer, transcript data of 14 tumour types from The Cancer Genome Atlas (TCGA) with more than five matched non-malignant and primary tumour samples were retrieved. Strikingly, 9 of 14 cancer

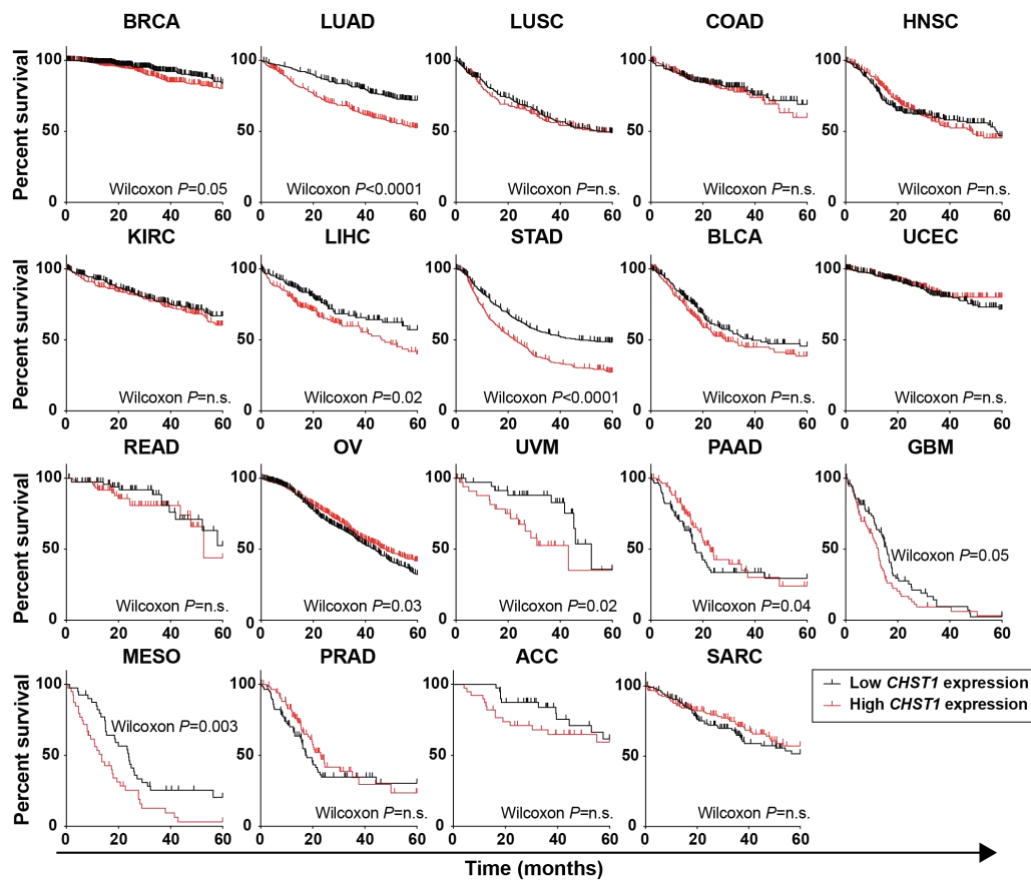


Figure 3.13 Kaplan-Meier 5-year overall survival curves of 19 cancer types based on *CHST1* expression. Patients were stratified based on median *CHST1* expression. Curated datasets for LUAD (lung adenocarcinoma), LUSC (lung squamous cell carcinoma), STAD (stomach carcinoma), and OV (ovarian serous cystadenocarcinoma) were retrieved from <https://kmplot.com>. TCGA datasets for BRCA(breast cancer), LIHC (liver hepatocellular carcinoma), KIRC (kidney renal clear cell carcinoma), BLCA (bladder urothelial carcinoma), HNSC (head and neck squamous cell carcinoma), PRAD (prostate adenocarcinoma), READ (rectum adenocarcinoma), COAD (colon adenocarcinoma), UVM (uveal melanoma), PAAD (pancreatic adenocarcinoma), GBM (glioblastoma multiforme), MESO (mesothelioma), UCEC (uterine corpus endometrial carcinoma), ACC (adrenocortical carcinoma), and SARC (sarcoma) were downloaded through the UCSC Xena tool (<https://xenabrowser.net>). Statistical analysis was performed using a Gehan-Breslow-Wilcoxon test.

types display significant upregulation of *CHST1* transcript levels compared to their non-malignant tissue counterparts (**Figure 3.12**). Next, it was assessed whether expression of *CHST1* can predict patient outcome in 19 different cancer types. publicly available data from TCGA and kmplot.com were used and cases were split into low and high expression levels based on median *CHST1* expression levels. Kaplan-Meier curves of 5-year overall survival revealed that *CHST1* expression is a predictor of poor patient outcome in 7 of 19 cancer types. These include lung adenocarcinoma (LUAD), stomach adenocarcinoma (STAD), breast invasive carcinoma (BRCA), liver hepatocellular carcinoma (LIHC), uveal melanoma (UVM), glioblastoma multiforme (GBM), and mesothelioma (MESO). (**Figure 3.13**) In two cancer types, ovarian serous cystadenocarcinoma (OV) and pancreatic adenocarcinoma (PAAD), *CHST1* expression correlated with favourable patient outcomes. Collectively, these results demonstrate that *CHST1* is upregulated in the majority of cancer types and can correlate with poorer patient outcomes in specific types of cancer. Given the numerous mechanisms of immune evasion used by tumors,^{141,142} which can vary based on cancer type, it is unsurprising that these effects are not uniform across all cancers.

3.3.8 Disrupting glycan sulfation levels on cancer cells downregulates Siglec ligands

Given the results presented above demonstrating that *CHST1* is upregulated in numerous cancer types and correlates with poorer survival rates in several cancers, it was wondered if upregulated *CHST1* expression gives rise to enhanced Siglec ligands on cancer cells. Sodium chlorate (NaClO₃) is an *in vitro* inhibitor of PAPS biosynthesis, and has been previously used to block the installation of sulfation on carbohydrates by CHSTs (**Figure 3.14.a**).^{124,143} To establish whether the increased binding of Siglecs to the CHST-transduced cells can be abolished by NaClO₃, sodium chlorate was treated on the *CHST1*-overexpressing U937 cells from 0 mM to 50 mM for two days to assess the binding of CD33-Fc and it was found a clear dose-dependent

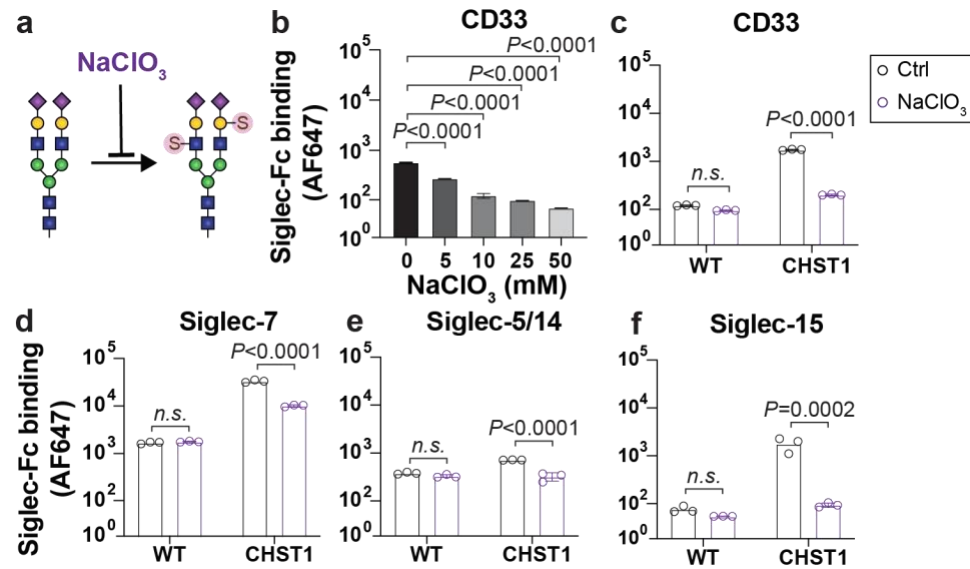


Figure 3.14 Inhibiting carbohydrate sulfation on cancer cells decreases Siglec ligands. (a) Illustration of NaClO₃ for inhibiting cellular carbohydrate sulfation (b) Titration of NaClO₃ in CD33 binding to the *CHST1*-overexpressing U937 cells. (c-f) The effect of NaClO₃ on the binding (c) CD33, (d) Siglec-7, (e) Siglec-5/14, and (f) Siglec-15 to the EV, *CHST1*-, and *CHST2*-overexpressing U937 cells. Data plots in b-f are presented as mean \pm SD. Statistical significance was calculated using a two-way ANOVA with Sidak's multiple comparisons test (b-f).

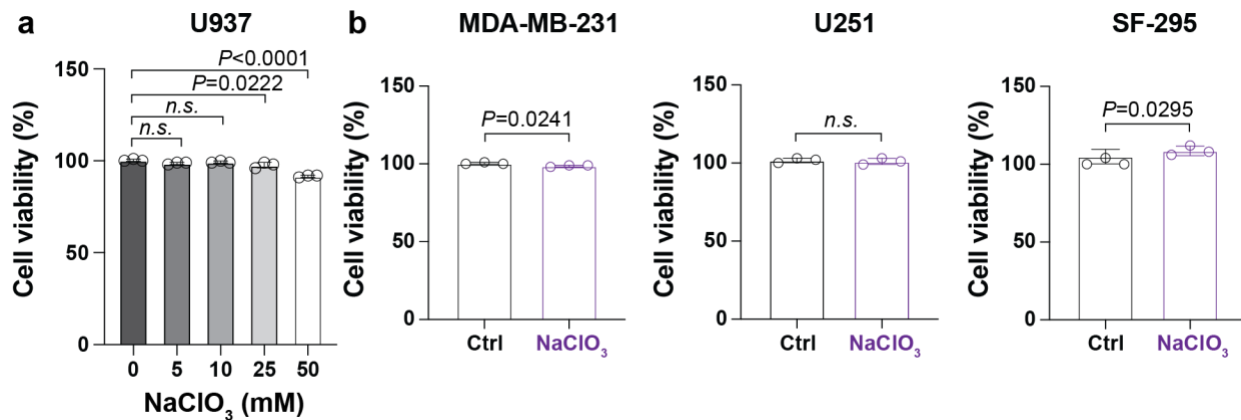


Figure 3.15 Assessing cell viability following NaClO₃ treatment.

(a) Titration of NaClO₃ to assess the effect of NaClO₃ on cell viability in U937. (b) Cell viability assay under 25 mM of NaClO₃ in MDA-MB-231, U251, and SF-295. Data plots are presented as mean \pm SD. Statistical significance was calculated using a one-way ANOVA with an unpaired student's t-test.

decrease in CD33-Fc binding (**Figure 3.14b**). Although 50 mM of NaClO₃ showed the largest decrease, 25 mM was chosen for subsequent experiments as there was minimal effect on cell viability at this concentration (**Figure 3.15**). It was additionally observed that the enhanced binding of CD33, Siglec-7, -5/14, and -15 on the *CHST1*-overexpressing cells was abrogated by 25 mM NaClO₃ treatment, providing further evidence for the critically importance role of sulfation in Siglec binding (**Figure 3.14c-f**). NaClO₃ treatment did not perturb the binding of these Siglecs to WT U937 cells, suggesting that these cells do not express sulfated glycans, which is consistent with our MS data (**Figure 3.1b**).

Table 3.2 RNA expression of CHSTs in the selected cancer cell lines.

Cell lines	CHST1	CHST2	CHST3	CHST4	CHST5	CHST6	CHST7	CHST8
MCF7	0.205	0	0.309	0.274	0.384	0.068	0.491	1.432
MDA-MB-231	0.067	1.832	0.354	0.531	0.046	0.016	2.326	0.06
SF-295	2.662	0.682	2.571	0	0.181	0.2	1.493	0.439
SNB-75	1.847	2.929	3.161	0.081	0.228	0.924	1.025	0.653
U251	2.451	1.828	3.629	0	0.105	0.257	2.135	0.106
K-562	0.061	2.617	0.064	0	0	0	0.143	0.094
A549	0.11	0.054	1.367	0.147	0.158	0.088	1.124	0.366

Next, the effect of NaClO₃ on a variety of cancer cell lines was evaluated to determine if Siglec binding was abrogated. Based on transcript levels of each *CHST* gene in the NCI-60 cancer cell library,⁵⁸ MDA-MB-231, U251, and SF-295 were selected as promising candidates because these cells express high levels of *CHST1* or *CHST2* (**Table 3.2**). Conversely, CCRF-CEM, MCF-7, A549, and MDA-MB-468 express low levels of *CHST1* and *CHST2*. Therefore, it was analyzed whether NaClO₃ sensitivity to Siglec-Fc binding correlates with expression of these *CHSTs* with

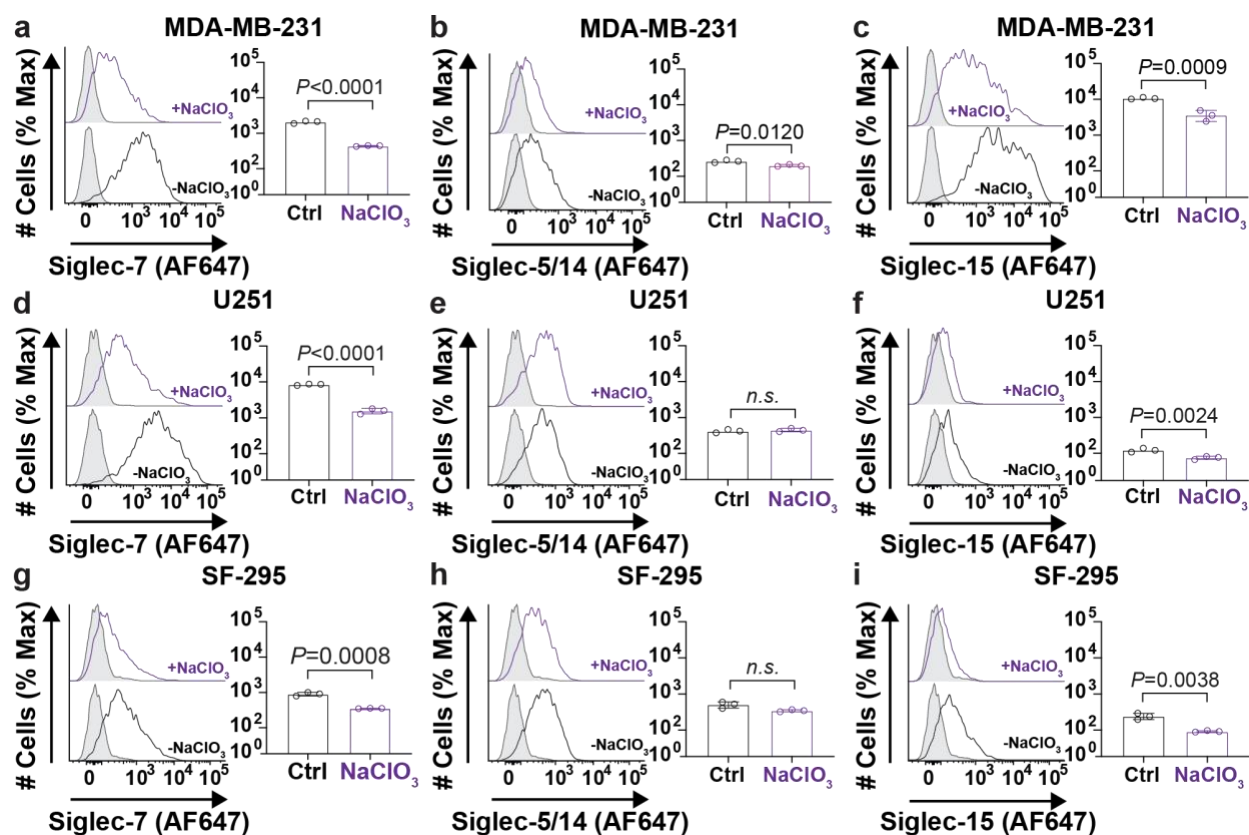


Figure 3.16. Inhibition of endogenous carbohydrate sulfation leads to the decrease of the binding of Siglec-7, 5/14, and 15 in a panel of cancer cell line. (a-i) The effects of inhibiting sulfation with NaClO₃ on Siglec-Fc binding were assessed in (a-c) MDA-MB-231, (d-f) U251, and (g-i) SF-295, with Siglec-7 (a,d,g), Siglec-5/14 (b,e,h) and Siglec-15 (c,f,i). Data plots in b-i are presented as mean \pm SD. Statistical significance was calculated using an unpaired t test (a-i).

these cancer cell lines, focusing on Siglec-7, -5/14, and -15 given that these Siglecs show robust binding to these cell lines. Viability of each cell was minimally affected by 25 mM NaClO₃ treatment (Figure 3.15). In the four high *CHST1*- or *CHST2*-expressing cancer cell lines, inhibiting endogenous sulfation led to decreases in Siglec binding in all cases except for Siglec-15 to MDA-MB-231 and Siglec-5/14 to U251 and SF-295 cells (Figure 3.16). These results reinforced our earlier data showing that the context in which carbohydrate sulfation is presented is important, with the correct ensemble of underlying glycans required to support sulfation as a mechanism for increasing Siglec ligands. In the low expressing *CHST1* and *CHST2* cells (A549, MDA-MB-468,

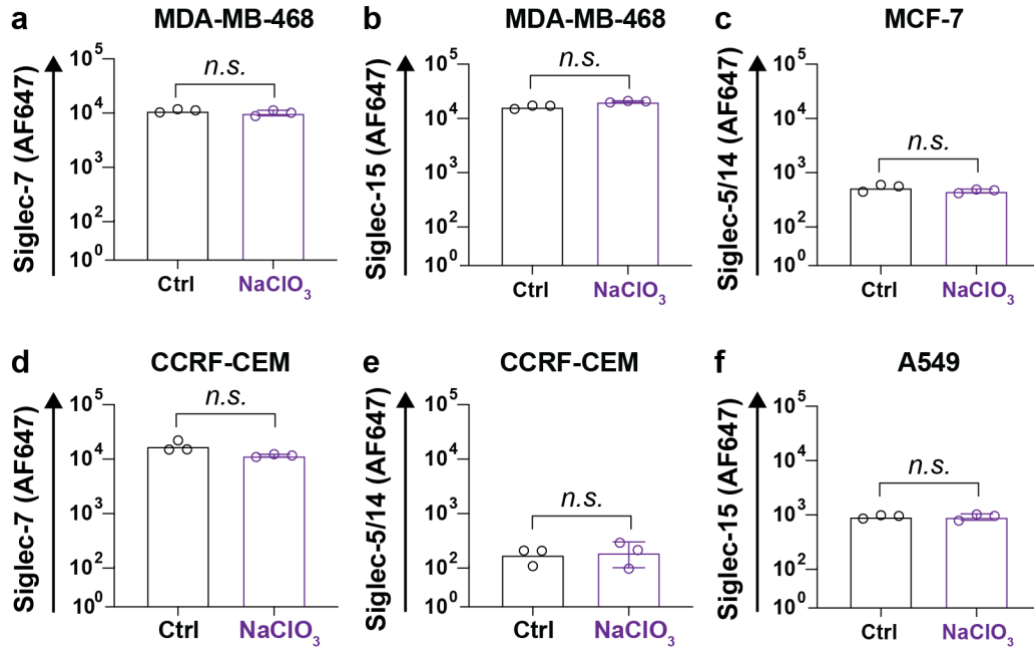


Figure 3.17. Assessing Siglec-Fc binding to cell lines expressing low levels of *CHST1* and *CHST2*. (a,b) MDA-MB-468, (c) MCF-7, (d,e) CCRF-CEM, and (f) A549 cells were probed by Siglec-Fc with and without 25 mM NaClO₃. Data plots are presented as mean \pm SD. Statistical significance was calculated using an unpaired two-tailed Student's *t*-test.

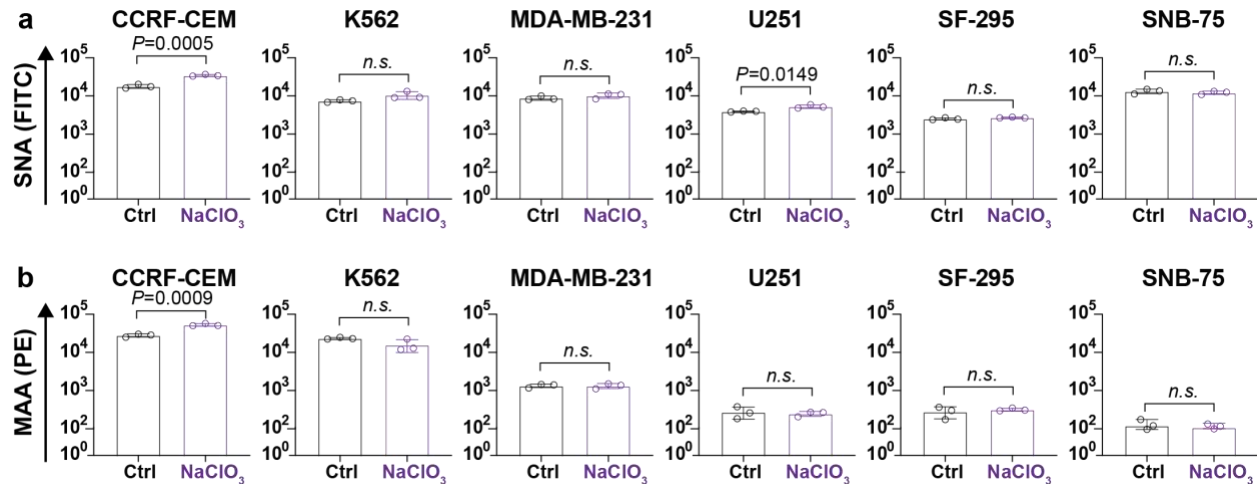


Figure 3.18. Treatment of cancer cell lines with NaClO₃ does not alter staining with SNA or MAA. CCRF-CEM, K562, MDA-MB-231, U251, SF-295, and SNB-75 were treated with NaClO₃ and probed by (a) SNA and (b) MAA, and measured by flow cytometry. Statistical significance was calculated using an unpaired *t*-test.

CCRF-CEM, and MCF-7), Siglec-7, Siglec-5/14 and Siglec-15 did not show any binding decrease after NaClO₃ treatment (**Figure 3.17**). Treatment with NaClO₃ did not result in a significant decrease to either SNA or MAA binding to CCRF-CEM, K562, MDA-MB-231, U251, SF-295, and SNB-75 cells (**Figure 3.18**). Taken together, expression levels of *CHST1* or *CHST2* in cancer cell lines correlate with sensitivity of Siglec ligands to inhibition of carbohydrate sulfation, strongly suggesting that enhanced carbohydrate sulfation is a mechanism by which cancer cells upregulate Siglec ligands.

3.4 Discussions

Sulfation on either 6-*O*-Gal or 6-*O*-GlcNAc has been known for a factor of enhancing the binding affinity between several Siglecs (hCD22, hSiglec-7, 8, 9) and sialosides, but this effect of sulfation on CD33, and Siglec-15 are recently found,¹¹⁹ but to my knowledge, this is the first demonstration of *CHST1*-dependent up-regulation of Siglec-5/14 ligands.⁹³ Most efforts on how sulfation affects Siglecs have focused on Siglec-8 where its direct dissociation constant were determined with SLe^X with different degree of sulfation on 6-*O*-Gal or 6-*O*-GlcNAc.⁹¹ Consistent with our study showing that disulfated 3'SLN is the tightest ligand for CD33 fragment (**Figure 3.8g**), disulfated sLe^X also presented the tightest binding with Siglec-8.⁹¹ In line with this evidence, sialylated keratan sulfates (KS) have been discovered as a strong ligand for Siglec-8 in a human airway.³⁹

KS is a type of proteoglycans which have LacNAc repeats with high degree of sulfation on 6-*O*-Gal or 6-*O*-GlcNAc.¹⁴⁴ It is supposed that KS biosynthesis is mediated by combined enzymatic activities of *KSGal6ST1* (*CHST1*), *GlcNAc6STs*, *B3GNTs*, and *B4GALTs*.¹⁴⁴ Of note, it was shown that either *CHST1* or *CHST2* KO in the mouse brain can lead to abrogation of KS expression in the brain.^{145,146} The length and the degree of sulfation in KS are highly heterogeneous

and numerous proteins carrying KS have been discovered.¹⁴⁷ Recently, *ST3Gal4* is shown to transfer CMP-Neu5Ac to the terminal of the KS.¹⁴⁸ KS can be categorized into type I (*N*-linked glycosylation), II (*O*-linked glycosylation), and III (*O*-mannose glycosylation) depending on a type of glycosylation from which KS is elongated. The type of KS also has organ-specific localization as shown below.¹⁴⁴ (**Figure. 3.19**)

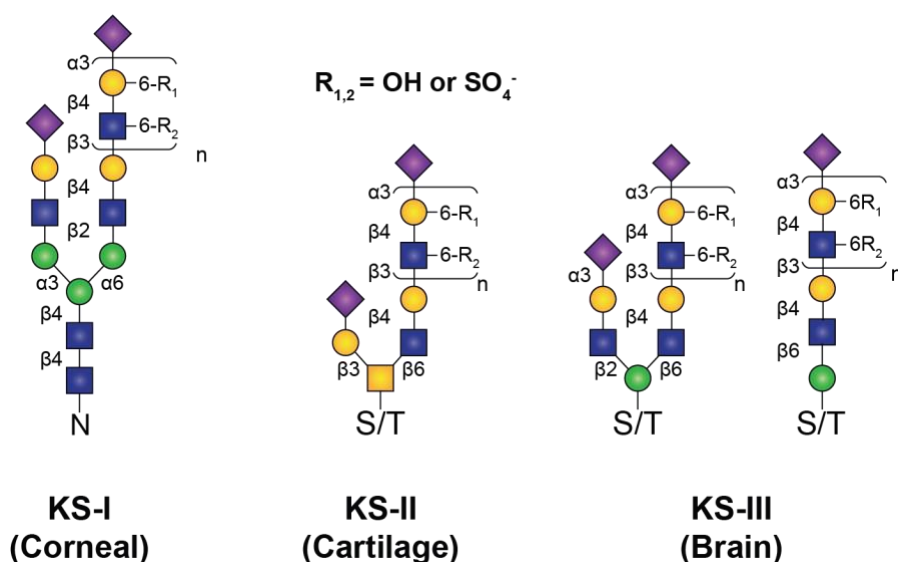


Figure. 3.19 KS can be extended from different types of glycosylation. The types of KS were categorized based on their core backbone. Each type of KS is known for having organ specificity.

Following publication of my work, the most recent updates on the interaction between Siglecs and KS have been made in the human and mouse brain. Homogenized cortex tissues from donors were used for blotting with Siglec-3 and -8.¹⁴⁸ Through the overlay staining with anti-RTPT ζ antibody and MS-based proteomics analysis, receptor protein tyrosine phosphatase zeta (RPTP ζ) is turned out to be a major KS carrier protein in the cortex of the brain, which is presenting high affinity ligands for both Siglec-3 and -8. The evidence was given that sialidase or keratanase I digestion demolished Siglec-3 and -8 binding to the protein, which is further validated with mice brains under either *ST3Gal4*^{-/-} or *CHST1*^{-/-}.¹⁴⁸ The earlier study in 2019 had elucidated RPTP ζ as

KS-carrying proteins in the brain with R10G antibody, but did not further investigate whether Siglecs can bind to the molecule or not.¹⁴⁹ Intriguingly, the loss of 5D4 reactive KS structures in the brains of the 5xFAD *CHST2*^{-/-} mice promoted microglial phagocytosis toward amyloid- β , which may indicate sialylated KS in the brain modulates the function of Siglecs in the brain.¹⁴⁵

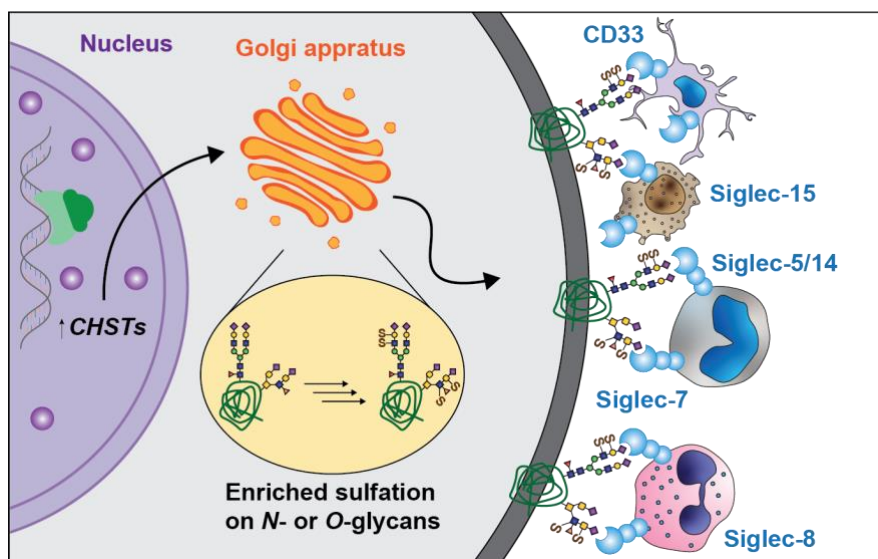


Figure. 3.20 The upregulated carbohydrate sialylation and sulfation in cancer may be able to exploit Siglecs for immune-evasion. Immune cells carrying sulfation-sensitive Siglecs may have high chance to be exploited by cancer cells upregulating CHSTs. This mechanism needs to be defined to predict cancer malignancy and develop therapies.

Most tumours with high *CHST1* expression have shown to correlate with unfavorable outcomes in the Kaplan-Meier 5-year overall survival curves. It has been demonstrated that most cancer cells protect themselves from immune clearance by upregulating cell-surface sialylation to exploit the function of Siglecs, which are immune checkpoint molecules.²¹ As sulfation on a variety of sialosides has been proved to enhance the interactions between Siglecs and their ligands, the synergetic effects of CHSTs and STs on Siglec-mediated immune-evasion in numerous cancers need to be investigated. (**Figure 3.20**) Of note, cancer cells expressing high levels of sialylated KS can manipulate the function of CD33 and Siglec-8 as sialylated KS was demonstrated ligands for

both Siglecs. Metastatic tumours in pancreatic, lung and liver have shown to present higher level of KS than primary tumours,¹⁵⁰ so the correlation between their immune-evasion mechanisms and Siglecs should be elucidated.

3.5 Conclusions

Our results demonstrate that overexpression of either *CHST1* or *CHST2* on several cancer cell lines by lentivirus transduction increases cellular carbohydrate sulfation, which enhances binding to CD22, CD33, Siglec-5/14, Siglec-7, Siglec-8, Siglec-9, and Siglec-15. (Table 3.3) Although enhanced carbohydrate sulfation increased the levels of Siglec ligands in all cells tested, cell type-specific patterns emerged, pointing to the importance of the underlying glycome for supporting sulfated glycans as Siglec ligands. By overexpressing both *CHST1* and *CHST2* in U937 cells to create disulfated glycans, significantly enhanced binding increases of numerous Siglecs (CD33, Siglec-7, Siglec-8, Siglec-5/14, and Siglec-15) were observed. This preference for disulfated glycans by human CD33 is supported by ESI-MSI affinity measurements, which

showed that 6,6'-S,S-3'SLN demonstrated a remarkable 28-fold increase in affinity relative to its non-sulfated counterpart. Interestingly, the binding enhancement of human CD33-Fc to *CHST1*- or *CHST2*-overexpressing cells was not conserved with mCD33-Fc, which has important

Table 3.3 Ligands for Siglecs

Human

Name	Preferred Type of Glycosylation	Ligands
Siglec-1	Gangliosides	
Siglec-2	N-glycans	
Siglec-3	N-glycans O-glycans	
Siglec-4	Gangliosides O-glycans	
Siglec-5/14	N-glycans O-glycans	
Siglec-7	N-glycans O-glycans	
Siglec-8	N-glycans O-glycans	
Siglec-9	N-glycans O-glycans	
Siglec-10	N-glycans	
Siglec-11/16	N-glycans	
Siglec-15	O-glycans	

implications for the study of CD33 as susceptibility loci in Alzheimer's disease. Finally, it was revealed that high expression of *CHST1* is found in many cancers and can lead to poor survival outcomes. The implications of upregulated *CHST1* expression in cancer on Siglec ligands was demonstrated by showing that NaClO₃ abrogates Siglec binding selectively in cancer cells that express high levels of *CHST1* or *CHST2*. Therefore, it is proposed that carbohydrate sulfation as a mechanism that cancer cells can exploit for enhancing interactions with Siglecs to promote immune-evasion.

3.6 Methods and materials

3.6.1 Cell culture

All cell lines were obtained from ATCC, maintained under sterile conditions, cultured with either DMEM or RPMI (Gibco) containing 10% FBS (Gibco), 50 U/mL Penicillin/Streptomycin (Gibco), and 100 mM HEPES (Gibco) and grown at 37 °C with 5% CO₂.

3.6.2 Cloning and transduction of *CHSTs*

The DNA constructs encoding *CHST1*, *CHST2*, *CHST4*, *CHST8*, *CHST9*, *Gal3ST2*, *Gal3ST3*, and *Gal3ST4* were synthesized from GeneArt Gene Synthesis (Thermo Fisher). Each gene was cloned to include 5' NheI site and 3' AgeI site, ligated to linearized RP172 vector¹³ with the same enzyme pair and transformed into NEB stable Competent *E. coli*. Colonies were grown in LB media containing 100 µg/mL of ampicillin (Fisher Scientific) and plasmids were isolated from these bacterial cells using GeneJET plasmid miniprep kit (Thermo Fisher). Sequences of *CHST*-encoding plasmids were verified through Sanger sequencing. Lentivirus was made for each *CHST* gene as described previously. Briefly, the RP172 plasmids containing the *CHST* genes were co-transfected with RP18 and RP19 helper plasmids into HEK293T cells after incubating with TransIT®-LT1 transfection reagent (Mirus Bio). After 3 days, viruses were harvested and

concentrated using Lenti-X Concentrator (Takara Bio) and stored at -80°C in 10 uL aliquots. 200,000 of WT U937 cells were transduced with 10 uL of each virus and cultured for 3 days. Subsequently, positively transduced cells were selected using 3.3 µg/mL of Zeocin (Thermo Fisher). After zeocin selection, mAmetrine positive populations were gated and used in the binding assays.

3.6.3 Construction of the RP173 vector

The mAmetrine fluorescent marker in RP172 was replaced with EGFP to create the RP173 vector. PCR was used to amplify the EGFP sequence, and it was cloned into the RP172 vector using 5' BstBI and 3' SalI restriction sites. The plasmid was transformed into NEB® Stable competent E. coli cells (NEB), and the isolated plasmids were verified through sanger sequencing.

3.6.4 Mutagenesis of the two O-glycosylation sites on CD33.

The original CD33-Fc DNA template¹⁰⁶ was amplified with a forward primer (5'-AGCAGCGCTAGCATGCCGCTGCTGCTACTGCTG-3') and a reverse primer (5'-GGAAAGATAACCAGTGGCTGGGTTCTGTGGAAC-3') to create the T239A mutation. The PCR product (to be used as a megaprimer) was gel-purified, and concentrated into 50 µL of elution buffer, using the GeneJET Gel Extraction Kit (Thermo Fisher). The CD33-Fc template was amplified again using 3 µL of the megaprimer and a reverse primer (5'-AGCAGCACCGGTATGAACCACTCCTGCTCTGG-3'). The resulting PCR product was gel-purified, digested with NheI and AgeI, and ligated into a linearized pcDNA5/FRT/V5-His-TOPO vector containing C-terminal hIgG-His₆-Strep-tag II. After transformation into DH5α competent cells (NEB), the colonies were selected on LB agar containing 100 µg/mL Ampicillin and plasmids from several colonies were isolated for Sanger sequencing. Next, the DNA of CD33-Fc with T239A was amplified with a forward primer (5'-

AGCAGCGCTAGCATGCCGCTGCTGCTACTGCTG-3') and a reverse primer (5'-AAGTACAGGTTCTCACCGGC-3') to mutate T260 into Ala. The resulting megaprimer was gel-purified and concentrated into 50 µL of the elution buffer. The purified megaprimer and a reverse primer (5'-AGCAGCACCGGTTCACTTCTCGAACT-3') were used to amplify CD33-Fc containing T239A. The final product was gel-purified, digested with NheI and AgeI, and ligated into a linearized pcDNA5/FRT/V5-His-TOPO vector. The ligated plasmids were transformed into DH5α competent cells and cultured on LB-agar containing 100 µg/mL Ampicillin. Several colonies were selected and cultured, and plasmids were isolated and sequenced by Sanger sequencing. The pcDNA5 plasmids containing CD33-Fc with T239A and T260A were transfected into CHO Lec-1 cells as previously described.¹⁰⁶

3.6.5 Staining method for Flow-cytometry assay

Cells on a 96-microwell plate were washed with PBS twice. Siglec-Fc were used at approximately 20 µg/mL and AF647 Goat anti-hIgG (Clone: HP6017, 0.4 µg/mL, Biolegend) were mixed in a 1:1 volume ratio right before staining, and 100 µL of the mixture was used for staining each well. After incubation on ice for 30 min, the stained cells were washed twice with PBS, resuspended in PBS, and analyzed using Fortessa X-20 Flow cytometer (BD Biosciences). Flow cytometric data were analyzed using FlowJo v10 software.

3.6.6 Glycosylation or Sulfation inhibitors

500,000 of the WT, CHST1, and CHST2 overexpressing U937 cells were plated onto 6-well plates and incubated at 37 °C in 2 mL of RPMI media containing 10% FBS, 50 U/mL of Penicillin/Streptomycin, and 100 mM of HEPES with kifunensine (2.5 µg/mL, Toronto Research Chemicals) or benzyl-α-GalNAc (2.0 mM, Sigma) or Genz123346 (5 µg/mL, Toronto Research

Chemicals) for 72 hours. The same number of the WT, CHST1, and C HST2 cells were treated with 25 mM of NaClO₃ for 48 hours to inhibit cellular carbohydrate sulfation.

3.6.7 Statistical analyses

All statistical analyses in this study were performed using GraphPad Prism version 7 software. For experiments comparing two groups, an unpaired two-tailed Student's *t*-test was used to evaluate statistical significance. For datasets with more than two sample groups, a one-way ANOVA with Tukey's post hoc test was performed. For group analysis, a two-way ANOVA with either Tukey's post hoc test or Sidak's multiple comparisons test was performed. A non-parametric Wilcoxon matched-pairs signed-rank test was used to assess statistical significance for *CHST1* transcript expression levels in matched non-malignant adjacent tissues and their primary tumour counterparts. Statistical significance of the 5-year overall survival curves between assigned case groups were evaluated by Gehan-Breslow-Wilcoxon test. Data were considered statistically significant when $P < 0.05$.

3.6.8 Liposome binding assay

EV, CHST1, CHST2, and CHST1/2-overexpressing U937 cells were resuspended in media, and 100,000 cells were added to a 96-well U-bottom plate in 200 μ L RPMI containing 10% FCS. Cells were centrifuged (300 rcf, 5 min) and the supernatant was discarded. The cell pellet was re-suspended in 50 μ L of fresh media and 50 μ L of media containing naked (without CD33 ligand) or CD33 ligand (CD33L) tagged liposomes was added to it. The final concentration of liposomes was 100 μ M (total concentration of lipid). The CD33L concentration on the liposomes was 3.33 mol %, all liposomes contained 0.1 mol % AF647-conjugated lipid and were prepared as described previously.¹⁰⁶ These suspensions were incubated for 60 min at 37 °C. Following this incubation, 100 μ L of media was added to each sample and they were centrifuged (300 rcf, 5 min).

The supernatant was discarded, and the cell pellet was suspended in a flow buffer and further analyzed by flow cytometry.

3.6.9 Measurement of cell viability

Adherent cell lines (MDA-MB-231, MDA-MB-468, MCF-7, SF-295, SNB-75) were trypsinized and resuspended into RPMI growth media (10% FBS, 50 U/mL penicillin/streptomycin, 100 mM HEPES). Non-adherent cell lines (U937, K562, and CCRF-CEM) were directly resuspended into the RPMI growth media. A half million of each cancer cell line was plated onto a 12-well plate with 2 mL the RPMI growth media containing 25 mM NaClO₃ and cultured at 37°C with 5% CO₂ for 48h. Subsequently, the adherent cell lines were incubated with PBS containing 1 mM EDTA for 5 min and carefully taken off from the surface by pipetting up and down. The non-adherent cell lines were moved into 15 mL tube. After washing the cells with PBS twice with centrifugation (300 rcf, 5 min), the cells were incubated with PBS containing 2 µg/mL propidium iodide (PI) on ice for 10 mins. The cell viability was measured by flow cytometry by assessing PI negative population.

Chapter 4

A development of a reporter system for screening small molecule drug candidates inducing the expression of hCD33m

4.1 Acknowledgement

We thank Dr. Samuel A. Hasson for advice on CD33 splicing reporter system. I thank Dr. Ghazaleh Eskandari-Sedighi for advice on hCD33m localization.

Section 4.3.1 and 4.3.2 were published in Bhattacharjee, Jung *et al.*, *Molecular Neurodegeneration*, 2021.

4.2 Introduction

Allelic difference within *CD33*, at rs12459419, affects a ratio of the long (hCD33M) and short (hCD33m) isoforms, which is associated with AD susceptibility. The rare allele (rs12459419T) decreases AD risk, which correlates with more exon-2 skipping events and, consequently, more hCD33m. Conversely, the common allele (rs12459419C) undergoes exon-2 skipping less frequently, resulting in more expression of hCD33M and less hCD33m, which is associated with increased LOAD risk.¹⁵¹ It was first hypothesized that a lower expression of hCD33M can be AD protective.^{63,70} Several studies supported this loss-of-function hypothesis. First, BV2 mouse microglia cell lines transiently transfected with hCD33M showed a reduced ability to phagocytose A β .⁶³ These results were in line with mCD33^{-/-} microglia being more phagocytic, and consequently mCD33^{-/-} mice accumulated less A β in the brain in an AD mouse model.⁶³ Nevertheless, it is important to note that our laboratory challenged these findings by showing that mCD33 and hCD33 are not functionally conserved.⁹² In parallel with this study in mouse microglia, it was found that human monocytes homozygous for the risk allele (C/C) were less phagocytic compared to monocytes heterozygous (C/T) or homozygous (T/T) for the rare CD33 allele.⁷⁵ These results were originally interpreted in the context of the loss-of-function hypothesis, but it is also possible that enhanced phagocytosis was the result of increased hCD33m expression. **(Figure 1.9)**

Another *CD33* allele, which is a null allele (rs201074739), was analyzed for AD susceptibility in a prospective review of GWAS data and, surprisingly, was found not to be AD protective.⁶⁵ This finding suggested the loss-of-function hypothesis may not fully explain how the CD33 locus can be AD protective. Directly testing a gain-of-function role for hCD33m presented challenges because human cells express both hCD33 isoforms, making it difficult to deconvolute

isoform-specific phenotypes. In addition, hCD33m is preferentially localized intracellularly, so it is difficult to detect its expression level without overexpressed conditions.^{79,81,152} A free Cys42 on hCD33m was speculated to be a factor for the intracellular localization of hCD33m, but it remains to be directly tested if the unpaired cysteine is responsible for retaining hCD33m inside cells. Overall, the notable differences in subcellular localization of the two hCD33 isoforms could be major determinant of their distinctive functions, but this has yet to be rigorously tested.

In work that I contributed to by our laboratory, CD33^{-/-} U937 cells overexpressing hCD33m showed increased phagocytosis toward A β aggregates compared to CD33^{-/-} U937 cells.¹³⁵ This gain-of-function role for hCD33m expression was also observed in transgenic primary mouse microglia expressing hCD33m. My contribution to this work was to create biochemical tools (monoclonal antibodies) and cell lines that helped support the conclusions. In parallel, another group independently demonstrated similar results in cultured BV-2 mouse microglia.¹⁵³ In summary, these results demonstrate a gain-of-function role for hCD33m, which may contribute to AD protection.

In this Chapter, new antibodies specific for hCD33m were developed and characterized, enabling detection of hCD33m at the protein level. Subsequently, *CD33* minigene constructs recapitulating the common or rare allele of rs12459419 were designed and expressed in CD33^{-/-} U937 cells. Different ratios of both isoforms at the protein level were observed for the C or T alleles. Moreover, this *CD33* minigene with the T allele endowed cells with a greater ability to phagocytose A β aggregates. The role of Cys42 on hCD33m in protein localization was also investigated and it was revealed that the unpaired Cys42 gives rise to the intracellular localization of hCD33m. Finally, three luciferase-based CD33 splicing reporter systems were devised and

tested, which may find utility in the future for screening compounds that can induce exon-2 skipping.

4.3 Results

4.3.1 Development of monoclonal antibodies that recognize hCD33m

An anti-CD33 antibody (clone HIM3-4) is reported to recognize the C2 domain of hCD33, but in our hands, it does not stain U937 cells overexpressing hCD33m.¹⁵⁴ (**Figure 4.1**) Conversely, another anti-CD33 antibody (clone WM53) recognizes the V-set domain and reliably detects hCD33M, but not hCD33m, expressing cells. For this reason, we pursued the development of monoclonal antibodies that are specific for hCD33m. Recombinant hCD33m was prepared using the Fc chimeric expression system discussed in Chapter 2. Briefly, hCD33m-Fc was cloned and transfected to WT CHO cells for protein expression. Supernatant (4L) was purified over a HisTrap and Protein-G column to obtain hCD33m-Fc. Digestion with TEV protease released hCD33m fragments from the Fc, and the TEV and Fc domain were removed by passing the reaction back through a nickel column. The pure hCD33m fragment (4 mg) was sent to a contract research organization (Pierce Proteomics) for custom antibody development. Hybridoma clones were initially screened by ELISA and sent back for validation with U937 cells overexpressing hCD33m. (**Figure 4.2a**). Of the 48 clones that bound CD33m by ELISA, two promising candidates (clone S503 and S823) were able to detect hCD33m by flow cytometry and both were identified to be of the mIgG1 isotype. (**Figure 4.2b, c**)

Upon further examining the specificity of S503, a small but significant amount of binding to hCD33M was observed. (**Figure 4.2b**) Interestingly, when cells were treated with 1 mM dithiotreitol (DTT), which reduces the interdomain disulfide bond between Cys36 and Cys169,

enhanced staining of hCD33M, but not hCD33m, was observed (**Figure 4.3**). These results suggest that the epitope of the S503 antibody is exposed upon reduction of the interdomain disulfide bond.

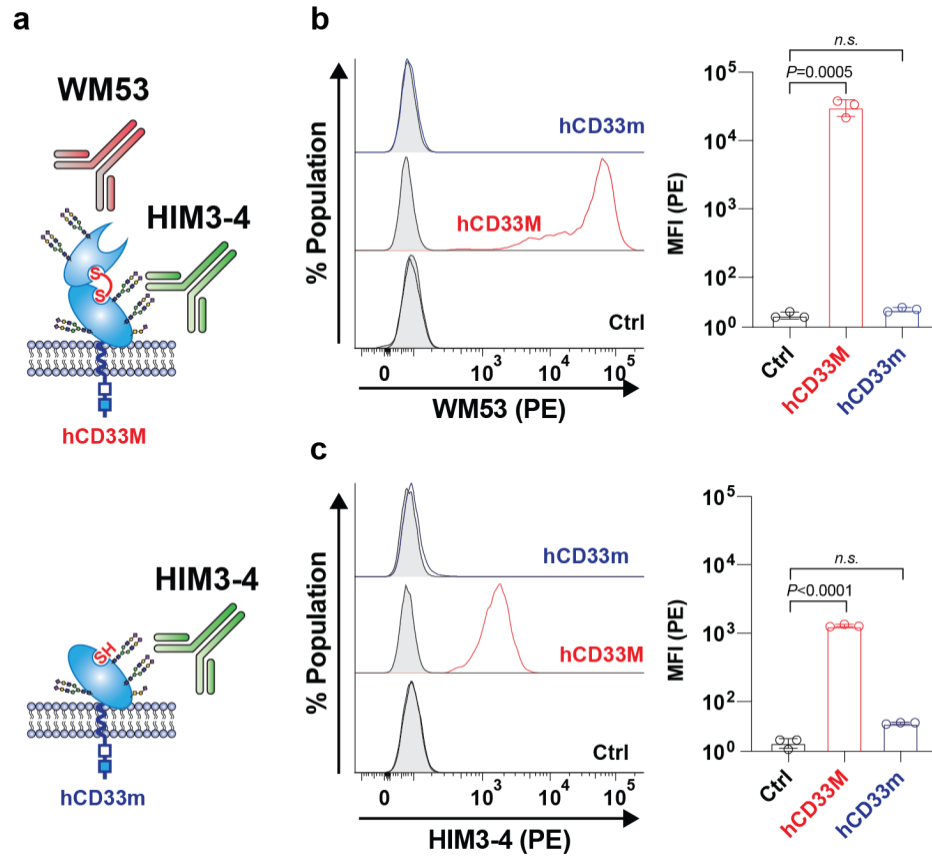


Figure 4.1 Antibody staining of the two isoforms of human CD33. (a) WM53 recognizes the V-set domain of CD33 and HIM3-4 recognizes the C2 domain. (b,c) CD33^{-/-} U937 cells were transduced with lentivirus expression hCD33M (red), hCD33m (blue), control (black) and stained with WM53 (b) and HIM3-4 (c). Data plots in b, and c are presented as mean \pm SD. Statistical significance was calculated using a one-way ANOVA with Tukey's multiple comparisons test.

The weak binding of S503 to hCD33M either represents very weak exposure of the S503 epitope on hCD33M, or a very small percentage of hCD33M that does not contain the intact disulfide bond.

Previously, it was reported that hCD33m is colocalized in peroxisomes.⁸⁰ In another study, detection of hCD33m on cancer cell lines and human acute leukemia cells from AML patients was not successful in either the intracellular or extracellular compartments.¹⁵⁵ To test the localization of hCD33m, transgenic mice expressing hCD33m on microglia were tested for staining with S503.

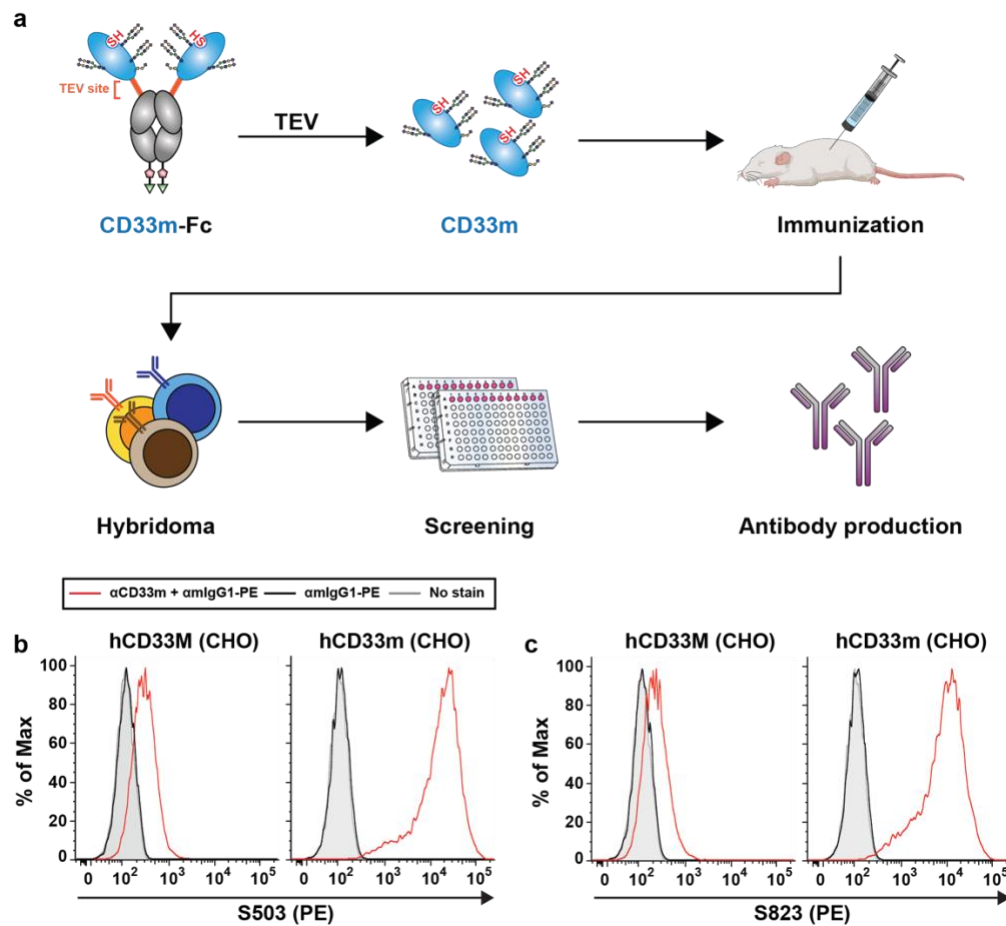


Figure 4.2 Development of two anti-hCD33m monoclonal antibodies. (a) A schematic for antibody production. CHO cells expressing hCD33M or hCD33m were screened against 48 clones. (b) S503 and (c) S823 recognize hCD33m-expressing cells, and both are mouse IgG1 isotype.

Primary microglia cells were isolated from adult (3-month old) sex-matched mouse brains. For cell surface staining, microglia expressing hCD33m did not show any positive staining. (**Figure 4.4a**) However, a small amount of staining with S503 was observed in permeabilized cells,

confirming that hCD33m prefers an intracellular location under conditions that do not represent overexpression. (Figure 4.4b)

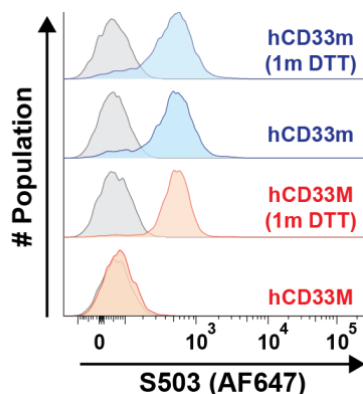


Figure 4.3 Effect of DDT on S503 and S823 staining of hCD33m. U937 cells expressing hCD33M or hCD33m were treated with DTT to reduce the interdomain disulfide bond on hCD33M. The DTT treated cells were stained with S503 carrying Alexa647 and the signal was measured by flow cytometry.

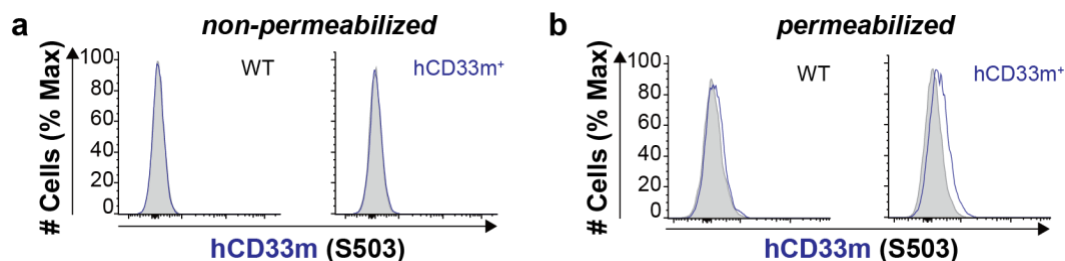


Figure 4.4 Intracellular staining of S503 reveals that hCD33m is localized inside mouse microglia. (a) Extracellular staining or (b) intracellular staining with S503 on isolated primary mouse microglia expressing hCD33m were done by flow cytometry. WT microglia was used as a control.

4.3.2 CD33 minigene constructs that faithfully mimics differential splicing

At mRNA level, it has been shown that the rs12459419 locus containing C or T affects the ratio of hCD33M and hCD33m transcripts.^{70,156} At the protein level, it was shown that hCD33M expression level in C allele is higher than T allele, but hCD33m protein expression levels were not evaluated due to the lack of a suitable anti-hCD33m antibody.¹⁵⁶ To evaluate how frequently the common (rs12459419C) or rare (rs12459419T) alleles undergo exon-2 skipping, two minigene

constructs were created with either C or T at the key rs12459419 locus. The minigene constructs included the two introns flanking exon-2 but without the downstream introns because they would make the constructs too large. **(Figure 4.5a)** The two minigene constructs were transduced into CD33^{-/-} U937 cells, which were created by Chris St. Laurent in our laboratory by CRISPR/Cas9.¹³⁵ To validate the expression levels of both isoforms, cells were stained with the antibodies WM53 or S503 and tested by flow cytometry. **(Figure 4.5b)** Higher expression of hCD33M was observed in the C allele than the T allele, which was matched by a corresponding 3-fold lower expression of hCD33m in the C allele. **(Figure 4.5c)** These results are consistent with the previously evaluated mRNA results^{70,156} and demonstrate for the first time at the protein level that the T allele has higher expression of hCD33m in company with lower expression of hCD33M.

It was expected that U937 cells expressing *CD33* minigene with the T allele should have enhanced phagocytosis ability compared to U937 cells expressing *CD33* minigene with the C allele due to the combined effects of less expression of hCD33M and more expression of hCD33m. To test this hypothesis, U937 cells expressing the *CD33* minigenes with either C or T allele were evaluated for their ability to phagocytose A β ₁₋₄₂ aggregates. Cells treated with Cytochalasin D, which inhibits phagocytosis,¹⁵⁷ served as a negative control. In flow cytometry analysis, the U937 cells expressing *CD33* minigene with the T allele had higher A β ₁₋₄₂ uptake. **(Figure 4.5d)** In parallel to this result, a colleague of mine in the laboratory demonstrated that primary mouse microglia and U937 expressing hCD33m phagocytose more A β ₁₋₄₂ aggregates than their WT control.¹³⁵ Overall, these findings suggest that higher expression of hCD33m in T allele contributes to enhanced phagocytosis.

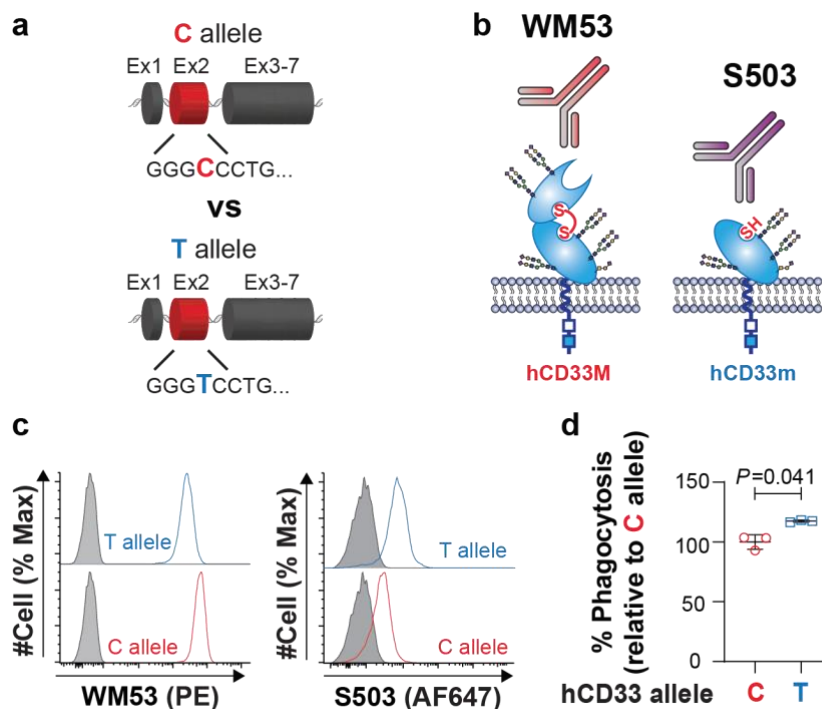


Figure 4.5 Recapitulation of alternative splicing at the rs12459419 locus with *CD33* minigene constructs. (a) Two *CD33* minigenes with the C or T allele were designed to evaluate how each allele affects the expression level of both CD33 isoforms. (b) WM53 detects the V set domain of hCD33M and S503 detects the C2 domain of hCD33m. (c) Flow cytometry analysis of the protein expression level of both isoforms was carried out in CD33^{-/-} U937 cells overexpressing either of *CD33* minigene constructs. The grey shaded histogram indicates an isotype control. (d) Flow cytometry-based phagocytosis assay with aggregated A β_{1-42} labeled with HiLyteTM Fluor 555 was performed on CD33^{-/-} U937 cells expressing CD33 minigene with the C or T allele. Statistical significance was calculated using a two-tailed unpaired Student's *t* test.

4.3.3 A role for the free cysteine residue on hCD33m in influencing its cellular localization

As demonstrated above, DTT treatment exposed the epitopes for S503 and S823, which likely took place through disrupting the interdomain disulfide bond between Cys36 and Cys169. (Figure 1.11) To test this more directly, Cys169 on hCD33M and the corresponding Cys42 on hCD33m were mutated to Ala. The expression levels of CD33 were measured with WM53 and S503. The C169A hCD33M mutant stained lower with WM53 but higher with S503, once again suggesting that the interdomain disulfide bond keeps the S503 epitope hidden. (Figure 4.6a,b)

The C42A hCD33m mutant was also expressed at higher levels on the cell surface. (**Figure 4.6b**) This latter result suggests either that the free Cys42 in hCD33m is involved in a disulfide bond with another protein, which thereby hinders access to the S503 epitope, or that the free Cys42 keeps hCD33m inside the cell. However, given that DTT did not affect S503 staining, as shown above (**Figure 4.3**), it favors the explanation that differences in expression levels of hCD33m between WT and C42A may be due to an altered cellular localization. Therefore, it was hypothesized that the Cys42 is important for retaining hCD33m inside the cell since WT hCD33m is localized intracellularly under physiological expression levels.⁸⁰ (**Figure 4.4c**)

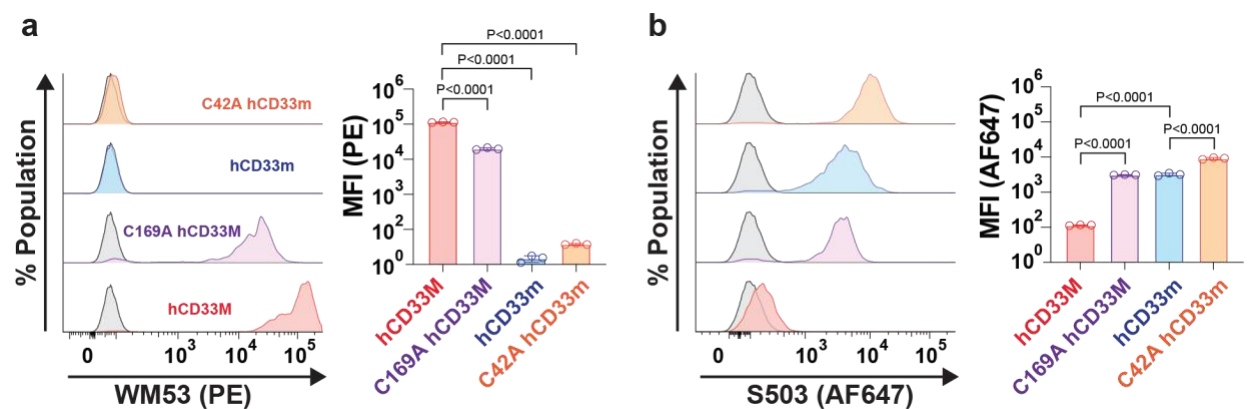


Figure 4.6 The effect of an interdomain disulfide bond on S503 recognition. Four constructs of hCD33M, C169A hCD33M, hCD33m, C42A hCD33m were introduced to CD33^{-/-} U937 cells and the effect of the mutation on antibody staining was assessed on (a) WM53 and (b) S503. Statistical significance was calculated using a one-way ANOVA with Tukey's multiple comparisons test.

Given that levels of protein expression clearly influence how much hCD33m reaches the cell surface, a strategy for achieving different expression levels of hCD33m was devised by altering the CD33 minigene system. To avoid overexpressed conditions, we inserted two stop codons (2xSTOP) in the middle of exon-2 to prevent expression of hCD33M, which still allows for hCD33m expression when exon-2 exclusion occurs. Accordingly, this 2xSTOP constructs were used and analyzed in the context of a C42A mutation. First, in the comparison between C and T

CD33 minigene constructs, the T allele *CD33* minigene had 3-fold higher S503 staining, indicating there was more exon-2 skipping. (**Figure 4.7**) Next, the effect of a C42A mutation was evaluated in both alleles. In each allele, a C42A mutation led to significantly higher cell surface S503 staining,

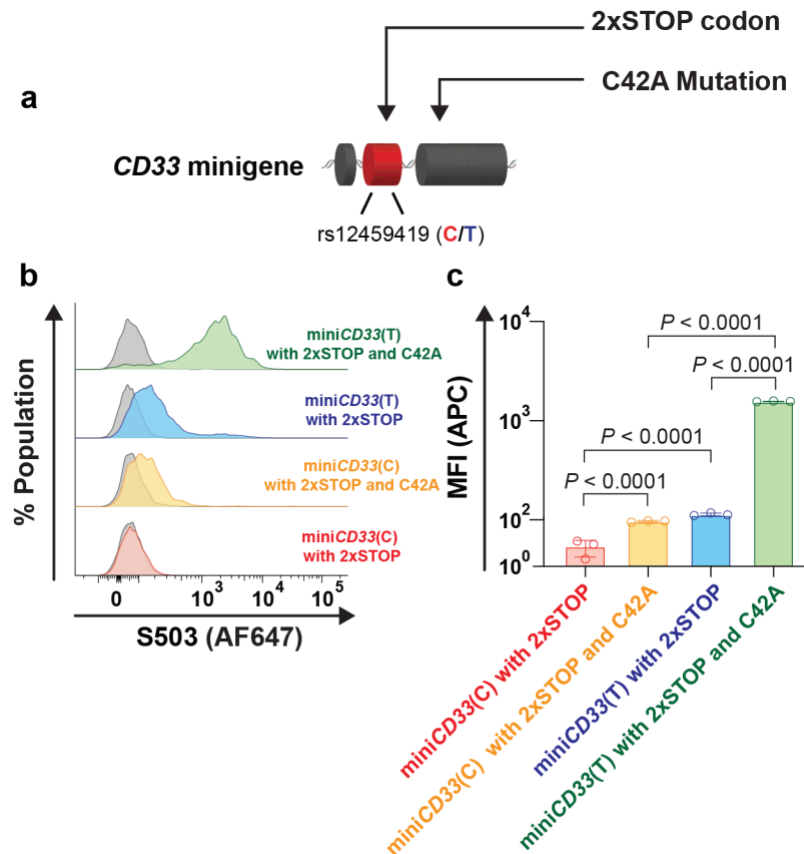


Figure 4.7 A novel system for studying rs12459419 and evaluation of C42A mutation. (a) Different expression levels of hCD33m can be achieved by inserting 2xSTOP at exon-2 to both rs12459419 alleles, to specifically block the expression of hCD33M. (b,c) The effect of C42A mutation on the cell surface expression of hCD33m was evaluated with the flow cytometry plot (b) and quantification (c) shown. Statistical significance was calculated using a one-way ANOVA with Tukey's multiple comparisons test.

which again supports the idea that the C42A mutation affects the localization of hCD33m. However, another possibility for these results is that the C42A mutation affects the total expression levels of hCD33m levels through a post-translational effect. To rule out this possibility, intracellular staining was carried out. Attempts to use S503 were not successful due to unexpectedly high background staining of the CD33^{-/-} U937 cells, therefore, another approach was

used, which involved genetically encoding a 3xFLAG-tag to the C-terminus. Extracellular staining with S503 confirmed that the C42A mutation led to 17-fold higher cell-surface expression than WT. (**Figure 4.8a**) However, this difference was only 1.2-fold difference in the intracellular staining between hCD33m WT and C42A mutation. (**Figure 4.8b,c**) These results support that the C42A mutation alters the trafficking of hCD33m, but not overall expression levels.

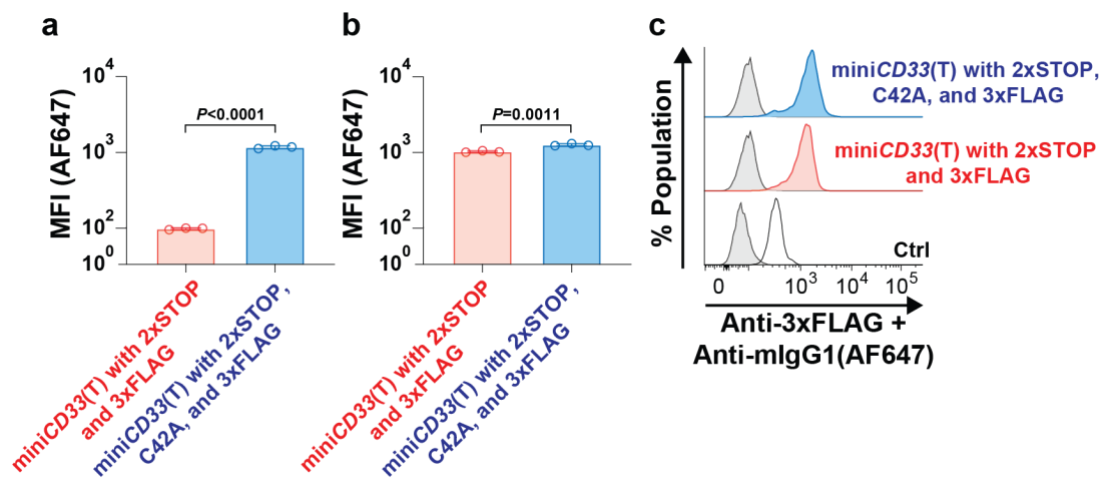


Figure 4.8 3xFLAG-tag helps to evaluate the total expression level of hCD33m. (a-c) Extracellular staining, and intracellular staining of WT and C42A *CD33* minigene constructs with the T allele, 2xSTOP, and C-terminal 3xFLAG-tag was shown in quantification (a,b), and flow cytometry plot (c). Statistical significance was calculated using a two-tailed unpaired Student's *t* test.

4.3.4 Development of an artificial system for transforming hCD33M to hCD33m

Inducing exon-2 skipping could be an attractive therapeutic strategy to leverage the beneficial effects of hCD33m on LOAD. To assess its therapeutic potential, a system for turning hCD33M expression into hCD33m expression in a targeted cell would be a good positive control and provide a proof-of-concept approach to motivate drug discovery. To enable this goal, a new strategy was developed whereby two LoxP sites were embedded into intron-1 and -2 of the *CD33* minigene, which is named based on a floxed exon-2: $CD33^{ex2(fl/ flx)}$. (**Figure 4.9a**) In this way, Cre recombinase would be able to remove the DNA fragment between the two LoxP sites, resulting in

exon-2 deletion. Lentivirus encoding the C allele of $CD33^{ex2(fl x/fl x)}$ was transduced to WT U937 cells. The overexpression of $CD33^{ex2(fl x/fl x)}$ was evident through increased WM53 staining.

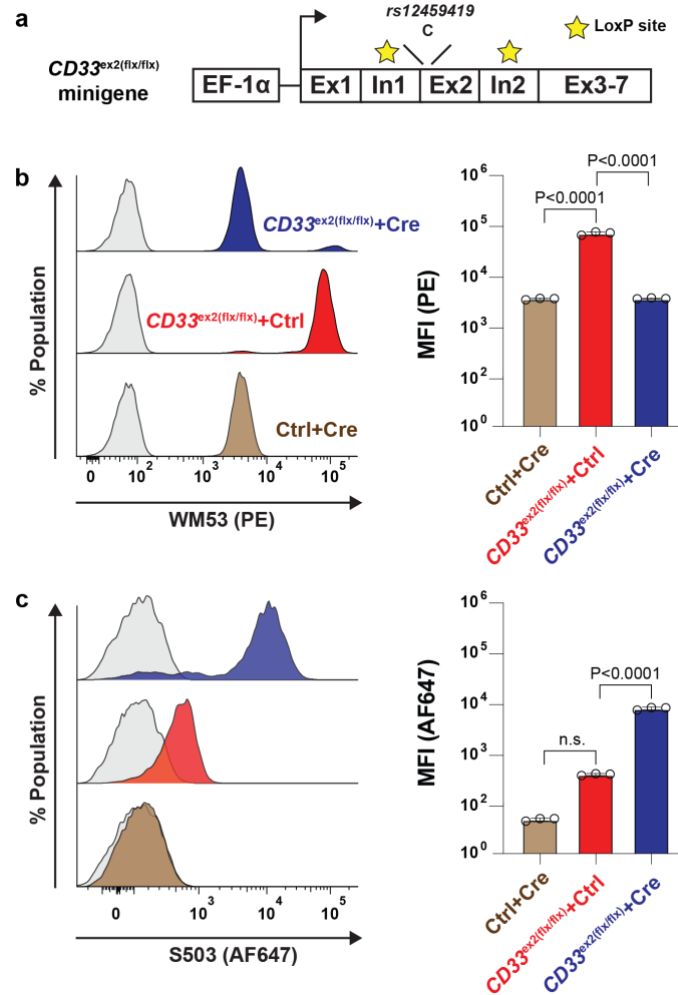


Figure 4.9 Transforming hCD33M into hCD33m through Cre recombination (a) DNA compositions of $CD33^{ex2(fl x/fl x)}$. Two LoxP sites were embedded in the middle of intron-1 and -2 for Cre recombination. WM53 staining (b) or S503 staining (c) on WT U937 cells with expressing 1) Control + Cre recombinase 2) Control + $CD33^{ex2(fl x/fl x)}$, and 3) $CD33^{ex2(fl x/fl x)}$ + Cre recombinase. Statistical significance was calculated using a one-way ANOVA with Tukey's multiple comparisons test.

(Figure 4.9b) After Cre recombinase was introduced to the cell expressing $CD33^{ex2(fl x/fl x)}$, WM53 staining levels returned to the WT levels, strongly suggesting the successful removal of Exon-2. Concurrently, a ten-fold enhancement in S503 staining was observed, which demonstrates that

CD33^{ex2(flx/flx)} was transformed into hCD33m expression. (**Figure 4.9c**) As 10% of exon-2 skipping occurs in the C allele, increasing this to 100% events is in line with the ten-fold increase in S503 staining. In the future, it can be envisioned using this strategy to switch CD33M expression into CD33m expression *in vivo* to study the therapeutic potential of exon-2 skipping drugs.

4.3.5 Development of cell-based reports of exon-2 skipping

Overall, current data support both a loss-of-function, stemming from less expression of hCD33M, and gain-of-function, stemming from increased expression of hCD33m, effects as mediators of the impact of CD33 on AD susceptibility.^{92,135,153} (**Figure 4.5**) Accordingly, enhancing exon-2 skipping is an exciting therapeutic strategy for LOAD. To recapitulate the biology of enhanced exon-2 skipping being AD protective, Pfizer initiated work to look for small molecules capable of inducing exon-2 skipping, which was done through a high-throughput cell-based screen.¹⁵⁶ Specifically, a Nanoluc-based reporter system was used that was based on the backbone sequence of CD33 including exon-1, intron-1, exon-2 with two stop codons (2xSTOP), intron-2, and truncated exon-3. (**Figure 4.10a**). Moreover, the truncated exon-3 contained a P2A self-cleavage site and Nanoluc Luciferase (NLuc) gene to monitor luminescence signal that reflects the degree of exon-2 skipping. To validate this reporter system, approximately 90 2'-O-methoxyethylribose antisense nucleotides (MOE-ASO) were screened for their ability to induce exon-2 skipping, with several promising candidates.¹⁵⁶ In a follow-up study, a HTS on small molecules was conducted using this cell reporter to discover several drug candidates capable of inducing exon-2 skipping.¹⁵⁸ 'Compound 1' was selected as the most promising lead compound, with a limited amount of follow-up testing. (**Figure 4.10b**) Specifically, Compound 1 was able to reduce cell surface hCD33M expression by 43%.¹⁵⁸ Similar to other studies, direct detection of

hCD33m at the protein levels was not carried out due to the unavailability of biochemical tools for detecting hCD33m.

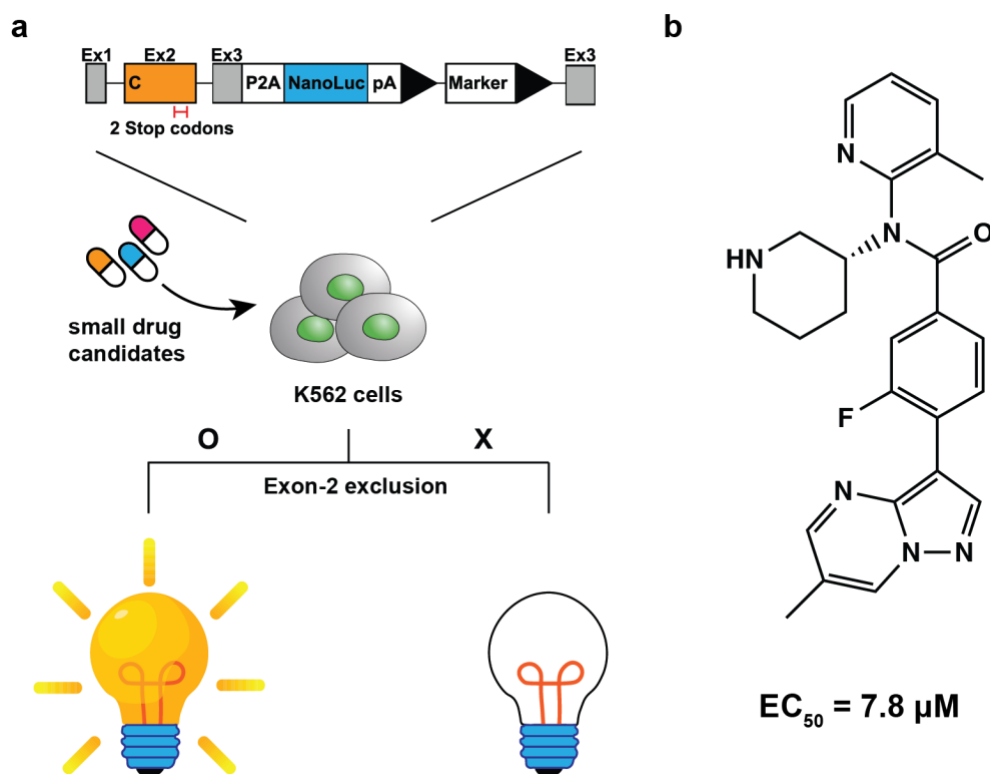


Figure 4.10 A reporter system for discovering drug candidates that induce exon-2 skipping. (a) A Schematic for a cell-based reporter to detect small molecules blocking exon-2 inclusion, which leads to the conversion from hCD33M to hCD33m. The insertion of double stop codons at the end of 3' exon-2 ensures that NLuc is only expressed when exon-2 skipped. (b) This reporting system was applied to a HTS, which led to the discovery of Compound 1.¹⁵⁸

Aiming to perform our own screening for small molecule drug candidates capable of inducing exon-2 skipping, several luciferase-based reporting systems were developed and validated. Our efforts in this area started before Pfizer published their paper identifying lead compounds identified from this assay. The first construct (Construct I) was similar to the one published by Pfizer.¹⁵⁶ (**Figure 4.11a**) Construct I is a *CD33* minigene, where a 2xSTOP system is introduced in the middle of exon-2 so that a luminescence signal is only generated upon

successful exon-2 skipping. A P2A self-cleavage site followed by a NLuc gene placed in frame within exon-3. To validate that Construct I recapitulates the biology of the rs12459419 locus, Construct I was created as both the C and T allele. Consistent with our previous work with the *CD33* minigene, Construct I with the T allele showed a 3.5-fold higher bioluminescent signal. (Figure 4.11b) This result indicates that even though this reporter is introduced into a different part of the genome, unlike the Pfizer construct that was knocked-in to the *CD33* locus,¹⁵⁶ the fidelity of splicing is maintained. To test maximum bioluminescent intensity when all pre-mRNA

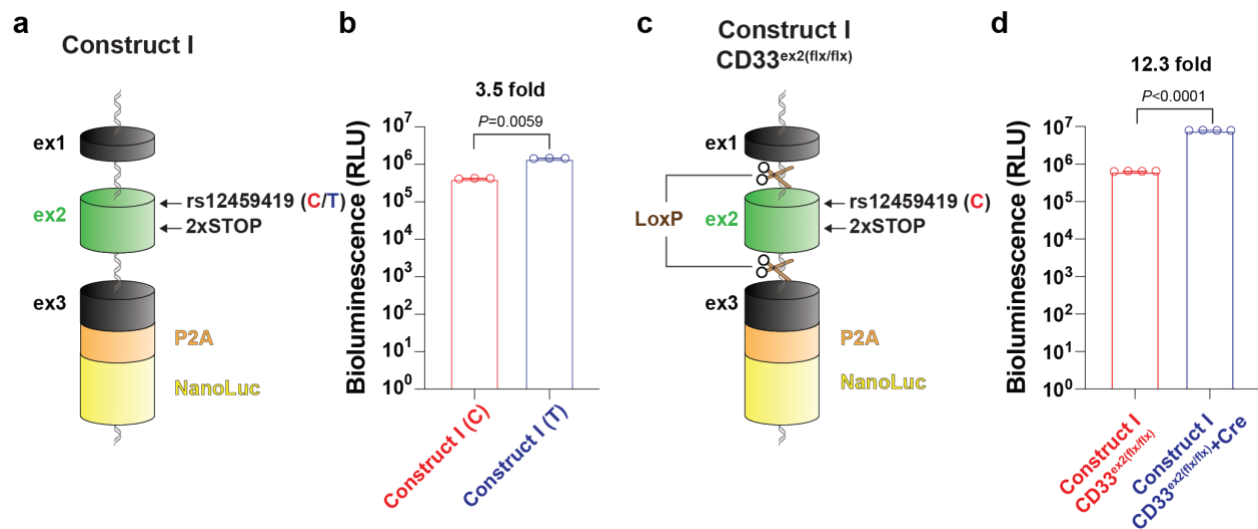


Figure 4.11 Validation of splicing reporter Construct I. (a) DNA compositions of Construct I is shown. Construct I with C or T allele were made. At the end of exon-3, A P2A self-cleavage site and NLuc gene were recombined. (b) Bioluminescent intensities for Construct I with C and T allele were shown in a bar graph. Construct I with T allele had 3.5-fold higher signal intensity than C allele. (c) DNA compositions of Construct I CD33^{ex2(flx/flx)}. (d) After Cre recombination, bioluminescent intensity went up to 12.3-fold. Statistical significance was calculated using a two-tailed unpaired Student's *t* test.

of *CD33* undergoes exon-2 skipping, two LoxP sites were introduced onto intron-1 and -2. (Figure 4.11c) After Cre recombination, which was accomplished by lentivirus-mediated expression of Cre recombinase, the bioluminescent intensity went up to 12.3-fold compared to a control. (Figure 4.11d) As rs12459419C undergoes exon-2 exclusion at approximately a 10% frequency, this 12.3-fold increase is very close to the expected 10-fold increase (10% to 100%) of exon-2 exclusion.

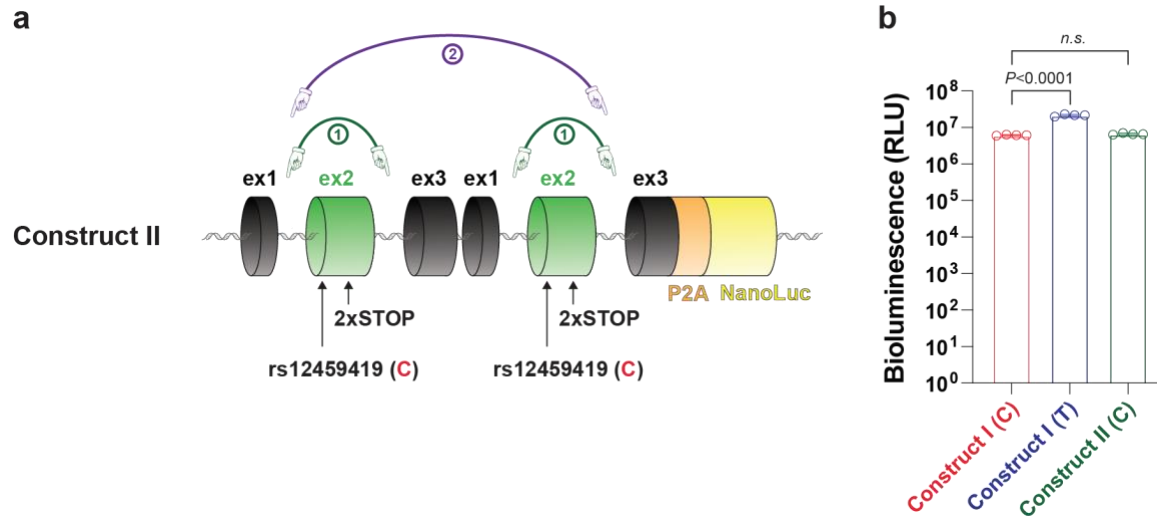


Figure 4.12 Validation of reporter splicing Construct II. (a) Construct II is composed of double cassettes, which contain CD33 splicing elements including exon-1, intron-1, exon-2, intron-2, and exon-3. Each cassette contains rs12459419C and 2xSTOP at exon-2. Two possible splicing mechanisms are shown. (b) Bioluminescent intensities for Construct I with the C and T alleles, and Construct II with the C allele were measured. Statistical significance was calculated using a one-way ANOVA with Tukey's multiple comparisons test.

Although Construct I worked well, there are two ways that it could be improved, which were addressed by developing Construct II and III. The first way that Construct I could be improved is by increasing the maximum signal that can be produced upon exon-2 skipping. Even for the C allele, exon-2 exclusion occurs at a 10% frequency.⁶⁴ Therefore, the maximum increase we would expect to see from full exon-2 skipping would be 10-fold. This was indeed what was observed above with the CD33^{ex2(flx/flx)} construct. (**Figure 4.11b,d**) To maximize the gap between basal signal and signal from complete exon-2 skipping, Construct II was devised by creating a duplicated a splicing cassette ranging from exon-1 to exon-3. (**Figure 4.12a**) It was hypothesized that if exon-2 exclusion occurs independently on each cassette, two exon-2 skipping events would be required to generate signal, hence, potentially lowering the basal signal down to 1% (10% multiplied by 10%) from the maximum signal. (**Figure 4.12a**) To test this idea, Construct II was run in parallel with Construct I with the C or T allele. Unexpectedly, the double cassette generated

a signal that was equal to Construct I with the C allele. (**Figure 4.12b**) This was likely due to an exclusion event from the 5' end of the first intron 1 to the 3' end of the second intron 2. Thus, this approach was not practical, yet this result is highly informative because it indicates that the number of nucleotides between the two splice sites can be significantly lengthened.

The second way where Construct I could be improved is by having an internal control that would rule out compounds that impact the overall CD33 expression levels. Several options were considered, with the most compelling use of two orthogonal luciferase signals; one that reports on exon-2 inclusion and one that reports on exon-2 exclusion. Therefore, Construct III was designed with these thoughts in mind, to leverage a ratiometric dual luciferase assay.¹⁵⁹ (**Figure 4.13a**) This reporter system has NLuc in the place of exon-2 and Firefly Luciferase (Fluc) after exon-3. Accordingly, the NLuc signal represents exon-2 inclusion. To prevent exon-2 inclusion from leading to an enriched signal from both luciferases, a STOP codon was inserted at the 3' end of NLuc gene. When exon-2 skipping occurs, the NLuc gene is spliced out, leading to a Fluc signal. Therefore, the ratio of the FLuc signal over the NLuc signal represents exon-2 skipping. To test this, Construct III with C or T allele was tested and it was found that in the FLuc bioluminescent signal, Construct III with T allele was higher than C allele (**Figure 4.13b**). However, the NLuc signal was not statistically significant (**Figure 4.13c**). The latter result was unexpected because the signal should have decreased by a factor of 7/9 (23% decrease) stemming from fraction of CD33M transcript product in the T allele (70%) versus the C allele (90%). The precision in our assay, clonal variability, or saturation levels of luminescence may have masked the ability to observe a 23% reduction. Nevertheless, FLuc/NLuc values between C and T allele were statistically different, where T allele had 35 % higher value. (**Figure 4.13d**) To test the maximum FLuc/NLuc value in

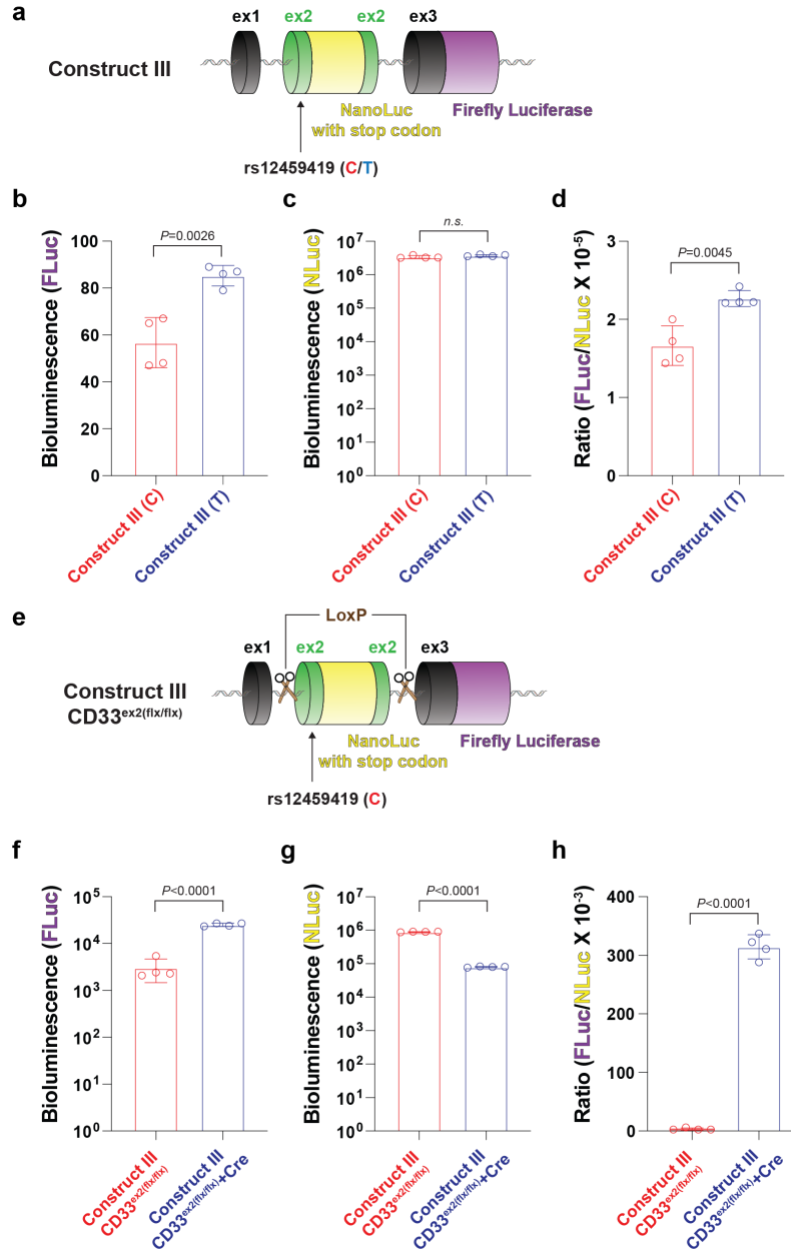


Figure 4.13 Validation of splicing reporter Construct III. (a) DNA elements of Construct III are shown. A NLuc gene was embedded in the middle of exon-2 and a FLuc gene was recombined at the end of exon-3. To keep splicing junction sequence for exon-2 exclusion, the first and last 20 nucleotide sequences of exon-2 were retained. (b-d) bioluminescent signals from (b) FLuc and (d) NLuc, and (d) the values of FLuc/NLuc for Construct III between C or T allele were shown. (e) Construct III CD33^{ex2(flx/flx)} was made by inserting two LoxP sites to intron-1 and -2 on Construct III. After Cre recombination, (f) FLuc and (g) NLuc signals, and (h) the values of FLuc/NLuc were compared between a control and Cre recombination. Statistical significance was calculated using a two-tailed unpaired Student's *t* test.

Construct III that would be expected based on 100% exon-2 exclusion, two LoxP sites were inserted to both intron-1 and -2 for Cre recombination. After Cre recombination, NLuc signal is expected to go down and FLuc signal should increase. As expected, Cre recombination led to both decreased NLuc signal and increased FLuc signal. **(Figure 4.13e-f)** The combined outcome of these two changes led to a FLuc/NLuc ratio change of 92-fold. Compared to the maximum 10-fold difference in Construct I, complete exon-2 skipping in Construct III yielded much higher signal sensitivity, which may be beneficial for screening small molecule drug candidates.

4.4 Discussion

A challenge in studying hCD33m has been the lack of biochemical tools to detect it. In parallel with our efforts, several other groups published monoclonal antibodies targeting hCD33m but these have not been made commercially available.^{155,160} The hCD33m-specific monoclonal antibody S503 developed in this work (clone S503) was shown to detect cell-surface hCD33m under overexpressed conditions. However, it still remains difficult to detect endogenous cell surface hCD33m expression in U937 cells, as well as cell surface expression of hCD33m in transgenic mouse microglia that do not represent overexpressed conditions¹⁵⁵ This is likely due to the preference of hCD33m for an intercellular localization. It is possible that Cys42 in hCD33m is disulfide bonded with another hCD33m or other proteins, and this blocked the S503 epitope. Consistent with this, C42A hCD33m had significantly increased S503 staining. However, unlike hCD33M, hCD33m did not show increased staining with S503 upon DTT treatment, suggesting that this may not be the same. Unfortunately, attempts to visualize hCD33m by microscopy with S503 have, to date, not been successful. We are currently testing if S823 has more versatility than S503 in different applications.

For this reason, the 3xFLAG-tag was introduced at the C-terminal of various CD33 constructs. Results with the 3xFLAG-tag system clearly demonstrated that the unpaired Cys42 is a key factor affecting the intracellular localization of hCD33m. In the future, the mechanism how Cys42 keeps hCD33m inside the cells needs to be elucidated. One prediction and testable hypothesis is that WT and C42A hCD33m may undergo different *N*-glycosylation processing, which could be easily identified. It is predicted that WT hCD33m has high-mannose type *N*-glycans, due to not passing through the entire secretory pathway, while C42A hCD33m has complex type *N*-glycans. As an extension, WT and C42A CD33m would be predicted to have differential sensitivity to Endo-H due to its specific for high mannose *N*-glycans.

To develop a system for directly assessing the expression levels of hCD33m without the confounding effects of hCD33M expression, a *CD33* minigene construct was used that contains the two key introns flanking exon-2 with the rs12459419 SNP (C or T) and two stop codons were introduced within exon-2. Consistent with a previous study,¹⁵⁶ the *CD33* minigene with C allele had lower levels of exon-2 exclusion than the T allele, leading to a lower expression level of hCD33m.⁷⁰ In the future, applying this different hCD33m expression system to WT U937 cells, instead of CD33^{-/-} U937 cells, will be a good strategy to assess the effect of hCD33m in the presence of endogenous hCD33M expression. It is noteworthy that in our work using transgenic mouse microglia, my colleague showed that the gain-of-function effect for hCD33m appeared to be dominant, since phagocytosis was enhanced to the same degree in microglia that expressed hCD33m compared to those that expressed both isoforms.¹³⁵

Cells expressing the *CD33* minigenes with either of the two alleles showed altered phagocytosis of A β ₁₋₄₂ aggregates. As the *CD33* T allele is AD-protective, the combined outcomes of lower expression of hCD33M, representing loss-of-function, and higher expression of hCD33m,

representing gain-of-function, may be the key factors for AD protection. Ongoing experiments in our laboratory are tackling this question in the context of a mouse model of AD and should provide clarity. These findings motivated the development of CD33 splicing reporter systems with the goal of using them in the future to screen for drugs capable of inducing exon-2 skipping. Construct I was conceptually very similar to the reporter system used and published previously by Pfizer. However, it was first shown that Construct I with the T allele had 3.5-fold higher a NLuc bioluminescent signal than the C allele. In the previous study, SRSF1 and PTBP1 were discovered as splicing factors involved in exon-2 inclusion. Interfering with SRSF1 gene expression led to increased hCD33m expression at mRNA level, which was reproduced when MOE-ASO blocked a specific site on exon-2 of *CD33*. In this regard, Construct III needs more improvements since the MOE-ASO binding site is around 80 nucleotides away from 3' end of exon-2 and Construct III only contains about 20 nucleotide junction sequences at 5' and 3' end of exon-2. Hence, additional nucleotides from 3' end of exon-2 covering the MOE-ASO acting sites that Pfizer validated can be added to Construct III. An issue in Construct III is that a FLuc signal is much less sensitive than a NLuc signal, so making FLuc represent exon-2 skipping is less optimal. It was expected that the long FLuc DNA sequence might have a negative effect on exon-2 skipping. However, based on results with Construct II, we now know that a longer DNA sequence can be inserted to exon-2. Therefore, efforts are underway to design Construct IV, which has the position of NLuc and FLuc swapped.

A concern during the development of CD33^{ex2(fl_x/fl_x)} system was that splicing factors involving in CD33 transcription may fail to recognize exon-2 after Cre recombination. This is because each LoxP site was embedded in the middle of intron-1, and -2, so residual nucleotides in both introns may not be spliced out. However, this was not the case as hCD33M could be

successfully turned into hCD33m by Cre recombination. In the future, the CD33^{ex2(fl/x/fl/x)} system can be combined with a Cre-ER^{T2} system in both cells and mice, where Tamoxifen leads to translocation of Cre-ER^{T2} from cytoplasm to nucleus, enabling Cre recombination. (**Figure 4.14**) This system will help to understand the benefits of inducing exon-2 skipping in the context of AD mouse model.

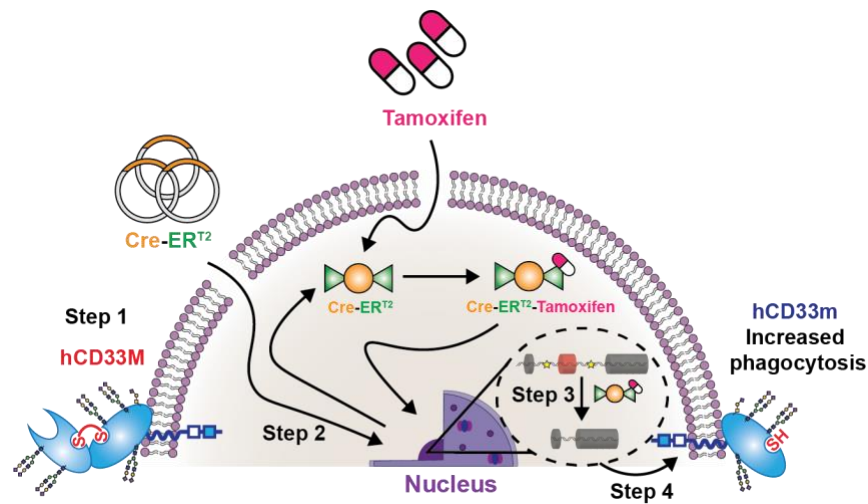


Figure 4.14 An inducible system for turning hCD33M to hCD33m. Cells transduced with CD33^{ex2(fl/x/fl/x)} express hCD33M on the surface. (*Step 1*) Transduction of Cre-ER^{T2} generates an environment where Tamoxifen can induce Cre recombination. (*Step 2*) Tamoxifen translocate Cre-ER^{T2} from cytoplasm to nucleus where Cre recombination takes place. (*Step 3*) The expression of hCD33M turns into the expression of hCD33m. (*Step 4*)

4.5 Conclusions

In this Chapter, *CD33* minigene constructs were used to recapitulate rs12459419. To detect hCD33m expression on the surface, two anti-hCD33m antibodies were developed. For the *CD33* minigene construct with the risk C allele, the expression level of hCD33M was higher than the protective T allele while the T allele has higher hCD33m expression. This phenotypical difference was shown to have altered phagocytosis ability in U937 cells, where the *CD33* minigene with the T allele had higher phagocytosis toward A β ₁₋₄₂ aggregates. To dissect a molecular mechanism how

hCD33m correlates with AD protection phenotype, its localization was studied, and it was discovered that Cys42 on hCD33m maintains its intracellular localization. Finally, new CD33 splicing reporter systems were developed. Moreover, to simulate complete exon-2 exclusion, an artificial construct with two loxP sites in the introns flanking exon-2 was created and validated.

4.6 Methods

4.6.1 Generation of hCD33m-specific hybridomas

The gene encoding hCD33m (amino acids 1–132) was cloned into pcDNA5/FRT/V5-His-TOPO® vector (Thermo Fisher Scientific), which contained a C-terminal hIgG1-Fc, TEV cleavage site on the N-terminal side of the Fc, as well as a C-terminal His₆ tag and a Strep-tag II.¹⁰⁶ This construct was co-transfected with a plasmid pOG44, encoding the Flp recombinase, into Chinese Ovary Hamster (CHO) Flp-In cells using the manufacturer's recommended protocol. Cells were grown in DMEM-F12 media containing 5% fetal bovine serum (FBS, Gibco), 1.0 mg/mL Hygromycin B, 100 U/mL of Penicillin & Streptomycin (P/S, Gibco) and 2.438 g/L Sodium Bicarbonate (Gibco). Selection with Hygromycin B (Thermo Fisher) took place for 10 days starting at 2 days post-transfection. For expression, one million hCD33m-Fc expressing CHO cells were added to a T175 flask with 50 mL of media (DMEM-F12, 5% FBS, 100 U/mL P/S, and 10 mM HEPES) and cultured for 12 days (10 days post-confluence). The supernatant was filtered through a sterile filtered unit (0.2 µm pore size, Fisher Scientific). The filtered supernatant was loaded into a 5 mL HisTrap excel column (GE healthcare) pre-equilibrated with 50 mL of equilibration buffer (20 mM Sodium Phosphate, 500 mM NaCl, pH 7.4). After washing the column with 75 mL of washing buffer (20 mM Sodium Phosphate, 500 mM NaCl, 20 mM Imidazole, pH 7.4), bound protein was eluted with 30 mL elution buffer (20 mM Sodium Phosphate, 500 mM NaCl, 500 mM Imidazole, pH 7.4) in 2.5 mL fractions. Eluted fractions containing hCD33m-Fc

were pooled, diluted 10-fold in 20 mM sodium phosphate buffer (pH 7.2), and loaded into a Protein-G column (GE healthcare) pre-equilibrated with the same phosphate buffer. hCD33m-Fc was eluted with 0.1 M Glycine buffer (pH 2.7) directly in 1 mL fractions into microcentrifuge tubes containing 30 μ L 1 M Tris-HCl (pH 8). The fractions containing protein were pooled and digested with 20 μ L of 4 mg/mL His₆-tagged TEV protease per 1 mL of 1 mg/mL protein solution at 37 °C for 3 h in a shaking incubator (150 rpm). The resultant reaction was put through a 1 mL HisTrap column to remove the bound Fc and TEV, while the hCD33m fragment came through in the flow through. The purified hCD33m was dialyzed three times into PBS and concentrated to 1 mg/mL using a 10 kDa cutoff Amicon ultra centrifugal filter unit (Millipore Sigma). In total, 4 mg of the CD33m fragment was prepared. An additional 2 mg of hCD33m-Fc was also saved for screening. The 4 mg of hCD33m and 2 mg of hCD33m-Fc were used for mouse immunization and monoclonal development (Thermo Fisher Scientific Pierce Proteomics). Supernatant from polyclonal hybridomas that were positive by ELISA for hCD33m-Fc were screened against CHO cells expressing hCD33M or hCD33m by flow cytometry using a PE-labelled anti-mouse IgG secondary antibody. Two hybridoma stocks uniquely stained hCD33m-expressing cells but not hCD33M-expressing cells. These two frozen hybridomas were thawed, expanded, and subcloned in RPMI media containing 20% FBS and 100 U/mL P/S by diluting to 1 cell per 3 wells in 96-well plates containing 220 μ L of media. After 10 days, 50 μ L of the supernatant from each well was screened against hCD33M-expressing, CD33m-expressing, and WT CHO cells by flow cytometry. A monoclonal hybridoma, designated S503, that produced antibody specific for hCD33m-expressing CHO cells was isolated.

4.6.2 Expression, labelling, and isotyping of hCD33m-specific antibodies

For antibody expression, one million hybridoma cells were added to a T175 flask containing 50 mL of RPMI containing 10% FBS, 100 U/mL P/S, and 10 mM HEPES and cultured for five days after post-confluence. Each supernatant was filtered (0.2 μ m pore size) and stored at 4 °C until ready for purification. Filtered supernatant (100 mL) was loaded onto a 1 mL Protein G column pre-equilibrated with 20 mM sodium phosphate buffer (pH 7.4). After washing the column with the 20 mM sodium phosphate buffer, anti-CD33m was eluted with 0.1 M glycine buffer (pH 2.7) in 1 mL fractions, directly into tubes containing 30 μ L of 1 M Tris-base buffer (pH 8). The eluent was dialyzed three times against phosphate buffered saline (pH 7.4) and stored at 4 °C. For fluorophore labeling, antibody was concentrated to 0.2 mg/mL. Prior to fluorophore labelling, 50 μ L of ammonium bicarbonate (1 M) was added to 950 μ L of the 0.2 mg/ml solution of antibody solution to increase the pH, followed by the addition of 6.64 μ L of 2.5 mM NHS-AF647 (Thermo Fisher Scientific), dissolved in DMSO, to achieve a 10-fold molar excess. The reaction was gently rocked for 1 h at 37 °C and immediately dialyzed three times in PBS. Mouse IgG1 isotype antibody (Thermo Fisher Scientific) was pre-dialyzed into the same PBS prior to labeling and then fluorophore labeled at the same time in order to achieve the same level of fluorophore modification. The ratio of absorption at A_{280} to A_{647} confirmed an antibody labelling ratio of appropriate 2.5 molecules of AF647 per antibody.

4.6.3 Isolation of adult mouse microglia

Adult mice were euthanized under CO₂. Media used for the primary microglia consisted of RPMI with 10% FBS, 100 U/mL Penicillin and 100 μ g/ml Streptomycin. Isolated brain samples were homogenized by 5 ml syringe plungers in media through 40 μ m corning filter units under sterile condition. Homogenized samples were centrifuged at 500 rcf for 5 min and the pellet was treated

with 3 ml of red blood cell lysis buffer (150 mM NH₄Cl, 9 mM NaHCO₃, and 0.1 mM EDTA). Following centrifugation at 300 rcf for 5 min, the pellet was dissolved in 3 ml of 30% Percoll (Percoll PLUS, GE Healthcare) and carefully layered on top of 70% Percoll and immediately centrifuged (650 rcf for 20 min). Immune cells were isolated from the border between the two layers, washed (300 rcf, 5 min) and resuspended in media.

4.6.4 Mutagenesis on C169A for hCD33M and C42A for hCD33m

DNA templates encoding hCD33M and hCD33m was amplified with a pair of forward primer (agcagcgctagcatgccgctgctgctactgctg) and reverse primer (cggggggtgtccctgctctgcggcccagga cacag). Subsequently, PCR products were gel-purified, and used for a mega forward primer with a reverse primer (AGCAGCACCGGTTCACTGGGTCCTGACCTCTG) to amplify the two DNA templates encoding hCD33M and hCD33m. The forward primer at the first PCR and the reverse primer at the second PCR were designed for containing 5' NheI and 3' AgeI sites, which were used for ligation.

4.6.5 Cloning of RP172 vectors

CD33 minigene constructs, Construct I, II, and III with C, T, and CD33^{ex2(fl/x/flx)} were synthesized from GeneArt Gene Synthesis (Thermo Fisher Scientific). (Appendix A) The constructs were designed to have 5' NheI and 3' AgeI sites when they were synthesized in commercial vectors. Subsequently, the *CD33* minigene constructs, Construct I, III, and III were ligated to a RP172 vector with 5' NheI and 3' AgeI restriction sites. The ligated vectors were transformed to NEB[®] Stable competent *E. coli* and promising colonies were selected on LB agar plates containing 100 µg/mL ampicillin. After culturing colonies in LB growth media containing 100 µg/mL Ampicillin, cells from each colony were miniprepmed, and sent for Sanger Sequencing.

4.6.6 Lentivirus production

Lentivirus was produced and titered with the following changes: 1 million HEK293T cells were plated in a 6-well plate on day 0. On Day 1, a transfection mix was made containing 625 ng RP18, 625 ng RP19, 1250 ng of transfer vector, 7.5 μ l TransIT®-LT1 Reagent (Mirus Bio), and 258 μ l Opti-MEM media (Gibco) and added to the HEK293T cells. 72 h post transfection, supernatant was harvested and concentrated using Lenti-X concentrator reagent (Takara Bio) as per manufacturer's instructions.

4.6.7 Lentiviral transduction of U937 cells

A quarter million U937 cells were plated in a 12 well plate in 500 μ l growth media. Lentivirus was added to each well at a MOI = 1 and incubated in a tissue culture incubator. 16 h post-transduction, media was topped up to 2 ml. 72 h following transduction, 300 μ g/ml zeocin was added to each well as a selection agent to kill untransduced cells. Cells were monitored daily for cell death, and every 2 days the media was changed and re-supplemented with 300 μ g/ml zeocin. After 1 week, cells were assayed on a flow cytometer to assess the purity of the selected population via the fluorescent marker mAmetrine. Cells were found to be > 97% mAmetrine positive after 1 week of zeocin selection. Phagocytosis assays were done 3 days after removing zeocin from the media.

4.6.8 Flow cytometry

All flow cytometry data was collected on a 5-laser Fortessa X-20 (BD Bioscience) and analyzed using FlowJo software (BD Biosciences). Cell sorting took place on an Aria III (BD Bioscience).

4.6.9 Phagocytosis assessment in U937 cells

U937 cells were grown up to $\sim 1 \times 10^6$ cells/mL density in a T75 flask and subsequently 100,000 cells were added to 96-well U-bottom plates in 100 μ l of media/well. To initiate the phagocytosis

assay, 100 μ L of media containing fluorescent A β ₁₋₄₂ (400 nM) or polystyrene beads (1200 dilution from the vendor stock) was added to the cells and incubated for 30 min at 37 °C. After incubation period cells were washed with media twice and resuspended in flow buffer containing 1 μ g/mL propidium iodide. Samples were analyzed by flow cytometry. Median fluorescence intensity (MFI) was used to assess extent of phagocytosis in A β ₁₋₄₂ treated samples while % of cells taking up at least one bead was counted for polystyrene beads. Cytochalasin-D (10 μ M) pre-treated samples were also carried out in parallel and subtract for fluorescence not associated with phagocytosis.

4.6.10 Intracellular staining

U937 cells were plated into a 96-well U-bottom plate and washed twice with PBS (300 rcf, 5 min, 4 °C). The cells were labelled with Zombie Red™ Fixable Viability Dye in PBS (1:3000 dilution, Biolegend) at RT for 20 min in a dark room. After washing with PBS containing 5% FBS twice (300 rcf, 5 min, 4 °C), permeabilization and fixation were done together using a commercial Fixation/Permeabilization Solution Kit (BD Bioscience) on ice for 20 min. The fixed and permeabilized cells were washed twice with 1x BD Perm/Wash™ Buffer (2,500 rcf, 5 min, 4 °C) and stained with 200-fold diluted anti-FLAG tag antibody (Clone: M2) in flow buffer at 4 °C for 20 min. After washing twice with flow buffer, the cells were stained with secondary anti-mouse IgG1 antibody with AF647 on ice for 20 min. Before the flow assay, each well was washed with flow buffer two times (2,500 rcf, 5 min, 4 °C) and transferred to flow tubes.

4.6.11 Nano-Glo® Luciferase Assay for Construct I

C4 U937 cells expressing Construct I_C or T were plated onto a white opaque 96-well microplate in 100 μ L media, which contained 100,000 cells per well. While the cells were incubated at RT for 5 mins, 20 μ L of Nano-Glo® Luciferase Assay Substrate (Promega) was added to 1 mL Nano-

Glo® Luciferase Assay Buffer (Promega). Subsequently, 100 µL of Nano-Glo® Luciferase Assay Buffer containing bioluminescence substrates was added to each well. After incubation for 10 min, the plate was loaded onto SpectraMax® ID5 plate reader and bioluminescence signal per well was recorded.

4.6.12 Nano-Glo® Dual-Luciferase Assay for Construct II

C4 U937 cells expressing Construct I_C or T were plated onto a white opaque 96-well microplate in 80 µL media, which contained 100,000 cells per well. Subsequently, ONE-Glo™ EX Reagent (Promega) was added to each well with equal volume to media and incubated on an orbital shaker (300 rpm) for minimum 3 min at RT. The plate was loaded onto SpectraMax® ID5 plate reader and firefly bioluminescence signal per well was recorded. Next, 80 µL NanoDLR™ Stop & Glo® Reagent (Promega) was added to each well and the reagents were incubated on an orbital shaker (600 rpm) for at least 3 min at RT. After additional 7-minute incubation on a bench, Nanoluc bioluminescence signal was measured.

4.6.13 Statistical analyses

For experiments with only two groups, a two-tailed Student's *t*-test was used to access statistical significance. When experiments were carried out in competition, a paired Student's *t*-test was used, while an unpaired Student's *t*-test was used in cases where experiments groups were analyzed independently. For experiments with more than two groups, a one-way ANOVA was carried out to determine if there were differences between the means. If applicable, the Tukey multiple comparisons post-test was then used to indicate significance between groups. N.S. means no statistical significance ($P > 0.05$).

Chapter 5

Conclusions and Future Directions

CD33 encodes two protein isoforms mediated by alternative splicing.^{70,79,81} rs12459419 is a SNP in *CD33* influencing the ratio of both *CD33* isoforms, which influences LOAD susceptibility.¹⁵¹ In this thesis, to better understand roles of hCD33M in the brain, the glycan ligands of hCD33M were identified through ESI-MS and flow cytometry-based approaches. Furthermore, through developing biochemical tools, we discovered how major and minor alleles of rs12459419 affect the expression level of the two *CD33* isoforms, which can affect cellular phagocytosis.

In Chapter 2, a new generation of CD33-Fc was developed and validated with flow cytometry and ESI-MS-based approaches in collaboration with the Klassen group.¹⁰⁶ For flow cytometry applications, a C-terminal Strep-tag II in CD33-Fc helped to form octameric presentation to facilitate high avidity. Through increasing the avidity, it was discovered that both α 2-3 and α 2-6 sialosides can be *trans* ligands for CD33 on U937 cells. For ESI-MS-based binding assay to determine K_d values, an Endo H sensitive glycoform of CD33-Fc was expressed Lec-1 CHO cells, which enabled a fragment of CD33 to be generated with a minimal number of glycoforms. In addition, an internal TEV recognition site was used to separate the CD33 fragments from the Siglec-Fc fusion protein scaffold. These two enzymatic reactions made the CD33 fragments suitable for determining dissociation constants by ESI-MS. Subsequently, the K_d values of the CD33 fragments towards 3'SLN and 6'SLN were determined. Of note, K_d values of CD33 for 3'SLN and 6'SLN were similar to each other, which indicates CD33 recognizes both α 2-3 and α 2-6 linkages. This was consistent with the results from the cell-based assay. For the CaR-ESI-MS application, a natural *N*-glycan library from U937 cells was generated for library screening with CD33-Fc. CaR-ESI-MS screening with CD33-Fc and the natural *N*-glycan library yielded 6 *N*-glycan compositions. Of note, composition representing disialylated biantennary *N*-glycans

were the most abundant ligand candidates for CD33-Fc. Consequently, through this novel CD33-Fc scaffold, ligands for CD33 were characterized. CD33 interacts with both α 2-3 and α 2-6 linked sialosides, which was supported by both ESI-MS and flow cytometry-based applications. In the future, cell-derived *O*-glycan or ganglioside libraries can be screened to complement the screening we performed on *N*-glycans. Ultimately, it will be of great interest to screen, through CaR-ESI-MS, *N*- or *O*-glycans from healthy and AD human brains.

In Chapter 3, U937 cells overexpressing eight members of CHSTs were generated through lentiviral transduction and assessed for binding against all human and mouse Siglec-Fc chimeric constructs. Of note, 6-*O*-sulfation on Gal, mediated by CHST1, greatly enhanced binding of CD33, Siglec-5, -7, -8, and -15 for human and Siglec-F, Siglec-E, and Siglec-15 for mouse. On the other hand, 6-*O*-sulfation on GlcNAc significantly strengthened the binding of CD22, Siglec-7, and -9 for human, and CD22 and Siglec-E for mouse. To dissect what types of glycosylation affected the binding of each Siglec-Fc construct, the biosynthesis of *N*-glycosylation or mucin type *O*-glycosylation were pharmaceutically perturbed. It was consistent that CD22 was sensitive to kifunensine treatment while Siglec-7 binding was abrogated by benzyl- α -GalNAc. CD33, Siglec-5/14, Siglec-8 and Siglec-15 showed sensitivity to both kifunensine and benzyl- α -GalNAc, which may indicate that these Siglecs can recognize sulfated *N*- and *O*-glycans. *CMAS*^{-/-} U937 cells overexpressing *CHST1* or *CHST2* were tested with Siglec-Fc and it was demonstrated that sialic acid is still essential for Siglec binding even though sulfation can enhance the interactions between Siglecs and sulfated sialosides. A novel *O*-glycosylation site on CD33 was discovered and mutated to facilitate ESI-MS-based K_d determination. Of note, disulfated 3'SLN was the highest affinity ligand for CD33, whose binding affinity was 28-fold stronger than 3'SLN. Consistent with these results, U937 cells expressing both CHST1 and CHST2 were shown to upregulate disulfated

3'SLN leading to more enhanced binding of CD33, Siglec-5/14, Siglec-7, Siglec-8, and Siglec-15 than U937 cells overexpressing CHST1 only. Sodium chlorate as an endogenous sulfation inhibitor was validated and was used for testing the effects of endogenous sulfation on several breast cancer and glioma cell lines. Notably, it was found that Siglec binding was affected by NaClO₃ treatments in several cancer cell lines, which supports that cancer cells can also take advantage of carbohydrate sulfation to enhance interactions with Siglecs. Consequently, this chapter supports carbohydrate sulfation as important biosynthetic mechanisms fine-tuning glycan ligands for Siglecs. Most recently, receptor protein tyrosine phosphatase zeta (RTP ζ) carrying sialylated keratan sulfates was identified as a CD33 ligand in the brain.¹⁴⁸ In the future, CHST1 U937 cells additionally expressing CHST2, CHST4, CHST5, or CHST6 will be assessed with several anti-keratan sulfate antibodies to confirm keratan sulfate expression levels. Subsequently, CD33 binding to those cell lines will be evaluated in conjunction with keratanase treatment to determine if sialylated keratan sulfates on the surface accounts for a majority of CD33 binding.

In Chapter 4, a developed anti-hCD33m monoclonal antibody (S503) helped to characterize levels of hCD33m expression in cells expressing a rs12459419C and T *CD33* allele. To test the difference between the C and T allele, a *CD33* minigene constructs were made and validated. Higher expression of hCD33m were observed in the T allele, which was first demonstrated at the protein level. Furthermore, phagocytosis assays supported that U937 cells expressing the *CD33* minigene with T allele were more phagocytic toward A β ₁₋₄₂ aggregates compared to cells expressing the *CD33* minigene with C allele. This finding supports a growing body of research showing that hCD33m is a gain-of-function protein.^{135,153} As hCD33m appears to prefer an intracellular localization, the effect of Cys42 on hCD33m localization was investigated. It was found that C42A hCD33m had higher cell-surface expression levels than WT

hCD33m, supporting the concept that Cys42 on hCD33m might be important in retaining hCD33m intracellularly. Given the beneficial roles of hCD33m, a version of CD33, called *CD33^{ex2(flx/flx)}*, was devised to turn hCD33M into hCD33m by Cre recombination. In the future, this *CD33^{ex2(flx/flx)}* construct can be applied in transgenic mice to induce isoform switching of hCD33M to hCD33m at specific timepoints within an AD mouse model, which can help to investigate different roles of both isoforms in AD pathology. Finally, to discover small molecules that can induce exon-2 skipping, three CD33 splicing reporter systems were developed and validated. These culminated in Construct III that leverages a dual luciferase system where one signal goes up and another goes down upon induction of exon-2 exclusion. In the future, such a ratiometric reporter should prove valuable in screening for small molecules capable of inducing exon-2 exclusion, which will have potential to treat AD.

References

- (1) Rosenblum, M. D.; Remedios, K. A.; Abbas, A. K. Mechanisms of human autoimmunity. *J Clin Invest* **2015**, *125* (6), 2228-2233. DOI: 10.1172/JCI78088 From NLM Medline.
- (2) Macauley, M. S.; Crocker, P. R.; Paulson, J. C. Siglec-mediated regulation of immune cell function in disease. *Nat Rev Immunol* **2014**, *14* (10), 653-666. DOI: 10.1038/nri3737 From NLM Medline.
- (3) Zou, Z.; Chastain, A.; Moir, S.; Ford, J.; Trandem, K.; Martinelli, E.; Cicala, C.; Crocker, P.; Arthos, J.; Sun, P. D. Siglecs facilitate HIV-1 infection of macrophages through adhesion with viral sialic acids. *PLoS One* **2011**, *6* (9), e24559. DOI: 10.1371/journal.pone.0024559 From NLM Medline.
- (4) Muller, J.; Lunz, B.; Schwab, I.; Acs, A.; Nimmerjahn, F.; Daniel, C.; Nitschke, L. Siglec-G Deficiency Leads to Autoimmunity in Aging C57BL/6 Mice. *J Immunol* **2015**, *195* (1), 51-60. DOI: 10.4049/jimmunol.1403139 From NLM Medline.
- (5) Varki, A. Since there are PAMPs and DAMPs, there must be SAMPs? Glycan "self-associated molecular patterns" dampen innate immunity, but pathogens can mimic them. *Glycobiology* **2011**, *21* (9), 1121-1124. DOI: 10.1093/glycob/cwr087 From NLM Medline.
- (6) Li, L.; Chen, Y.; Sluter, M. N.; Hou, R.; Hao, J.; Wu, Y.; Chen, G. Y.; Yu, Y.; Jiang, J. Ablation of Siglec-E augments brain inflammation and ischemic injury. *J Neuroinflammation* **2022**, *19* (1), 191. DOI: 10.1186/s12974-022-02556-1 From NLM Medline.
- (7) Duan, S.; Paulson, J. C. Siglecs as Immune Cell Checkpoints in Disease. *Annu Rev Immunol* **2020**, *38*, 365-395. DOI: 10.1146/annurev-immunol-102419-035900 From NLM Medline.
- (8) Movsisyan, L. D.; Macauley, M. S. Structural advances of Siglecs: insight into synthetic glycan ligands for immunomodulation. *Org Biomol Chem* **2020**, *18* (30), 5784-5797. DOI: 10.1039/d0ob01116a From NLM Medline.
- (9) Saunderson, S. C.; Dunn, A. C.; Crocker, P. R.; McLellan, A. D. CD169 mediates the capture of exosomes in spleen and lymph node. *Blood* **2014**, *123* (2), 208-216. DOI: 10.1182/blood-2013-03-489732 From NLM Medline.
- (10) Sun, S.; Zhang, H. Identification and Validation of Atypical N-Glycosylation Sites. *Anal Chem* **2015**, *87* (24), 11948-11951. DOI: 10.1021/acs.analchem.5b03886 From NLM Medline.
- (11) Klaas, M.; Dubock, S.; Ferguson, D. J. P.; Crocker, P. R. Sialoadhesin (CD169/Siglec-1) is an extended molecule that escapes inhibitory cis-interactions and synergizes with other macrophage receptors to promote phagocytosis. *Glycoconj J* **2023**. DOI: 10.1007/s10719-022-10097-1 From NLM Publisher.
- (12) Paul R Crocker, J. C. P., Ajit Varki. Siglecs and their roles in the immune system . *Nature Reviews Immunology* **2007**, *7*. DOI: <https://doi.org/10.1038/nri2056>.
- (13) van de Weijer, M. L.; Bassik, M. C.; Luteijn, R. D.; Voorburg, C. M.; Lohuis, M. A.; Kremmer, E.; Hoebe, R. C.; LeProust, E. M.; Chen, S.; Hoelen, H.; et al. A high-coverage shRNA screen identifies TMEM129 as an E3 ligase involved in ER-associated protein degradation. *Nat Commun* **2014**, *5*, 3832. DOI: 10.1038/ncomms4832 From NLM Medline.
- (14) Angata, T. Siglecs that Associate with DAP12. *Adv Exp Med Biol* **2020**, *1204*, 215-230. DOI: 10.1007/978-981-15-1580-4_9 From NLM Medline.
- (15) Harduin-Lepers, A.; Vallejo-Ruiz, V.; Krzewinski-Recchi, M. A.; Samyn-Petit, B.; Julien, S.; Delannoy, P. The human sialyltransferase family. *Biochimie* **2001**, *83* (8), 727-737. DOI: 10.1016/s0300-9084(01)01301-3 From NLM Medline.

- (16) Bowles, W. H. D.; Gloster, T. M. Sialidase and Sialyltransferase Inhibitors: Targeting Pathogenicity and Disease. *Front Mol Biosci* **2021**, *8*, 705133. DOI: 10.3389/fmolb.2021.705133 From NLM PubMed-not-MEDLINE.
- (17) Park, S. S. Post-Glycosylation Modification of Sialic Acid and Its Role in Virus Pathogenesis. *Vaccines (Basel)* **2019**, *7* (4). DOI: 10.3390/vaccines7040171 From NLM PubMed-not-MEDLINE.
- (18) Brinkman-Van der Linden, E. C.; Varki, A. New aspects of siglec binding specificities, including the significance of fucosylation and of the sialyl-Tn epitope. Sialic acid-binding immunoglobulin superfamily lectins. *J Biol Chem* **2000**, *275* (12), 8625-8632. DOI: 10.1074/jbc.275.12.8625 From NLM Medline.
- (19) Khatua, B.; Bhattacharya, K.; Mandal, C. Sialoglycoproteins adsorbed by *Pseudomonas aeruginosa* facilitate their survival by impeding neutrophil extracellular trap through siglec-9. *J Leukoc Biol* **2012**, *91* (4), 641-655. DOI: 10.1189/jlb.0511260 From NLM Medline.
- (20) Lim, J.; Sari-Ak, D.; Bagga, T. Siglecs as Therapeutic Targets in Cancer. *Biology (Basel)* **2021**, *10* (11). DOI: 10.3390/biology10111178 From NLM PubMed-not-MEDLINE.
- (21) Rodrigues, E.; Macauley, M. S. Hypersialylation in Cancer: Modulation of Inflammation and Therapeutic Opportunities. *Cancers (Basel)* **2018**, *10* (6). DOI: 10.3390/cancers10060207 From NLM PubMed-not-MEDLINE.
- (22) Johnson, D. B.; Nebhan, C. A.; Moslehi, J. J.; Balko, J. M. Immune-checkpoint inhibitors: long-term implications of toxicity. *Nat Rev Clin Oncol* **2022**, *19* (4), 254-267. DOI: 10.1038/s41571-022-00600-w From NLM Medline.
- (23) Meng, C.; Sasmal, A.; Zhang, Y.; Gao, T.; Liu, C. C.; Khan, N.; Varki, A.; Wang, F.; Cao, H. Chemoenzymatic Assembly of Mammalian O-Mannose Glycans. *Angew Chem Int Ed Engl* **2018**, *57* (29), 9003-9007. DOI: 10.1002/anie.201804373 From NLM Medline.
- (24) Wang, S.; Chen, C.; Guan, M.; Liu, D.; Wan, X. F.; Li, L. Terminal Epitope-Dependent Branch Preference of Siglecs Toward N-Glycans. *Front Mol Biosci* **2021**, *8*, 645999. DOI: 10.3389/fmolb.2021.645999 From NLM PubMed-not-MEDLINE.
- (25) Blixt, O.; Head, S.; Mondala, T.; Scanlan, C.; Huflejt, M. E.; Alvarez, R.; Bryan, M. C.; Fazio, F.; Calarese, D.; Stevens, J.; et al. Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc Natl Acad Sci U S A* **2004**, *101* (49), 17033-17038. DOI: 10.1073/pnas.0407902101 From NLM Medline.
- (26) Yu, H.; Gonzalez-Gil, A.; Wei, Y.; Fernandes, S. M.; Porell, R. N.; Vajn, K.; Paulson, J. C.; Nycholat, C. M.; Schnaar, R. L. Siglec-8 and Siglec-9 binding specificities and endogenous airway ligand distributions and properties. *Glycobiology* **2017**, *27* (7), 657-668. DOI: 10.1093/glycob/cwx026 From NLM Medline.
- (27) Muller, J.; Obermeier, I.; Wohner, M.; Brandl, C.; Mrotzek, S.; Angermuller, S.; Maity, P. C.; Reth, M.; Nitschke, L. CD22 ligand-binding and signaling domains reciprocally regulate B-cell Ca²⁺ signaling. *Proc Natl Acad Sci U S A* **2013**, *110* (30), 12402-12407. DOI: 10.1073/pnas.1304888110 From NLM Medline.
- (28) Sylvie D.Freeman, S. r., Elizabeth K.Barber, Paul R.Crocker. Characterization of CD33 as a New Member of the Sialoadhesin Family of Cellular Interaction Molecules. *blood* **1995**, *85* (8), 2005-2012. DOI: 10.1182/blood.V85.8.2005.bloodjournal8582005.
- (29) Crocker, P. R.; Varki, A. Siglecs in the immune system. *Immunology* **2001**, *103* (2), 137-145. DOI: 10.1046/j.0019-2805.2001.01241.x From NLM Medline.

- (30) Courtney, A. H.; Puffer, E. B.; Pontrello, J. K.; Yang, Z. Q.; Kiessling, L. L. Sialylated multivalent antigens engage CD22 in trans and inhibit B cell activation. *Proc Natl Acad Sci U S A* **2009**, *106* (8), 2500-2505. DOI: 10.1073/pnas.0807207106 From NLM Medline.
- (31) Gray, M. A.; Stanczak, M. A.; Mantuano, N. R.; Xiao, H.; Pijnenborg, J. F. A.; Malaker, S. A.; Miller, C. L.; Weidenbacher, P. A.; Tanzo, J. T.; Ahn, G.; et al. Targeted glycan degradation potentiates the anticancer immune response in vivo. *Nat Chem Biol* **2020**, *16* (12), 1376-1384. DOI: 10.1038/s41589-020-0622-x From NLM Medline.
- (32) Kayvon Pedram, D. J. S., Gabrielle S. Tender, Natalia R. Mantuano, Jason J. Northey, Kevin J. Metcalf, Simon P. Wisnovsky, Nicholas M. Riley, Giovanni C. Forcina, Stacy A. Malaker, Angel Kuo, Benson M. George, Caitlyn L. Miller, Kerriann M. Casey, José G. Vilches-Moure, Valerie M. Weaver, Heinz Laübli, Carolyn R. Bertozzi. Design of a mucin-selective protease for targeted degradation of cancer-associated mucins. *BioRxiv* **2022**. DOI: 10.1101/2022.05.20.492748
- bioRxiv.
- (33) Carpenter, E. P.; Beis, K.; Cameron, A. D.; Iwata, S. Overcoming the challenges of membrane protein crystallography. *Curr Opin Struct Biol* **2008**, *18* (5), 581-586. DOI: 10.1016/j.sbi.2008.07.001 From NLM Medline.
- (34) Martin, J.; Sawyer, A. Tech news. *Biotechniques* **2019**, *66* (4), 167-170. DOI: 10.2144/btn-2019-0030 From NLM Medline.
- (35) Czajkowsky, D. M.; Hu, J.; Shao, Z.; Pleass, R. J. Fc-fusion proteins: new developments and future perspectives. *EMBO Mol Med* **2012**, *4* (10), 1015-1028. DOI: 10.1002/emmm.201201379 From NLM Medline.
- (36) Levin, D.; Golding, B.; Strome, S. E.; Sauna, Z. E. Fc fusion as a platform technology: potential for modulating immunogenicity. *Trends Biotechnol* **2015**, *33* (1), 27-34. DOI: 10.1016/j.tibtech.2014.11.001 From NLM Medline.
- (37) Fishman, J. B.; Berg, E. A. Protein A and Protein G Purification of Antibodies. *Cold Spring Harb Protoc* **2019**, *2019* (1). DOI: 10.1101/pdb.prot099143 From NLM Medline.
- (38) Buchwalow, I.; Samoilova, V.; Boecker, W.; Tiemann, M. Non-specific binding of antibodies in immunohistochemistry: fallacies and facts. *Sci Rep* **2011**, *1*, 28. DOI: 10.1038/srep00028 From NLM Medline.
- (39) Gonzalez-Gil, A.; Porell, R. N.; Fernandes, S. M.; Wei, Y.; Yu, H.; Carroll, D. J.; McBride, R.; Paulson, J. C.; Tiemeyer, M.; Aoki, K.; et al. Sialylated keratan sulfate proteoglycans are Siglec-8 ligands in human airways. *Glycobiology* **2018**, *28* (10), 786-801. DOI: 10.1093/glycob/cwy057 From NLM Medline.
- (40) Gieseke, F.; Mang, P.; Viebahn, S.; Sonntag, I.; Kruchen, A.; Erbacher, A.; Pfeiffer, M.; Handgretinger, R.; Muller, I. Siglec-7 tetramers characterize B-cell subpopulations and leukemic blasts. *Eur J Immunol* **2012**, *42* (8), 2176-2186. DOI: 10.1002/eji.201142298 From NLM Medline.
- (41) Crocker, P. R.; Gordon, S. Properties and distribution of a lectin-like hemagglutinin differentially expressed by murine stromal tissue macrophages. *J Exp Med* **1986**, *164* (6), 1862-1875. DOI: 10.1084/jem.164.6.1862 From NLM Medline.
- (42) Gonzalez-Gil, A.; Schnaar, R. L. Siglec Ligands. *Cells* **2021**, *10* (5). DOI: 10.3390/cells10051260 From NLM Medline.
- (43) Macauley, M. S.; Kawasaki, N.; Peng, W.; Wang, S. H.; He, Y.; Arlian, B. M.; McBride, R.; Kannagi, R.; Khoo, K. H.; Paulson, J. C. Unmasking of CD22 Co-receptor on Germinal Center B-

cells Occurs by Alternative Mechanisms in Mouse and Man. *J Biol Chem* **2015**, 290 (50), 30066-30077. DOI: 10.1074/jbc.M115.691337 From NLM Medline.

(44) Rillahan, C. D.; Paulson, J. C. Glycan microarrays for decoding the glycome. *Annu Rev Biochem* **2011**, 80, 797-823. DOI: 10.1146/annurev-biochem-061809-152236 From NLM Medline.

(45) Park, H.; Jung, J.; Rodrigues, E.; Kitova, E. N.; Macauley, M. S.; Klassen, J. S. Mass Spectrometry-Based Shotgun Glycomics for Discovery of Natural Ligands of Glycan-Binding Proteins. *Anal Chem* **2020**, 92 (20), 14012-14020. DOI: 10.1021/acs.analchem.0c02931 From NLM Medline.

(46) Bochner, B. S.; Alvarez, R. A.; Mehta, P.; Bovin, N. V.; Blixt, O.; White, J. R.; Schnaar, R. L. Glycan array screening reveals a candidate ligand for Siglec-8. *J Biol Chem* **2005**, 280 (6), 4307-4312. DOI: 10.1074/jbc.M412378200 From NLM Medline.

(47) Campanero-Rhodes, M. A.; Childs, R. A.; Kiso, M.; Komba, S.; Le Narvor, C.; Warren, J.; Otto, D.; Crocker, P. R.; Feizi, T. Carbohydrate microarrays reveal sulphation as a modulator of siglec binding. *Biochem Biophys Res Commun* **2006**, 344 (4), 1141-1146. DOI: 10.1016/j.bbrc.2006.03.223 From NLM Medline.

(48) Padler-Karavani, V.; Hurtado-Ziola, N.; Chang, Y. C.; Sonnenburg, J. L.; Ronaghy, A.; Yu, H.; Verhagen, A.; Nizet, V.; Chen, X.; Varki, N.; et al. Rapid evolution of binding specificities and expression patterns of inhibitory CD33-related Siglecs in primates. *FASEB J* **2014**, 28 (3), 1280-1293. DOI: 10.1096/fj.13-241497 From NLM Medline.

(49) Gao, C.; Wei, M.; McKittrick, T. R.; McQuillan, A. M.; Heimbürg-Molinaro, J.; Cummings, R. D. Glycan Microarrays as Chemical Tools for Identifying Glycan Recognition by Immune Proteins. *Front Chem* **2019**, 7, 833. DOI: 10.3389/fchem.2019.00833 From NLM PubMed-not-MEDLINE.

(50) Blixt, O.; Collins, B. E.; van den Nieuwenhof, I. M.; Crocker, P. R.; Paulson, J. C. Sialoside specificity of the siglec family assessed using novel multivalent probes: identification of potent inhibitors of myelin-associated glycoprotein. *J Biol Chem* **2003**, 278 (33), 31007-31019. DOI: 10.1074/jbc.M304331200 From NLM Medline.

(51) Miyazaki, K.; Sakuma, K.; Kawamura, Y. I.; Izawa, M.; Ohmori, K.; Mitsuki, M.; Yamaji, T.; Hashimoto, Y.; Suzuki, A.; Saito, Y.; et al. Colonic epithelial cells express specific ligands for mucosal macrophage immunosuppressive receptors siglec-7 and -9. *J Immunol* **2012**, 188 (9), 4690-4700. DOI: 10.4049/jimmunol.1100605 From NLM Medline.

(52) Kimura, N.; Ohmori, K.; Miyazaki, K.; Izawa, M.; Matsuzaki, Y.; Yasuda, Y.; Takematsu, H.; Kozutsumi, Y.; Moriyama, A.; Kannagi, R. Human B-lymphocytes express alpha2-6-sialylated 6-sulfo-N-acetyllactosamine serving as a preferred ligand for CD22/Siglec-2. *J Biol Chem* **2007**, 282 (44), 32200-32207. DOI: 10.1074/jbc.M702341200 From NLM Medline.

(53) Choi, H. Y.; Park, H.; Hong, J. K.; Kim, S. D.; Kwon, J. Y.; You, S.; Do, J.; Lee, D. Y.; Kim, H. H.; Kim, D. I. N-glycan Remodeling Using Mannosidase Inhibitors to Increase High-mannose Glycans on Acid alpha-Glucosidase in Transgenic Rice Cell Cultures. *Sci Rep* **2018**, 8 (1), 16130. DOI: 10.1038/s41598-018-34438-z From NLM Medline.

(54) Patsos, G.; Hebbe-Viton, V.; Robbe-Masselot, C.; Masselot, D.; San Martin, R.; Greenwood, R.; Paraskeva, C.; Klein, A.; Graessmann, M.; Michalski, J. C.; et al. O-glycan inhibitors generate aryl-glycans, induce apoptosis and lead to growth inhibition in colorectal cancer cell lines. *Glycobiology* **2009**, 19 (4), 382-398. DOI: 10.1093/glycob/cwn149 From NLM Medline.

(55) Zhao, H.; Przybylska, M.; Wu, I. H.; Zhang, J.; Siegel, C.; Komarnitsky, S.; Yew, N. S.; Cheng, S. H. Inhibiting glycosphingolipid synthesis improves glycemic control and insulin

sensitivity in animal models of type 2 diabetes. *Diabetes* **2007**, *56* (5), 1210-1218. DOI: 10.2337/db06-0719 From NLM Medline.

(56) Zhao, Z.; Li, C.; Tong, F.; Deng, J.; Huang, G.; Sang, Y. Review of applications of CRISPR-Cas9 gene-editing technology in cancer research. *Biol Proced Online* **2021**, *23* (1), 14. DOI: 10.1186/s12575-021-00151-x From NLM PubMed-not-MEDLINE.

(57) Stanley, P. What Have We Learned from Glycosyltransferase Knockouts in Mice? *J Mol Biol* **2016**, *428* (16), 3166-3182. DOI: 10.1016/j.jmb.2016.03.025 From NLM Medline.

(58) Wang, Y.; Ju, T.; Ding, X.; Xia, B.; Wang, W.; Xia, L.; He, M.; Cummings, R. D. Cosmc is an essential chaperone for correct protein O-glycosylation. *Proc Natl Acad Sci U S A* **2010**, *107* (20), 9228-9233. DOI: 10.1073/pnas.0914004107 From NLM Medline.

(59) Narimatsu, Y.; Joshi, H. J.; Yang, Z.; Gomes, C.; Chen, Y. H.; Lorenzetti, F. C.; Furukawa, S.; Schjoldager, K. T.; Hansen, L.; Clausen, H.; et al. A validated gRNA library for CRISPR/Cas9 targeting of the human glycosyltransferase genome. *Glycobiology* **2018**, *28* (5), 295-305. DOI: 10.1093/glycob/cwx101 From NLM Medline.

(60) Linenberger, M. L. CD33-directed therapy with gemtuzumab ozogamicin in acute myeloid leukemia: progress in understanding cytotoxicity and potential mechanisms of drug resistance. *Leukemia* **2005**, *19* (2), 176-182. DOI: 10.1038/sj.leu.2403598 From NLM Medline.

(61) Sutherland, M. K.; Yu, C.; Lewis, T. S.; Miyamoto, J. B.; Morris-Tilden, C. A.; Jonas, M.; Sutherland, J.; Nesterova, A.; Gerber, H. P.; Sievers, E. L.; et al. Anti-leukemic activity of lintuzumab (SGN-33) in preclinical models of acute myeloid leukemia. *Mabs* **2009**, *1* (5), 481-490. DOI: 10.4161/mabs.1.5.9288 From NLM Medline.

(62) Walter, R. B. The role of CD33 as therapeutic target in acute myeloid leukemia. *Expert Opin Ther Targets* **2014**, *18* (7), 715-718. DOI: 10.1517/14728222.2014.909413 From NLM Medline.

(63) Griciuc, A.; Serrano-Pozo, A.; Parrado, A. R.; Lesinski, A. N.; Asselin, C. N.; Mullin, K.; Hooli, B.; Choi, S. H.; Hyman, B. T.; Tanzi, R. E. Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. *Neuron* **2013**, *78* (4), 631-643. DOI: 10.1016/j.neuron.2013.04.014 From NLM Medline.

(64) Eskandari-Sedighi, G.; Jung, J.; Macauley, M. S. CD33 isoforms in microglia and Alzheimer's disease: Friend and foe. *Mol Aspects Med* **2022**, 101111. DOI: 10.1016/j.mam.2022.101111 From NLM Publisher.

(65) Estus, S.; Shaw, B. C.; Devanney, N.; Katsumata, Y.; Press, E. E.; Fardo, D. W. Evaluation of CD33 as a genetic risk factor for Alzheimer's disease. *Acta Neuropathol* **2019**, *138* (2), 187-199. DOI: 10.1007/s00401-019-02000-4 From NLM Medline.

(66) Bertram, L.; Lange, C.; Mullin, K.; Parkinson, M.; Hsiao, M.; Hogan, M. F.; Schjeide, B. M.; Hooli, B.; Divito, J.; Ionita, I.; et al. Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. *Am J Hum Genet* **2008**, *83* (5), 623-632. DOI: 10.1016/j.ajhg.2008.10.008 From NLM Medline.

(67) Hollingworth, P.; Harold, D.; Sims, R.; Gerrish, A.; Lambert, J. C.; Carrasquillo, M. M.; Abraham, R.; Hamshere, M. L.; Pahwa, J. S.; Moskvina, V.; et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* **2011**, *43* (5), 429-435. DOI: 10.1038/ng.803 From NLM Medline.

(68) Naj, A. C.; Jun, G.; Beecham, G. W.; Wang, L. S.; Vardarajan, B. N.; Buross, J.; Gallins, P. J.; Buxbaum, J. D.; Jarvik, G. P.; Crane, P. K.; et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* **2011**, *43* (5), 436-441. DOI: 10.1038/ng.801 From NLM Medline.

- (69) Karch, C. M.; Jeng, A. T.; Nowotny, P.; Cady, J.; Cruchaga, C.; Goate, A. M. Expression of novel Alzheimer's disease risk genes in control and Alzheimer's disease brains. *PLoS One* **2012**, *7* (11), e50976. DOI: 10.1371/journal.pone.0050976 From NLM Medline.
- (70) Malik, M.; Simpson, J. F.; Parikh, I.; Wilfred, B. R.; Fardo, D. W.; Nelson, P. T.; Estus, S. CD33 Alzheimer's risk-altering polymorphism, CD33 expression, and exon 2 splicing. *J Neurosci* **2013**, *33* (33), 13320-13325. DOI: 10.1523/JNEUROSCI.1224-13.2013 From NLM Medline.
- (71) Nomura, S.; Tandon, N. N.; Nakamura, T.; Cone, J.; Fukuhara, S.; Kambayashi, J. High-shear-stress-induced activation of platelets and microparticles enhances expression of cell adhesion molecules in THP-1 and endothelial cells. *Atherosclerosis* **2001**, *158* (2), 277-287. DOI: 10.1016/s0021-9150(01)00433-6 From NLM Medline.
- (72) Ramsborg, C. G.; Papoutsakis, E. T. Global transcriptional analysis delineates the differential inflammatory response interleukin-15 elicits from cultured human T cells. *Exp Hematol* **2007**, *35* (3), 454-464. DOI: 10.1016/j.exphem.2006.11.013 From NLM Medline.
- (73) Li, X.; Shen, N.; Zhang, S.; Liu, J.; Jiang, Q.; Liao, M.; Feng, R.; Zhang, L.; Wang, G.; Ma, G.; et al. CD33 rs3865444 Polymorphism Contributes to Alzheimer's Disease Susceptibility in Chinese, European, and North American Populations. *Mol Neurobiol* **2015**, *52* (1), 414-421. DOI: 10.1007/s12035-014-8880-9 From NLM Medline.
- (74) Raj, T.; Ryan, K. J.; Replogle, J. M.; Chibnik, L. B.; Rosenkrantz, L.; Tang, A.; Rothamel, K.; Stranger, B. E.; Bennett, D. A.; Evans, D. A.; et al. CD33: increased inclusion of exon 2 implicates the Ig V-set domain in Alzheimer's disease susceptibility. *Hum Mol Genet* **2014**, *23* (10), 2729-2736. DOI: 10.1093/hmg/ddt666 From NLM Medline.
- (75) Bradshaw, E. M.; Chibnik, L. B.; Keenan, B. T.; Ottoboni, L.; Raj, T.; Tang, A.; Rosenkrantz, L. L.; Imboywa, S.; Lee, M.; Von Korff, A.; et al. CD33 Alzheimer's disease locus: altered monocyte function and amyloid biology. *Nat Neurosci* **2013**, *16* (7), 848-850. DOI: 10.1038/nn.3435 From NLM Medline.
- (76) Tortora, F.; Rendina, A.; Angiolillo, A.; Di Costanzo, A.; Aniello, F.; Donizetti, A.; Febbraio, F.; Vitale, E. CD33 rs2455069 SNP: Correlation with Alzheimer's Disease and Hypothesis of Functional Role. *Int J Mol Sci* **2022**, *23* (7). DOI: 10.3390/ijms23073629 From NLM Medline.
- (77) Saha, S.; Siddiqui, S. S.; Khan, N.; Verhagen, A.; Jiang, W.; Springer, S.; Ghosh, P.; Varki, A. Controversies about the subcellular localization and mechanisms of action of the Alzheimer's disease-protective CD33 splice variant. *Acta Neuropathol* **2019**, *138* (4), 671-672. DOI: 10.1007/s00401-019-02065-1 From NLM Medline.
- (78) Zhao, L. CD33 in Alzheimer's Disease - Biology, Pathogenesis, and Therapeutics: A Mini-Review. *Gerontology* **2019**, *65* (4), 323-331. DOI: 10.1159/000492596 From NLM Medline.
- (79) Hernandez-Caselles, T.; Martinez-Esparza, M.; Perez-Oliva, A. B.; Quintanilla-Cecconi, A. M.; Garcia-Alonso, A.; Alvarez-Lopez, D. M.; Garcia-Penarrubia, P. A study of CD33 (SIGLEC-3) antigen expression and function on activated human T and NK cells: two isoforms of CD33 are generated by alternative splicing. *J Leukoc Biol* **2006**, *79* (1), 46-58. DOI: 10.1189/jlb.0205096 From NLM Medline.
- (80) Siddiqui, S. S.; Springer, S. A.; Verhagen, A.; Sundaramurthy, V.; Alisson-Silva, F.; Jiang, W.; Ghosh, P.; Varki, A. The Alzheimer's disease-protective CD33 splice variant mediates adaptive loss of function via diversion to an intracellular pool. *J Biol Chem* **2017**, *292* (37), 15312-15320. DOI: 10.1074/jbc.M117.799346 From NLM Medline.
- (81) Perez-Oliva, A. B.; Martinez-Esparza, M.; Vicente-Fernandez, J. J.; Corral-San Miguel, R.; Garcia-Penarrubia, P.; Hernandez-Caselles, T. Epitope mapping, expression and post-translational

modifications of two isoforms of CD33 (CD33M and CD33m) on lymphoid and myeloid human cells. *Glycobiology* **2011**, *21* (6), 757-770. DOI: 10.1093/glycob/cwq220 From NLM Medline.

(82) Bhattacharjee, A.; Daskhan, G. C.; Bains, A.; Watson, A. E. S.; Eskandari-Sedighi, G.; St Laurent, C. D.; Voronova, A.; Macauley, M. S. Increasing phagocytosis of microglia by targeting CD33 with liposomes displaying glycan ligands. *J Control Release* **2021**, *338*, 680-693. DOI: 10.1016/j.jconrel.2021.09.010 From NLM Medline.

(83) Laszlo, G. S.; Estey, E. H.; Walter, R. B. The past and future of CD33 as therapeutic target in acute myeloid leukemia. *Blood Rev* **2014**, *28* (4), 143-153. DOI: 10.1016/j.blre.2014.04.001 From NLM Medline.

(84) Okerblom, J. J.; Schwarz, F.; Olson, J.; Fletes, W.; Ali, S. R.; Martin, P. T.; Glass, C. K.; Nizet, V.; Varki, A. Loss of CMAH during Human Evolution Primed the Monocyte-Macrophage Lineage toward a More Inflammatory and Phagocytic State. *J Immunol* **2017**, *198* (6), 2366-2373. DOI: 10.4049/jimmunol.1601471 From NLM Medline.

(85) Gotz, J.; Bodea, L. G.; Goedert, M. Rodent models for Alzheimer disease. *Nat Rev Neurosci* **2018**, *19* (10), 583-598. DOI: 10.1038/s41583-018-0054-8 From NLM Medline.

(86) Janus, C.; Westaway, D. Transgenic mouse models of Alzheimer's disease. *Physiol Behav* **2001**, *73* (5), 873-886. DOI: 10.1016/s0031-9384(01)00524-8 From NLM Medline.

(87) Radde, R.; Duma, C.; Goedert, M.; Jucker, M. The value of incomplete mouse models of Alzheimer's disease. *Eur J Nucl Med Mol Imaging* **2008**, *35 Suppl 1*, S70-74. DOI: 10.1007/s00259-007-0704-y From NLM Medline.

(88) Cao, H.; Crocker, P. R. Evolution of CD33-related siglecs: regulating host immune functions and escaping pathogen exploitation? *Immunology* **2011**, *132* (1), 18-26. DOI: 10.1111/j.1365-2567.2010.03368.x From NLM Medline.

(89) Brinkman-Van der Linden, E. C.; Angata, T.; Reynolds, S. A.; Powell, L. D.; Hedrick, S. M.; Varki, A. CD33/Siglec-3 binding specificity, expression pattern, and consequences of gene deletion in mice. *Mol Cell Biol* **2003**, *23* (12), 4199-4206. DOI: 10.1128/MCB.23.12.4199-4206.2003 From NLM Medline.

(90) Tsai, T. Y.; Huang, M. T.; Sung, P. S.; Peng, C. Y.; Tao, M. H.; Yang, H. I.; Chang, W. C.; Yang, A. S.; Yu, C. M.; Lin, Y. P.; et al. SIGLEC-3 (CD33) serves as an immune checkpoint receptor for HBV infection. *J Clin Invest* **2021**, *131* (11). DOI: 10.1172/JCI141965 From NLM Medline.

(91) Propster, J. M.; Yang, F.; Rabbani, S.; Ernst, B.; Allain, F. H.; Schubert, M. Structural basis for sulfation-dependent self-glycan recognition by the human immune-inhibitory receptor Siglec-8. *Proc Natl Acad Sci U S A* **2016**, *113* (29), E4170-4179. DOI: 10.1073/pnas.1602214113 From NLM Medline.

(92) Bhattacharjee, A.; Rodrigues, E.; Jung, J.; Luzentales-Simpson, M.; Enterina, J. R.; Galleguillos, D.; St Laurent, C. D.; Nakhaei-Nejad, M.; Fuchsberger, F. F.; Streith, L.; et al. Repression of phagocytosis by human CD33 is not conserved with mouse CD33. *Commun Biol* **2019**, *2*, 450. DOI: 10.1038/s42003-019-0698-6 From NLM PubMed-not-MEDLINE.

(93) Jung, J.; Enterina, J. R.; Bui, D. T.; Mozaneh, F.; Lin, P. H.; Nitin; Kuo, C. W.; Rodrigues, E.; Bhattacharjee, A.; Raeisimakiani, P.; et al. Carbohydrate Sulfation As a Mechanism for Fine-Tuning Siglec Ligands. *ACS Chem Biol* **2021**, *16* (11), 2673-2689. DOI: 10.1021/acscchembio.1c00501 From NLM Medline.

(94) Hane, M.; Chen, D. Y.; Varki, A. Human-specific microglial Siglec-11 transcript variant has the potential to affect polysialic acid-mediated brain functions at a distance. *Glycobiology* **2021**, *31* (3), 231-242. DOI: 10.1093/glycob/cwaa082 From NLM Medline.

- (95) Maley, F.; Trimble, R. B.; Tarentino, A. L.; Plummer, T. H., Jr. Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Anal Biochem* **1989**, *180* (2), 195-204. DOI: 10.1016/0003-2697(89)90115-2 From NLM Medline.
- (96) Wilkinson, H.; Saldova, R. Current Methods for the Characterization of O-Glycans. *J Proteome Res* **2020**, *19* (10), 3890-3905. DOI: 10.1021/acs.jproteome.0c00435 From NLM Medline.
- (97) Zhang, Q.; Li, Z.; Song, X. Preparation of Complex Glycans From Natural Sources for Functional Study. *Front Chem* **2020**, *8*, 508. DOI: 10.3389/fchem.2020.00508 From NLM PubMed-not-MEDLINE.
- (98) Ishibashi, Y.; Nakasone, T.; Kiyohara, M.; Horibata, Y.; Sakaguchi, K.; Hijikata, A.; Ichinose, S.; Omori, A.; Yasui, Y.; Imamura, A.; et al. A novel endoglycoceramidase hydrolyzes oligogalactosylceramides to produce galactooligosaccharides and ceramides. *J Biol Chem* **2007**, *282* (15), 11386-11396. DOI: 10.1074/jbc.M608445200 From NLM Medline.
- (99) Wang, C.; Wen, Y.; Yang, M.; Huang, L.; Wang, Z.; Fan, J. High-sensitivity quantification of glycosphingolipid glycans by ESI-MS utilizing ozonolysis-based release and isotopic Girard's reagent labeling. *Anal Biochem* **2019**, *582*, 113355. DOI: 10.1016/j.ab.2019.113355 From NLM Medline.
- (100) Song, X.; Ju, H.; Lasanajak, Y.; Kudelka, M. R.; Smith, D. F.; Cummings, R. D. Oxidative release of natural glycans for functional glycomics. *Nat Methods* **2016**, *13* (6), 528-534. DOI: 10.1038/nmeth.3861 From NLM Medline.
- (101) Glish, G. L.; Vachet, R. W. The basics of mass spectrometry in the twenty-first century. *Nat Rev Drug Discov* **2003**, *2* (2), 140-150. DOI: 10.1038/nrd1011 From NLM Medline.
- (102) Kitova, E. N.; El-Hawiet, A.; Schnier, P. D.; Klassen, J. S. Reliable determinations of protein-ligand interactions by direct ESI-MS measurements. Are we there yet? *J Am Soc Mass Spectrom* **2012**, *23* (3), 431-441. DOI: 10.1007/s13361-011-0311-9 From NLM Medline.
- (103) El-Hawiet, A.; Shoemaker, G. K.; Daneshfar, R.; Kitova, E. N.; Klassen, J. S. Applications of a catch and release electrospray ionization mass spectrometry assay for carbohydrate library screening. *Anal Chem* **2012**, *84* (1), 50-58. DOI: 10.1021/ac202760e From NLM Medline.
- (104) Vafa, O.; Gilliland, G. L.; Brezski, R. J.; Strake, B.; Wilkinson, T.; Lacy, E. R.; Scallon, B.; Teplyakov, A.; Malia, T. J.; Strohl, W. R. An engineered Fc variant of an IgG eliminates all immune effector functions via structural perturbations. *Methods* **2014**, *65* (1), 114-126. DOI: 10.1016/j.ymeth.2013.06.035 From NLM Medline.
- (105) Kapust, R. B.; Tozser, J.; Copeland, T. D.; Waugh, D. S. The P1' specificity of tobacco etch virus protease. *Biochem Biophys Res Commun* **2002**, *294* (5), 949-955. DOI: 10.1016/S0006-291X(02)00574-0 From NLM Medline.
- (106) Rodrigues, E.; Jung, J.; Park, H.; Loo, C.; Soukhtehzari, S.; Kitova, E. N.; Mozaneh, F.; Daskhan, G.; Schmidt, E. N.; Aghanya, V.; et al. A versatile soluble siglec scaffold for sensitive and quantitative detection of glycan ligands. *Nat Commun* **2020**, *11* (1), 5091. DOI: 10.1038/s41467-020-18907-6 From NLM Medline.
- (107) O'Gorman, S.; Fox, D. T.; Wahl, G. M. Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **1991**, *251* (4999), 1351-1355. DOI: 10.1126/science.1900642 From NLM Medline.
- (108) Pedro, L.; Van Voorhis, W. C.; Quinn, R. J. Optimization of Electrospray Ionization by Statistical Design of Experiments and Response Surface Methodology: Protein-Ligand Equilibrium Dissociation Constant Determinations. *J Am Soc Mass Spectrom* **2016**, *27* (9), 1520-1530. DOI: 10.1007/s13361-016-1417-x From NLM Medline.

- (109) Chuang, G. Y.; Boyington, J. C.; Joyce, M. G.; Zhu, J.; Nabel, G. J.; Kwong, P. D.; Georgiev, I. Computational prediction of N-linked glycosylation incorporating structural properties and patterns. *Bioinformatics* **2012**, *28* (17), 2249-2255. DOI: 10.1093/bioinformatics/bts426 From NLM Medline.
- (110) North, S. J.; Huang, H. H.; Sundaram, S.; Jang-Lee, J.; Etienne, A. T.; Trollope, A.; Chalabi, S.; Dell, A.; Stanley, P.; Haslam, S. M. Glycomics profiling of Chinese hamster ovary cell glycosylation mutants reveals N-glycans of a novel size and complexity. *J Biol Chem* **2010**, *285* (8), 5759-5775. DOI: 10.1074/jbc.M109.068353 From NLM Medline.
- (111) Wang, F.; Wang, X.; Yu, X.; Fu, L.; Liu, Y.; Ma, L.; Zhai, C. High-level expression of endo-beta-N-acetylglucosaminidase H from *Streptomyces plicatus* in *Pichia pastoris* and its application for the deglycosylation of glycoproteins. *PLoS One* **2015**, *10* (3), e0120458. DOI: 10.1371/journal.pone.0120458 From NLM Medline.
- (112) Shental-Bechor, D.; Levy, Y. Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc Natl Acad Sci U S A* **2008**, *105* (24), 8256-8261. DOI: 10.1073/pnas.0801340105 From NLM Medline.
- (113) Rillahan, C. D.; Macauley, M. S.; Schwartz, E.; He, Y.; McBride, R.; Arlian, B. M.; Rangarajan, J.; Fokin, V. V.; Paulson, J. C. Disubstituted Sialic Acid Ligands Targeting Siglecs CD33 and CD22 Associated with Myeloid Leukaemias and B Cell Lymphomas. *Chem Sci* **2014**, *5* (6), 2398-2406. DOI: 10.1039/C4SC00451E From NLM PubMed-not-MEDLINE.
- (114) Christensen, S.; Egebjerg, J. Cloning, expression and characterization of a sialidase gene from *Arthrobacter ureafaciens*. *Biotechnol Appl Biochem* **2005**, *41* (Pt 3), 225-231. DOI: 10.1042/BA20040144 From NLM Medline.
- (115) Gut, H.; King, S. J.; Walsh, M. A. Structural and functional studies of *Streptococcus pneumoniae* neuraminidase B: An intramolecular trans-sialidase. *FEBS Lett* **2008**, *582* (23-24), 3348-3352. DOI: 10.1016/j.febslet.2008.08.026 From NLM Medline.
- (116) Miles, L. A.; Hermans, S. J.; Crespi, G. A. N.; Gooi, J. H.; Doughty, L.; Nero, T. L.; Markulic, J.; Ebner, A.; Wroblewski, B.; Oehlrich, D.; et al. Small Molecule Binding to Alzheimer Risk Factor CD33 Promotes Abeta Phagocytosis. *iScience* **2019**, *19*, 110-118. DOI: 10.1016/j.isci.2019.07.023 From NLM PubMed-not-MEDLINE.
- (117) Barkal, A. A.; Brewer, R. E.; Markovic, M.; Kowarsky, M.; Barkal, S. A.; Zaro, B. W.; Krishnan, V.; Hatakeyama, J.; Dorigo, O.; Barkal, L. J.; et al. CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. *Nature* **2019**, *572* (7769), 392-396. DOI: 10.1038/s41586-019-1456-0 From NLM Medline.
- (118) Habuchi, O. Carbohydrate(KeratanSulfateGal-6) Sulfotransferase1(CHST1). *SpringerJapan* **2014**, 989-995. DOI: 10.1007/978-4-431-54240-7_42.
- (119) Bull, C.; Nason, R.; Sun, L.; Van Coillie, J.; Madriz Sorensen, D.; Moons, S. J.; Yang, Z.; Arbitman, S.; Fernandes, S. M.; Furukawa, S.; et al. Probing the binding specificities of human Siglecs by cell-based glycan arrays. *Proc Natl Acad Sci U S A* **2021**, *118* (17). DOI: 10.1073/pnas.2026102118 From NLM Medline.
- (120) Hemmerich, S.; Rosen, S. D. Carbohydrate sulfotransferases in lymphocyte homing. *Glycobiology* **2000**, *10* (9), 849-856. DOI: 10.1093/glycob/10.9.849 From NLM Medline.
- (121) Fukuda, M.; Hiraoka, N.; Akama, T. O.; Fukuda, M. N. Carbohydrate-modifying sulfotransferases: structure, function, and pathophysiology. *J Biol Chem* **2001**, *276* (51), 47747-47750. DOI: 10.1074/jbc.R100049200 From NLM Medline.

- (122) Torii, T.; Fukuta, M.; Habuchi, O. Sulfation of sialyl N-acetylglucosamine oligosaccharides and fetuin oligosaccharides by keratan sulfate Gal-6-sulfotransferase. *Glycobiology* **2000**, *10* (2), 203-211. DOI: 10.1093/glycob/10.2.203 From NLM Medline.
- (123) Sakaguchi, H.; Kitagawa, H.; Sugahara, K. Functional expression and genomic structure of human N-acetylglucosamine-6-O-sulfotransferase that transfers sulfate to beta-N-acetylglucosamine at the nonreducing end of an N-acetylglucosamine sequence. *Biochim Biophys Acta* **2000**, *1523* (2-3), 269-276. DOI: 10.1016/s0304-4165(00)00136-7 From NLM Medline.
- (124) Bistrup, A.; Bhakta, S.; Lee, J. K.; Belov, Y. Y.; Gunn, M. D.; Zuo, F. R.; Huang, C. C.; Kannagi, R.; Rosen, S. D.; Hemmerich, S. Sulfotransferases of two specificities function in the reconstitution of high endothelial cell ligands for L-selectin. *J Cell Biol* **1999**, *145* (4), 899-910. DOI: 10.1083/jcb.145.4.899 From NLM Medline.
- (125) Xia, G.; Evers, M. R.; Kang, H. G.; Schachner, M.; Baenziger, J. U. Molecular cloning and expression of the pituitary glycoprotein hormone N-acetylgalactosamine-4-O-sulfotransferase. *J Biol Chem* **2000**, *275* (49), 38402-38409. DOI: 10.1074/jbc.M007821200 From NLM Medline.
- (126) Hiraoka, N.; Misra, A.; Belot, F.; Hindsgaul, O.; Fukuda, M. Molecular cloning and expression of two distinct human N-acetylgalactosamine 4-O-sulfotransferases that transfer sulfate to GalNAc beta 1-->4GlcNAc beta 1-->R in both N- and O-glycans. *Glycobiology* **2001**, *11* (6), 495-504. DOI: 10.1093/glycob/11.6.495 From NLM Medline.
- (127) Honke, K.; Tsuda, M.; Koyota, S.; Wada, Y.; Iida-Tanaka, N.; Ishizuka, I.; Nakayama, J.; Taniguchi, N. Molecular cloning and characterization of a human beta-Gal-3'-sulfotransferase that acts on both type 1 and type 2 (Gal beta 1-3/1-4GlcNAc-R) oligosaccharides. *J Biol Chem* **2001**, *276* (1), 267-274. DOI: 10.1074/jbc.M005666200 From NLM Medline.
- (128) Suzuki, A.; Hiraoka, N.; Suzuki, M.; Angata, K.; Misra, A. K.; McAuliffe, J.; Hindsgaul, O.; Fukuda, M. Molecular cloning and expression of a novel human beta-Gal-3-O-sulfotransferase that acts preferentially on N-acetylglucosamine in N- and O-glycans. *J Biol Chem* **2001**, *276* (26), 24388-24395. DOI: 10.1074/jbc.M103135200 From NLM Medline.
- (129) Seko, A.; Hara-Kuge, S.; Yamashita, K. Molecular cloning and characterization of a novel human galactose 3-O-sulfotransferase that transfers sulfate to gal beta 1-->3galNAc residue in O-glycans. *J Biol Chem* **2001**, *276* (28), 25697-25704. DOI: 10.1074/jbc.M101558200 From NLM Medline.
- (130) Bai, X.; Brown, J. R.; Varki, A.; Esko, J. D. Enhanced 3-O-sulfation of galactose in Asn-linked glycans and Maackia amurensis lectin binding in a new Chinese hamster ovary cell line. *Glycobiology* **2001**, *11* (8), 621-632. DOI: 10.1093/glycob/11.8.621 From NLM Medline.
- (131) Bensing, B. A.; Li, Q.; Park, D.; Lebrilla, C. B.; Sullam, P. M. Streptococcal Siglec-like adhesins recognize different subsets of human plasma glycoproteins: implications for infective endocarditis. *Glycobiology* **2018**, *28* (8), 601-611. DOI: 10.1093/glycob/cwy052 From NLM Medline.
- (132) Ribeiro, J. P.; Pau, W.; Pifferi, C.; Renaudet, O.; Varrot, A.; Mahal, L. K.; Imberty, A. Characterization of a high-affinity sialic acid-specific CBM40 from Clostridium perfringens and engineering of a divalent form. *Biochem J* **2016**, *473* (14), 2109-2118. DOI: 10.1042/BCJ20160340 From NLM Medline.
- (133) Ali, S. R.; Fong, J. J.; Carlin, A. F.; Busch, T. D.; Linden, R.; Angata, T.; Areschoug, T.; Parast, M.; Varki, N.; Murray, J.; et al. Siglec-5 and Siglec-14 are polymorphic paired receptors that modulate neutrophil and amnion signaling responses to group B Streptococcus. *J Exp Med* **2014**, *211* (6), 1231-1242. DOI: 10.1084/jem.20131853 From NLM Medline.

(134) Heinz Läubli, K. K., Cijo George Vazhappilly, Rachel Matar, Maxime Merheb, Shoib Sarwar Siddiqui. Tools to study and target the Siglec–sialic acid axis in cancer

. *The FEBS Journal* **2020**. DOI: <https://doi.org/10.1111/febs.15647>.

(135) Bhattacharjee, A.; Jung, J.; Zia, S.; Ho, M.; Eskandari-Sedighi, G.; St Laurent, C. D.; McCord, K. A.; Bains, A.; Sidhu, G.; Sarkar, S.; et al. The CD33 short isoform is a gain-of-function variant that enhances Abeta1-42 phagocytosis in microglia. *Mol Neurodegener* **2021**, *16* (1), 19. DOI: 10.1186/s13024-021-00443-6 From NLM Medline.

(136) Redelinghuys, P.; Antonopoulos, A.; Liu, Y.; Campanero-Rhodes, M. A.; McKenzie, E.; Haslam, S. M.; Dell, A.; Feizi, T.; Crocker, P. R. Early murine T-lymphocyte activation is accompanied by a switch from N-Glycolyl- to N-acetyl-neuraminic acid and generation of ligands for siglec-E. *J Biol Chem* **2011**, *286* (40), 34522-34532. DOI: 10.1074/jbc.M111.243410 From NLM Medline.

(137) Wasim, L.; Buhari, F. H. M.; Yoganathan, M.; Sicard, T.; Ereno-Orbea, J.; Julien, J. P.; Treanor, B. N-Linked Glycosylation Regulates CD22 Organization and Function. *Front Immunol* **2019**, *10*, 699. DOI: 10.3389/fimmu.2019.00699 From NLM Medline.

(138) Malaker, S. A.; Pedram, K.; Ferracane, M. J.; Bensing, B. A.; Krishnan, V.; Pett, C.; Yu, J.; Woods, E. C.; Kramer, J. R.; Westerlind, U.; et al. The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. *Proc Natl Acad Sci U S A* **2019**, *116* (15), 7278-7287. DOI: 10.1073/pnas.1813020116 From NLM Medline.

(139) V Piller, F. P., M Fukuda. Biosynthesis of truncated O-glycans in the T cell line Jurkat. Localization of O-glycan initiation. *Journal of Biological Chemistry* **1990**, *265* (16). DOI: [https://doi.org/10.1016/S0021-9258\(19\)38842-8](https://doi.org/10.1016/S0021-9258(19)38842-8).

(140) Rodrigues, E.; Macauley, M. S. Targeted self-destruction. *Nat Chem Biol* **2020**, *16* (12), 1281-1283. DOI: 10.1038/s41589-020-0612-z From NLM Medline.

(141) Stanczak, M. A.; Siddiqui, S. S.; Trefny, M. P.; Thommen, D. S.; Boligan, K. F.; von Gunten, S.; Tzankov, A.; Tietze, L.; Lardinois, D.; Heinzelmann-Schwarz, V.; et al. Self-associated molecular patterns mediate cancer immune evasion by engaging Siglecs on T cells. *J Clin Invest* **2018**, *128* (11), 4912-4923. DOI: 10.1172/JCI120612 From NLM Medline.

(142) Vinay, D. S.; Ryan, E. P.; Pawelec, G.; Talib, W. H.; Stagg, J.; Elkord, E.; Lichtor, T.; Decker, W. K.; Whelan, R. L.; Kumara, H.; et al. Immune evasion in cancer: Mechanistic basis and therapeutic strategies. *Semin Cancer Biol* **2015**, *35* Suppl, S185-S198. DOI: 10.1016/j.semcancer.2015.03.004 From NLM Medline.

(143) Baeuerle, P. A. H., W. B. Chlorate — a Potent Inhibitor of Protein Sulfation in Intact Cells. *Biochemical and Biophysical Research Communications* **1986**, *141* (2). DOI: [https://doi.org/10.1016/S0006-291X\(86\)80253-4](https://doi.org/10.1016/S0006-291X(86)80253-4).

(144) Caterson, B.; Melrose, J. Keratan sulfate, a complex glycosaminoglycan with unique functional capability. *Glycobiology* **2018**, *28* (4), 182-206. DOI: 10.1093/glycob/cwy003 From NLM Medline.

(145) Zhang, Z.; Takeda-Uchimura, Y.; Foyez, T.; Ohtake-Niimi, S.; Narentuya; Akatsu, H.; Nishitsuji, K.; Michikawa, M.; Wyss-Coray, T.; Kadomatsu, K.; et al. Deficiency of a sulfotransferase for sialic acid-modified glycans mitigates Alzheimer's pathology. *Proc Natl Acad Sci U S A* **2017**, *114* (14), E2947-E2954. DOI: 10.1073/pnas.1615036114 From NLM Medline.

(146) Hoshino, H.; Foyez, T.; Ohtake-Niimi, S.; Takeda-Uchimura, Y.; Michikawa, M.; Kadomatsu, K.; Uchimura, K. KSGal6ST is essential for the 6-sulfation of galactose within keratan sulfate in early postnatal brain. *J Histochem Cytochem* **2014**, *62* (2), 145-156. DOI: 10.1369/0022155413511619 From NLM Medline.

- (147) Puri, S.; Coulson-Thomas, Y. M.; Gesteira, T. F.; Coulson-Thomas, V. J. Distribution and Function of Glycosaminoglycans and Proteoglycans in the Development, Homeostasis and Pathology of the Ocular Surface. *Front Cell Dev Biol* **2020**, *8*, 731. DOI: 10.3389/fcell.2020.00731 From NLM PubMed-not-MEDLINE.
- (148) Gonzalez-Gil, A.; Porell, R. N.; Fernandes, S. M.; Maenpaa, E.; Li, T. A.; Li, T.; Wong, P. C.; Aoki, K.; Tiemeyer, M.; Yu, Z. J.; et al. Human brain sialoglycan ligand for CD33, a microglial inhibitory Siglec implicated in Alzheimer's disease. *J Biol Chem* **2022**, *298* (6), 101960. DOI: 10.1016/j.jbc.2022.101960 From NLM Medline.
- (149) Narentuya; Takeda-Uchimura, Y.; Foyez, T.; Zhang, Z.; Akama, T. O.; Yagi, H.; Kato, K.; Komatsu, Y.; Kadomatsu, K.; Uchimura, K. GlcNAc6ST3 is a keratan sulfate sulfotransferase for the protein-tyrosine phosphatase PTPRZ in the adult brain. *Sci Rep* **2019**, *9* (1), 4387. DOI: 10.1038/s41598-019-40901-2 From NLM Medline.
- (150) Leiphrakpam, P. D.; Patil, P. P.; Remmers, N.; Swanson, B.; Grandgenett, P. M.; Qiu, F.; Yu, F.; Radhakrishnan, P. Role of keratan sulfate expression in human pancreatic cancer malignancy. *Sci Rep* **2019**, *9* (1), 9665. DOI: 10.1038/s41598-019-46046-6 From NLM Medline.
- (151) Malik, M.; Chiles, J., 3rd; Xi, H. S.; Medway, C.; Simpson, J.; Potluri, S.; Howard, D.; Liang, Y.; Paumi, C. M.; Mukherjee, S.; et al. Genetics of CD33 in Alzheimer's disease and acute myeloid leukemia. *Hum Mol Genet* **2015**, *24* (12), 3557-3570. DOI: 10.1093/hmg/ddv092 From NLM Medline.
- (152) Gbadamosi, M. O.; Shastri, V. M.; Hylkema, T.; Papageorgiou, I.; Pardo, L.; Cogle, C. R.; Doty, A.; Loken, M. R.; Meshinchi, S.; Lamba, J. K. Novel CD33 antibodies unravel localization, biology and therapeutic implications of CD33 isoforms. *Future Oncol* **2021**, *17* (3), 263-277. DOI: 10.2217/fon-2020-0746 From NLM Medline.
- (153) Butler, C. A.; Thornton, P.; Brown, G. C. CD33M inhibits microglial phagocytosis, migration and proliferation, but the Alzheimer's disease-protective variant CD33m stimulates phagocytosis and proliferation, and inhibits adhesion. *J Neurochem* **2021**, *158* (2), 297-310. DOI: 10.1111/jnc.15349 From NLM Medline.
- (154) Laszlo, G. S.; Harrington, K. H.; Gudgeon, C. J.; Beddoe, M. E.; Fitzgibbon, M. P.; Ries, R. E.; Lamba, J. K.; McIntosh, M. W.; Meshinchi, S.; Walter, R. B. Expression and functional characterization of CD33 transcript variants in human acute myeloid leukemia. *Oncotarget* **2016**, *7* (28), 43281-43294. DOI: 10.18632/oncotarget.9674 From NLM Medline.
- (155) Godwin, C. D.; Laszlo, G. S.; Wood, B. L.; Correnti, C. E.; Bates, O. M.; Garling, E. E.; Mao, Z. J.; Beddoe, M. E.; Lunn, M. C.; Humbert, O.; et al. The CD33 splice isoform lacking exon 2 as therapeutic target in human acute myeloid leukemia. *Leukemia* **2020**, *34* (9), 2479-2483. DOI: 10.1038/s41375-020-0755-7 From NLM Medline.
- (156) van Bergeijk, P.; Seneviratne, U.; Aparicio-Prat, E.; Stanton, R.; Hasson, S. A. SRSF1 and PTBP1 Are trans-Acting Factors That Suppress the Formation of a CD33 Splicing Isoform Linked to Alzheimer's Disease Risk. *Mol Cell Biol* **2019**, *39* (18). DOI: 10.1128/MCB.00568-18 From NLM Medline.
- (157) Ribes, S.; Ebert, S.; Regen, T.; Agarwal, A.; Tauber, S. C.; Czesnik, D.; Spreer, A.; Bunkowski, S.; Eiffert, H.; Hanisch, U. K.; et al. Toll-like receptor stimulation enhances phagocytosis and intracellular killing of nonencapsulated and encapsulated *Streptococcus pneumoniae* by murine microglia. *Infect Immun* **2010**, *78* (2), 865-871. DOI: 10.1128/IAI.01110-09 From NLM Medline.
- (158) Chappie, T. A.; Abdelmessih, M.; Ambroise, C. W.; Boehm, M.; Cai, M.; Green, M.; Guilmette, E.; Stepan, C. M.; Stevens, L. M.; Wei, L.; et al. Discovery of Small-Molecule CD33

Pre-mRNA Splicing Modulators. *ACS Med Chem Lett* **2022**, *13* (1), 55-62. DOI: 10.1021/acsmchemlett.1c00396 From NLM PubMed-not-MEDLINE.

(159) Gibbons, A. E.; Luker, K. E.; Luker, G. D. Dual Reporter Bioluminescence Imaging with NanoLuc and Firefly Luciferase. *Methods Mol Biol* **2018**, *1790*, 41-50. DOI: 10.1007/978-1-4939-7860-1_4 From NLM Medline.

(160) Godwin, C. D.; Laszlo, G. S.; Fiorenza, S.; Garling, E. E.; Phi, T. D.; Bates, O. M.; Correnti, C. E.; Hoffstrom, B. G.; Lunn, M. C.; Humbert, O.; et al. Targeting the membrane-proximal C2-set domain of CD33 for improved CD33-directed immunotherapy. *Leukemia* **2021**, *35* (9), 2496-2507. DOI: 10.1038/s41375-021-01160-1 From NLM Medline.

Appendix A.

DNA sequences of all constructs used in this thesis

A.1 CD33

atgccgctgctgctactgctgcccctgctgtgggcagggggccctggctatggatccaaatttctggctgcaagtgcaggagtcagtgcggt
acaggaggggtttgtgctcctcgtgccctgcacttttccatcccataccctactacgacaagaactccccagttcatggttactggttcgg
gaaggagccattatatccagggactctccagtggccacaaacaagctagatcaagaagtacaggaggagactcagggcagattccgcctc
cttggggatcccagtaggaacaactgctccctgagcatcgtagacgccaggaggaggataatggttcatacttcttcggatggagagag
gaagtaccaaatacagttacaaatctcccagctctctgtgcatgtgacagacttgaccacaggcccaaatcctcatccctggcactctag
aaccggccactccaaaaacctgacctgctctgtgtcctgggcctgtgagcagggaacacccccgatcttctcctggtgtcagctgcccc
acctccctgggccccaggactactcactcctcgggtgctcataatcaccccacggccccaggaccacggcaccaacctgacctgtcaggtg
aagttcgctggagctggtgtgactacggagagaaccatccagctcaacgtcacctatgttccacagaaccaacaactggtatctttccagg
agatggctcagggaacaagagaccagagcaggagtgttcatggggccattggaggagctggtgttacagccctgctcgtctttgtctc
tgctcatcttctcatagtgaagaccacaggaggaaagcagccaggacagcagtgggcaggaatgacaccaccctaccacagggtc
agcctccccgaaacaccagaagaagtccaagttacatggccccactgaaacctcaagctgttcaggtgccgccctactgtggagatggat
gaggagctgcattatgcttccctcaactttcatgggatgaatccttccaaggacacctccaccgaataactcagaggtcaggaccagtgga

A.2 CD33-Fc

atgccgctgctgctactgctgcccctgctgtgggcagggggccctggctatggatccaaatttctggctgcaagtgcaggagtcagtgcggt
acaggaggggtttgtgctcctcgtgccctgcacttttccatcccataccctactacgacaagaactccccagttcatggttactggttcgg
gaaggagccattatatccagggactctccagtggccacaaacaagctagatcaagaagtacaggaggagactcagggcagattccgcctc
cttggggatcccagtaggaacaactgctccctgagcatcgtagacgccaggaggaggataatggttcatacttcttcggatggagagag
gaagtaccaaatacagttacaaatctcccagctctctgtgcatgtgacagacttgaccacaggcccaaatcctcatccctggcactctag
aaccggccactccaaaaacctgacctgctctgtgtcctgggcctgtgagcagggaacacccccgatcttctcctggtgtcagctgcccc
acctccctgggccccaggactactcactcctcgggtgctcataatcaccccacggccccaggaccacggcaccaacctgacctgtcaggtg
aagttcgctggagctggtgtgactacggagagaaccatccagctcaacgtcacctatgttccacagaaccaacaactggtatctttccagg

agatggctcagggaaacaagagaccagagcaggagtgggtcataccggtgagaacctgtacttccagggggacaaaactcacacatgcc
caccgtgccagcacctgaagccgccggggcctcctcagtcttcttcccccaaaaccaaggacacctcatgatctcccgacccc
tgaggtcacatgcgtgggtggacgtgagcgccgaagacctgaggtcaagttcaactgggtacgtggacggcgtggaggtgcataatgc
caagacaaagccgcgggaggagcagtacaacagcacgtaccgtgtggtcagcgtcctcaccgtcctgcaccaggactggctgaatggc
aaggagtacaagtgaaggtctcaacaaagccctcccatcctcatcgagaaaacctctcaaagccaaagggcagccccgcgaacc
acaggtgtacacctgccccatcacgggaggagatgaccaagaaccaggtcagcctgacctgcctgggtcaaaggttctatcccagcga
catcgccgtggagtgggagagcaatgggcagccggagaacaactacaagaccacgcctcccgctgtggactccgacggctccttctct
ctacagcaagctcaccgtggacaagagcaggtggcagcaggggaacgtcttctcatgtccgtgatgcacgaggctctgcacaaccacta
cacgcagaagagcctctcctgtctccgggtaaacaccatcaccatcaccattgggtccacccccagttcgagaagtga

A.3 *CD33* minigene (C allele)

atgccgctgctgctactgctgccccctgctgtgggcaggtgagtggtgtggtgggagaggggtgtcgggctggggccgagctgacctcgttt
ccccacagggggcctggctatggatccaaatttctggctgcaagtgcaggagtcagtgacggtacaggaggggttgtgcgtcctcgtgcc
tgcacttttccatcccataccctactacgacaagaactccccagttcatggttactggtccgggaaggagccattatatccagggactctcc
agtggccacaaacaagctagatcaagaagtacaggaggagactcagggcagattccgcctccttggggatcccagtaggaacaactgct
ccctgagcatcgtagacgccaggaggaggataatggttcatacttcttcggatggagagaggaagtaccaatacagttacaaatctccc
cagctctctgtgcatgtgacaggtgaggcacaggcttcagaagtggccgcaagggaagttcatgggtactgcagggcagggctgggatg
ggacctggtactgggaggggtttaggggtaaagcctgtcgtgcttagcgggggagcttgaccagaggttgatcttctcaggccctcac
ctggacctccctcctgattctgcatcccccttcttctcctactagacttgaccacaggcccaaatcctcatccctggcactctagaacccg
gccactccaaaaacctgacctgctctgtgtcctgggcctgtgagcagggaaacccccgatcttctcctggtgtcagctgccccacctcc
ctgggccccaggactactcactcctcggtgctcataatccccacggccccaggaccacggcaccaacctgacctgtcaggtgaagttc
gctggagctggtgtgactacggagagaacctccagctcaacgtcacctatgtccacagaaccaacaactggatatcttccaggagatgg
ctcagggaaacaagagaccagagcaggagtgggtcatggggccattggaggagctgggttacagccctgctcgtctttgtctctgcctca
tcttctcatagtgaagaccacaggaggaaagcagccaggacagcagtgggcaggaatgacaccacccctaccacagggtcagcctcc

ccgaaacaccagaagaagtccaagttacatggccccactgaaacctcaagctgttcaggtgccgccccctactgtggagatggatgaggag
ctgcattatgcttccctcaactttcatgggatga

A.4 *CD33* minigene (T allele)

atgccgctgctgctactgctgccccctgctgtgggcaggtgagtggtgtgtgggagaggggtgtcgggctgggcccagctgacctcgttt
ccccacaggggtcctggctatggatccaaatttctggctgcaagtgcaggagtcagtacggtacaggaggggtttgtcgtcctcgtgcctt
gcacttttccatcccataccctactacgacaagaactccccagttcatggttactggtccgggaaggagccattatatccagggactctcc
agtggccacaaacaagctagatcaagaagtacaggaggagactcagggcagattccgcctccttggggatcccagtaggaacaactgct
ccctgagcatcgtagacgccaggaggaggataatggttcatacttcttcggatggagagaggaagtaccaaatacattacaaatctccc
cagctctctgtgcatgtgacaggtgaggcacaggcttcagaagtggccgcaagggaagttcatgggtactgcagggcagggctgggatg
ggaccttggtactgggaggggttaggggtaaagcctgtcgtgcttagcgggggagcttgaccagaggtgatcttctcagggcctcac
ctggacctccctcctgattctgcatcccccttctcctcactagacttgaccacaggcccaaaatcctcatccctggcactctagaacccg
gccactccaaaaacctgacctgctctgtgtcctgggcctgtgagcagggaaacacccccgatcttctcctggttgatcagtgccccacctcc
ctgggccccaggactactcactcctcggtgctcataatcaccccacggccccaggaccacggcaccaacctgacctgtcaggtgaagttc
gctggagctggtgtgactacggagagaaccatccagctcaacgtcacctatgttccacagaaccaacaactggtatctttccaggagatgg
ctcagggaacaagagaccagagcaggagtgttcatggggccattggaggagctggtgttacagccctgctcgtctttgtctctgcctca
tcttctcatagtgaagaccacaggaggaaagcagccaggacagcagtgggcaggaatgacaccaccctaccacagggtcagcctcc
ccgaaacaccagaagaagtccaagttacatggccccactgaaacctcaagctgttcaggtgccgccccctactgtggagatggatgaggag
ctgcattatgcttccctcaactttcatgggatgaatcctccaaggacacctccaccgaatactcagaggtcaggaccagtgatga

A.5 *CD33^{ex2(flx/flx)}* minigene

gctagcatgccgctgctgctactgctgccccctgctgtgggcaggtgagtggtgtgtgggagaggggtgtcggataactcgtatagcatac
attatacgaagttagctgggcccagctgacctcgtttccccacaggggccctggctatggatccaaatttctggctgcaagtgcaggagtc
agtacggtacaggaggggtttgtcgtcctcgtgccctgcacttttccatcccataccctactacgacaagaactccccagttcatggttact

ggttcgggaaggagccattatatccagggactctccagtggccacaaacaagctagatcaagaagtacaggaggagactcagggcaga
ttccgcctccttggggatcccagtaggaacaactgctccctgagcatcgtagacgccaggaggaggataatggttcatacttcttcggatg
gagagaggaagtaccaaatacagttacaaatctccccagctctctgtgcatgtgacaggtgaggcacaggcttcagaagtggccgcaagg
gaagttcatgggtactgcagggcagggctgggatgggacctgggtactgggaggggttaggggtaaagcctataacttcgtatagcatac
attatacgaagttatgtcgtgcttagcgggggagcttgaccagaggttgatcttctctcaggccctcacctggacctccctcctgattctgcat
ccctcttttctcactagacttgaccacaggcccaaaatcctcatccctggcactctagaacccggccactccaaaaacctgacctgctc
tgtgtcctgggcctgtgagcagggaaacacccccgatcttctcctggtgtcagctgccccacctccctgggcccaggactactcactcct
cgggtctcataatcacccacggccccaggaccacggcaccaacctgacctgtcaggtgaagttcgctggagctggtgtgactacggaga
gaacctccagctcaacgtcacctatgtccacagaaccaacaactggtatctttccaggagatggctcagggaaacaagagaccagagc
aggagtgggtcatggggccattggaggagctgggtttacagccctgctcgtctttgtctctgcctcatcttctcatagtgaagaccacagg
aggaaagcagccaggacagcagtgggcaggaatgacccccacctaccacagggtcagcctccccgaaacaccagaagaagtccaag
ttacatggccccactgaaacctcaagctgttcaggtgcccccctactgtggagatggatgaggagctgcattatgcttccctcaacttcatg
ggatgaatccttccaaggacacctccaccgaatactcagaggtcaggaccagtgaaaccgt

A.6 Construct I (C allele)

atgccgtgctgctactgctgccctgctgtgggcaggtgagtggtgtggggagaggggtgtcgggctgggcccagctgacctcggtt
ccccacagggggccctggctatggatccaaatttctggctgcaagtgcaggagtcagtgacggtacaggagggttgtgctcctcgtgcc
tgcactttctgatgattccatcccatacctactacgacaagaactccccagttcatggttactggttccgggaaggagccattatatccagg
actctccagtggccacaaacaagctagatcaagaagtacaggaggagactcagggcagattccgcctccttggggatcccagtaggaaca
actgctccctgagcatcgtagacgccaggaggaggataatggttcatacttcttcggatggagagaggaagtaccaaatacagttacaaa
tctccccagctctctgtgcatgtgacaggtgaggcacaggcttcagaagtggccgcaagggaagttcatgggtactgcagggcagggctg
ggatgggacctgggtactgggaggggttaggggtaaagcctgtcgtgcttagcgggggagcttgaccagaggttgatcttctctcaggcc
ctcacctggacctccctcctgattctgcatccctcttttctcctcactagacttgaccacaggcccaaaatcctcatccctggcactctagaa
ccggccactccaaaaacctgacctgctctgtgtcctgggcctgtgagcagggaaacacccccgatcttctcctggtgtcagctgccccac

ctccctgggccccaggactactcactcctcgggtgctcataatcaccacggccccaggaccacggcaccaacctgacctgtcaggtgaa
gttcgctggagctgggtgtgactacggagagaacctccagctcaacgtcaccgccaccaacttctccctgctgaagcaggccggcgacgt
ggaggagaacccccggccccgtcttcacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttga
acagggaggtgtgtccagtttgttcagaatctcgggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagat
cgacatccatgtcatcatcccgtatgaaggtctgagcggcgaccaaattggccagatcgaaaaattttaaggtggtgtaccctgtggatg
atcatcactttaaggtgatcctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaag
gcatcgccgtgttcgacggcaaaaagatcactgtaacagggaccctgtggaacggcaaaaaattatcgacgagcgccctgatcaaccccg
acggctccctgctgttccgagtaacctcaacggagtgaccggctggcggtgtgcgaacgcattctggcgtaa

A.7 Construct I (T allele)

atgccgtgctgctactgctgccctgctgtgggcaggtgagtggtgtggggagaggggtgtcgggctgggccgagctgacctcggtt
ccccacaggggtcctggctatggatccaaatttctggctgcaagtgcaggagtcagtgcaggtacaggaggggttctgctcctcgtgccct
gcactttctgatgattccatcccatacctactacgacaagaactcccagttcatggttactggttccgggaaggagccattatattccaggga
ctctccagtggccacaaacaagctagatcaagaagtagcaggaggagactcagggcagattccgctccttggggatcccagtaggaacaa
ctgctccctgagcatcgtagacgccaggaggaggataatgggtcatacttcttcgatggagagaggaagtaccaaatacagttacaaat
ctccccagctctctgtgcatgtgacaggtgaggcacaggcttcagaagtggccgcaagggaagttcatgggtactgcagggcagggtgg
gatgggaccctgggtactgggaggggttaggggtaagcctgtcgtgcttagcgggggagcttgaccagaggttgatcttctcaggccct
cacctggacctccctcctgattctgcatccctctttctcctactagacttgaccacaggcccaaaatcctcatccctggcactctagaacc
cggccactccaaaaacctgacctgctgtgtcctgggcctgtgagcagggaacacccccgatcttctcctggtgtcagctgccccacct
ccctgggccccaggactactcactcctcgggtgctcataatcaccacggccccaggaccacggcaccaacctgacctgtcaggtgaagt
tcgctggagctgggtgtgactacggagagaacctccagctcaacgtcaccgccaccaacttctccctgctgaagcaggccggcgacgtgg
aggagaacccccggccccgtcttcacactcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaac
aggaggtgtgtccagtttgttcagaatctcgggggtgtccgtaactccgatccaaaggattgtcctgagcggtgaaaatgggctgaagatc
gacatccatgtcatcatcccgtatgaaggtctgagcggcgaccaaattggccagatcgaaaaattttaaggtggtgtaccctgtggatgat

catcactttaaggtgatcctgcactatggcacactggtaatcgacgggggttacgccgaacatgatcgactatttcggacggccgtatgaagg
catcgccgtgttcgacggcaaaaagatcactgtaacagggaccctgtggaacggcaaaaaattatcgacgagcgctgatcaaccccgga
cggctccctgctgttccgagtaacctcaacggagtgaccggctggcggctgtgcgaacgcattctggcgtaa

A.8 Construct I (*CD33*^{ex2(flX/flX)})

atccgctgctgctactgctgccctgctgtgggcaggtgagtggtgtggggagaggggtgtcgataacttcgtatagcatacattatac
gaagtatgctgggcccagctgaccctcgtttccccacaggggccctggctatggatccaaattctggctgcaagtgcaggagtcagtgc
ggtacaggaggggttgtgctcctcgtgccctgcactttctgatgattccatcccataccctactacgacaagaactccccagttcatggttact
ggttccgggaaggagccattatatccagggactctccagtggccacaaacaagctagatcaagaagtacaggaggagactcagggcaga
ttccgcctccttggggatcccagtaggaacaactgctccctgagcatcgtagacgccaggaggaggataatggttcatacttcttccggatg
gagagaggaagtacaaatacagttacaaatctcccagctctctgtgcatgtgacaggtgaggcacaggcttcagaagtggccgcaagg
gaagttcatgggtactgcagggcagggtgggatgggaccctggctactgggaggggttaggggtaaagccataacttcgtatagcatac
attatacgaagttatgtcgtgcttagcgggggagcttgaccagaggtgatcttctctcagggccctcacctggaccctccctcctgattctgcat
cccctcttctcctcactagacttgaccacaggcccaaatcctcatccctggcactctagaacccggccactccaaaaacctgacctgctc
tgtgtcctgggcctgtgagcagggaacacccccgatcttctcctggtgtcagctgccccacctccctgggccccaggactactcactcct
cgggtgctcataatcacccacggccccaggaccacggcaccaacctgacctgtcaggtgaagttcgtgagctggtgtgactacggaga
gaaccatccagctcaacgtcaccgccaccaacttctccctgctgaagcaggccggcgacgtggaggagaaccccgccccgtcttcaca
ctcgaagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacaggaggtgtgtccagtttgttcagaa
tctcgggggtgtccgtaactccgatccaaaggattgtcctgagcgggtgaaaatgggctgaagatcgacatccatgtcatcatcccgatgaag
gtctgagcggcgaccaaattgggccagatcgaaaaattttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatg
gcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagat
cactgtaacagggaccctgtggaacggcaaaaaattatcgacgagcgctgatcaaccccgacggctccctgctgttccgagtaacctc
aacggagtgaccggctggcggctgtgcgaacgcattctggcgtaa

A.9 Construct II (C allele)

gctagcatgccgctgctgctactgctgcccctgctgtgggcaggtgagtggtgtggggagaggggtgtcgggctgggccgagctgacc
ctcgtttcccccacaggggccttgctatggatccaaatttctggctgcaagtgcaggagtcagtgacggtacaggagggttgtgcgtctc
gtgccctgcactttctgatgattccatcccataccctactacgacaagaactccccagttcatggttactggttccgggaaggagccattatc
cagggactctccagtggccacaaacaagctagatcaagaagtacaggaggagactcagggcagattccgcctccttggggatcccagta
ggaacaactgctccctgagcatcgtagacgccaggaggaggataatggttcatacttcttcggatggagagaggaagtaccaaatacag
ttacaaatctccccagctctctgtgcatgtgacaggtgaggcacaggcttcagaagtggccgcaagggaagttcatgggtactgcagggca
gggctgggatgggaccctgggtactgggaggggttaggggtaaagcctgtcgtgcttagcgggggagcttgaccagaggttgatcttctc
caggccctcacctggaccctccctcctgattctgcatcccctcttctcctcactagactgaccacaggcccaaatcctcatccctggcac
tctagaacccggccactccaaaaacctgacctgctctgtgctcctgggcctgtgagcagggaacacccccgatcttctcctggtgtcagctg
ccccacctccctgggccccaggactactcactcctcgggtgctcataatcacccacggccccaggaccacggcaccaacctgacctgtc
aggtgaagttcgtggagctggtgtgactacggagagaacctccagctcaacgtcaccccgtgctgctactgctgcccctgctgtgggc
aggtgagtggtgtggggagaggggtgtcgggctgggcccagctgaccctcgtttcccccacaggggccttggtatggatccaaatttct
ggctgcaagtgcaggagtcagtgacggtacaggaggggttgtgcgtcctcgtgccctgcactttctgatgattccatcccataccctactacg
acaagaactccccagttcatggttactggttccgggaaggagccattatccagggactctccagtggccacaaacaagctagatcaaga
agtacaggaggagactcagggcagattccgcctccttggggatcccagtaggaacaactgctccctgagcatcgtagacgccaggagga
gggataatggttcatacttcttcggatggagagaggaagtaccaaatacagttacaaatctccccagctctctgtgcatgtgacaggtgagg
cacaggcttcagaagtggccgcaagggaagttcatgggtactgcagggcagggtgggatgggaccctgggtactgggaggggttagg
ggtaaagcctgtcgtgcttagcgggggagcttgaccagaggtgacttctctcaggccctcacctggaccctccctcctgattctgcatccc
ctcttctcctcactagacttgaccacaggcccaaatcctcatccctggcactctagaacccggccactccaaaaacctgacctgctctgtg
tcttgggcctgtgagcagggaacacccccgatcttctcctggtgtcagctgccccacctccctgggccccaggactactcactcctcgggt
gctcataatcacccacggccccaggaccacggcaccaacctgacctgtcaggtgaagttcgtggagctggtgtgactacggagagaa
ccatccagctcaacgtcacgccaccaacttctccctgctgaagcaggccggcgacgtggaggagaacccccggccccgtcttcacactcg

aagatttcgttggggactggcgacagacagccggctacaacctggaccaagtccttgaacaggagggtgtgtccagtttgttcagaatctc
gggggtgtccgtaactccgatccaaaggattgtcctgagcgggtgaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggctt
gagcggcgaccaaattgggccagatcgaaaaattttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcac
actggtaatcgacgggggttacccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcact
gtaacaggggaccctgtggaacggcaaaaaattatcgacgagcgctgatcaaccccgacggctccctgctgttccgagtaaccatcaac
ggagtgaccggctggcggctgtgcgaacgcattctggcgtaaaccggt

A.10 Construct III (C allele)

atgccgctgtgctactgtgccccctgtgtgggcaggtgagtggtgtggggagaggggtgtcgggctgggccgagctgacctcgttt
ccccacagggggccctggctatggatccaaatgtcttcacactcgaagatttcgttggggactggcgacagacagccggctacaacctggac
caagtccctgaacaggagggtgtgtccagtttgttcagaatctcggggtgtccgtaactccgatccaaaggattgtcctgagcgggtgaaat
gggctgaagatcgacatccatgtcatcatcccgtatgaaggctgagcggcgaccaaattgggccagatcgaaaaattttaaggtggtgta
ccctgtggatgatcatcactttaaggtgatcctgcactatggcacactggtaatcgacgggggttacgccgaacatgatcgactatttcggacg
gccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacaggggaccctgtggaacggcaaaaaattatcgacgagcgct
gatcaaccccgacggctccctgctgttccgagtaaccatcaacggagtaccggctggcggctgtgcgaacgcattctggcgtgatgaca
gctctctgtgcatgtgacaggtgaggcacaggcttcagaagtggccgcaagggaagttcatgggtactgcagggcagggctgggatggg
acctgtgtactgggaggggttaggggtaaagcctgtcgtgcttagcgggggagcttgaccagaggtgatcttctcaggccctcacctg
gacctccctcctgattctgcatcccccttttctcctcactagacttgaccacaggcccaaaatcctcatccctggcactctagaacccggcc
actccaaaaacctgacctgctctgtcctgggcctgtgagcagggaacacccccgatcttctcctggtgtcagctgccccacctccctgg
gccccaggactactcactcctcgggtgctcataatcacccacggccccaggaccacggcaccaacctgacctgtcaggtgaagttcgctg
gagctggtgtgactacggagagaaccatccagctcaacgtcacccaagacgcaaaaacataaagaaaggcccgccgcccattctatccg
ctggaagatggaaccgctggagagcaactgcataaggctatgaagagatacgccctggttctggaacaattgcttttacagatgcacatat
cgaggtggacatcacttacgtgagtacttcgaaatgtccgttcgggtggcagaagctatgaacgatatgggctgaatacaaatcacagaat
cgtcgtatgcagtgaaaactctcttcaattctttatgccgggtgttgggcgcgttatttatcgagttgcagttgcgcccgcaacgacattataa

tgaacgtgaattgctcaacagtatgggcatttcgcagcctaccgtggtgttcgtttccaaaaaggggtgcaaaaaatttgaacgtgcaaaaa
aagctcccaatcatccaaaaattattatcatggattctaaaacggattaccagggatttcagtcgatgtacacgttcgtcacatctcatctacct
cccggttttaataacgattttgtgccagagtccttcgatagggaagacaattgcactgatcatgaactcctctggatctactggctctgcct
aaaggtgtcgtctgcctcatagaactgcctgcgtgagattctcgcagccagagatcctattttgcaatcaaatattccggatactgcgat
tttaagtgtgttcattccatcacgggtttggaatgtttactacactcggatatttgatatgtggatttcgagtcgtcttaatgtatagattgaagaa
gagctgtttctgaggagccttcaggattacaagattcaaagtgcgtgctggtgccaaacctattctccttcttcgcaaaaagcactctgattga
caaatacatttatctaatttacagaaattgcttctggtggcgctccctctctaaggaagtcggggaagcgggtgccaaagaggtccatctg
ccaggtatcaggcaaggatatgggctcactgagactacatcagctattctgattacacccgaggggggatgataaacggggcgcggtcggta
aagtggtccatttttgaagcgaaggttggtgatctggataccgggaaaacgctgggcgttaatcaaagaggcgaactgtgtgtgagaggtc
ctatgattatgtccggttatgtaacaatccggaagcgaccaacgccttgattgacaaggatggatggctacattctggagacatagcttactg
ggacgaagacgaacacttctcatcgttgaccgcctgaagtctctgattaagtacaaaaggctatcaggtggctcccgtgaattggaatccat
cttgctccaacaccccaacatcttcgacgcaggtgtcgcaggtcttcccgacgatgacgccgggtgaacttcccgcccgctgtgttttga
gcacggaaagacgatgacggaaaaagagatcgtggattacgtcgccagtcaagtaacaaccgcgaaaaagtgcgcggaggagttgtgt
ttgtggacgaagtaccgaaaggtcttaccgaaaaactgcagcaagaaaaatcagagagatcctcataaaggccaagaagggcggaag
atcgccgtgtga

A.11 Construct III (T allele)

atgccgctgctgctactgctgcccctgctgtgggcaggtgagtggtgtggggagaggggtgtcgggctgggcccagctgacctcggtt
ccccacaggggtcctggctatggatccaaatgtcttcacactcgaagatttcgttggggactggcgacagacagccggctacaacctggac
caagtccttgaacagggaggtgtgtccagttgttcagaatctcggggtgtccgtaactccgatccaaaggattgtcctgagcgggtgaaaat
gggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgagcggcgaccaaattgggccagatcgaaaaattttaaggtggtga
ccctgtggatgatcatcactttaaggtgatcctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacg
gccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacagggaccctgtggaacggcaaaaaattatcgacgagcgcct
gatcaaccccgacggctccctgctgttccgagtaacctcaacggagtgaccggctggcggctgtgcgaacgcattctggcgtgatgaca

gctctctgtgcatgtgacaggtgagggcacaggcttcagaagtggccgcaagggaagttcatgggtactgcagggcagggctgggatggg
accctgggtactgggagggggttaggggtaaagcctgtcgtgcttagcgggggagcttgaccagaggtgatcttctcaggcctcacctg
gacctccctcctgattctgcatccccctttctcctcactagacttgaccacaggcccaaatcctcatccctggcactctagaacccggcc
actccaaaacctgacctgctctgtgctctgggcctgtgagcaggggaacacccccgatcttctcctgggtgtcagctgccccacctccctgg
gccccaggactactcactcctcgggtgctcataatcacccacggccccaggaccacggcaccaacctgacctgtcaggtgaagttcgtg
gagctgggtgactacggagagaacctccagctcaacgtcacccaagacgcaaaaacataaagaaggccccggcgccattctatccg
ctggaagatggaaccgctggagagcaactgcataaggctatgaagagatacgccctgggtcctggaacaattgcttttacagatgcacatat
cgaggtggacatcacttacgtgagtacttcgaaatgtccgttcgggtggcagaagctatgaacgatatgggctgaatacaaatcacagaat
cgtcgtatgcagtgaaaactctcttcaattctttatgccgggtgtggcgcggtatttatcggagttgcagttgcgcccgcgaacgacattataa
tgaacgtgaattgtcaacagtatgggcatttcgcagcctaccgtgggtgttcgttccaaaagggggtgcaaaaaatttgaacgtgcaaaaa
aagctcccaatcatcaaaaaattattatcatggattctaaaacggattaccagggatttcagtcgatgtacacgttcgtcacatctcatctacct
ccccgttttaataacgattttgtgccagagtccttcgatagggacaagacaattgcactgatcatgaactcctctggatctactggtctgcct
aaagggtgctcctcatagaactgcctgcgtgagattctcgcgatccagagatcctattttggcaatcaaatattccggatactgcgat
tttaagtgtgttcattccatcacggttttggaatgtttactacactcgatatttgatatgtggatttcgagtcgtttaatgtatagattgaagaa
gagctgtttctgaggagcctcaggattacaagattcaaagtgcgtgctgggtgccaaccctattctccttctcgccaaaagcactctgattga
caaatagatttatctaatttacagaaattgcttctgggtggcgctccccctcttaaggaagtcggggaagcgggtgccaagagggtccatctg
ccaggtatcaggcaaggatatgggctcactgagactacatcagctattctgattacacccgagggggatgataaacggggcgcggtcggtgta
aagttgtccatttttgaagcgaagggtgtggatctggataccgggaaaacgctgggcgttaataaagaggcgaactgtgtgtgagaggtc
ctatgattatgtccggttatgtaacaatccggaagcgaccaacgccttgattgacaaggatggatggctacattctggagacatagcttactg
ggacgaagacgaacacttctcatcgttgaccgcctgaagtctctgattaagtacaaaggctatcaggtggctcccgtgaattggaatccat
cttgcctcaacaccccaacatcttcgacgcaggtgtcgcaggtcttcccgacgatgacgccggtgaacttccgcccgcgtgtgttttggga
gcacggaaaagacgatgacggaaaaagagatcgtggattacgtgccagtcagtaacaaccgcgaaaaagtgcgcggaggaggtgtgt

ttgtggacgaagtaccgaaaggtcttaccgaaaactcgacgcaagaaaaatcagagagatcctcataaaggccaagaaggcggaag
atcgccgtgtga

A.12 Construct III (*CD33^{ex2(fl/flu)}*)

gctagcatgccgctgctgctactgctgccccctgctgtgggcaggtgagtgctgtggggagaggggtgtcggataactcgtatagcatac
attatacgaagttatgctgggccgagctgacctcggttccccacaggggccctggctatggatccaaatgtcttcacactcgaagattcggt
ggggactggcgacagacagccggctacaacctggaccaagtccttgaaacaggaggtgtgtccagttgtttcagaatctcgggggtgtccg
taactccgatccaaaggattgtcctgagcgggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggctgagcggcgac
caaatgggccagatcgaaaaattttaaggtggtgtaccctgtggatgatcatcactttaaggatcctgcactatggcacactggaatcg
acgggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacaggga
ccctgtggaacggcaacaaaattatcgacgagcgccgtgatcaaccccgacggctccctgctgttccgagtaacctcaacggagtaccg
gctggcggtgtgcaacgcattctggcggtgatgacagctctctgtgcatgtgacaggtgaggcacaggcttcagaagtggccgcaaggg
aagttcatgggtactgcagggcagggctgggatgggacctggtactgggaggggttaggggtaaaagcctataacttcgtatagcatac
ttatacgaagttatgtcgtgcttagcgggggagcttgaccagaggtgatcttctctcaggccctcacctggacctccctcctgattctgcatc
ccctcttctcctcactagacttgacctacaggcccaaaatcctcatccctggcactctagaacccggccactccaaaaacctgacctgctct
gtgtcctgggcctgtgagcaggggaacacccccgatcttctcctggtgtcagctgccccacctccctgggccccaggactactcactctc
ggtgtcataatcacccacggccccaggaccagggcaccaacctgacctgtcaggtgaagttcgctggagctggtgtgactacggagag
aacatccagctcaacgtcagaagacgccaaaaacataaagaaaggccggcgccattctatccgctggaagatggaaccgctggagag
caactgcataaggctatgaagagatacgccctggttctggaacaattgcttttacagatgcacatatcgaggtggacatcacttacgctgagt
acttcgaaatgtccgttcgggtggcagaagctatgaaacgatatgggctgaatacaaatcacagaatcgctgtatgcagtgaaaactctctca
attctttatgccggtgttggcgcggtatttatcgaggtgcagttgcgcccgcgaacgacatttataatgaacgtgaattgctcaacagtatgg
gcatttcgcagcctaccgtggtgttcgtttccaaaaaggggttgcaaaaaatttgaaacgtgcaaaaaagctcccaatcatcaaaaaattatt
atcatggattctaaaacggattaccagggatttcagtcgatgtacacgttcgtcacatctcatctacctcccggttttaataacgattttgtgc
cagagtccttcgatagggacaagacaattgcactgatcatgaactcctctggatctactggtctgcctaaagggtgtcgtctgcctcatagaac

tgctgcgtgagattctcgcatgccagagatcctatTTTTGGCAATCAAATCATTCCGGATACTCGGATTTAAGTGTGTTCCATTCCATCAGGTT
TTGGAATGTTTACTACACTCGGATTTGATATGTGGATTCGAGTCGTCTTAATGTATAGATTGAAGAAGAGCTGTTTCTGAGGAGCCTTCAGGAT
TACAAGATTCAAAGTGCCTGCTGGTGCCAACCTATTCTCTTCTCGCCAAAAGCACTCTGATTGACAAATACGATTATCTAATTTACACGAA
ATTGCTTCTGGTGGCGCTCCCTCTCTAAGGAAGTCGGGAAGCGGTTGCCAAGAGGTTCCATCTGCCAGGTATCAGGCAAGGATATGGGCT
CACTGAGACTACATCAGCTATTCTGATTACACCCGAGGGGGATGATAAACGGGCGCGGTCGGTAAAGTTGTTCCATTTTTGAAGCGAAGGT
TGTGGATCTGGATACGGGAAAACGCTGGGCGTTAATCAAAGAGGCGAAGTGTGTGTGAGAGGTCCTATGATTATGTCCGGTTATGTAACAA
TCCGGAAGCGACCAACGCCTTGATTGACAAGGATGGATGGCTACATTCTGGAGACATAGCTTACTGGGACGAAGACGAACACTTCTCATCG
TTGACCGCCTGAAGTCTCTGATTAAGTACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCCATCTTGTCCAACACCCCAACATCTTGA
CGCAGGTGTCGCAGGTCTCCCGACGATGACGCCGGTGAAGTCCCGCCGCGTGTGTTTTGGAGCACGGAAAGACGATGACGGAAAA
AGAGATCGTGGATTACGTCGCCAGTCAAGTAACAACCGCAAAAAGTGCgcggaggagttgtgtttgtggacgaagtaccgaaaggctt
accggaactcgacgcaagaaaaatcagagagatcctataaaggccaagaagggcggaaagatcgccgtgtgaaccggt