Protein-sphingolipid interactions and their role in immunity and disease

Maria Shamin

King's College, University of Cambridge

Supervised by Dr Janet Deane

Cambridge Institute for Medical Research (CIMR)

May 2021

This dissertation is submitted for the degree of Doctor of Philosophy

To my mother.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution. In accordance with the Clinical Medicine and Clinical Veterinary Medicine guidelines, this thesis does not exceed 60,000 words excluding figure legends, tables, appendices and bibliography.

Summary

Protein-sphingolipid interactions and their role in immunity and disease, Maria Shamin

Sphingolipids are a diverse class of lipids involved in fundamental cellular processes. In addition to their role in membrane architecture, sphingolipids can act as bioactive molecules, regulate membrane protein function and be presented to the immune system as lipid antigens by CD1 family molecules. The critical importance of sphingolipids for cell function is highlighted by the catastrophic diseases that result from defects in sphingolipid metabolism.

Lipid loading onto CD1d is facilitated by the lysosomal lipid transfer protein Saposin B, but the lipid loading mechanism is unknown. I used a range of highly sensitive biochemical assays to probe the interaction between CD1d and SapB. Interestingly, no direct interaction was detectable. Importantly, this suggests that additional proteins may promote their interaction.

In addition to their role as antigens, sphingolipids have roles in immunity by regulating the trafficking and function of immune receptors. Krabbe disease, a neurodegenerative disease caused by loss of function of the lysosomal sphingolipid hydrolase GALC, is accompanied by immune defects possibly driven by phagocytic cell dysfunction. I investigated how GALC deficiency alters the phenotype and function of the phagocytic cell line THP-1. This study was compromised by substantial off-target effects caused by genetic manipulation of THP-1 cells. Characterising these off-target changes identified perturbations in sphingolipid metabolism pathways, confirming that these cell lines are not appropriate for further studies.

The lysosomal enzyme acid ceramidase (AC) hydrolyses ceramide to sphingosine, thus regulating the levels of pro-apoptotic and pro-survival lipids. This role in apoptosis means that regulation of AC activity is of interest for cancer treatment. Although the structure of AC was recently determined, it is still unclear how it accesses membrane-embedded lipid substrates. I attempted to co-crystallise AC with Saposin A lipoprotein nanodiscs to understand this process. Although this system did not prove suitable for this project, I was able to determine the crystal structure of a new saposin oligomer highlighting even greater diversity of tertiary arrangement of these important lipid transfer proteins. I next asked whether AC processes galactosylceramide to psychosine, the cytotoxic GALC substrate accumulating in Krabbe disease, contributing to Krabbe disease pathology by increasing oligodendrocyte cell death. Surprisingly, I found that GALC knockout oligodendrocytes were not more

susceptible to cell death than wild-type oligodendrocytes, calling into question whether psychosine accumulation is the direct cause of oligodendrocyte cell death and demyelination in Krabbe disease. The sphingolipid-interacting adhesion protein PTPRM is lost at the surface of GALC-deficient oligodendrocytes, suggesting that specific sphingolipid-PTPRM interactions cause the mistrafficking of this adhesion protein in Krabbe disease oligodendrocytes, potentially contributing to demyelination. I probed the interaction of PTPRM with sphingolipids through X-ray crystallography studies and fluorescence-based trafficking assays.

Acknowledgements

My biggest thanks go to Dr Janet Deane, who has been an incredible supervisor and mentor. She has been extraordinarily generous with her time, advice, knowledge and support and taught me many scientific and writing skills as well as important life lessons that will be valuable far beyond my PhD. It has been such a pleasure to work with someone who is so passionate about their work and I feel very lucky to have spent four years working alongside her.

I would like to thank all past and present members of the Deane lab for making work even better, and especially for staying connected during the pandemic. My special thanks go to Eve Caroe, who has done the preliminary work leading to Chapter 6 and who has been such a great friend in the lab. The work of Dr Sam Spratley and Dr Stuart Fawke was also crucial for several of the experiments described in this thesis.

My thanks are extended to the Graham group in the Department of Pathology. Throughout my PhD, Dr Stephen Graham has consistently shared ideas, advice, and materials as well as generous help with crystallography experiments. Tomasz Benedyk, a friend and PhD student in the Graham group, has shared some skills that were instrumental to the work described in Chapter 3. I would also like to thank Dr Robin Antrobus and Dr Jack Houghton from the proteomics facility at CIMR for their kind help with my first proteomics experiment. Also at CIMR, Prof Randy Read kindly helped with a crystallography problem described in Chapter 5, and Dr David Gershlick helped me set up the trafficking assay detailed in Chapter 6.

Our collaborations with the group of Prof Paul Lehner at the Cambridge Institute of Therapeutic Immunology & Infectious Disease and with Prof Ian Brierley and Dr Katy Brown in the Department of Pathology made the work in Chapter 4 possible. Thanks are due to Dr Lidia Duncan for sharing key materials and protocols, to Dr James Williamson for running many mass spectrometry experiments, to Prof Ian Brierley for teaching me how to prepare samples for RNA sequencing in his lab, where I had a lot of fun and doughnuts, and to Dr Katy Brown for analysing the RNA sequencing data.

I also want to thank to my mentors at UCL who encouraged me to embark on this adventure. In particular, my undergraduate course director Prof Richard Milne for his contagious love of viruses and the thoughtfulness with which he guided me onto this path, and my undergraduate project supervisor Dr Nicholas Peters for believing in me from day one and giving me the most encouragement I have ever received. I have been very lucky to have had many exceptional teachers in my life and each of

them also deserves a thanks, because I would not be here without the enthusiasm and care that they put into their work.

I have been lucky to have been part of a great cohort of students, and I am really appreciative of the many hours of mutual support and complaining at the pub with Georgia, Skye, Helena and Martin, especially at the beginning of this journey. I am also very grateful to Matthew, Eve and Janeska for the many tea breaks we have had over the years.

My family also deserves a huge thanks. I want to thank my mother Valéria for teaching me to dream, and to work to make it happen. My father Slava, for teaching me to love practical work, and Lena, for her warm support for everything I do. Jacques, for wanting to understand every experiment I did. My sisters and brothers, for always surprising me.

Finally, I wish to thank the very special people who have brought me much happiness outside the lab in the past few years. Zoé, Emilia, Victoria, Quenton and Liangliang have all enriched my life during this time in more ways than I could describe. I am very deeply grateful for the friendship, love and support they have all given me. Luiza, Cecilia, Nikki and Marianna were the best housemates I could have hoped for. And, of course, I want to give a huge and loving thank you to Wolfy, for the hundred ways in which they made every day brighter, and for the endless love and support they gave me, on good and bad days, to help make this happen.

Table of contents

List of figures

List of tables

Abbreviations

4MβDG	4-methylumbelliferyl-β-D-galactopyranoside
AC	acid ceramidase
AC27	acid ceramidase (residues 27-395)
BAP	biotin acceptor peptide
BFP	blue fluorescent protein
BMDM	bone marrow derived macrophages
BMP	bis(monoacylglycero)phosphate
$BS(PEG)_5$	bis(sulfosuccinimidyl)suberate
C1P	ceramide-1-phosphate
Casp-3	caspase-3
CNS	central nervous system
CV	column volume
DDM	n-Dodecyl-beta-Maltoside
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
DSSO	disuccinimidyl sulfoxide
DTT	dithiothreitol
ECD	extracellular domain
EGFR	epidermal growth factor receptor
Endo H	Endoglycosidase H
EphA2	Ephrin A2
ER	endoplasmic reticulum
EV	empty vector
FBS	fetal bovine serum
Fn 3-4	fibronectin III domains 3-4
Gal-3S	D-Galactose 3-sulfate
Gal-4S	D-Galactose 4-sulfate
GALC	β-galactocerebrosidase, also known as β-galactosylceramidase
GalCer	galactosylceramide
Gb3	globotriaosylceramide

GCase	glucosylceramidase
GluCer	glucosylceramide
GM2A	GM2-activator protein
GPCR	G protein-coupled receptor
GPI	glycosylphosphatidylinositol
gRNA	guide RNA
H_2-DCF-DA	2',7'-dichlorodihydrofluorescein diacetate
H_6	hexahistidine
IFN-γ	interferon γ
IL	interleukin
iNKT	invariant natural killer T
KO	knock-out
LAPTM4B	Lysosome Associated Protein Transmembrane 4B
LBP	lipopolysaccharide-binding protein
LC-MS	liquid chromatography-mass spectrometry
LDAO	lauryldimethylamine oxide
LPS	lipopolysaccharide
MWCO	molecular weight cut-off
NKT	natural killer T
NMR	nuclear magnetic resonance
NPC1/2	Niemann–Pick C1/2 protein
PAM	protospacer adjacent motif
PBMC	peripheral blood mononuclear cells
PBS	phosphate-buffered saline
PBST	phosphate-buffered saline, Tween-20
PC	phosphatidylcholine
PEI	polyethylenimine
PMA	phorbol 12-myristate 13-acetate
PMP	plasma membrane profiling
PMSF	phenylmethylsulfonyl fluoride
PNS	peripheral nervous system
ProtA	Protein A
PSAP	prosaposin

PTPRM	Protein Tyrosine Phosphatase Receptor Type μ
qPCR	quantitative polymerase chain reaction
RMSD	root mean square deviation
RNAseq	RNA sequencing
ROS	reactive oxygen species
RUSH	retention using selective hooks
S1P	sphingosine-1-phosphate
S1PR	sphingosine-1-phosphate receptor
Sap	saposin A-D
SBP	Streptavidin-binding peptide
SDK1/2	sidekick-1/-2
SDS-PAGE	sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SEC-MALS	size exclusion chromatography–multi-angle light scattering
SMA-PME	spinal muscular atrophy with progressive myoclonic epilepsy
SMALP	styrene maleic acid lipid particle
TBE	Tris-Borate EDTA buffer
TBS	Tris-buffered saline
TCR	T-cell receptor
TEAB	triethylammonium bicarbonate buffer
TEV	Tobacco Etch Virus
TLC	thin layer chromatography
TLR	Toll-like receptor
TMT	Tandem mass tag
TUNEL	Terminal deoxynucleotidyl transferase dUTP nick end labeling
TX-100	Triton X-100
UGT8	UDP-galactose:ceramide galactosyltransferase 8
WCP	whole cell proteomics
WT	wild-type
α-GalCer	α-galactosylceramide
β_2m	β-2-microglobulin

1 Introduction

1.1 Sphingolipids in normal cell function

Sphingolipids are a very diverse class of lipids and an essential component of all eukaryotic cell membranes. Ceramide, the building block for complex sphingolipids, consists of a sphingoid backbone N-acylated with fatty acid (Fig. 1.1A). Complex sphingolipid species are formed by addition of various polar groups to the C1-hydroxyl position of ceramide (Fig. 1.1B). The major classes of complex sphingolipids are sphingomyelin, formed by addition of a phosphorylcholine head group; glycosphingolipids, formed by addition of mono-, di- or oligosaccharides to ceramide; and gangliosides, formed by addition of sialic acid residues to the carbohydrate head groups of glycosphingolipids. The diversity of sphingolipid species is further enhanced by differences in the length and degree of saturation the fatty acyl backbone, as well as variations in the sphingoid backbone[1,2]. These variations determine the biological functions of sphingolipids, underscoring the diversity of their roles. An additional layer of complexity is added by the compartmentalisation and dynamic regulation of individual sphingolipid species.

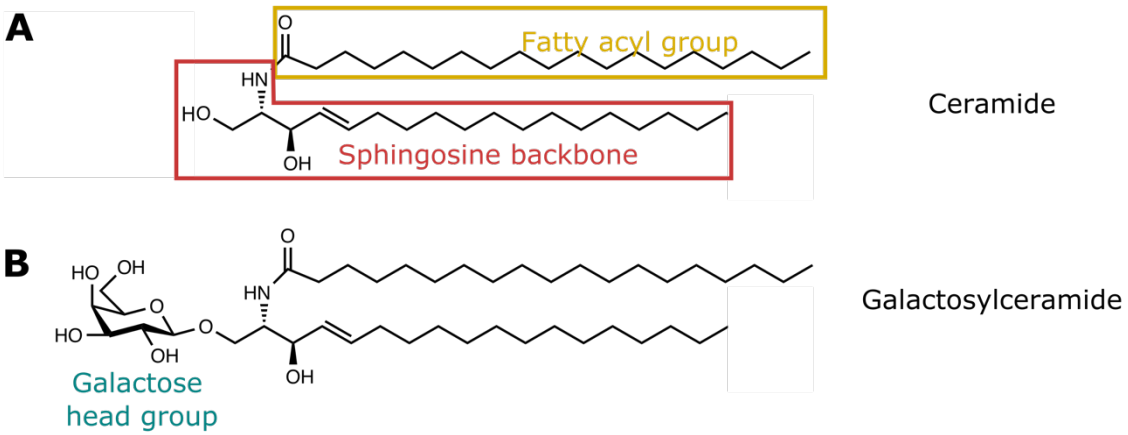

Figure 1.1. Sphingolipid structure
A. Skeletal formula of ceramide highlighting the sphingosine backbone and fatty acyl group. B. Skeletal formula of galactosylceramide.

1.1.1 The lipid raft concept: organisation of proteins and lipids into membrane microdomains

The primary function of lipids is to form a physical barrier between individual cells and their environment and segregate organelles from the cytosol. Sphingolipids contribute to this fundamental function and modulate the biophysical properties of membranes[3]. The majority of membrane lipids have kinked unsaturated hydrocarbon chains, adopting liquid-disordered membrane phase behaviour

characterised by low order and high fluidity. Sphingolipids have long, saturated or *trans*-monounsaturated hydrocarbon chains forming a tall and narrow cylinder shape with polar head groups capable of forming hydrogen bonds. These properties allow sphingolipids to pack together at higher densities than unsaturated membrane lipids and adopt an ordered, gel-like phase in membranes. Complex sphingolipids such as sphingomyelin and glycosphingolipids are enriched in the outer leaflet of the plasma membrane and favourably interact with cholesterol, which fluidises the gel-like phase to form a liquid-ordered phase characterised by dense packing, high order and high fluidity[3]. These properties give the plasma membrane resistance to mechanical stress[3]. The favourable interaction of cholesterol with complex sphingolipids is explained by the “umbrella model”. The hydroxyl polar head group of cholesterol is too small to shield the bulky, hydrophobic sterol rings from water. Tight packing of cholesterol with complex sphingolipids covers the sterol rings under the "umbrella" formed by large polar head groups[4].

The preferential interaction between cholesterol and saturated glycolipids including complex sphingolipids has been hypothesised to promote the formation of tightly packed membrane domains with distinct lipid and protein compositions from the bulk membrane, named lipid rafts[5]. The "raft hypothesis" has generated considerable controversy since its formulation because these domains were initially observed in model membranes and systems, including detergent-insoluble membrane fractions posited to be lipid rafts, but these could not be visualised in live cells[6]. However, accumulating evidence for the existence of membrane rafts in cells comes from recent studies using fluorescent lipid and protein probes in live cell super-resolution microscopy experiments combined with studies on membrane blebs with a lipid and protein composition similar to the plasma membrane, known as giant plasma membrane vesicles. These studies show that segregated membrane domains enriched in cholesterol, sphingolipids and specific proteins exist but these are highly dynamic and very small (tens of nanometres), making direct observation challenging[7].

Lipid rafts appear to facilitate the recruitment of specific proteins and the exclusion of others. Several factors targeting proteins to lipid rafts have been identified: post-translational modifications such as glycosylphosphatidylinositol (GPI) anchoring and palmitoylation (reversible covalent attachment of GPI or a fatty acid molecule, respectively), as well as transmembrane domain length and surface area[7-9]. Lipid rafts were originally thought to serve as a platform for cargo protein sorting and initiation of vesicular transport from the *trans* Golgi network to the plasma membrane[10]. The evidence for a central role of lipid rafts for this process is still inconclusive[11]. However, lipid rafts have been implicated in protein recycling from endosomes to the plasma membrane. The evidence supporting this is that

shortening of the transmembrane domain of a plasma membrane lipid raft protein is sufficient to exclude it from lipid rafts and results in its mistargeting to the lysosome[12].

A well-documented role for lipid rafts is the assembly of cell signalling platforms facilitating signal transduction. A number of transmembrane receptors, and in particular immune receptors, have been shown to associate with lipid rafts upon activation. Super-resolution microscopy of live B-cells showed that clustering of the B-cell receptor facilitates the stabilisation of an ordered lipid domain, thereby bringing antigen-bound B-cell receptors in close proximity with raft-associated activating signalling proteins and away from raft-excluded inhibitory regulators, resulting in downstream B-cell receptor signalling[13]. Evidence suggests that innate pattern recognition receptors of the Toll-like receptor (TLR) family also localise to lipid rafts upon stimulation. TLR4 is activated by bacterial lipopolysaccharide (LPS) binding to the TLR4 co-receptor MD-2. LPS is initially extracted by LPS-binding protein (LBP) and is transferred to the raft-associated, GPI-anchored protein CD14. Upon LPS binding, CD14 has been shown to recruit TLR4 to lipid rafts, transferring LPS to the TLR4:MD-2 complex and bringing TLR4 in proximity to activator proteins[14,15]. Reduction of cholesterol, sphingomyelin or glucosylceramide (GluCer) levels reduced the response of macrophages to LPS, indicating that these lipids are necessary for LPS-induced TLR signalling[14,16-18]. Köberlin *et al*. observed that perturbations to sphingolipid metabolism led to changes in TLR4 internalisation after LPS stimulation[19]. Accumulation of cholesterol in macrophages causes enhanced signalling of TLR4 but also TLR2, TLR7 and TLR9, suggesting the function of these receptors may also be regulated by cholesterol-enriched lipid rafts[20].

A distinction has been made between lipid rafts and ceramide enriched lipid domains, named ceramide-rich platforms. The head group of sphingomyelin can be cleaved by sphingomyelinases to form ceramide within lipid rafts at the plasma membrane. Sphingomyelinases can be activated by a number of stimuli such as the engagement of TNF receptor superfamily members including TNF-α receptor 1, the T-cell co-stimulatory receptor CD40 and the death receptors Fas and CD95, as well as stress stimuli including ultraviolet light and oxidative stress[21,22]. Ceramide has distinct biophysical properties from sphingomyelin and glycosphingolipids. Due to its small polar head group, ceramide is thought to displace cholesterol from lipid rafts, thereby decreasing membrane fluidity within lipid rafts and potentially modifying their protein content[21,23]. Ceramide is known to increase membrane permeability and introduce negative curvature[21]. It is also unique among sphingolipids in its capacity to spontaneously translocate to the inner leaflet of the plasma membrane, where it can interact with cytosolic proteins[21]. Finally, ceramide promotes the clustering of small nanodomains to form larger

microdomains in model membranes[24]. This facilitates the clustering of receptors and signalling molecules[25-27].

It is unclear if distinct receptors preferentially associate with sphingolipid domains of different composition. Considering the huge variety of sphingolipids, lipid rafts are likely to be heterogeneous and serve different functions. Our understanding of the roles of individual sphingolipid species and other glycolipid classes in lipid rafts is still very limited due to the lack of appropriate tools to study these differences.

1.1.2 The ceramide/sphingosine-1-phosphate cell death rheostat

Ceramide, sphingosine and their phosphorylated derivatives sphingosine-1-phosphate (S1P) and ceramide-1-phosphate (C1P) are interconvertible sphingolipid species regulating a cell death rheostat (Fig. 1.2). Ceramide has a well-documented role in mediating cell death. Although the molecular mechanisms of ceramide-mediated cell death are not fully understood, they are thought to involve mitochondrial membrane perturbation and/or triggering of cell death signalling in a cell-type, cellular compartment and ceramide species-dependent manner[2,28]. Cell death signalling may be promoted by the formation of ceramide-rich platforms activating signal transduction by death receptors as well as spontaneous flipping of ceramide to the cytoplasmic leaflet where it can initiate cell death signalling cascades[21]. Ceramide phosphorylation by ceramide kinase produces C1P, a pro-inflammatory and pro-survival lipid[2].

Ceramide catabolism forms sphingosine, which also possesses pro-apoptotic properties[28]. Sphingosine can in turn be converted to S1P by sphingosine kinases. S1P is an anti-apoptotic immune signalling molecule which promotes cell survival and migration via its action on G protein-coupled S1P receptors (S1PR). S1P is produced intracellularly and exported to the outer leaflet of the plasma membrane by transporters including ATP-binding cassette transporters and spinster homolog 2 (*SPNS2*)[29], resulting in autocrine activation of S1PRs on the same cell. Structural evidence indicates that S1P accesses the binding site of $S1PR_1$ by lateral diffusion within the plasma membrane[30]. Additionally, S1P can be extracted from the plasma membrane by lipid transfer proteins such as Apolipoprotein M. Lipoprotein complexes can then diffuse in the extracellular environment for autocrine and paracrine action[31]. S1P is more hydrophilic than other sphingolipids and has therefore been proposed to diffuse freely in the aqueous environment[32]; however, the majority of plasma S1P is associated with lipoproteins[33]. S1PR-independent intracellular actions of S1P have also been documented but their physiological relevance is still unclear[31].

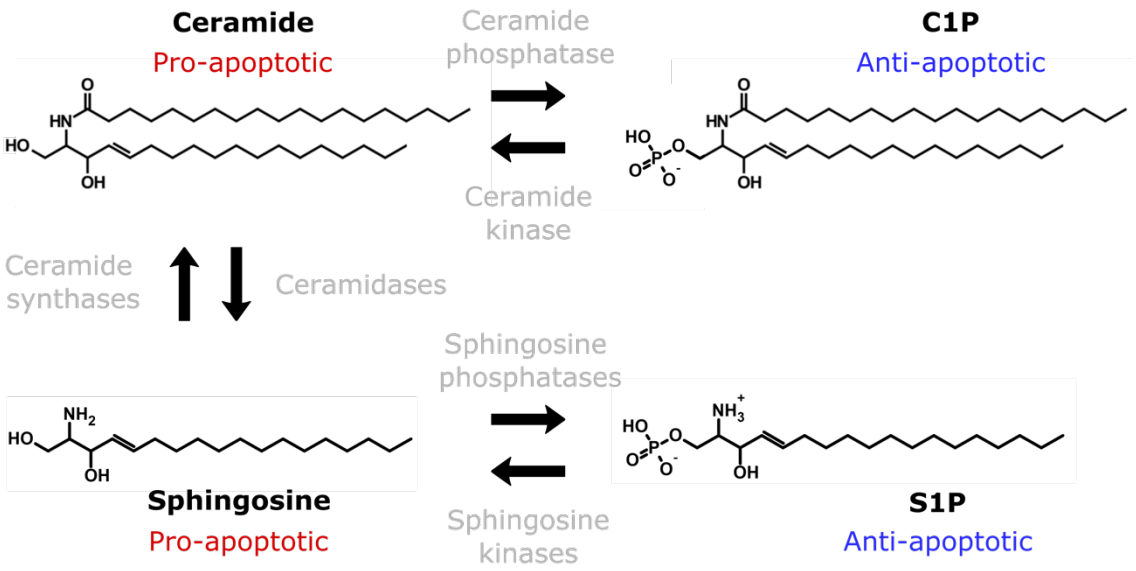

Figure 1.2. The sphingolipid cell death rheostat
Cell death and survival is regulated by the levels of pro-apoptotic ceramide and sphingosine and anti-apoptotic ceramide-1-phosphate (C1P) and sphingosine-1-phosphate (S1P). The interconversion of these sphingolipids is mediated by the action of ceramidases, ceramide synthases, kinases and phosphatases.

The levels of ceramide, sphingosine, C1P and S1P controlling the cell death rheostat is dynamically regulated by the action of sphingomyelinases, ceramide synthases, ceramidases, sphingosine and ceramide kinases, phosphatases and S1P lyases (Fig. 1.2). One of these players, acid ceramidase, will be described further in section 1.6.

1.1.3 Specific interactions between proteins and sphingolipids

An emerging concept in the field is that membrane-embedded sphingolipids can directly interact with membrane proteins. The first example of a sequence-specific interaction is the sphingolipid-binding motif VXXTLXXIY in the transmembrane helix of the vesicular transport protein p24. This motif mediates specific binding to a single species of sphingomyelin (d18:1/18:0) via binding of the transmembrane helix of p24 to both the lipid head group and the hydrocarbon chains of sphingomyelin. Sphingomyelin binding promotes the homodimerisation of p24 and regulates vesicular trafficking[34]. Subsequently, several transmembrane receptors containing a similar motif were confirmed to bind sphingolipids[35,36]. Among these, the lysosomal ceramide exporter Lysosome Associated Protein Transmembrane 4B (LAPTM4B) was found to bind ceramide. Ceramide binding induces a conformational change in the transmembrane region of LAPTM4B and promotes its dimerisation with the leucine transporter 4F2hc/LAT1, resulting in mTORC activation[37]. Sequence analysis also indicates that TLR3 and TLR5 possess a sphingolipid-binding motif within their

transmembrane domain[38], although whether it mediates specific binding to sphingolipids and the functional consequences of sphingolipid binding have not yet been investigated.

Sphingolipids, and in particular gangliosides, have been found to modulate the function of G protein-coupled receptors (GPCRs) through specific interactions. The ganglioside GM1 is thought to modulate the function of the serotonin$_{1A}$ receptor through binding to a sphingolipid-binding motif[39,40]. The epidermal growth factor receptor (EGFR) is maintained in its resting state in the absence of ligand by two separate lateral interactions with the head group of the ganglioside GM3: the first through a membrane-proximal lysine residue[41], and the second through terminal *N*-acetylglucosamine residues on EGFR N-linked glycans[42], demonstrating that protein glycosylation can contribute to sphingolipid binding. Besides regulating transmembrane protein oligomerisation and function, it is conceivable that specific interactions with sphingolipids may regulate the stability of transmembrane proteins or target them to lipid rafts, although these mechanisms have not yet been demonstrated.

Sphingolipid-protein interactions can also mediate intercellular binding. Circulating phagocytes exit the bloodstream and access inflammatory sites in tissues by a complex process involving binding of surface sialofucosylated glycans attached to proteins and lipids including gangliosides to the protein E-selectin expressed on the surface of endothelial cells[43]. In the nervous system, the myelin-associated glycoprotein facilitates the wrapping of the myelin sheath around neuronal axons by specifically binding to the gangliosides GD1a and GT1b on the apposed neuronal membrane[44]. Moreover, sulfatide exposed on the surface of apoptotic cells facilitates their phagocytosis by macrophages, probably via an interaction between a macrophage scavenger receptor and the sulfatide head group[45]. Several pathogens exploit sphingolipids as cellular entry receptors. For example, the human immunodeficiency virus envelope protein gp120 can bind galactosylceramide (GalCer) and globotriaosylceramide (Gb3) for viral entry[46] and *Shigella* targets activated CD4$^+$ T-cells through binding of the gangliosides GM1 and GM3 on the activated T-cell surface[47].

1.2 Sphingolipid synthesis and transport to the plasma membrane

The synthesis of sphingolipids is a complex stepwise and compartmentalised process, involving regulated sphingolipid transport and numerous synthetic enzymes working in a sequential manner. While the sphingolipid synthetic pathway is well understood, our understanding of sphingolipid trafficking pathways is still being refined. *De novo* synthesis of sphingolipids begins with formation of ceramide through a series of reactions on the cytosolic face of the endoplasmic reticulum (ER). Ceramide can translocate to the ER lumen, where it is used to synthesise GalCer by UDP-

galactose:ceramide galactosyltransferase 8 (UGT8)[48]. GalCer is transported to the Golgi apparatus by vesicular transport, where it can be sulfated to form sulfatide. Alternatively, ceramide is transported to the Golgi apparatus, where it is converted to sphingomyelin or GluCer. Ceramide transport from the ER to the *trans* Golgi network for sphingomyelin synthesis is effected through the action of ceramide transport protein (*CERT*). The C-terminal START domain of CERT extracts ceramide from ER membranes and transports it to *trans* Golgi regions at ER–Golgi membrane contact sites[49,50]. Ceramide is also thought to be transported to the *cis* Golgi network by vesicular transport for GluCer synthesis[51]. GluCer is then transported to the *trans* Golgi by the lipid transfer protein 4-phosphate adaptor protein 2 (*FAPP2*) for synthesis of complex glycosphingolipids and subsequent vesicular transport to the plasma membrane[52].

1.3 Sphingolipid degradation

Sphingolipid degradation and recycling is essential for the maintenance of sphingolipid homeostasis. The major sphingolipid degradative pathway occurs in the lysosome, where sphingolipids are subjected to sequential degradation by lysosomal sphingolipid hydrolases with the help of the lipid transfer proteins GM2 activator (GM2A) and saposins. Sphingolipids are trafficked to lysosomal compartments through vesicular trafficking pathways including endocytic, phagocytic and autophagic pathways[53,54]. The limiting membrane of the lysosome is protected from hydrolysis by the glycocalyx, a dense layer of oligosaccharides on the luminal side of lysosomal membrane glycoproteins[55]. The generation of sphingolipid-enriched intraluminal vesicles by inner budding of endosomal membranes renders these lipids accessible to lysosomal hydrolases[53]. Intraluminal vesicle membranes have high curvature, high concentration of the anionic phospholipid bis(monoacylglycero)phosphate (BMP) and low concentration of cholesterol. These conditions, combined with the low pH of the lysosome, promote the activity of sphingolipid hydrolases and saposins[53].

Sphingosine, the final product of sphingolipid catabolism, can exit the lysosome and enter the sphingolipid salvage pathway by re-acylation of sphingosine to form ceramide on the cytosolic face of the ER. The salvage pathway contributes to 50% to 90% of sphingolipid biosynthesis[56]. The mechanisms of sphingosine efflux from the lysosome and delivery to the ER have not yet been identified. The lysosomal cholesterol exporter NPC1 has been proposed to function as a sphingosine exporter[57].

In addition to the lysosomal sphingolipid degradation pathway, sphingolipid degrading enzymes are present in numerous cellular compartments, including the plasma membrane, ER, Golgi, nucleus and

mitochondria. Several of these enzymes, for example sphingomyelinases, are known to be responsive to various stimuli, thus dynamically regulating the levels of individual sphingolipid species in a localised manner[1].

1.3.1 Defective sphingolipid degradation: sphingolipidoses

Defective sphingolipid degradation due to loss-of-function mutations in lysosomal sphingolipid hydrolases or their respective helper proteins causes a subset of lysosomal storage diseases called sphingolipidoses (Fig. 1.3). These are severe and currently incurable autosomal recessive disorders. Loss of enzyme function results in the accumulation of primary sphingolipid substrates and/or their derivatives in specific cell types and tissues. The accumulation of undigested substrate is primarily lysosomal, but their accumulation has been observed throughout the cell, including at the plasma membrane[58,59]. While each sphingolipidosis is associated with unique pathologies, several pathogenic cascades are shared by lysosomal storage diseases, including altered cellular trafficking and degradation, oxidative stress, calcium homeostasis, inflammation and activation of cell death pathways[60,61]. The precise mechanism by which sphingolipids initiate cellular dysfunction remains unclear, but may include disruption of membrane integrity, architecture, and functional domains. In addition to the primary substrate, deacylated forms of the accumulated sphingolipid accumulate in a number of sphingolipidoses[62]. These lysolipids have detergent-like properties and are thought to damage membrane integrity, contributing to the pathogenesis of sphingolipidoses.

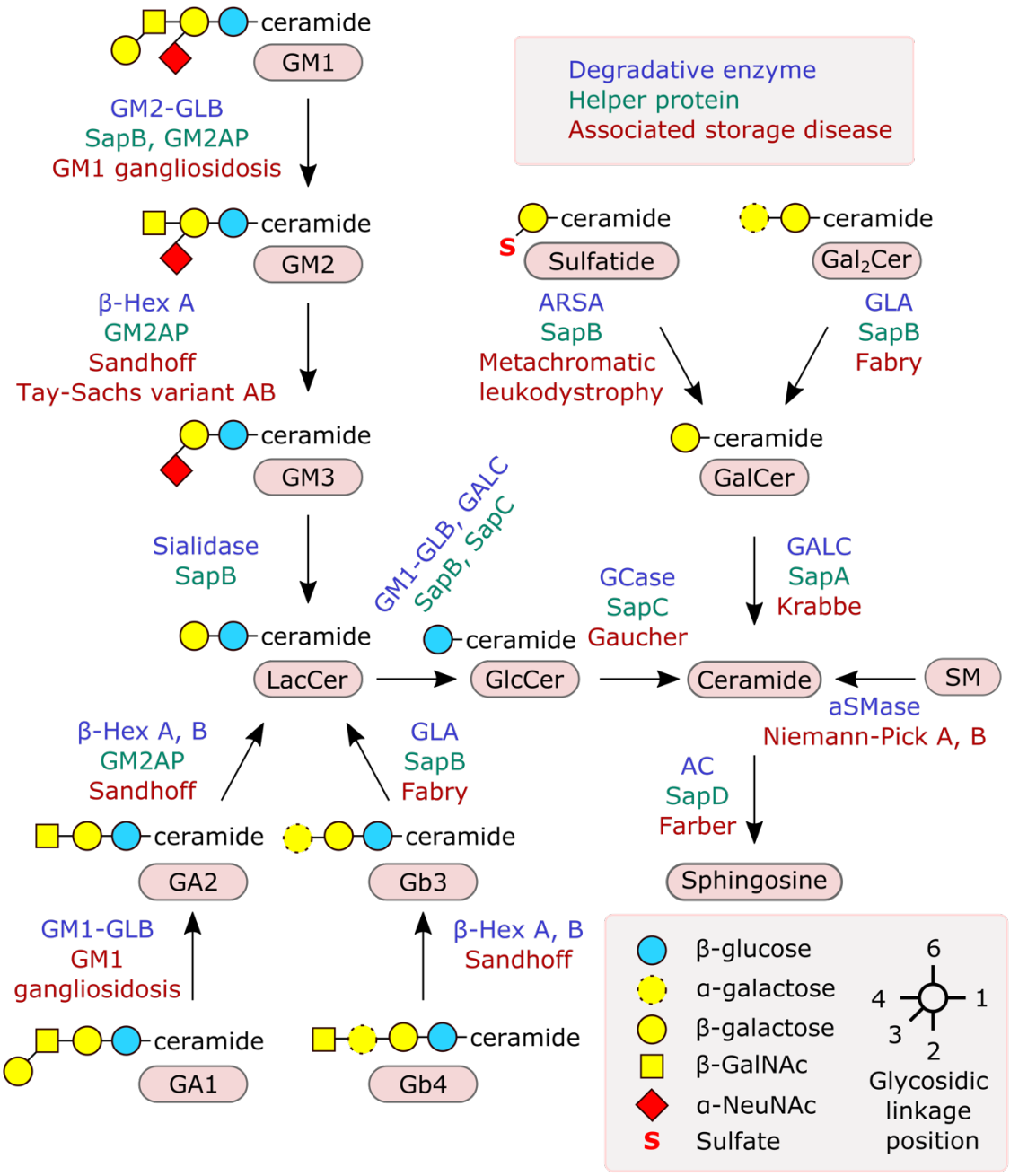

Figure 1.3. Lysosomal storage diseases caused by defective glycosphingolipid catabolism
Hereditary diseases (red) caused by genetic defects of individual hydrolase enzymes (blue) or activator proteins (green) in the glycosphingolipid and ganglioside degradation pathway. Abbreviations: N-Acetyl-β-D-galactosamine (β-GalNAc); α-N-acetylneuraminic acid (α-NeuNAc); lactosylceramide (LacCer); glucosylceramide (GlcCer), galactosylceramide (GalCer), digalactosylceramide (Gal_2Cer); sphingomyelin (SM); β-galactosidase (GLB); β-hexosaminidase (β-Hex); α-galactosidase (GLA); arylsulfatase A(ARSA); glucosylceramidase (GCase); β-galactocerebrosidase (GALC); acid ceramidase (AC); acid sphingomyelinase (aSMase); GM2 activator protein (GM2AP).

1.4 Saposins

Lipids are embedded within membranes[63] and thus are not directly accessible to soluble sphingolipid degrading enzymes. Lysosomal lipid transfer proteins, namely saposins and GM2A provide sphingolipid hydrolases access to their substrates. There are four saposin proteins named Saposin A, B, C, D (SapA-D). Each saposin functions in conjunction with specific hydrolases to facilitate the degradation of different sphingolipids and the loss of saposin function phenotypically resembles the loss of the associated hydrolases[64-69] (Fig. 1.3). Saposins are expressed ubiquitously and conserved across vertebrates[70]. All four saposins are the products of the proteolytic cleavage of a single precursor glycoprotein, prosaposin (PSAP). PSAP is folded and glycosylated in the ER and transported to endolysosomal compartments by the sorting receptor Sortilin[71]. Sortilin also delivers acid aspartyl protease cathepsin D[72], which cleaves PSAP to form SapA, SapB, SapC and SapD in the lysosome[73].

Several lysosomal lipid transfer proteins have been identified besides saposins. Interestingly, acid sphingomyelinase (*SMPD1*) does not require an external lipid transfer protein but possesses an N-terminal saposin homology domain that allows it to directly access its substrate sphingomyelin[74]. GM2A is the essential co-factor for the degradation of the ganglioside GM2 by β-hexosaminidase A[75,76]. Finally, Niemann–Pick C2 (NPC2) solubilises cholesterol for transfer to the lysosomal cholesterol exporter Niemann-Pick C1 (NPC1)[77]. These additional lipid transfer proteins will not be discussed further since my work has focused on the saposins.

1.4.1 Structural features of saposins

Saposins are small (approximately 80 amino acids, 8-12 kDa) non-enzymatic proteins. Each saposin possesses 6 conserved cysteines and a conserved N-glycosylation site at the same position[78]. Although individual saposins have low sequence identity (< 35%), they adopt a common fold consisting of four amphipathic α-helices (Fig. 1.4A). The N- and C-terminal α-helices α1 and α4 form a stem stabilised by two disulfide bonds. The central α-helices α2 and α3 form a hairpin motif maintained by a single disulphide bond. Two flexible hinge loops separate the stem and hairpin motifs, allowing saposins to open and close in a jack-knife manner. The internal face of saposins is hydrophobic, allowing for lipid binding (Fig. 1.4B). Several X-ray crystallography and nuclear magnetic resonance (NMR) structures demonstrating different conformational states are available for each saposin[79-86] (Fig. 1.4C-K). Saposin monomers can adopt a closed, globular conformation to shield the internal hydrophobic surface from the aqueous environment (Fig. 1.4C, G and J). In their open form, saposins can dimerise, enclosing a hydrophobic cavity (Fig. 1.4D, F, H, K). This cavity can be used to solubilise lipids, as illustrated by a crystal structure of SapB enclosing the hydrophobic moiety of a lipid molecule (Fig. 1.4F). SapA also

presents lipids substrates to its cognate enzyme as a dimer, although these dimers possess different conformations (Fig. 1.4D). Although the precise mechanism of lipid extraction from membranes is not known, the interaction of saposins with membranes is likely to require opening of the saposins to expose the internal hydrophobic face, as observed in an NMR structure of detergent micelle-associated SapC in its open conformation (Fig. 1.4I).

1.4.2 The liftase and solubiliser models of saposin function

There are two proposed mechanisms for how saposins assist in lipid presentation to hydrolases. In the solubiliser model, saposins extract lipids from membranes and present them as soluble protein-lipid complexes to cognate enzymes in solution (Fig. 1.5A). The liftase model proposes that saposins bind to membranes and destabilise them, thereby providing membrane-associated enzymes access to lipid substrates (Fig. 1.5B). SapA and SapB are thought to be solubilisers, whereas SapC and SapD are proposed to function as liftases. The solubiliser model has been demonstrated by a crystal structure of SapA in complex with its soluble cognate enzyme obtained by our group[80]. SapB is also proposed to function as a solubiliser, based on its capacity to extract radiolabelled lipids from liposomes[87] and a crystal structure of SapB enclosing a single lipid molecule[82] (Fig. 1.4F). The liftase mechanism has been observed for SapC. Atomic force microscopy combined with fluorescence microscopy and Förster resonance energy transfer microscopy showed that SapC binds and remodels lipid bilayers, recruiting its cognate enzyme glucosylceramidase (GCase) and promoting its activity and/or access to substrate[88,89]. Similarly, SapD associates with membranes and enhances the activity of its cognate enzyme at the membrane interface[90]. *In vitro* studies have showed that SapD increases the permeability of large unilamellar vesicles[91] and breaks down large lipid vesicles into smaller vesicles at very high concentrations[92]. The physiological relevance and implications of these findings are still unclear but suggest that SapD introduces membrane perturbations and functions as a liftase.

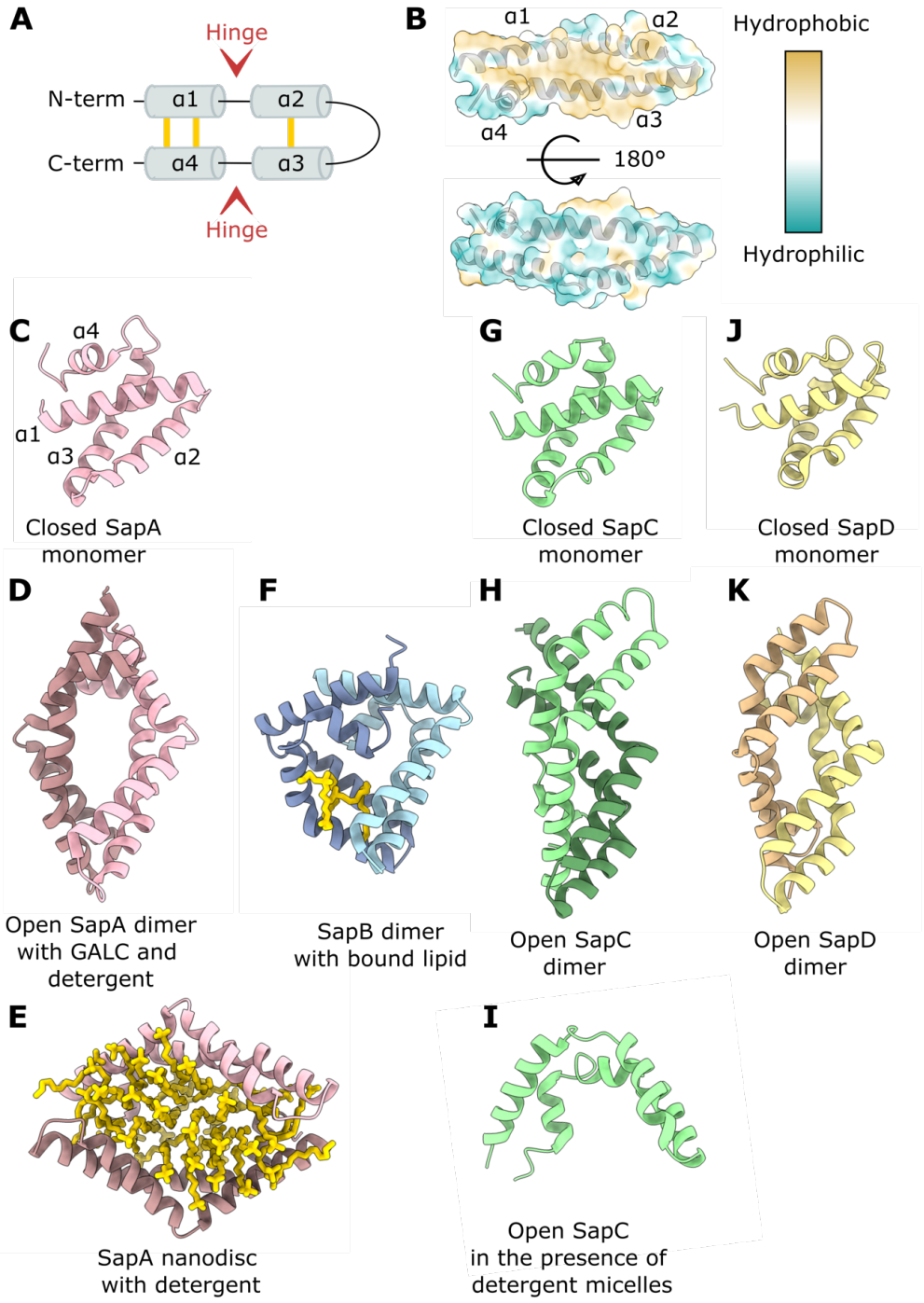

Figure 1.4. Structure of saposins

A. Topology diagram of the saposin fold. Disulphide bonds are shown in yellow. B. Ribbon and transparent surface representation of an open SapA monomer (PDB ID: 4DDJ) with the surface coloured based on residue hydrophobicity. (C-K) Ribbon representation of published saposin structures. C. Crystal structure of a closed human SapA monomer (PDB ID: 2DOB). D. Crystal structure of GALC and LDAO-bound murine SapA (PDB ID: 5NXB). E. Crystal structure of a human SapA-LDAO nanodisc (PDB ID: 4DDJ). Detergent molecules are shown in stick representation in yellow. F. Crystal structure of phosphatidylethanolamine (PE)-bound human SapB (PDB ID: 1N69). PE is shown in stick representation in yellow. G. Crystal structure of a closed human SapC monomer (PDB ID: 2GTG). H. Crystal structure of a human SapC dimer (PDB ID: 2QYP). I. NMR structure of human SapC in the presence of SDS micelles (PDB ID: 1SN6). J. Crystal structure of a closed human SapD monomer (PDB ID: 3BQP). K. Crystal structure of a murine SapD dimer (PDB ID: 5U85).

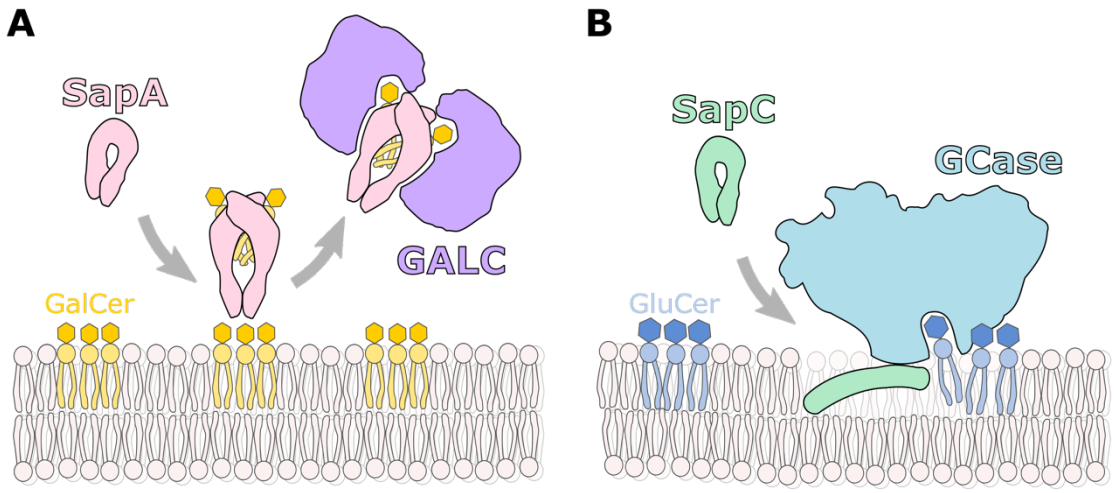

Figure 1.5. Solubiliser and liftase models of saposin function
A. Schematic diagram illustrating the solubiliser model of saposin function. SapA (pink) extracts GalCer (yellow) and presents it to the hydrolase GALC (purple) in solution. Adapted figure, courtesy of Dr Janet Deane. B. Schematic diagram illustrating the liftase model of saposin function. SapC (green) binds and perturbs lipid bilayers to promote the activity of glucosylceramidase (GCase; sky blue) on its substrate glucosylceramide (GluCer; dark blue) at the membrane surface.

1.4.3 Saposin nanodiscs for structural studies

The lipid binding properties of saposins have been employed to form stable and uniform lipoprotein nanodiscs suitable for NMR[93] and electron microscopy[94] studies of membrane proteins (Fig. 1.4E). SapA has been found to assemble with a very wide range of lipids to form nanodiscs[95]. Interestingly, all saposins can form nanodiscs[95], regardless of whether they are thought to function according to the solubiliser or liftase model. This suggests that each saposin has a specific mode of action when presenting substrates to its cognate enzyme, but all saposins also share the ability to form lipoprotein discs *in vitro*. It is unknown whether lipoprotein disc formation occurs at physiological concentrations in the lysosome.

1.5 β-galactocerebrosidase (GALC)

The lysosomal enzyme β-galactocerebrosidase, also known as β-galactosylceramidase (GALC) is essential for the normal catabolism and recycling of galactosphingolipids. GALC cleaves the terminal galactose head group of glycosphingolipids, catalysing the catabolism of GalCer to ceramide and galactose, psychosine (galactosylsphingosine) to sphingosine and galactose, and lactosylceramide to GluCer and galactose (Fig. 1.6A). Access to these substrates is provided by SapA (Fig. 1.6B and 1.5A)[80]. GALC is particularly important for myelin maintenance, and loss-of-function mutations in GALC or SapA cause the neurodegenerative sphingolipidosis Krabbe disease[64,65,96].

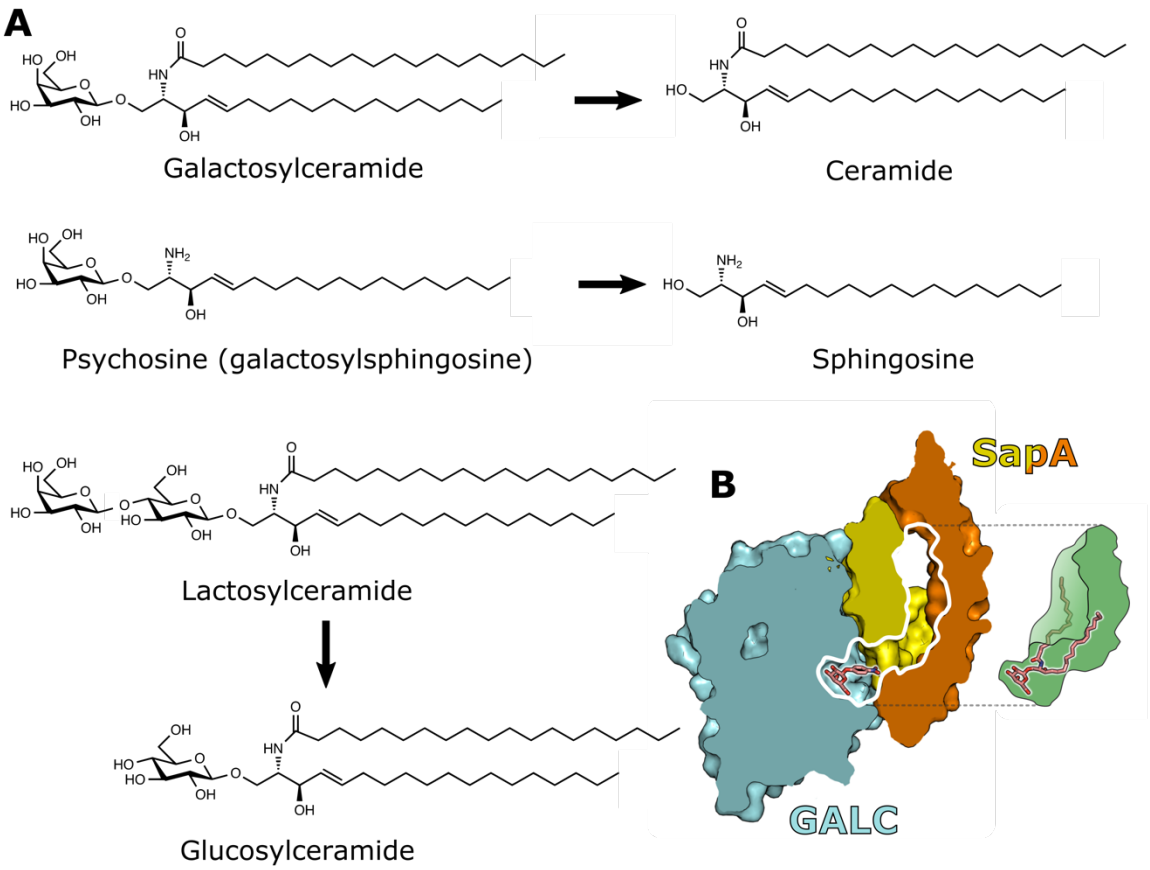

Figure 1.6. GALC hydrolyses galactosphingolipids
A. Skeletal formulae of GALC substrates and lipid products. GALC converts galactosylceramide to ceramide, psychosine to sphingosine and lactosylceramide to glucosylceramide. B. Figure courtesy of Dr Janet Deane. The mechanism of GALC function is illustrated by a cross-section through the GALC-SapA structure (PDB ID: 5NXB) revealing an open channel stretching from the GALC active site into the SapA hydrophobic cavity. The GALC surface and cut-through is shown in cyan and the SapA dimer surface is shown in yellow and orange. For clarity, the second GALC monomer is not shown. Substrate binding to the active site pocket is modelled based on the previously determined GalCer analogue-bound GALC structure (PDB ID: 4CCC). The illustration of the hydrophobic cavity (green) shows that the substrate GalCer (stick representation, in pink) can fit into this cavity, bridging the channel from the GALC active site to hydrophobic SapA core.

1.5.1 Krabbe disease

Krabbe disease, also known as globoid cell leukodystrophy, is a sphingolipidosis caused by loss-of-function mutations in *GALC*. Krabbe disease is characterised by demyelination of neurons in the central and peripheral nervous system (CNS and PNS), resulting in neurodegeneration and death. Krabbe disease is an autosomal recessive disorder with a prevalence of 1 in 100,000, although this varies between countries and communities. Approximately 250 pathogenic mutations in *GALC* have been identified; the most common is a 30 kilobase (kb) gene deletion. In the majority of cases, symptoms appear before 12 months of age (infantile-onset disease) and results in death in the first two years of life. A minority of patients develop a less severe form of Krabbe disease with late or adult

onset and slower disease progression[96]. Hematopoietic stem cell transplantation administered prior to the onset of symptoms is currently the only treatment extending the lifespan and reducing the symptoms of Krabbe disease patients[97]. There are several animal models of Krabbe disease, including the Twitcher mouse phenotypically resembling human Krabbe disease due to a spontaneously occurring mutation in *GALC*[98]. Canine[99], feline[100], primate[101] and ovine[102] models of Krabbe disease are also available.

1.5.2 Krabbe disease pathogenesis

Myelin is formed by tight wrapping of myelinating cell membranes to form an insulating, multilayered myelin sheath around neuronal axons. Oligodendrocytes are the myelinating cells of the CNS; Schwann cells perform this function in the PNS. The myelin sheath is primarily composed of lipids (70-80% lipids, 15-30% proteins by weight) and is particularly enriched in the sphingolipid species sphingomyelin, GalCer, and its sulfated derivative sulfatide[103]. The formation and stability of myelin is dependent on close packing and tight regulation of interactions between neurons and the myelin sheath, as well as between myelin layers[104]. These interactions include protein-protein interactions, such as between oligodendrocyte neurofascin and the axonal proteins contactin-1 and Caspr[105]; protein-lipid interactions, for example between myelin-expressed Myelin-associated glycoprotein and the gangliosides GD1a and GT1b in the neuronal membrane[44]; as well as potential *trans* interactions between GalCer and sulfatide head groups on apposed myelin sheath surfaces[106,107]. The mechanisms leading to demyelination in Krabbe disease have not yet been fully elucidated. The primary substrate of GALC, GalCer, does not accumulate at high levels in Krabbe disease. Instead, psychosine, the deacylated form of GalCer (Fig. 1.6A), has been found to accumulate in the brain, liver and kidney of Krabbe disease patients. Psychosine is a cytotoxic lysolipid normally undetectable in the tissues of healthy patents. The cytotoxic properties of psychosine prompted the formulation of the "psychosine hypothesis" proposing that accumulation of psychosine in myelinating cells causes their death and subsequent demyelination and neurodegeneration[108-110]. Multiple mechanisms of psychosine cytotoxicity have been proposed, including disruption of membrane domains and membrane integrity[111] and direct activation of cell stress and death signalling pathways[112]. Unfortunately, many of the studies on the cytotoxic activity of psychosine have relied on exogenous feeding of psychosine, which may not accurately reflect the localisation and concentration of psychosine in the disease state. In Chapter 6, I explore how accumulation of galactosphingolipids alters protein trafficking in GALC-deficient oligodendrocytes.

1.5.3 Immune dysfunction in Krabbe disease

Krabbe disease is accompanied by immune dysregulation. A hallmark of Krabbe disease is the formation of giant, multinucleated phagocytic cells called globoid cells in the brain. Globoid cells are thought to be derived from microglia, the brain-resident macrophages, and/or infiltrating peripheral macrophages. Globoid cell formation has been hypothesized to be driven by chronic inflammation caused by myelin damage or storage of undigested GALC substrates following phagocytosis of myelin debris[113]. However, globoid cells have been reported in the spinal cord of foetuses with Krabbe disease prior to demyelination[114,115], and globoid cells and neuroinflammation have been found in the absence of demyelination in a mouse model of Krabbe disease[116], suggesting that globoid cell formation may be an early event directly caused by the loss of GALC and galactosphingolipid accumulation in phagocytic cells.

Furthermore, elevated levels of pro-inflammatory cytokines including TNF-α have been observed in the brain of the Twitcher mouse model and rhesus macaques with Krabbe disease[117,118]. An intriguing hypothesis is that globoid cells or aberrantly activated immune cells drive Krabbe disease pathogenesis by producing pro-inflammatory cytokines such as TNF-α, which can cause oligodendrocyte cell death[119,120]. Peripheral blood mononuclear cells (PBMCs) isolated from Krabbe disease patients have been found to produce higher levels of TNF-α than healthy controls[121], suggesting that peripheral immune cells may be abnormally activated in Krabbe disease. A major type of PBMC involved in TNF-α production are monocytes. These are circulating phagocytic cells that can differentiate to macrophages[122].

Thus, some evidence suggests that loss of GALC in Krabbe disease may affect phagocytic cell function, potentially through abnormal accumulation of galactosphingolipids in these cells. As discussed in section 1.1, the importance of sphingolipids for immune receptor function is becoming increasingly appreciated. GluCer, lactosylceramide, GM3 and (iso)Gb3 have been found to be abundant in monocyte/macrophages[123], but the presence of GalCer has not been previously reported in these cells. However, quantification of GalCer by lipidomics is particularly challenging because GalCer is an isobaric structural isomer of GluCer, requiring additional procedures for the separation of these lipid species prior to mass spectrometry. It is therefore possible that GalCer is present in these cells but indistinguishable from GluCer using conventional glycosphingolipid analysis techniques. Even small amounts of a lipid species may still be functionally relevant, especially if concentrated in functional domains. Whether galactosphingolipids are present and important in phagocytic cells has not yet been directly investigated. This question is explored in Chapter 4.

1.6 Acid ceramidase (AC)

Acid ceramidase (AC; *ASAH1*) catalyses the final step of sphingolipid degradation, the hydrolysis of ceramide to sphingosine in lysosomal compartments. This role makes AC a key enzyme for the maintenance of sphingolipid homeostasis. Loss-of-function mutations in *ASAH1* result in ceramide accumulation, causing the lysosomal storage disease Farber disease. AC also plays a key role in apoptosis by regulating the balance between pro-apoptotic ceramide and anti-apoptotic S1P (discussed in section 1.1.2), and up-regulation of AC expression has been implicated in several cancers. Furthermore, AC is essential for embryonic development, as demonstrated by the fact that AC knockout in mice leads to early embryonic death at the 4-cell stage[124,125].

AC is a glycoprotein synthesised in the ER and modified by N-linked glycosylation at five sites[126]. AC is transported from the *trans* Golgi network to endosomal compartments by mannose-6-phosphate receptor-dependent trafficking[126]. AC activity is restricted to endolysosomal compartments through several mechanisms. Formation of catalytically active, mature AC requires autoproteolysis of the peptide bond between Cys143 and Ile142, forming a disulfide-bonded heterodimer composed of two subunits. This autoproteolysis event is triggered by the low pH of endolysosomal compartments[127,128]. Moreover, the optimal pH for AC catalytic activity is pH 4.5. AC is a soluble enzyme and its interaction with substrate-containing membranes is promoted by low pH, the presence of the anionic lipid BMP and the high membrane curvature associated with intraluminal vesicles[90]. Interestingly, AC is also capable of reverse activity *in vitro*, converting fatty acid and sphingosine to ceramide[129,130]. The forward and reverse activity are regulated by differences in optimal pH and lipid environment, suggesting that AC may be recruited to different compartments for ceramide synthesis under certain conditions[130].

1.6.1 Role of saposins in AC function

SapD has been shown to enhance the activity of AC[90,131,132]. SapD does not promote the recruitment of AC to membranes, but appears to stimulate its activity at the membrane interface[90], supporting the hypothesis that SapD functions as a liftase to promote AC activity. However, it is unclear whether AC is entirely dependent on SapD to access lipids. *In vitro* activity assays show reduced activity (3-20%)[90,128] of AC on liposome-embedded ceramide in the absence of SapD. Loss-of-function mutations in SapD have not been identified in humans, but a SapD knock-out mouse model is viable. SapD knock-out mice do not present Farber disease symptoms or massive accumulation of ceramide as seen in Farber disease patients, and show no phenotypic abnormalities until adulthood[133]. This could be

because SapA and SapC, which have been shown to enhance AC activity on liposome-embedded ceramide to a lesser extent than SapD[90], are able to help AC *in vivo*. Alternatively, AC could be able to function without the help of saposins, albeit less efficiently.

1.6.2 Structural features of AC

Previous work by our group led to the determination of a low-resolution crystal structure of AC showing negatively charged loops around the active site of AC likely to be necessary for its interaction with membranes and activity on membrane-embedded substrate. In the course of the project described in Chapter 5, a crystal structure of AC was published by Gebai *et al*.[128]. This structure allowed the authors to propose a catalytic mechanism of AC based on molecular docking of the substrate ceramide and confirmed the presence of hydrophobic patches necessary for AC function around the deep active site (Fig. 1.7). However, many questions on the mechanism of AC function remained unanswered. It is unclear how substrates access the active site of AC from a lipid bilayer. This process may require conformational changes triggered by interactions with the membrane and partial penetration of the hydrophobic loops of AC into the lipid bilayer (Fig. 1.8A). Moreover, the exact mechanism by which SapD, and perhaps other saposins promote AC activity needs to be clarified (Fig. 1.8B). The interaction of AC with membranes is explored in Chapter 5.

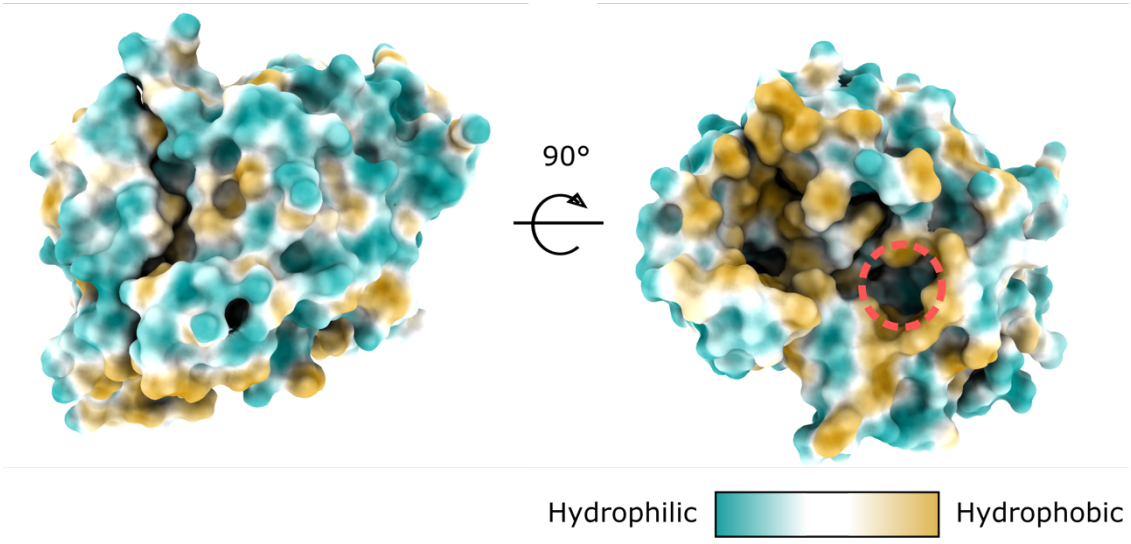

Figure 1.7. The activity of AC is dependent on hydrophobic loops around its active site
Surface representation of the structure of AC (PDB ID: 5U7Z) coloured by hydrophobicity, showing that the active site of AC (circled with red dashes) is surrounded by hydrophobic loops. The view in the panel on the right is rotated by 90° along the horizontal axis

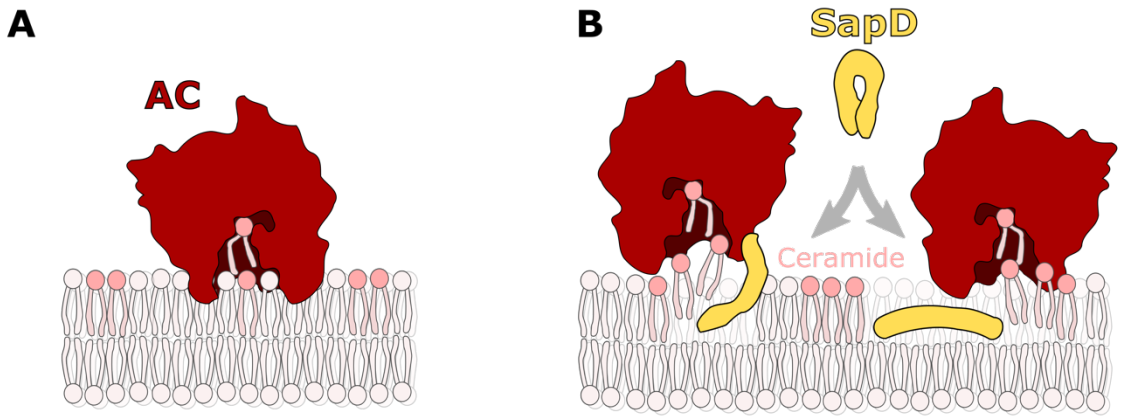

Figure 1.8. The mechanism of ceramide access to the active site of AC is unknown
Schematic representations of potential mechanisms of substrate access to the active site of AC. A. The association of AC with membranes may drive partial insertion of AC (red) into the lipid bilayer and induce conformational changes facilitating ceramide (pink) access to the active site of AC. B. SapD promotes AC activity at the membrane interface. This may be done by association of SapD (yellow) with the lipid bilayer, perturbing the membrane surface to facilitate access of ceramide to the active site of AC.

1.6.3 Farber disease

Loss-of-function mutations in *ASAH1*, the gene encoding AC, cause Farber disease, a lysosomal storage disease characterised by ceramide accumulation in multiple tissues. This causes destructive inflammation resulting in subcutaneous nodules, joint contractures, and voice hoarseness, often accompanied by neurological and respiratory complications and spleen and liver enlargement. Farber disease is extremely rare, with only 200 reported cases and a predicted prevalence of less than 1 in 1,000,000. There are approximately 60 known pathologic loss-of-function mutations in *ASAH1*[134]. Severe cases of Farber disease result in death before the age of 2-3 years. Patients with mild and attenuated forms of Farber disease have an increased lifespan but suffer from similar symptoms. There is little correlation between residual AC activity, disease severity and response to treatment. Furthermore, a very small number of people have been diagnosed with distinct symptoms corresponding to a separate disease named spinal muscular atrophy with progressive myoclonic epilepsy (SMA-PME). It is not clear what differentiates SMA-PME from Farber disease, especially since patients sharing the same mutation can present different diseases[134]. There is currently no cure for AC deficiency. Hematopoietic stem cell transplantation extends the lifespan of patients and resolves inflammation, alleviating peripheral symptoms but not neurological symptoms[135]. Enzyme replacement therapy and gene therapy are currently being investigated as therapeutic approaches for AC deficiencies [134].

1.6.4 Involvement of AC in cancer

AC up-regulation is associated with several types of cancers, and in particular with prostate cancer where it has been found to be overexpressed in 40-80% of tumours[136]. Radiation-induced AC expression is also associated with resistance to radiation therapy[137,138]. It is therefore actively investigated as a target for cancer therapy[139]. The non-competitive AC inhibitor carmofur is used clinically but is associated with severe toxicity[140]. Recently published structures of AC[128] as well as carmofur-bound AC[141] give hope for structure-based design of improved AC inhibitors.

1.6.5 Involvement of AC in other sphingolipidoses

Recent evidence points to an involvement of AC in sphingolipidoses besides Farber disease. Gaucher disease is caused by loss of GCase and subsequent accumulation of GluCer and its deacylated form glucosylsphingosine in macrophages and neurons. Recent studies indicate that AC can deacylate accumulated GluCer to glucosylsphingosine in cells deficient in GCase (Fig. 1.9A). Irreversible inhibition of GCase[142] or GCase knock-out[143] causes a massive accumulation of glucosylsphingosine, and this accumulation is prevented by inhibition of AC with carmofur. Moreover, AC deficient fibroblasts from Farber disease patients do not accumulate glucosylsphingosine upon GCase inhibition[142,144], and glucosylsphingosine accumulation returns following transfection of AC[142]. Similarly, Fabry disease fibroblasts deficient in the enzyme α-galactosidase A accumulate lyso-Gb3, the deacylated form of the main α-galactosidase A substrate Gb3 (Fig. 1.9B), and this accumulation is prevented by inhibition of AC by carmofur[142]. This indicates that AC can accommodate the large head groups of glycosphingolipids and catalyse their deacylation. I investigated whether AC can convert GalCer to psychosine (Fig. 1.9C), thus contributing to the pathogenesis of Krabbe disease in Chapter 5.

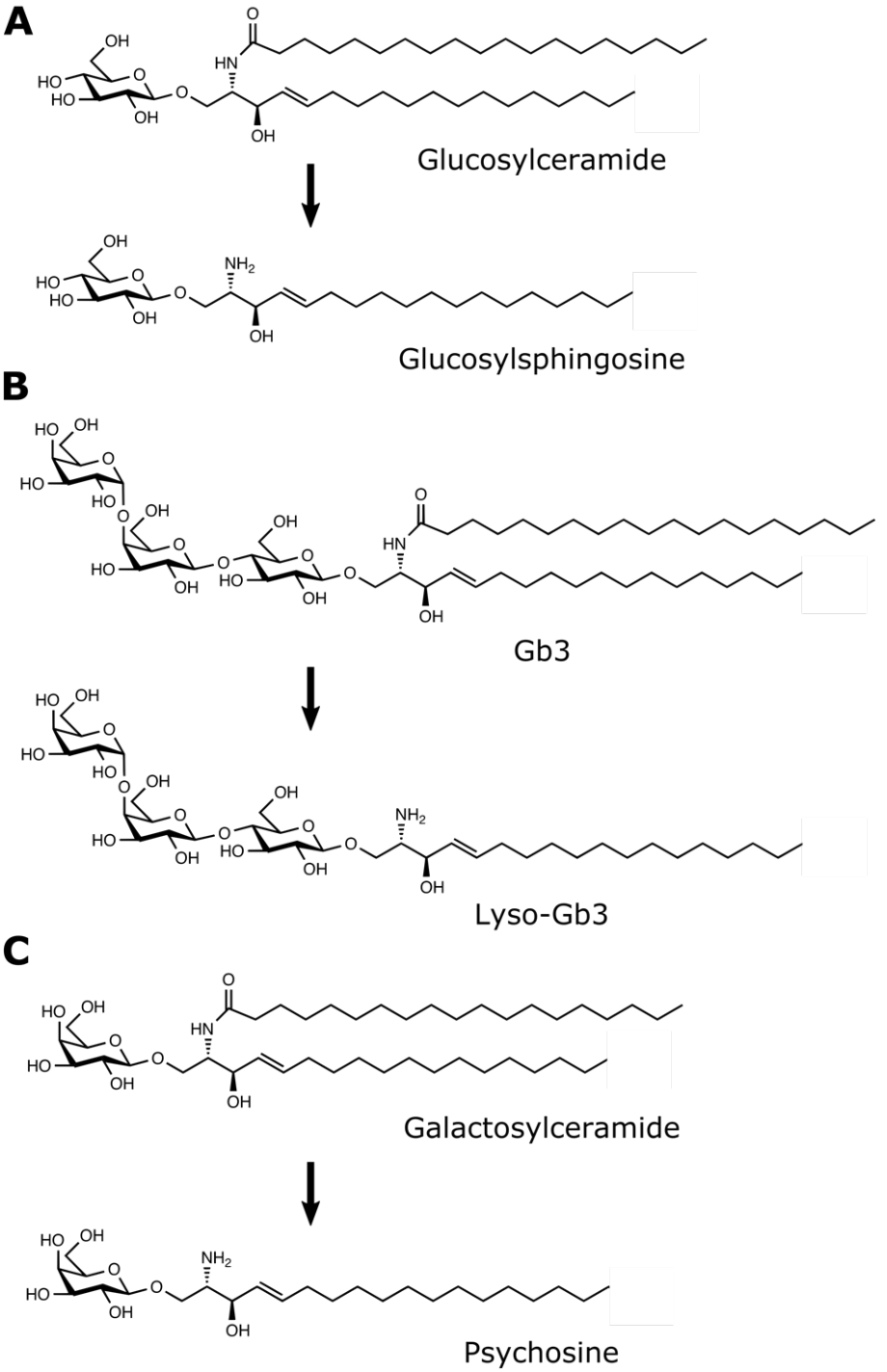

Figure 1.9. AC may deacylate glycosphingolipid substrates
Skeletal formulae of potential AC glycosphingolipid substrates and products. A. Evidence suggests that AC can deacylate glucosylceramide to glucosylsphingosine. B. Evidence suggests that AC may also deacylate Gb3 to lyso-Gb3. C. AC may deacylate galactosylceramide to psychosine.

1.7 Sphingolipid presentation to the immune system

1.7.1 CD1 molecules

In an analogous way to peptide presentation by MHC molecules, endogenous and microbial lipids are presented to the immune system by CD1 molecules and recognised by specialised T-cells. This allows the immune system to detect and respond to the lipidome of cells in states of homeostasis, infection or under stress. CD1 molecules are primarily expressed on antigen-presenting cells[145,146] and have structural similarity to MHC class-I, including association with β-2-microglobulin (β_2m), but unlike MHC molecules, CD1 molecules are non-polymorphic and the antigen-binding grooves of CD1 molecules are hydrophobic, allowing the binding of most classes of lipids including sphingolipids, phospholipids, glycerolipids and lysolipids[147]. There are five CD1 isoforms expressed in humans, named CD1a, CD1b, CD1c, CD1d and CD1e. CD1 molecules are synthesised in the ER, where they associate with β2m and endogenous ER lipids before trafficking to the plasma membrane via the secretory pathway. Each isoform then samples different environments for lipids. CD1a remains predominantly at the cell surface. CD1c and CD1d recycle between the plasma membrane and the early and late endocytic compartments via interactions with adaptor protein-2, whereas CD1b cycles between the plasma membrane and late endosomal/lysosomal compartments via adaptor protein-3-dependent trafficking[148]. CD1e is cleaved in the lysosome to form a soluble isoform that does not present lipids to T-cells but may function as a lipid transfer protein to promote loading of lipids onto other CD1 molecules[149]. CD1 isoforms differ in their expression profile, lipid-binding cavity size, their lipid binding profile and the T-cell subsets recognising them. Of the five CD1 family members, CD1d is best characterised and is the only CD1 isoform present in mice. CD1d is recognised by the T-cell receptor (TCR) of natural killer T (NKT) cells.

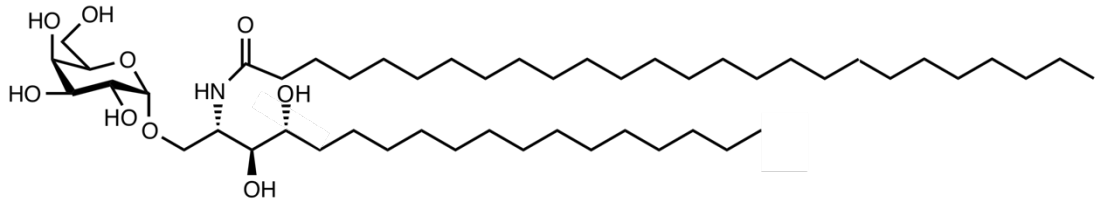

Figure 1.10. Skeletal formula of the NKT cell agonist α-galactosylceramide

1.7.2 NKT cells

NKT cells recognising CD1d are divided into two groups: invariant NKT (iNKT) cells with a semi-invariant T cell receptor recognising the synthetic lipid α-Galactosylceramide (α-GalCer), and more diverse type II NKT cells. α-GalCer is a synthetic saturated glycosphingolipid derived from a marine sponge (Fig.

1.10). The α-glycosidic linkage between the galactose head group and the ceramide moiety of α-GalCer is found in bacteria[150] but not in humans, where GalCer is synthesised with a β-glycosidic linkage (Fig. 1.1B).

iNKT cells have been well characterised due to the availability of α-GalCer/CD1d tetramers as a tool to identify and isolate α-GalCer-responsive cells. iNKT cells are innate-like T-cells involved in the orchestration of immune responses[151]. These cells possess NK cell markers such as NK1.1 and are able to rapidly produce large quantities of various cytokines including interleukin (IL)-4 and interferon γ (IFN-γ) in response to TCR stimulation. Alternatively, iNKT cells can be activated by stimulation by pro-inflammatory cytokines such as IL-12 even in the absence of TCR stimulation. These responsive properties allow iNKT cells to initiate, coordinate and suppress immune responses[151,152]. iNKT cells are important in response to infection by pathogens such as *Streptococcus pneumoniae*[153]. The importance of CD1d-dependent NKT cell activation in response to viral infection is highlighted by the fact that CD1d expression is down-regulated by a number of viruses including herpes simplex virus type 1 and human immunodeficiency virus[154]. iNKT cells also play a critical role in autoimmune disease, transplant immunology, and cancer and are therefore the subject of extensive research to understand and modulate their function for therapeutic purposes.

Despite being more numerous than iNKT cells in humans, type II NKT cells are less well characterised due to their diversity and the lack of tools to identify, isolate and study these cells. However, they are also known to play important roles in various pathological contexts including infection, cancer and autoimmunity. For example, sulfatide-reactive NKT cells play a protective role in a mouse model of multiple sclerosis[155].

1.7.3 Structural features of CD1d

Over one hundred structures of CD1d have been deposited. CD1d has a similar structure to MHC molecules, with an antigen-binding groove formed by two α-helices lying over an antiparallel β-sheet (Fig. 1.11A). The antigen-binding groove of CD1d is highly hydrophobic to accommodate a lipid molecule, enclosing it away from the solvent and allowing the polar head group to protrude for TCR recognition (Fig. 1.11B). Molecular dynamics simulations show that the portal to the lipid binding groove of CD1d possesses a high degree of plasticity[156], which is likely to be necessary for lipid antigen loading.

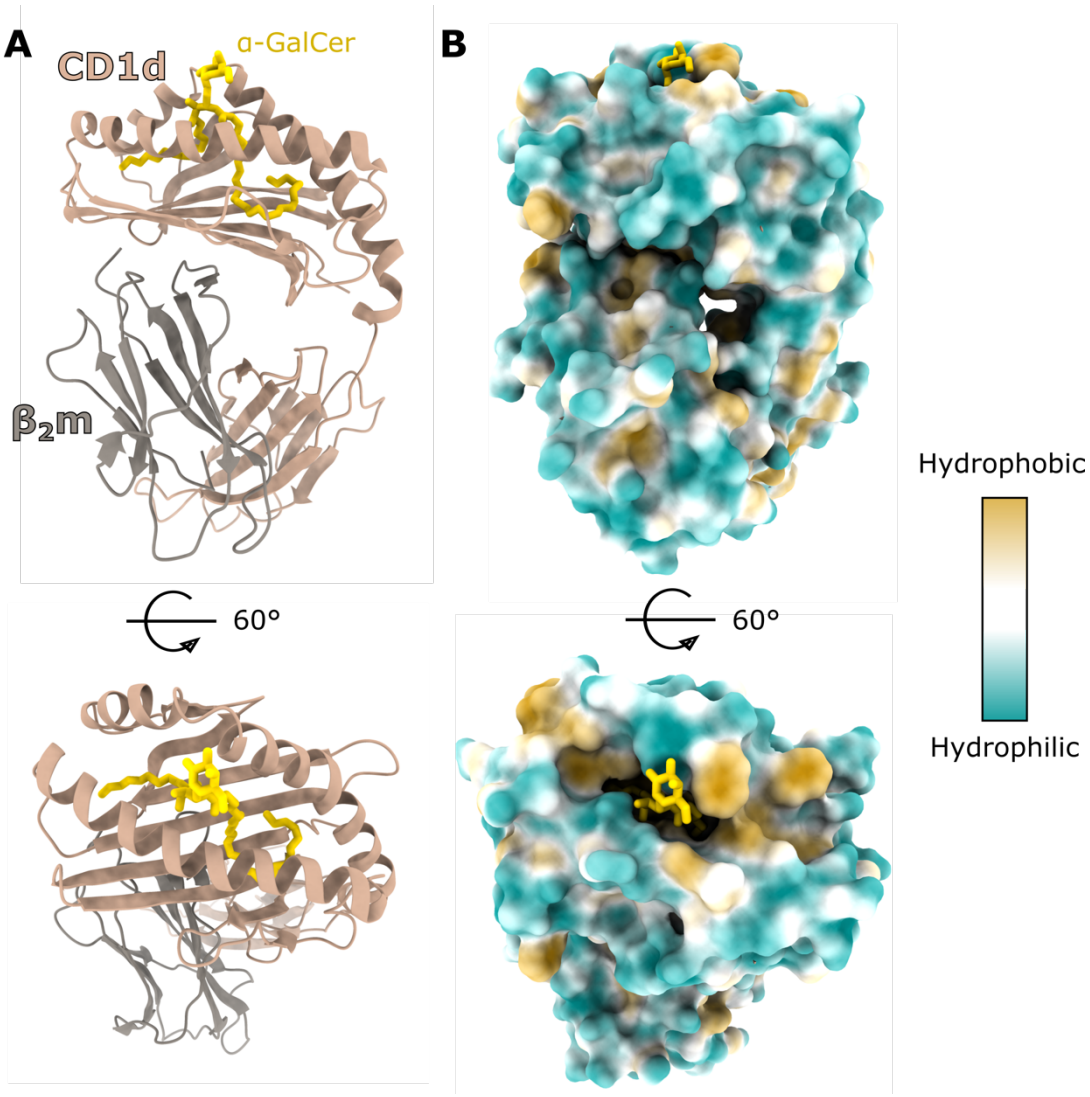

Figure 1.11. CD1d possesses a tight hydrophobic lipid-binding groove
A. Ribbon representation of a crystal structure of human CD1d-β_2m with the lipid α-GalCer in stick representation in yellow (PDB ID: 1ZT4). The view in the bottom panel is rotated by 60° about the horizontal axis. B. Surface representation of the structure shown in A, coloured by hydrophobicity, showing that the hydrocarbon tails of α-GalCer are enclosed within a tight hydrophobic groove.

1.7.4 Role of lipid transfer proteins

Lipids are embedded within membranes and thus are not directly accessible to the lumenal lipid-binding groove of CD1 molecules. The loading of lipid antigen onto CD1 molecules therefore requires the action of lipid transfer proteins for efficient antigen presentation. The loading of lipids onto CD1d can occur in the ER, at the cell surface and in endocytic compartments. Relatively little is known about how lipid specificity is determined, and studies using secreted or recycling, surface-cleavable human CD1d found differences in the lipid repertoire bound to CD1d, reflecting the lipids present in the compartments through which it had trafficked[157-159]. Different lipid transfer proteins are present in the different cellular locations where lipid antigen loading occurs and are likely to determine the lipid

repertoire bound to CD1 molecules[160]. Interestingly, the late endocytic/lysosomal compartment is the site of both glycolipid catabolism and lipid loading of CD1d and the resident lipid transfer proteins function in both these processes. Importantly, the antigenicity of glycolipids can be regulated by processing of their glycan head goups prior to presentation to T-cells, highlighting the complex interplay between these two pathways[161,162].

The evidence for saposin-mediated lipid loading of CD1 molecules comes from animal knockout studies, cell-based assays and *in vitro* assays. Knockout of the saposin precursor PSAP in mice results in the dramatic loss of presentation of endogenous iNKT cell ligands by CD1d[163]. The presentation of exogenous lipid antigens, such as α-GalCer, is impaired in PSAP knockout or knockdown cells[163-166]. *In vitro* assays have identified that all the saposins, as well as the other lysosomal lipid transfer proteins, GM2A and NPC2, are capable of transferring lipids onto CD1 molecules[163,167]. However, some lipid transfer proteins appear to be more efficient than others at loading specific lipid antigens onto different CD1 molecules[165,168,169]. Specificity within the saposin family in this process has been explored via *in vitro* assays identifying SapB as the dominant saposin for the loading of lipid antigens onto CD1d. Specifically, incubation of CD1d with SapB and α-GalCer has been shown to enhance the stimulation of iNKT cells[165,166] and SapB can mediate lipid binding to CD1d in T-cell independent assays[165] suggesting a direct role for this specific saposin in lipid loading of CD1d. The ability of SapB to facilitate and enhance lipid exchange on CD1d has also been monitored *in vitro* via iNKT TCR binding to CD1d following incubation with lipids alone or lipids plus SapB[166]. These *in vitro* assays implicate SapB as a "lipid editor" that facilitates the loading and unloading of lipid antigens onto CD1d. The liftase/solubiliser models are relevant for the proposed mechanisms by which saposins load lipids onto CD1 molecules[169]. SapB is thought to behave as a solubiliser, forming protein-lipid complexes that enhance lipid loading of CD1d, while SapC appears to recruit CD1c to membranes and facilitate lipid loading of CD1c at the membrane surface.

Due to the hydrophobic properties of lipids, their exposure to aqueous solutions is thermodynamically unfavourable. Therefore, lipid transfer between hydrophobic lipid-binding cavities of proteins requires direct interaction of the protein components to provide continuous shielding of the lipid from the aqueous phase[170]. This hypothesis is supported by the fact that other known lipid transfer proteins directly interact with the protein receiving the lipid[77,171,172]. SapB and CD1d are thus expected to participate in a direct interaction facilitating lipid transfer from one protein to the other. This interaction is investigated in Chapter 3.

1.8 Aims

The interaction of proteins with sphingolipid molecules and membrane domains is emerging as an understudied fundamental mechanism regulating protein and cell function. Although a number of these interactions have come to light in recent years, our understanding of the molecular basis of these interactions is still limited. Expanding our knowledge of these interactions is essential in order to design therapies, manipulate cellular pathways and make predictions about processes involving sphingolipids. In this study, I aimed to explore protein-sphingolipid interactions critical for immunity and disease. Specifically, I aimed to:

1. Understand how SapB mediates lipid loading onto CD1d.

Using structural and biochemical methods, I aimed to characterise the interaction between CD1d and SapB facilitating the transfer of lipids between these two proteins.

2. Explore the impact of GALC deficiency on the membrane proteome of phagocytic cells.

I aimed to investigate the effect of GALC knockout in a monocytic cell line using CRISPR-Cas9. I set out to use proteomics and RNA sequencing approaches to determine whether the expression, localisation and trafficking of plasma membrane proteins were affected by loss of GALC and subsequent accumulation of galactosphingolipids in monocytes.

3. Understand the mechanism of AC function and its role in Krabbe disease.

Using structural, biochemical and cell biology methods, I aimed to understand how AC accesses membrane-embedded substrates and whether it processes GalCer to psychosine, the cytotoxic lipid accumulating in Krabbe disease, thus contributing to the pathogenesis of Krabbe disease.

4. Investigate whether protein-sphingolipid interactions contribute to the loss of an adhesion protein down-regulated at the surface of GALC-deficient oligodendrocytes.

Following on previous work in our lab, I aimed to use structural methods to probe the interaction of Protein Tyrosine Phosphatase Receptor Type M with sulfatide.

2 Methods

2.1 Molecular biology

2.1.1 DNA amplification by PCR

Primer stocks were diluted in water to a working concentration of 5 µM. Template DNA was diluted to 10 ng/µL in water. PCR reactions were assembled as in Table 2.1. A hot-start PCR protocol was executed using a thermal cycler; 95 °C for 2 minutes, followed by 30 cycles of denaturation (95 °C, 20 seconds), annealing (55 °C, 10 seconds) and extension (70 °C, 20 seconds/kb). PCR products were visualised by agarose gel electrophoresis (section 2.1.2). Excess primers, template and dNTPs were removed using a Monarch PCR & DNA Cleanup Kit (NEB) according to manufacturer's instructions.

Table 2.1. PCR components for DNA amplification

Component	Volume (µL)
10x KOD buffer (Novagen)	5
$MgSO_4$ (25 mM)	3
dNTPs (2.5 mM)	5
Template DNA (10 ng/µL)	1
Forward primer (5 µM)	4
Reverse primer (5 µM)	4
MQW	27.6
KOD hot-start DNA polymerase (Novagen)	0.4

2.1.2 Analysis of DNA by agarose gel electrophoresis

1% agarose gels were made by melting 1 g of agarose powder in 100 mL of 1X Tris-Borate EDTA (TBE) buffer by microwave heating prior to the addition of 10 µL SYBR Safe DNA stain (Invitrogen). DNA samples were mixed with 6X DNA loading dye (NEB) before being resolved by horizontal electrophoresis in TBE buffer (70V, 140 mA). DNA was then visualised with visible blue light.

2.1.3 Restriction digestion of DNA

Insert and vector DNA were digested with enzymes and buffers supplied by NEB. Per 50 µL reaction, 1 µg DNA was incubated with 1 µL of each enzyme in CutSmart buffer (NEB) at 37 °C for 5-15 minutes (Time-Saver qualified enzymes) or 60 minutes. 1 µL alkaline phosphatase (Promega) was added to vector reactions and incubated at 37 °C for 1 hour. Cut vector DNA was analysed by agarose gel electrophoresis (section 2.1.2) and extracted from excised gel slices using a Monarch DNA Gel Extraction Kit (NEB), as directed by the manufacturer. 20 units of DpnI was added to digested inserts

and incubated at 37 °C for 1 hour followed by PCR clean-up purification using a Monarch PCR & DNA Cleanup Kit (NEB) as directed by the manufacturer.

2.1.4 DNA ligation

Cut DNA fragments were mixed in a 6:1 molar ratio of insert to vector (~100 ng total DNA), combined with 0.5 µL T4 DNA ligase in T4 ligase buffer (NEB) in a total volume of 10 µL. Control reactions without insert or vector were always performed. Ligation reactions were incubated at room temperature for 2 hours or overnight at 16 °C. 2 µL of the ligation reaction were then transformed into chemically competent *E. coli* cells (section 2.1.6). 200 µL of transformed cells were plated and plasmid DNA prepared as in section 2.1.7. Each construct was sequenced in full by Sanger sequencing (GENEWIZ, Cambridge).

2.1.5 Preparation of chemically competent *E. coli* cells

Liquid cultures (100 mL) of *E. coli* were grown to an optical density at 600 nm of 0.25 to 0.3 and chilled on ice for 15 minutes. Cells were pelleted by centrifugation (4000*g*, 4 °C, 10 minutes), resuspended in 30 mL sterile ice-cold 0.1 M $CaCl_2$ and incubated on ice for a further 30 minutes. Cells were then pelleted again and resuspended in 5 mL sterile 0.1 M $CaCl_2$ with 15% v/v glycerol. Chemically competent cells were stored at -80 °C in 50 µL aliquots until required.

2.1.6 Transformation of chemically competent *E. coli* cells with plasmid DNA

DH5α competent *E. coli* cells were used for transformation of most constructs. Aliquots of chemically competent cells were thawed on ice. 100 ng plasmid DNA was added to cells, incubated on ice for 30 minutes then heat-shocked (50 seconds, 42 °C). Cells were chilled on ice for 2 minutes and recovered in 400 µL SOC medium in a shaking incubator (37 °C, 1 hour, 230 rpm). After recovery, selection was applied by plating 50-200 µL of transformed cells on LB agar supplemented with the appropriate antibiotic and incubated overnight at 37 °C. Lentiviral expression vectors were transformed into Stbl3 competent *E. coli* cells (Thermo Fisher Scientific). Transformation was performed as described above, however, recovery and plate incubation were performed at 30 °C.

2.1.7 Preparation of plasmid DNA

Single colonies of transformed cells were picked and grown overnight in 2XTY in the presence of appropriate antibiotic, at 37 °C for DH5α cells or 30 °C for Stbl3 (230 rpm). Bacterial cultures were pelleted by centrifugation (4 000*g*, 10 minutes) and plasmid DNA prepared using commercially

available variations of the alkaline lysis method. For small scale preparations of <20 µg DNA, a PureLink HiPure Miniprep kit (Invitrogen) was used to isolate plasmid DNA from 5 mL overnight cultures of *E. coli*. For large scale preparations, a PureLink HiPure plasmid Midiprep kit (Invitrogen) was used to isolate plasmid DNA from 100 mL overnight cultures of *E.coli*. Plasmid DNA was stored at -20°C.

2.2 Plasmid construction, cloning and mutagenesis

2.2.1 Cloning of CD1d and β_2m into PB-T-H_6

The vector PB-T-PAF was kindly provided by Prof James Rini (University of Toronto)[173]. In this vector, a tetracycline response element controls the expression of the target protein construct with an ER secretion peptide (eventually cleaved from the secreted protein) and an N-terminal Protein A tag. A modified version of this vector retaining the N-terminal secretion signal but with the Protein A fusion removed was constructed by Dr Janet Deane. Codon-optimised cDNA was synthesized (GeneArt) encoding human β_2m (residues 21-119, Uniprot P61769) and the extracellular domain of human CD1d (residues 20-301, Uniprot P15813) with a C-terminal hexahistiding tag (CD1d-H_6). CD1d DNA was amplified with the following primers:
forward 5'-GCACTAGTGAAGTTCCCCAGAGACTGTTCC-3';
reverse 5'- CGGGCGCGCCTCAGTGATGGTGATGGTGGTGCTTGC-3'.
β_2m DNA was amplified with the following primers:
forward 5'- GCACTAGTATCCAGCGGACCCCTAAGATC-3';
reverse 5'-CAGGCGCGCCTCACATGTCGCGGTCCCAC-3'.
Both inserts were digested with *Spe*I-HF and *Asc*I (NEB) and individually ligated to digested PB-T-PAF vector, to produce PB-T-CD1d-H_6 and untagged PB-T-β_2m.

2.2.2 Cloning of SapB into pHLsec

SapB cloning into the bacterial expression vector pET-15b was done by Dr Janet Deane. Codon-optimised cDNA was synthesized (GeneArt) encoding the segment of human PSAP corresponding to SapB (residues 195-274, Uniprot P07602) and subcloned into pET-15b using *Nco*I and *Xho*I restriction endonuclease sites to produce untagged protein.

For subcloning of SapB into the mammalian expression vector pHLsec-H_6, SapB was amplified by PCR using the following primers: forward 5'-GAAACCGGTGATGTTTGTCAGGATTG-3', reverse 5'-GTTTTTGTGATGAAGTGGGTACCAAGC-3', introducing *Age*I and *Kpn*I restriction enzyme sites for restriction cloning into pHLsec-H_6.

2.2.3 Cloning of CD1d and HLA-A2 into pEGFP-N1

CD1d and HLA-A2 were cloned into the pEGFP-N1 expression vector (a kind gift from Dr Stephen Graham) by Gibson assembly cloning. DNA encoding full-length CD1d preceded by a Kozak sequence was optimised for expression in mammalian cells and synthesised by GeneArt with 20 bp overlaps corresponding to the pEGFP-N1 vector. DNA encoding full-length HLA-A2 was generously provided by Louise Boyle. 20 bp overlaps corresponding to the pEGFP-N1 vector were added by PCR using primers designed with the NEBuilder assembly tool. Gibson assembly was performed using the NEBuilder HiFi DNA Assembly Master Mix (NEB, E2621S).

2.2.4 Cloning of AC-H_6 and AC-BAP-H_6

A construct encoding human AC without its secretion sequence (residues 22-395) was cloned from cDNA by Samantha Spratley. The sequence encoding AC was amplified using the following primers:
forward: 5'-GAAACTAGTCAGCACGCGCCGCCGTGG-3';
reverse 5'-GTGCTTAAGCCAACCTATACAAGGGTCAGGGC-3'.
Amplified insert DNA encoding AC and the vectors PB-T-H_6 and PB-T-BAP-H_6 was digested with SpeI-HF (NEB) and BfrI (Roche). AC was ligated to PB-T-H_6 and PB-T-BAP-H_6 as described in section 2.1.4.

2.2.5 Cloning of AC27 into PB-T-PAF

AC (residues 27-295) was cloned into the vector PB-T-PAF[173] (described in section 2.2.1) by Dr Janet Deane, forming the construct encoding ProtA-AC27.

2.2.6 Cloning of guide RNAs targeting GALC into a lentiviral vector

In order to clone the guide RNA (gRNA) target sequences into the lentiviral vector pKLV-U6gRNA-EF(BbsI)-PGKpuro2ABFP (a kind gift from the Lehner lab), the following DNA oligos pairs were used:
gRNA1 top: 5'-CACCGAAAGTGGAAATAGGTGGTGA-3';
gRNA1 bottom: 5'-AAACTCACCACCTATTTCCACTTTC-3';
gRNA2 top: 5'-CACCGGTGGTGATGGGCAGACAAC-3';
gRNA2 bottom: 5'-AAACGTTGTCTGCCCATCACCACC-3'.
DNA oligo pairs were phosphorylated and annealed as follows. 1 µL top oligo (100 µM), 1 µL bottom oligo (100 µM) and 5 units (U) T4 PNK (NEB) were incubated in T4 Ligation Buffer (NEB) at 37 °C for 30 min, followed by 5 min at 95 °C and gradual cooling to 25 °C at 5 °C/min. The annealed oligos were diluted 250-fold and cloned into the backbone vector with the following digestion-ligation reaction: 100 ng vector was incubated with 2 µL diluted oligo duplex, 1 µL *Bpi*I and 0.5 µL T7 ligase in Tango buffer with 500 µM dithiothreitol (DTT) and 500 µM ATP in a 20 µL reaction. The reaction was cycled

six times through the following steps: 37 °C for 5 min, 23 °C for 5 min. 11 µL of the ligation product was treated with 1.5 µL PlasmidSafe exonuclease (Lucigen) in PlasmidSafe Buffer with 1 mM ATP at 37 °C for 30 min. 1 µL of the final product was transformed into Stable Competent *E. coli* (NEB) as described in section 2.1.6. Plasmid DNA was isolated and sequenced as described in section 2.1.7. The insertion of the gRNA target sequence into the vector was verified by sequencing using a sequencing primer for the U6 promoter (5'-GACTATCATATGCTTACCGT-3').

2.2.7 Cloning of GALC rescue vectors

The *GALC* gene was introduced into a puromycin-resistant lentiviral pCW57 vector for rescue of GALC protein expression in GALC KO MO3.13 cells by Dr Stuart Fawke. For GALC expression in GALC KO THP-1 cells, silent mutations were introduced into GALC to escape gRNA targeting, and GALC mutant inserts were cloned into neomycin-resistant pCW57. pCW57-MCS1-P2A-MCS2 (Neo) was a gift from Adam Karpf (Addgene plasmid # 89180 ; http://n2t.net/addgene:89180 ; RRID:Addgene_89180). Mutagenesis was performed by PCR-driven overlap extension following a published protocol[174]. For each rescue GALC construct, overlapping segments of the GALC gene are amplified from template GALC plasmid using two flanking primers (named A and D) marking the 5' ends of both strands and two internal primers (B and C) introducing the mutations of interest creating overlapping nucleotide sequences. AB and CD products are amplified in a first PCR reaction, then used as template in a second PCR reaction, where denatured products hybridize and full-length mutant GALC is amplified. For GALC Rescue 1, the DNA sequence 5'-GAA ATA GGT GGT-3' was mutated to 5'-GAG ATC GGC GGC-3'. The following primers were used:

A: 5'-CAGGGACAGCAGAGATCCACTTTGGC-3'

B: 5'-GCCATCTGTTGTCTGCCCATCGCCGCCGATCTCCACTTTTAAAATATGCAAAGAGG-3'

C: 5'-CCTCTTTGCATATTTTAAAAGTGGAGATCGGCGGCGATGGGCAGACAACAGATGGC-3'

D: CGTTTCCCGGAACCACGCCCAG-3'

For the GALC Rescue 2 construct, the DNA sequence 5'-GAT GGG CAG ACA-3' was mutated to 5'-GAC GGC CAA ACC-3'. The following primers were used:

A: 5'-CAGGGACAGCAGAGATCCACTTTGGC-3'

B: 5'-GCTCAGTGCCATCTGTGGTTTGGCCGTCACCACCTATTTCCACTTTTAAAATATGC-3'

C: 5'-GCATATTTTAAAAGTGGAAATAGGTGGTGACGGCCAAACCACAGATGGCACTGAGC-3'

D: 5'-CGTTTCCCGGAACCACGCCCAG-3'

PCR reactions were performed with KOD following the protocol detailed in section 2.1.1 with an annealing temperature gradient between 47 °C and 61 °C. Amplification products were gel extracted (section 2.1.3) and diluted to 40 ng/µL. AB and CD fragments were combined and used as template

(80 ng) in a standard KOD PCR reaction with primers A and D. Amplified hybridisation products AD were gel extracted and purified. These products and the target vector pcW57(Neo) were digested with SalI-HF and NheI-HF (NEB) and ligated with a 5:1 molar ratio of insert to vector as described in section 2.1.4. Ligation products were transformed into Stbl3 *E. coli* and verified by sequencing (sections 2.1.6-7).

2.3 Mammalian cell culture

2.3.1 HEK293F cells

2.3.1.1 Maintenance

HEK293F cells were maintained in FreeStyle 293 media (Gibco) in polycarbonate flasks (125 mL flask for cultures up to 40 mL, 500 mL flasks for 150 mL cultures, and 1 L flasks for 300 mL cultures) in an incubation shaker (37 °C, 8% CO_2, 130 rpm) and passaged at a confluency of $2x10^6$ cells/mL (every 2-3 days). Viable cells were counted after Trypan blue staining using a haemocytometer. The appropriate volume of culture was removed, and cells were diluted to a concentration of $0.2x10^6$ cells/mL in fresh media.

2.3.1.2 Cell freezing

$60x10^6$ cells were spun down (60*g*, 5 min) and resuspended in 6 mL Freestyle293 media supplemented with 10% dimethyl sulfoxide (DMSO). 1 mL aliquots, each containing $10x10^6$ cells, were dispensed into sterile cryovials quickly transferred to a CoolCell LX freezing container (BioCision) for cooling to -80 °C at 1 °C/min. Frozen cells were stored in liquid nitrogen.

2.3.1.3 Cell thawing

Cryovials containing $10x10^6$ cells in 1 mL were thawed in a water bath at 37°C and quickly diluted in 30 mL warm FreeStyle293 media in a polycarbonate flask.

2.3.2 THP-1 cells

2.3.2.1 Maintenance

THP-1 cells were maintained in THP-1 medium: RPMI-1640 Medium with L-glutamine and sodium bicarbonate (Sigma, R8758), supplemented with 10% fetal bovine serum (FBS) and 50 µM 2-mercaptoethanol. Cells were maintained in T25, T75 or T175 flasks lying in a horizontal position at 37 °C in humidified air with 5% CO_2. The absence of mycoplasma contamination in THP-1 cell cultures was regularly verified using the MycoAlert Mycoplasma Detection Kit (Lonza).

THP-1 cells were passaged before they reached a density of $0.8x10^6$ cells/mL (every 2-3 days). Cells were thoroughly resuspended, and a cell suspension aliquot was mixed 1:1 with Trypan blue (Sigma) to assess cell viability. Viable cells were counted using a haemocytometer. A fraction of the cell suspension was removed from the flask, and the cell suspension was supplemented with fresh THP-1 medium for a cell density of $0.1\text{-}0.2x10^6$ cells/mL.

2.3.2.2 Cell freezing

For long-term cell storage, THP-1 cells were frozen in aliquots of $2.5x10^6$ cells. Cells were pelleted by centrifugation (125*g*, 5 min), the culture medium was removed by aspiration and replaced with 500 µL THP-1 freezing medium (90% (v/v) FBS, 10% (v/v) DMSO). Cells were rapidly transferred to a sterile cryovial, which was slowly cooled to -80 °C at the rate of -1 °C/minute using a CoolCell freezing container (BioCision). Frozen cells were then transferred to liquid nitrogen within 24 hours.

2.3.2.3 Cell thawing

Frozen aliquots of THP-1 cells were warmed with agitation using a water bath at 37 °C. Once defrosted, cells were rapidly transferred to a 15 mL Falcon tube containing 10 mL THP-1 recovery medium (RPMI-1640 Medium with L-glutamine and sodium bicarbonate supplemented with 20% FBS and 50 µM 2-mercaptoethanol). Cells were pelleted by centrifugation (125*g*, 5 min) and the medium containing DMSO was aspirated and replaced with 5 mL THP-1 recovery medium. The cell suspension was then transferred to a T25 flask and incubated at 37 °C. Cell density was monitored every day until cells started doubling every 24 hours. Cells were then split with THP-1 medium (containing 10% FBS).

2.3.2.4 Differentiation

THP-1 cells were differentiated by addition of 25 nM phorbol 12-myristate 13-acetate (PMA) to THP-1 medium[175]. For activity assays, cells were seeded at $0.4x10^6$ cells/mL in a T75 flask. For immunofluorescence microscopy, $0.5x10^6$ WT cells or $0.012x10^6$ Cas9, mixed population EV and GALC KO cells were seeded on glass coverslips in a 24-well plate. Cells were exposed to PMA for 48 hours. The medium was then removed and replaced with THP-1 medium without PMA for a recovery period of 24 hours. Differentiated cells were then used in subsequent experiments.

2.3.2.5 LPS stimulation

Differentiated THP-1 cells were stimulated with lipopolysaccharide (LPS) from *E. coli* O111:B4 (Sigma). Differentiated THP-1 cells were incubated with LPS at 100 ng/mL in THP-1 medium overnight.

2.3.3 HEK-293T cells

2.3.3.1 Maintenance

HEK293T cells were maintained in HEK293T medium: Dulbecco's Modified Eagle Medium (DMEM; Sigma) with 10% FBS and 2 mM L-glutamine. Cells were maintained at 37°C in humidified air with 5% CO2.

HEK293T cells were passaged before 100% confluency was reached (approximately every 2-3 days). For cell passage, the medium was aspirated, cells washed in phosphate-buffered saline (PBS) and detached with trypsin-EDTA (Sigma). On cell detachment, trypsin was deactivated with 4 volume equivalents of HEK293T medium and the cell suspension split and transferred to a new flask containing HEK293T medium. For example, for confluent HEK293T cells in a T75 flask, 2 mL of trypsin was applied to washed cells for 2 minutes at 37 °C and deactivated with 8 ml of HEK293T medium. The cell suspension was then split 1:8 into a new T75 flask and made up to 10 mL with HEK293T medium.

2.3.3.2 Lentivirus production

$7x10^6$ HEK293T cells were seeded in 30 mL HEK293T medium in a 150 mm dish. The following day, the media was changed to 8 mL OptiMEM (Gibco) and cells were transfected using FuGENE 6 (Promega). 60 µL FuGENE 6 was added to 1 mL OptiMEM, mixed by brief vortexing, and incubated for 5 min at room temperature. 6.5 µg pCMV-dR8.91, 3.5 µg pMD-VSVG and 10 µg of the lentiviral vector of interest (20 ng DNA in total) were then combined with FuGENE 6. The solution was briefly vortexed and incubated for 20 minutes at room temperature to allow complex formation before dropwise addition to the cells. 6 hours after transfection, the medium was replaced with 30 mL fresh HEK293T medium. Lentivirus-containing medium was harvested 48 hours post-transfection, filtered through a 0.45 µm membrane, and used immediately for viral transduction or stored in 1 mL aliquots at -80 °C.

2.3.4 Oligodendrocyte cells

2.3.4.1 Maintenance

Undifferentiated MO3.13 were maintained in DMEM with 10% FBS and 4 mM L-glutamine. Cells were maintained in T75 flasks at 37 °C in humidified air with 5% CO_2. Cells were passaged before they reached 80% confluency (every 2-4 days). Cells were washed twice with 5 mL PBS. 2 mL trypsin-EDTA (Gibco) was added to the cells and incubated for 5 minutes. Trypsin was neutralised with 8 mL DMEM and cells were split 1:5-1:8.

2.3.4.2 Differentiation

MO3.13 oligodendrocytes were differentiated by changing the media to differentiation media (DMEM, 1% FBS, 100 nM PMA) one day after seeding. Cells were grown in this differentiation media for 3 or 7 days before being used in subsequent assays.

2.4 Protein expression, purification and preparation

2.4.1 Stable H_6-tagged CD1d-β_2m expression in HEK293F cells

2.4.1.1 Transfection

For production of an inducible, stable co-expression cell line, HEK293F cells were grown to a density of $1x10^6$ cells/mL in 30 mL and quadruple-transfected with the following plasmids. One plasmid contains a neomycin resistance gene and a transposon encoding the tetracycline transactivator inducer (PB-RN, 5 µg), a helper plasmid (PBase, 5 µg) transiently expresses the PB transposase which integrates the transposons into the host cell genome, and two plasmids containing a transposon encoding the target proteins CD1d-H_6 and β_2m under the control of a tetracycline response element (PB-T-CD1d-H_6 and PB-T-β_2m, 12.5 µg each). All plasmids were added to 0.6 mL OptiProSFM (Thermo Fisher Scientific). 35 µL FreeStyle MAX transfection reagent (Thermo Fisher Scientific) was added to 0.6 mL OptiProSFM in a separate tube. DNA and transfection reagent mixtures were mixed and incubated at room temperature for 10 min before dropwise addition to the cells. Cells were grown for 2 days before starting antibiotic selection.

2.4.1.2 Antibiotic selection

Cells were maintained in 30 mL FreeStyle media supplemented with 500 µg/mL geneticin (neomycin analogue). Every 2-3 days, cells were pelleted (60*g*, 5 minutes) and media was replaced by fresh media with geneticin. Drug selection was stopped after two weeks. Cultures were expanded to 150 mL (cell density was always maintained above $0.2x10^6$ cells/mL) and aliquots were frozen.

2.4.1.3 Induction of protein expression

Cells were diluted to a density of $0.5x10^6$ cells/mL in 150 mL FreeStyle293 media. Recombinant protein expression was induced by adding doxycycline at 2 µg/mL.

2.4.1.4 Protein harvest

Expression media containing secreted recombinant protein was collected over a 2-month period from the start of induction. Every 2-3 days, one third of the culture volume was harvested and cells were pelleted (60*g*, 5 min). Supernatant containing protein was retained. The pellet was resuspended in the same volume of fresh media with 2 µg/mL doxycycline and returned to the culture flask. After one week, the culture volume was expanded to 300 mL. Subsequently, the cell pellet was discarded after harvesting the media. Media containing protein was cleared by centrifugation (4000*g*, 10 min), then filtered through a 0.2 µm filter and stored at 4°C. When harvested media was to be stored for longer than two weeks, it was supplemented with 0.02% sodium azide to prevent microbial growth and filtered through a 0.2 µm filter a second time immediately prior to purification.

2.4.2 Purification of H_6-tagged CD1d-β_2m

2.4.2.1 Nickel affinity purification

CD1d-H_6 was purified from conditioned media by nickel affinity chromatography. Conditioned media was incubated with 1 mL washed Ni-NTA agarose beads (Qiagen) per 100 mL media on an end-over-end rotator for 1 hour at 4 °C. The beads were then applied to a gravity column and washed in 20 column volumes (CV) PBS pH 7.4. CD1d-H_6 was eluted using PBS containing 300 mM imidazole in 500 µL fractions. Imidazole was diluted out from the protein buffer by multiple cycles of dilution in PBS and concentration in a centrifugal concentrator (10K molecular weight cut-off (MWCO); Amicon) until the imidazole concentration was below 10 mM.

2.4.2.2 Size exclusion chromatography

CD1d-β_2m was further purified by size exclusion chromatography on a Superdex 200 column (GE Healthcare). The column was equilibrated in SEC buffer (150 mM NaCl, 50 mM Tris pH 7.4) for crystallisation experiments or cross-linking buffer (20 mM HEPES pH 7.0, 150 mM NaCl) for interaction assays. CD1d-β_2m was concentrated to 2 mg/mL and stored at 4 °C for up to one month.

2.4.3 Expression of deglycosylated H_6-tagged CD1d-β_2m

CD1d-β_2m expressing cells were treated with the α-mannosidase I inhibitor kifunensine to produce protein with high mannose glycans readily cleaved by Endoglycosidase H (Endo H). 1 mM kifunensine stock (Cayman chemicals, 10009437) was prepared in warm, sterile water, filtered and stored at 4 °C for up to 3 weeks. 5 µM kifunensine was added to cells at the same time as protein expression was induced with doxycycline. When media was replaced, fresh media was supplemented with 5 µM

kifunensine as well as 2 µg/mL doxycycline. Cells were discarded after 3 weeks of kifunensine treatment because there was significant cell death after this point.

2.4.4 Purification of deglycosylated H_6-tagged CD1d-β2m

CD1d-β_2m expressed in cells treated with kifunensine was purified by nickel affinity purification with Ni-NTA agarose (Qiagen) and eluted with 100 mM citrate pH 4. CD1d-β_2m protein was buffer exchanged into 100 mM citrate pH 5.5 using repeated rounds of dilution and concentration in a centrifugal concentrator and concentrated to 2 mg/mL.

For small-scale optimisation of Endo H treatment, 20 µg CD1d-β_2m was incubated with a dilution series of Endo H (1000-62.5 units, NEB, P0702L) at room temperature or 37 °C, for 3 hours or overnight. As a positive control, CD1d-β_2m was denatured by incubation with 0.5% SDS, 40 mM DTT at 100 °C for 10 minutes prior to Endo H treatment at 37 °C overnight. After Endo H treatment, all samples were analysed by SDS-PAGE on a Coomassie-stained 4-12% Bis-Tris gel run in MOPS buffer. For preparative Endo H treatment, CD1d-β_2m (at 2 mg/mL in 100 mM citrate pH 5.5) was incubated with 6.25 µL (3125 NEB units) Endo H per milligram of CD1d for 3 hours at room temperature. A 20 µL sample was taken for SDS-PAGE analysis. Precipitate was removed by centrifugation (13 000g, 2 min), and 20 µL samples of the cleared solution and pellet were taken for SDS-PAGE analysis. Soluble CD1d-β_2m was further purified and separated from Endo H by size exclusion chromatography on a Superdex 200 10/300 column (GE Healthcare) equilibrated with 20 mM citrate pH 6, 150 mM NaCl and fractions containing CD1d-β_2m only were pooled, concentrated and used in crystallisation trials.

2.4.5 Expression of untagged saposins in *E. coli*

pET15b expression plasmids encoding untagged human SapB or mouse SapA were transformed into chemically competent ORIGAMI *E. coli* cells following the protocol detailed in section 2.1.6. Ampicillin and Tetracycline selection was maintained at all times. A single transformant colony was used to inoculate a 5 mL 2 x TY starter culture incubated overnight (37 °C, 200 rpm). The next day, 1 L of 2 x TY medium was inoculated with the starter culture and incubated (37 °C, 200 rpm) until it reached an absorbance at 600 nm of 0.6-0.8. Protein expression was induced by addition of 1 mM isopropyl β-D-1- thiogalactopyranoside. Following induction, incubation continued at 21 °C with shaking at 200 rpm overnight, prior to pelleting cells by centrifugation (5 000*g*, 30 min, 4 °C). Bacterial pellets were used immediately or stored at -20 °C.

2.4.6 Purification of untagged saposins

2.4.6.1 Mechanical lysis of bacterial cells

Bacterial pellets from 1 L of culture were fully resuspended in 30 mL ice-cold anion-exchange buffer (50 mM Tris pH 7.4, 25 mM NaCl) and lysed by passing twice through a hydraulic cell disruptor (Constant Systems TS) at 25 kpsi pressure, 8 °C. Insoluble cell wall and membrane fractions were pelleted by centrifugation (40 000*g*, 30 min, 4 °C) and discarded.

2.4.6.2 Heat treatment of bacterial lysate

Cleared lysate was heat-treated in a boiling water bath (100 °C, 5 min, repeated four times) to precipitate heat-sensitive proteins. Denatured, precipitated *E. coli* proteins were pelleted by centrifugation (40 000*g*, 30 min, 10 °C) and discarded. The supernatant contained heat-resistant proteins including saposins. The supernatant was supplemented with 20 µg/mL DNAse I, 2.5 mM $MgCl_2$, and 0.5 mM $CaCl_2$ to digest contaminating DNA. The solution was dialysed against 5 L of anion-exchange buffer using SnakeSkin dialysis tubing (3500 MWCO; Thermo Scientific).

2.4.6.3 Anion-exchange (AEX) chromatography

A 5 mL HiTrap Q Sepharose column (GE) was equilibrated with 10 CV anion exchange buffer (50 mM Tris pH 7.4, 25 mM NaCl) before loading the crude saposin extract using a peristaltic pump at 5 mL/min. The column was then washed with 10 CV anion exchange buffer A and connected to an FPLC (ÄKTÄ purifier, GE) for elution of bound proteins with a 0-50% gradient of 50 mM Tris pH 7.4, 1 M NaCl over 20 CV. 2 mL fractions were collected and A280 was continuously recorded during elution. Fractions corresponding to A280 peaks were analysed using 4-12% gradient Bis-Tris SDS-PAGE (section 2.6.2.2). All fractions containing saposins were pooled and concentrated to a volume of 5 mL using Vivaspin 20 centrifugal concentrators (3 000 MWCO, Sartorius).

2.4.6.4 Size exclusion chromatography

A HiLoad 16/600 Superdex 75 column (GE) was equilibrated with 1.5 CV size exclusion buffer (50 mM Tris pH 7.4, 150 mM NaCl) prior to injection of 5 mL concentrated AEX elution. Isocratic elution with size exclusion buffer was performed at 1 mL/min over 1.2 CV. 2 mL fractions were collected and A280 was continuously recorded during elution. Fractions corresponding to the main A280 peak were pooled and concentrated to 15-20 mg/mL using Vivaspin 20 centrifugal concentrators (3 000 MWCO, Sartorius). Purity was assessed by 4-12% gradient Bis-Tris SDS-PAGE (section 2.6.2.2). Purified saposins were stored at 4 °C and stable for at least six months.

2.4.7 Loading of proteins with α-GalCer

Endogenous lipids co-purified with CD1d and SapB were exchanged for the lipid α-GalCer (Avanti Lipids) based on a protocol used to obtain crystal structures of α-GalCer-loaded CD1d[176,177]. α-GalCer was resuspended in DMSO at 1 mg/mL and dissolved by heating the solution at 80 °C for 10 min, with brief sonication and vortexing every 3 minutes. CD1d-β_2m and SapB were incubated with a 3-fold molar excess of α-GalCer in PBS overnight at room temperature, in a volume such that the DMSO concentration did not exceed 1%. DMSO was then diluted out of the buffer by several cycles of dilution and concentration using centrifugal concentrators prior to concentration of the proteins for size exclusion chromatography. α-GalCer-loaded proteins were separated from free α-GalCer by size exclusion chromatography with a Superdex 200 10/300 column equilibrated in 20 mM HEPES pH 7.0, 150 mM NaCl. α-GalCer-loaded proteins were used within 24 hours after size exclusion chromatography to avoid significant hydrolysis of the lipid.

2.4.8 SapA nanodisc preparation

The protocol for SapA nanodisc preparation was adapted from Frauenfeld *et al*.[94] and Chien *et al.*[93]. Egg phosphatidylcholine (PC), 14:0 bis(monomyristoylglycero)phosphate (S,R) (BMP) and cholesterol (Chol) from ovine wool were purchased from Avanti Polar Lipids. The appropriate quantity of lipids in chloroform was mixed in a 2 mL glass vial and dried under an argon stream to form a lipid film. The lipid film was resuspended by vortexing vigorously in 104 mM citrate pH 4.0, 156 mM NaCl to make a solution with 5 mg/mL lipids. The lipid solution was incubated in a water bath at 37 °C for one hour, with vigorous vortexing every 15 min. This was followed by 10 cycles sonication in a sonicating water bath for 30 s and vortexing for 15 s. n-Dodecyl-beta-Maltoside (DDM) from a 5% (w/v) stock was then added for a final concentration of 0.2% DDM, 100 mM citrate pH 4, 150 mM NaCl, and mixed by pipetting. The solution remained cloudy. 385 µg mouse SapA was added for each milligram of lipid and mixed by pipetting. The solution was incubated at 37 °C for one hour, with mixing every 15 min. AC buffer (100 mM citrate pH 4.0, 150 mM NaCl) was added to make a final volume of 550 µL. The solution was incubated at room temperature for a further 10 min, then spin-filtered and loaded onto a Superdex 200 10/300 column (GE Healthcare) equilibrated with AC buffer. Fractions containing nanodiscs were concentrated in centrifugal concentrators (4 mL 3K MWCO, then 500 µL 5K MWCO; Amicon and Vivaspin). Nanodiscs were used in experiments within 24 hours. 1 mg lipids yielded approximately 400-600 µg nanodiscs.

2.4.9 Expression of H_6-tagged SapB in HEK293F cells

H_6-tagged SapB was expressed in HEK293F cells by polyethylenimine (PEI)-based transient transfection. 300 µg DNA encoding SapB-H_6 was diluted in 5 mL sterile PBS and mixed by brief vortexing. 450 µg PEI was added and the solution was mixed by brief vortexing. The mixture was incubated at room temperature for 20 minutes, then added dropwise to 300 mL HEK293F cells at a density of $1x10^6$ cells/mL. The cells were then incubated with shaking (37 °C, 8% CO_2, 130 rpm). After 4 days, the media was harvested as described in section 2.4.1.4.

2.4.10 Purification of SapB-H_6

SapB-H_6 was purified from conditioned media by nickel affinity purification followed by elution with 300 mM imidazole in PBS as described in section 2.4.2.1. SapB-H_6 was further purified by size exclusion chromatography with a Superdex 75 column (GE Healthcare) in cross-linking buffer (20 mM HEPES pH 7.0, 150 mM NaCl).

2.4.11 SapB-H_6 tag removal

Prior to crystallisation trials (but not cross-linking assays), the tag of SapB-H_6 was removed by incubation of SapB (1 mg/mL) with washed Carboxypeptidase A-agarose beads (1.2 U/mL; Sigma C1261) at 25 °C for 3 hours or overnight with shaking at 170 rpm. This was followed by incubation with washed Ni-NTA resin (Qiagen) on an end-over-end rotator to remove tagged protein (30 min, 4 °C). The supernatant containing untagged SapB was separated from Carboxypeptidase A-agarose and Ni-NTA agarose by passing through a gravity flow column and washing the beads several times with 250 µL cross-linking buffer. Remaining tagged SapB was eluted from Ni-NTA agarose with 300 mM imidazole in PBS. Flow-through, wash and eluted fractions were analysed by SDS-PAGE on a 4-12% Bis-Tris gel in MOPS-SDS running buffer. Flow-through and wash fractions containing pure untagged SapB in cross-linking buffer were pooled and concentrated in a centrifugal concentrator (3K MWCO; Sartorius).

2.4.12 Stable AC-H_6 and AC-BAP-H_6 expression in HEK293F cells

Production of inducible cell lines stably expressing AC-H_6 and AC-BAP-H_6 was performed as described in section 2.4.1, except that HEK293F cells were co-transfected with only 3 plasmids: PB-RN (5 µg), PBase (5 µg), and PB-T-AC-H_6 or PB-T-AC-BAP-H_6 (25 µg).

2.4.13 Purification of H_6-tagged AC

H_6-tagged AC (AC-H_6 and AC-BAP-H_6) was purified by nickel affinity purification with Ni-NTA agarose (Qiagen) and eluted with 100 mM citrate pH 4. Purified AC-BAP-H_6 was then buffer exchanged into PBS for subsequent biotinylation (section 2.4.15). Alternatively, AC-BAP-H_6 was eluted with 300 mM imidazole in PBS and imidiazole was diluted out by multiple cycles of dilution in PBS and concentration in a centrifugal concentrator (10K MWCO; Amicon).

AC-H_6 and AC-BAP-H_6 were concentrated to 2 mg/mL and stored at 4 °C in 100 mM citrate pH 4 and PBS, respectively. AC is a lysosomal protein and is therefore very resistant to proteolysis, remaining stable for several months at 4 °C.

2.4.14 AC-H_6 tag removal

The hexahistidine tag on the C-terminus of AC-H_6 was removed with Carboxypeptidase A in 20 mM Tris pH 7.4, 300 mM NaCl following the protocol described in section 2.4.11.

2.4.15 AC-BAP-H_6 biotinylation

AC-BAP-H_6 was biotinylated following a published method[178]. 40 µM (2 mg/mL) AC-BAP-H_6 was mixed with 5 µM $MgCl_2$, 2 mM ATP, 1 µM GST-BirA (previously purified by Dr Janet Deane) and 150 µM D-Biotin in 1 mL PBS. The sample was incubated on a gently shaking platform (30 °C, 2 h). The same amount of GST-BirA andD-Biotin were added to the sample for further incubation (30 °C, 2 h). GST-BirA was removed by incubating the sample with 25 µL glutathione sepharose 4B resin (GE Healthcare) with mixing (room temperature, 30 min). Beads were removed by centrifugation (4000*g*, 3 min). Biotin and ATP were diluted 200-fold by multiple cycles of dilution and concentration.

2.4.15.1 Gel-shift assay

The extent of biotinylation was measured by a gel-shift assay. Two samples with 5 µL of 10 µM (0.5 mg/mL) AC and a control sample of 5 µL PBS were mixed with 10 µL 2X SDS-PAGE loading buffer and heated at 95 °C for 5 minutes in a PCR bloc. The samples were cooled to room temperature. 5 µL of 1.75 mg/mL streptavidin in PBS were added to one AC sample and the control sample, leaving a control AC sample without streptavidin. Streptavidin was provided in excess, so that all biotinylated AC would bind streptavidin. After incubating at room temperature for 5 min, the samples were run on a 12% SDS-PAGE gel. The gel was stained with Coomassie stain and imaged with Odyssey (LI-COR Biosciences). The unbiotinylated α subunit of AC was used as a loading control. The intensity of the bands was calculated with Image Studio Lite (LI-COR Biosciences). The proportion of biotinylated AC

was calculated by dividing the ratio of unbiotinylated β subunit to α subunit in the sample mixed with streptavidin by the ratio of total β to α in the sample without streptavidin.

2.4.16 Purification of ProtA-AC27

The cell line stably expressing ProtA-AC27 was established by Dr Janet Deane using the PiggyBac system described in section 2.4.1.

2.4.16.1 Small-scale optimisation

ProtA-AC27 was captured from 3 mL conditioned media by incubation with 75 µL IgG Sepharose 6 Fast Flow beads (GE Healthcare, 17-0969-01) for 2 hours at 4 °C. Beads were spun down (500*g*, 2 min), the flow-through was removed and beads were washed twice with 1 mL TEV reaction buffer (50 mM Tris pH 7.4, 150 mM NaCl). After protein capture and wash, five different conditions for TEV protease digestion were tested in small scale experiments. 15 µL of protein-loaded beads were transferred to four PCR tubes. For the last test experiment, 1 mL of optimal TEV cleavage buffer (50 mM Tris pH 8, 0.1 M NaCl) was added to 15 µL loaded beads. The beads were spun down (500*g*, 2 min) and excess media was removed. The 50% bead slurry in optimal TEV cleavage buffer was transferred to a PCR tube. EZCut TEV Protease (BioVision, 7847-1000) was added to each tube for a final concentration of 0.1 or 0.2 mg/mL. The reaction was incubated at 37 °C or 4 °C for 2 hours or overnight. The supernatant, containing untagged AC27, was then separated from the beads by centrifugation (1000*g*, 1 min) through a Pierce microspin column (Thermo Scientific), then beads were washed four times with 50 µL TEV reaction buffer.

2.4.16.2 Large-scale preparation

ProtA-AC27 was captured from 500 mL conditioned media by incubation with 2.75 mL IgG Sepharose 6 Fast Flow beads. After ProtA-AC27 capture, IgG beads were washed with 5 CV TEV reaction buffer (50 mM Tris pH 7.4, 10 mM NaCl). TEV was then added to a 50% bead slurry for a final TEV concentration of 0.1 mg/mL, with 0.02% sodium azide to prevent bacterial growth. The mixture was incubated with gentle shaking (80 rpm) overnight at 34 °C. The supernatant containing untagged AC27 was separated from the beads using a gravity column. The beads were washed until all protein was eluted, as indicated by null absorbance at 280 nm of the wash fractions. Flow-through and wash fractions were combined, then H_6-tagged TEV was separated from AC27 by incubation with 1 mg Ni-NTA resin for 1 hour at room temperature. The flow-through was collected. Beads were washed with 30 mM imidazole in PBS. Flow-through and wash fractions, containing AC27, were combined, concentrated and the buffer was exchanged to AC buffer (100 mM citrate pH 4, 150 mM NaCl) by

several cycles of dilution and concentration in a centrifugal concentrator (10K MWCO; Amicon). AC27 was further purified by size exclusion chromatography using a Superdex 200 16/600 column equilibrated in AC buffer.

2.4.17 AC autoproteolysis

To allow autoproteolysis of AC into its active, mature form, AC was incubated in AC buffer (100 mM citrate pH 4, 150 mM NaCl) at 37 °C for 27 hours.

2.4.18 Transient expression of PTPRM Fn3-4-H_6 in HEK293F cells

PTPRM Fn3-4 (residues 478-723) was cloned into the pHLsec vector with a C-terminal H_6 tag by Eve Caroe and recombinantly expressed in HEK293F cells by transient transfection with PEI. 300 µg DNA encoding PTPRM Fn3-4-H_6 was diluted in 5 mL sterile PBS and mixed by brief vortexing. 450 µg PEI was added and the solution was mixed by brief vortexing. The mixture was incubated at room temperature for 20 minutes, then added dropwise to 300 mL HEK293F cells at a density of $1x10^6$ cells/mL. The cells were then incubated with shaking (37 °C, 8% CO2, 130 rpm). After 4 days, the media was harvested as described in section 2.4.1.4.

2.4.19 Purification of H_6-tagged PTPRM Fn3-4

H_6-tagged PTPRM Fn3-4 was purified from conditioned media by nickel affinity purification as described in section 2.4.2.1, but using 20 mM imidazole in TBS, pH 7.4 as wash buffer and 300 mM imidazole in TBS, pH 7.4 as elution buffer. PTPRM Fn3-4 was further purified by size exclusion chromatography with a Superdex 75 16/600 column (GE Healthcare) in 20 mM Tris pH 7.4, 150 mM NaCl.

2.5 X-ray crystallography: protein structure determination and analysis

2.5.1 Protein crystallisation

Proteins were concentrated with Corning Spin-X UF centrifugal concentrators. Crystallization experiments were set up using a Mosquito LCP robot (LCP Labtech) in 96-well nanolitre-scale sitting drops (200 nL protein plus 200 nL of precipitant) equilibrated at 20 °C against 80 µL reservoirs of precipitant. For most crystallisation experiments, the following crystallisation screens were set up: Morpheus (Molecular Dimensions), JCSG-plus (Molecular Dimensions), PACT Premier (Molecular Dimensions) and Index (Hampton Research). Plates were stored and imaged with Rock Imager (Formulatrix).

2.5.1.1 CD1d

Glycosylated CD1d, unglycosylated CD1d and α-GalCer loaded glycosylated CD1d was mixed with unglycosyated SapB, α-GalCer loaded unglycosylated SapB or glycosylated SapB at various concentrations and molar ratios detailed in Chapter 3. Co-crystallisation trials were set up immediately after mixing of the two components.

2.5.1.2 AC and nanodiscs

AC-H_6 (4.5 mg/mL) and SapA-PC nanodiscs (3.5 mg/mL) were mixed in equimolar quantities and immediately used to set up crystal trays, or incubated at room temperature for 75 minutes, then concentrated further to the solubility limit of the mixture prior to setting up co-crystallisation trials.

320 µL AC27 (4.1 mg/mL, 100 µM) was incubated with 320 µL psychosine (2 mM) in the same buffer (100 mM citrate pH 4, 150 mM NaCl) for one hour. The mixture was concentrated until AC27 was at a concentration of 4.1 mg/mL based on protein absorbance at 280 nm prior to setting up co-crystallisation trials.

2.5.1.3 PTPRM Fn3-4

All crystallisation experiments were set up with PTPRM Fn3-4 at 11.5 mg/mL unless indicated otherwise. Co-crystallisation experiments with galactose were performed by adding a 5-fold molar excess of galactose (2 mM) to concentrated PTPRM from a 0.1 M stock of galactose in water immediately prior to setting up co-crystallisation trials.

2.5.2 Optimisation of crystallisation conditions

When crystals, microcrystals, or microcrystalline precipitate were obtained, a grid search strategy was undertaken to optimise these crystals by manually setting up 96-well optimisation plates centred around the successful crystallisation condition with incremental changes in precipitant concentration, pH and/or additive concentration in the reservoir solution in each well.

2.5.3 Crystal microseeding

Microseeding of PTPRM Fn3-4 crystals was performed following a published protocol[179]. A stabilising solution was made with 50% original crystallisation condition (20% PEG 3350, 0.2 M ammonium nitrate), 35% water, 15% PTPRM Fn3-4 protein solution (at 9.1 mg/mL). Microcrystals were grown by

Eve Caroe in a drop containing 400 nL PTPRM Fn3-4 at 11.3 mg/mL with 200 nL reservoir solution containing 22% PEG 3350, 0.22 M ammonium nitrate. These microcrystals were resuspended in 50 uL stabilising solution. Seed stock was made by vortexing the microcrystal solution with a Teflon bead (Seed Bead, Hampton research) for 90 s. The seed solution was serially diluted 1 in 4 in stabilising solution to obtain a range of seeding solution concentrations. A 96-well crystallisation plate was set up with dilutions of the original reservoir solution (50%-72% dilution of 20% PEG 3350, 0.2 M ammonium nitrate) across columns. Drops were dispensed with 200 nL PTPRM Fn3-4 at 9.1 mg/mL and 200 nL seed stock with different dilutions of seed stock dispensed on each row.

2.5.4 Crystal soaks

PTPRM Fn3-4 crystals were soaked with galactosphingolipid headgroup analogues prior to harvest. 0.5 M stocks were made with galactose or D-Galactose 4-sulfate (Gal-4S) sodium salt (Sigma 90572) in water. A soak solution was made with 10 mM galactose or Gal-4S in reservoir solution. Crystals were transferred to a drop of soak solution and incubated for 10 to 30 minutes. Non-cryoprotective conditions were supplemented with 20% glycerol in soak solution prior to harvest. Crystals were then harvested using nylon loops and flash frozen in liquid nitrogen.

2.5.5 SDS-PAGE analysis of crystals

Crystal contents were analysed by SDS-PAGE analysis of crystals. Several crystals were harvested from the crystallisation drop and washed by transferring to a drop of reservoir solution to remove uncrystallised protein components. Crystals were then transferred to a second drop of reservoir solution and dissolved in SDS-PAGE loading buffer. Alternatively, large crystals from which data had been collected at the synchrotron were recovered into SDS-PAGE loading buffer from the crystal-mounting loop. Individual crystal components were analysed by SDS-PAGE followed by Coomassie staining.

2.5.6 X-ray data collection

Crystals were cryoprotected in reservoir solution supplemented with 20% v/v glycerol (unless the crystallisation condition was already cryprotective, as is the case in the Morpheus screen) and flash-cooled by plunging into liquid nitrogen. Diffraction data were recorded at Diamond Light Source on beamlines I03, I04 and I04-1. Diffraction data were indexed and integrated using the automated data processing pipelines at Diamond Light Source[180] implementing XIA2 DIALS[181], XIA2 3dii, autoPROC or autoPROC and STARANISO[182] depending on the extent of anisotropic diffraction, then scaled and merged using AIMLESS[183].

2.5.7 Molecular replacement

2.5.7.1 CD1d

The model for molecular replacement was generated using the published structure of CD1d-β_2m bound to lysophosphatidylcholine[184] (PDB ID: 3U0P, chains A and B), using phenix.sculptor[185,186] to remove the lipid ligand from the model and revert the glycosylation site mutations present in this structure. Molecular replacement was performed with Phaser-MR[187] and was followed by rigid body refinement using phenix.refine[188]. Molecular replacement into the highly anisotropic data (crystal form (iii) described in Table 3.2) was carried out using both the isotropic and anisotropically processed data, resulting in the same solution. Maps were inspected using Coot[189].

2.5.7.2 SapA assembly

The molecular replacement model for the SapA assembly structure was the open human SapA monomer (PDB ID: 4DDJ). The protein chain was extracted from the deposited model and the anisotropic B-factors of this high-resolution structure were converted to isotropic factors (B-factors = 30) using the PDB file editor in phenix[185]. The SapA assembly structure was solved by molecular replacement using four copies of the SapA model using Phaser-MR[187]. The resulting model was partially incorrect, so the fourth SapA chain was deleted from this solution using the PDB file editor in phenix[185] and re-placed manually by symmetry with the other 3 chains with the Secondary Structure Matching (SSM) superpose tool in Coot[189], with kind help from Randy Read.

2.5.7.3 PTPRM Fn3-4

The PTPRM Fn3-4 structure was solved by molecular replacement with four copies of the unpublished PTPRM Fn3-4 structure (chain A) previously obtained by Eve Caroe, Dr Stephen Graham and Dr Janet Deane using Phaser-MR[187].

2.5.8 Structure refinement

2.5.8.1 SapA assembly

The model was refined by rigid body refinement with phenix.refine[188]. Mismatches in the mouse vs human SapA sequence were mutated manually in Coot[189]. The model was further refined by XYZ coordinate refinement in reciprocal space with non-crystallographic symmetry and reference model restraints to minimise over-fitting of the data. The model was then refined by molecular dynamics simulation and structure manipulation in ISOLDE[190]. Where possible, obvious Ramachandran outliers,

rotamers outliers and non-Proline *cis* peptide bonds were corrected in ISOLDE. The model was then refined in phenix.refine. This cycle was repeated once. Model geometry was evaluated with MolProbity[191] and ISOLDE validation tools throughout the refinement process.

2.5.8.2 PTPRM Fn3-4

After molecular replacement, PTPRM Fn3-4 was refined in phenix.refine[188]. The first round of rigid body refinement corrected the erroneously assigned space group *P* 3_1 2 1 to *P* 3_2 2 1. Missing parts of the model were built in Coot[189]. The model was refined by multiple cycles of molecular dynamics simulation and structure manipulation in ISOLDE[190] followed by refinement in phenix.refine. Ramachandran outliers, rotamer outliers and non-Proline *cis* peptide bonds were corrected in ISOLDE. The following settings were used for refinement in phenix.refine: XYZ coordinate refinement in reciprocal space with individual B-factors and occupancies refinement, non-crystallographic symmetry restraints, X-ray/stereochemistry weight and X-ray/ADP weight optimisation and Ramachandran restraints. N-linked glycans were built in Coot using the carbohydrate building tool[192]. Model geometry was evaluated with MolProbity[191] and ISOLDE validation tools throughout the refinement process.

2.5.9 Structure analysis

Structure alignments were performed with the Matchmaker tool in ChimeraX[193]. Potential hydrogen bonds were identified in ChimeraX. Potential hydrophobic interactions were determined by identifying pairs of hydrophobic residues closer than 0.4 Å (Wan der Waals radius overlap ≥ -0.4 Å) with the Contacts tool in ChimeraX. Structural figures were rendered with ChimeraX. Surface residue hydrophobicity colouring was performed using the Molecular lipophilicity potential tool[194,195]. Structure colouring by conservation was done in Chimera[196].

2.5.10 Protein sequence alignment

Profile-profile alignment was performed using the Fold and Function Assignment System (FFAS) server[197] and structure based-alignment using the Dali server[198]. Furin cleavage sites were found using the ELM server[199]. Multiple sequence alignments were performed using MULTALIN[200] and visualised with Jalview[201].

2.6 Protein biochemistry

2.6.1 Protein quantification

2.6.1.1 Measuring UV absorbance at 280 nm (A280)

Concentrations of purified proteins containing at least one tyrosine or tryptophan were determined by measuring the A280 of a 1.5 µL sample using a NanoDrop One spectrophotometer (Thermo). Prior to measurement the instrument was blanked in appropriate buffer. Concentration was calculated by using the theoretical extinction coefficient at 280 nm and optical density provided by ExPASy protparam.

2.6.1.2 BCA microplate assay

A blank solution (lysis buffer diluted 1/5) and a set of BSA concentration standards (2000, 1000, 500, 250, 125, 62.5 µg/mL) in blank solution were prepared. Cell lysates were diluted 1/5 prior to analysis. 25 µL of each standard and diluted sample were transferred to a 96-well flat bottom transparent microplate in duplicate. 200 µL of Pierce BCA protein assay working reagent (Thermo Fisher) was added to each well using a dispensing pipette and the plate was incubated at 37 ˚C for 30 min prior to measuring the absorbance at 570 nm using a Spectramax M5 plate reader. Concentrations of protein samples were calculated from the BSA standard curve using a quadratic fit.

2.6.2 Analysis of proteins by polyacrylamide gel electrophoresis

2.6.2.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Gels were poured using Mini-PROTEAN glass plates and casting equipment (Bio-Rad). Gel composition is described in Table 2.2. Protein samples were denatured by heating (95 ˚C, 5 min) in SDS-PAGE loading buffer (20 mM DTT, 40 mM Tris, 2 % w/v SDS, 4% glycerol, 0.01% w/v bromophenol blue) prior to separation in SDS running buffer (20 mM Tris-HCl, 154 mM glycine, 0.1% w/v SDS) using a BioRad Mini-PROTEAN electrophoresis cell at room temperature (40 mA/gel, 200 V, < 30 W, 50 min).

2.6.2.2 Gradient Bis-Tris polyacrylamide gel electrophoresis

Pre-cast 4-12% Bis-Tris polyacrylamide gradient gels (NuPAGE) were also used for protein electrophoresis with 1X MES-SDS running buffer (NuPAGE) or, where indicated, 1X MOPS-SDS running buffer (NuPAGE) as directed by the manufacturer (125 mA/gel, 200 V, < 30 W, 35 min). Protein was denatured and reduced as above prior to separation.

2.6.2.3 Visualising separated proteins

SDS-PAGE gels were incubated at room temperature overnight with 10 mL InstantBlue Coomassie based reagent (Expedeon) on a rocking platform. Before scanning, gels were rinsed in water to remove excess stain. Gels were scanned as uncompressed TIF files, 600 dpi.

Table 2.2. SDS-PAGE gel compositions

Components	10% resolving gel	12% resolving gel	15% resolving gel	Stacking gel
Water	3.80 mL	3.20 mL	2.20 mL	2.04 mL
30% Acrylamide solution	3.40 mL	4.00 mL	5.00 mL	510 µL
Resolving gel buffer (1.5 M Tris-HCl, pH 8.8)	2.60 mL			-
Stacking gel buffer (1 M Tris-HCl, pH 6.8)	-			390 µL
20% SDS	50 µL			15 µL
Tetramethylethylenediamine	20 µL			10 µL
10% Ammonium persulfate	100 µL			30 µL

2.6.2.4 SYPRO Ruby staining

For sensitive protein detection, 1.5 mm 4-12% Bis-Tris gels were stained with SYPRO Ruby (Thermo Fisher) after SDS-PAGE electrophoresis, following the rapid protocol described in the product manual: the gel was fixed twice in 150 mL of fix solution (50% methanol, 7% acetic acid) with agitation on an orbital shaker at room temperature for 15 min each time. The fix solution was replaced with 90 mL SYPRO Ruby stain. The gel was microwaved for 45 s at full power, agitated for 30 s, microwaved for 45 s and incubated with shaking at room temperature for 5 min. The gel was then reheated for 45 s in the microwave then agitated at room temperature for 23 min, for a total stain time of 30 min. The gel was transferred to a clean container and washed in 150 mL of wash solution (10% methanol, 7% acetic acid) with agitation at room temperature for 30 min. The gel was then rinsed in MQW for 5 min twice prior to imaging using a G:BOX imaging system (Syngene).

2.6.3 Western blot

2.6.3.1 Cell lysis

For analysis of TNF-α in THP-1 cells, $3x10^6$ cells were washed in PBS 3 times, lysed in 300 µL RIPA buffer (50 mM Tris HCl, PH 7.4, 150 mM NaCl, 1% Triton X-100 (TX-100), 0.5% Sodium deoxycholate, 0.1% SDS, 1 mM EDTA, Complete protease inhibitor cocktail (Roche), 1 mM phenylmethylsulfonyl fluoride

(PMSF)) and incubated at 4 °C for 30 min on a rotating wheel. The lysate was cleared by centrifugation (12 000*g*, 20 min). Protein concentration was determined by BCA assay as described in section 2.6.1.2 and protein concentration was equalised across samples.

For analysis of cleaved Caspase-3 (Casp-3) in WT and GALC KO MO3.13 oligodendrocytes, cells were seeded in 6 well plates and differentiated on the next day. As a positive control, WT oligodendrocytes were treated with 2 µM staurosporine (Cell Guidance Systems) for 24 hours before cell lysis. 3 or 7 days after differentiation, cells were washed twice in PBS, lysed in 100 µL lysis buffer (10 mM Tris pH 8, 1 mM EDTA, 1% TX-100, 140 mM NaCl, 1 mM PMSF) and incubated at 4 °C for one hour on a rotating wheel. The lysate was cleared by centrifugation (12 000*g*, 10 min). Protein concentration was determined by BCA assay as described in section 2.6.1.2 and protein concentration was equalised across samples.

2.6.3.2 Electrophoretic separation and transfer of proteins

Protein was denatured in SDS-PAGE loading buffer and 60 µL of protein lysate were loaded onto a gel for protein separation by SDS-PAGE as described in section 2.6.2.1, alongside 5 µL of PageRuler Plus pre-stained protein ladder (Thermo Scientific). Proteins were transferred onto a nitrocellulose membrane in wet transfer buffer (25 mM Tris, 192 mM glycine, 20% v/v ethanol) in a Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad) at 100 V for 1 hour.

2.6.3.3 Immunoblotting

Antibodies used for immunoblotting are documented in Table 2.3. All antibodies were diluted in 5% w/v dried milk powder in PBST (PBS, 0.1% v/v Tween-20). All incubation steps were performed on a rocking platform. Wash steps were performed with 5% w/v milk in PBST. For Western blot analysis of cleaved Casp-3, PBS was replaced by TBS in all steps.

Membranes were blocked for one hour in 5% w/v milk in PBS at room temperature. Membranes were then incubated with primary antibody overnight at 4 °C and washed (5 min, 3 times) prior to incubation with infrared fluorophore–conjugated secondary antibody on a rocking platform (1 hour, room temperature). Membranes were washed with 5% milk in PBST (5 minutes, twice), then washed once in PBST (15 min), rinsed in PBS and dried. Immuno-reactive bands were visualised with an Odyssey CL-x infrared imager (LI-COR Biosciences). Band intensities were quantified using Image Studio-Lite (LI-COR Biosciences).

Table 2.3. Antibodies used for Western blot experiments

Target	Source	Dilution	Provider and ID
Primary Antibodies			
β_2m	Rabbit polyclonal	1:10000	Dako A0072 (provided by Louise Boyle)
GFP	Rabbit polyclonal	1:5000	Sigma G1544
TNF-α	Mouse monoclonal	1:500	R&D Systems MAB610
TNF-α	Rabbit polyclonal	1:1000	Cell Signalling Technology 3707S
Cleaved Casp-3	Rabbit polyclonal	1:1000	Cell Signalling Technology 9661
GAPDH	Mouse monoclonal	1:1000	Ambion AM43000
Secondary Antibodies			
Mouse IgG IRdye800	Goat polyclonal	1:10000	Li-COR Biosciences
Rabbit IgG IRdye680	Goat polyclonal	1:10000	Li-COR Biosciences

2.6.4 Equilibrium binding assay

Ni-NTA agarose beads were loaded with a saturating amount of CD1d-β_2m, and complete capture of CD1d-β_2m was verified by measuring the absorbance of the supernatant at 280 nm. Saturated beads were washed three times in equilibrium assay buffer (50 mM HEPES pH 7.0, 150 mM NaCl, 0.05% digitonin, 2.5 mM $CaCl_2$, 2.5 mM $MgCl_2$). Serial dilutions (1:1) of CD1d-β_2m-saturated beads were made in equilibrium assay buffer supplemented with unloaded Ni-NTA beads to equalize the bead volume across all samples. SapB was added at a final concentration of 680 nM in equilibrium assay buffer. The reaction was incubated with shaking at room temperature for 1 hour. The supernatant containing unbound SapB was separated from the beads by centrifugation (800*g*, 2 min) followed by separation with a Micro-Spin column (Pierce) (800*g*, 2 min) to remove all beads. The supernatant was analysed by SDS-PAGE followed by staining with SYPRO Ruby (Life Technologies) as described in section 2.6.2.4. For analysis of bead samples, these were boiled in SDS-PAGE loading dye, separated by SDS-PAGE and detected by Coomassie staining.

2.6.5 Cross-linking assay

Amine-reactive cross-linking agents disuccinimidyl sulfoxide (DSSO) and PEGylated bis(sulfosuccinimidyl)suberate ($BS(PEG)_5$) were dissolved in DMSO and added to 10 µM CD1d-β_2m alone, 20 µM SapB alone or 10 µM CD1d + 20 µM SapB with final cross-linker concentrations of 0, 200 µM or 1 mM in cross-linking buffer. The reaction was incubated for 30 min at room temperature and terminated by incubation with 20 mM Tris pH 8 for 15 min. The reaction products were analysed by SDS-PAGE on a 4-12% Bis-Tris gradient gel followed by Coomassie staining.

2.6.6 GFP co-immunoprecipitation

2.6.6.1 Expression of GFP-tagged constructs

80% confluent HeLa cells in 12 mL media in a 100 mm dish were transfected with 12 µg vector (pEGFP-N1 empty vector, CD1d-EGFP or HLA-A2-EGFP) pre-mixed with 36 µL FuGene 6 in Opti-MEM. 15 hours after transfection, the media was replaced with fresh media containing 200 U/mL interferon γ (IFN-γ; a kind gift from Louise Boyle). 48 hours after IFN-γ stimulation, cells were scraped into a falcon tube and washed 3 times by centrifugation (220*g*, 5 min, 4 °C) and resuspension in PBS.

2.6.6.2 Cell lysis

Cells were lysed for 30 minutes in 1 mL ice cold lysis buffer (50 mM citrate pH 6, 150 mM NaCl, 0.5 mM EDTA, 0.5% NP-40, 1x Pierce protease inhibitor tablet (Pierce, A32963)) by pipetting cells up and down every 10 minutes. The lysate was cleared by centrifugation (20 000*g*, 10 min, 4 °C). Total protein concentration in the supernatant was measured by BCA assay as described in section 2.6.1.2 and protein concentration was equalized across samples.

2.6.6.3 Immunoprecipitation

20 µL GFP-Trap_A (ChromoTek, gta-20) slurry pre-washed 3 times in wash buffer (50 mM citrate pH 6, 150 mM NaCl, 0.5 mM EDTA) were added to lysates. The lysates were incubated with resin on an end-over-end rotator for 30 min at 4 °C. Beads were pelleted by centrifugation (2500*g*, 4 °C, 2 min) and washed twice with 1 mL wash buffer. Beads were resuspended in 45 µL 2x SDS-PAGE loading buffer and boiled at 95 °C for 5 min. Beads were pelleted by centrifugation (20 000 *g*, room temperature, 2 min) and discarded. The supernatant was stored at -20 °C prior to WB (section 2.6.3) or mass spectrometry (section 2.9.8) experiments.

2.6.7 SEC-MALS analysis of SapA nanodiscs

Size exclusion chromatography - multi-angle light scattering (SEC-MALS) was performed at room temperature. 100 µL of SapA-PC nanodiscs was injected onto a Superdex 200 increase 10/300 GL column (GE Healthcare) pre-equilibrated in 100 mM citrate pH 4, 150 mM NaCl at a flow rate of 0.5 mL/min. The static light scattering, differential refractive index, and the UV absorbance at 280 nm were measured in-line by DAWN 8+ (Wyatt Technology), Optilab T-rEX (Wyatt Technology), and Agilent 1260 UV (Agilent Technologies) detectors. The molar mass was calculated using ASTRA 6 software (Wyatt Technology).

2.6.8 Nanodisc pull-down assay with biotinylated AC

10 µL of Pierce high capacity strepatividin agarose beads were washed twice by resuspending in 1 mL PBS and spinning at (500*g*, 5 min). 50 µg of biotinylated AC in 150 µL PBS was bound to the beads by incubating on a rotating wheel at room temperature for 1 hour. The beads were spun down (500*g*, 2 min) and washed twice with 150 µL pull-down buffer (50 mM citrate pH 4.0, 20 mM NaCl). 150 µL pull-down buffer, 160 µg SapA in 150 µL pull-down buffer, or nanodiscs containing 160 µg SapA (after buffer exchange into 150 µL pull-down buffer) were then added to the beads and incubated on a rotating wheel at room temperature for 1 hour. Beads were spun down (2500*g*, 1 min) and washed 3 times with 150 µL pull-down buffer. AC was eluted by boiling the beads in 20 µL 1X SDS-PAGE buffer. Eluted protein was run on a 15% SDS-PAGE gel and protein was stained with Coomassie stain.

2.7 Enzyme activity assays

2.7.1 GALC activity assay

2.7.1.1 Lysate preparation

4-10x10^6 THP-1 cells were pelleted by centrifugation (300*g*, 5 min, 4 °C) and washed twice with ice-cold PBS. Alternatively, adherent MO3.13 cells or differentiated THP-1 in a confluent T75 flask were gently washed three times with ice-cold PBS. For suspension cells, cell pellets were resuspended in 1 mL ice-cold lysis buffer (10 mM Tris pH 8, 1 mM EDTA, 1% TX-100, 140 mM NaCl, 1 mM PMSF). For adherent cells, 1 mL ice-cold lysis buffer was added to the cells and cells were scraped into an Eppendorf tube. Lysates were incubated for 1 hour at 4 °C on an end-over-end rotator, then cleared by centrifugation (12 000*g*, 10 min) and cleared lysates were transferred to a new Eppendorf tube. Protein concentration in each sample was determined with a bicinchoninic acid assay (Pierce) and protein concentration was equalised across samples. For each sample, two tubes containing 300-1000 µg protein in 1 mL lysate were aliquoted, to obtain one experimental sample and one “no antibody” control sample.

2.7.1.2 GALC immunoprecipitation

Lysate components non-specifically interacting with protein A resin were pre-cleared by incubation of the lysate with 40 µL washed Protein A slurry on an end-over-end rotator for 1 hour at 4 °C. Protein A beads and non-specifically bound proteins were removed by centrifugation (12 000*g*, 5 min, 4 °C). Pre-cleared lysate was incubated with 50 µL monoclonal antibody against GALC and 40 µL washed Protein A slurry overnight at 4 °C. GALC antibody was omitted from the “no antibody” control samples. This antibody was previously generated via a collaboration with Dr Karsten Skjødt, validated by our lab and

published in Spratley *et al.*[202]. Beads were pelleted by centrifugation (1000*g*, 5 min) and the supernatant was discarded. Beads were washed with 500 µL lysis buffer three times.

2.7.1.3 Activity assay

Beads were incubated with 400 µL pre-warmed activity assay buffer (20 mM sodium acetate, 150 mM NaCl, 0.1% NP-40, pH 4.6) with 0.3 mM 4-Methylumbelliferyl-β-D-galactopyranoside (4MβDG) at 37 °C, protected from light in an Infors shaker (210 rpm). At three time points, 110 µL of fully resuspended sample was mixed with 27.5 µL Stopping Buffer (2.4 M NaOH, 1.9 M Glycine, pH 13.6). Beads were pelleted by centrifugation (20 000*g*, 5 mins) and 125 µL of supernatant was transferred to a black half-volume 96-well plate. Fluorescence was measured using a Clariostar Plus (BMG) or SpectraMax M5 (Molecular Devices) microplate reader with excitation at 365 nm and emission at 445 nm.

2.7.2 AC Activity assay

2.7.2.1 Ceramide solubilisation

Several detergent mixtures were tested for ceramide solubilisation. Dried ceramide was resuspended by vigorous vortexing in 20 mM citrate pH 4, 150 mM NaCl. Various detergents were then added. The following detergent compositions used in published assays were tested:

- 0.125% TX-100 + 0.05% Tween 20 + 0.2% Sodium cholate[203];
- 0.5% TX-100 + 0.2 % Tween 20 + 0.2 % NP-40 + 0.8 % Na Cholate[204];
- 1% Na Cholate[205];
- 1% TX-100.

The mixture was sonicated in a sonicating water bath for 5 min. Solubilisation efficiency was visually assessed based on the cloudiness of the ceramide solution after this process.

2.7.2.2 Activity assay with AC after H_6 tag removal

C12 ceramide (d18:1) (Avanti, 860512P) and lauric acid (Sigma W261416) stocks were made in chloroform, aliquoted and dried under an argon stream to form a lipid film for storage. 50 nmoles (24 µg) of dried ceramide was resuspended in activity assay buffer without detergent (20 mM citrate pH 4, 150 mM NaCl) by vortexing, incubated at 37 °C for 1 hour with vortexing every 15 minutes, and liposomes were formed by 10 cycles of sonication for 1 min in a sonicator bath and vortexing for 10 s. TX-100 was then added for a final concentration of 1% (v/v). The lipid mixture was vortexed for 1 min and sonicated for 5 min. 0.65 nmoles (30 µg) AC was added to the lipids and mixed by pipetting gently. The activity assay reaction contained 4.8 µM AC, 370 µM ceramide and 1% (v/v) TX-100 in a final

volume of 135 µL. AC buffer (100 mM citrate pH 4) was added to the negative control instead of AC. The reaction was incubated at 37 °C for 4 hours and stopped by addition of 1 mL 2:1 chloroform/methanol. The chloroform phase was transferred to a fresh vial and dried under an argon stream. Lipids were resuspended in 30 µL chloroform.

2.7.2.3 Activity assay with AC27

50 nmoles (24 µg) of dried ceramide was resuspended in activity assay buffer with detergent (20 mM citrate pH 4, 150 mM NaCl, 1% (v/v) TX-100), and subjected to 4 cycles of heating in a water bath at 85 °C for 30 s, sonication for 30 s and vortexing for 10 s. The lipid solution was cooled prior to the addition of 0.6 nmoles (25 µg) AC27, mixed by pipetting gently. The activity assay reaction contained 3 µM AC, 250 µM ceramide and 1% (v/v) TX-100 in a final volume of 200 µL. Besides these differences, the assay was performed as described above.

2.7.2.4 Thin layer chromatography (TLC)

Silica-coated glass plates (Sigma, Z185310) or polyester plates (Sigma, Z122777) were developed in 80:10:1 (v/v/v) chloroform:methanol:water. The plate was pre-washed by developing the blank plate and air-dried for 5 min before running the samples. In order to maximise saturation of the chamber with solvent, the tank was lined with filter paper and resealed swiftly after opening. Lipids extracted from the activity assay reaction (30 µL) were manually spotted on the TLC plate using a 10 µL Hamilton syringe, alongside 10 µg of lipid standards.

2.7.2.5 Lipid detection

After chromatogram development, plates were air-dried for 5 min, sprayed with 10% cupric sulfate in 8% phosphoric acid and allowed to dry for 30 minutes. Plates were charred in an oven (150 °C, 15 min) and imaged within 30 minutes using the epi-white setting on a ChemiDoc (Biorad) system.

2.8 Generation of GALC knockout and rescue THP-1 cell lines

2.8.1 Guide RNA design

Target sequences for *GALC* gene editing by CRISPR/Cas9 were selected by Dr Janet Deane and Dr Stuart Fawke using the Dharmacon CRISPR Design Tool. No off-target effects were predicted for the 20-nucleotide sequences targeted by gRNA1 (5'-AAAGTGGAAATAGGTGGTGA-3') and gRNA2 (5'-GGTGGTGATGGGCAGACAAC-3'). These target sequences were cloned into the pSpCas9 BB vector facilitating the co-expression of gRNA with Cas9 by Dr Stuart Fawke.

2.8.2 Electroporation

THP-1 cells were electroporated with pSpCas9 BB vectors encoding Cas9 and gRNA targeting GALC following a protocol provided by Leah Hunter and Colin Crump (Department of Pathology, University of Cambridge). $2x10^6$ WT THP-1 cells were pelleted (1000*g*, 5 min) and resuspended in ice-cold THP-1 media. 400 µL of cell suspension at $5x10^6$ cells/mL was transferred to a 0.4 cm Gene Pulser cuvette (Bio-Rad). Cells were electroporated with 2.5 µg pSpCas9 BB vector in a Gene Pulser XCell system (Bio-Rad) at 250 V, 1000 µF. Immediately after electroporation, 500 µL warm THP-1 media was added to the cuvette. Cells were then transferred to a 6-well plate and supplemented with 2 mL warm media. After 24 h, cells were pelleted (1000*g*, 5 min) and resuspended in fresh media with 2 µg/mL puromycin for antibiotic selection. After two days, the media was changed to puromycin-free media and cells were allowed to recover for 2 weeks.

2.8.3 Lentiviral transduction of THP-1-Cas9 cells with gRNAs

THP-1 cells stably expressing FLAG-tagged *Streptococcus pyogenes* Cas9 (Cas9 THP-1) were a gift from the Lehner lab. This cell line was created using a lentiviral vector with Blasticidin resistance followed by clonal selection. Lentivirus stocks of the vectors encoding gRNA1 and gRNA2 (section 2.2.6) were produced as described in section 2.3.3.2. For lentiviral transduction of gRNAs, $0.5x10^6$ Cas9 THP-1 cells were seeded in THP-1 media into each well of three separate 24-well plates. As the lentivirus concentration was unknown, a range of volumes (0.025 to 1 mL) of each lentivirus stock (gRNA1, gRNA2 and EV) was added across each plate. THP-1 media was added to the wells for a final volume of 1.25 mL per well. Two additional wells were seeded as untransduced controls. The plates were centrifuged at 230*g* for 20 minutes to sediment lentivirus particles, then incubated at 37 °C in humidified air with 5% CO_2 for 24 hours.

2.8.4 Antibiotic selection

Cells from each plate were split 1:1 into two 24-well plates. The volume in each well was adjusted to 1.25 mL with THP-1 media. 2 µg/mL puromycin was added to each well of the second plate. Cells were incubated at 37 °C for 3 days. Successful antibiotic selection was confirmed by the death of puromycin-treated untransduced cells, as assessed by trypan blue staining.

2.8.5 Assessment of transduction efficiency

To determine which virus concentration results in the infection of 30% of cells, expression of the Blue Fluorescent Protein (BFP) fluorescent marker from lentiviral vectors was analysed by flow cytometry in cells grown in the absence of antibiotic selection. To this end, cells in suspension were pooled with any adhered cells detached by incubation with 200 µL trypsin for 5 min. Cells were pelleted (350*g*, 5 min), resuspended in 1 mL PBS and transferred to a polystyrene tube. Cells were analysed on a BD LSRFortessa cell analyser (20 000 events per sample, medium speed). BFP expression was detected using a 405 450/50 filter and cell autofluorescence was detected using a 488 530/30 filter. The lentivirus concentration yielding 30% infection in unselected cells was determined, and puromycin-selected cells transduced with this virus concentration were carried forward. Suspension cells were pooled with adherent cells detached with trypsin, pelleted (125*g*, 5 min) and washed with PBS three times. These mixed knockout cell populations were expanded, frozen and stored in liquid nitrogen.

2.8.6 Clonal selection

Clonal selection was done by limiting dilution. For each mixed population, suspension and detached adherent cells were pooled, counted, pelleted (200*g*, 5 min) and diluted to 5 cells/mL in fresh THP-1 media supplemented 1:1 with conditioned media from the flask of origin. 100 µL of diluted cells were distributed to each well of a 96-well plate. Clonal cell populations were then gradually expanded to 20 mL. Four clones per sample were frozen and carried forward for knockout validation.

2.8.7 Assessment of GALC gene editing

Genomic DNA was extracted from $5x10^6$ cells from each clonal population using a Pure Link Genomic DNA Mini Kit following the manufacturer's instructions. A region of DNA (862 bp) including the *GALC* editing site was amplified by PCR using the following primers:
forward: 5'-GCCCTACTTGCCCAATGATTGGCACAGAAAG-3';
reverse: 5'-CAAAACTGTCCCTATGCTCTGTCCTGTATATATAG-3'.
A 50 µL reaction containing 250-500 ng genomic DNA, 500 nM forward primer, 500 nM reverse primer and 0.25 µL GoTaq polymerase in GoTaq buffer with 200 µM dNTPs was incubated at 95°C for 5 minutes followed by 30 cycles of denaturation (95 °C, 30 s), annealing (64.7 °C, 30 s), and extension (72 °C, 1 min) and a final extension (72 °C, 5 min). The PCR product was purified using the QIAquick PCR Purification Kit (Qiagen). Sanger sequencing was done by GENEWIZ using the sequencing primer 5'-CACCATAGAATTTATTAGCCTCA-3'. Assessment of *GALC* gene editing was done by sequencing trace decomposition using the online tool TIDE (https://tide.nki.nl/)[206]. EV was used as a control. Indel size

range was set to the maximum (0-50). TIDE analysis was performed on two independent samples from each cell line.

2.8.8 Generation of GALC rescue cell lines

Lentivirus stocks encoding GALC Rescue 1 and GALC Rescue 2 recombinant constructs (section 2.2.7) were produced as in section 2.3.3.2. GALC KO1 and GALC KO2 THP-1 cells were seeded in 6-well plates for a final cell density of $0.2x10^6$ cells per well in 2 mL THP-1 media. 0-800 µL GALC Rescue 1 or GALC Rescue 2 lentivirus stock, respectively, was added to each well. After 48 h, geneticin was added to each well for a final concentration of 400 µg/mL. Cells were maintained under antibiotic selection for 4 days and cultures were expanded as cell density increased. Some untransduced control cells remained alive after 4 days of antibiotic selection, therefore antibiotic dosage was increased to 800 µg/mL geneticin for 4 days, at which point all untransduced cells were dead. The number of infection events was approximated based on the number of geneticin-resistant cells after 10 days, assuming a doubling rate of 24 hours. The culture produced by transduction of 5-10% of cells was selected as the rescue cell line.

2.8.9 Rescue GALC expression

Rescue GALC expression was induced by addition of 25-1000 µg/L doxycycline in the culture media. Cells were maintained in media with doxycycline for a minimum of 6 days prior to validation of GALC rescue by GALC activity assays (section 2.7.1) and downstream qPCR experiments.

2.8.10 Transduction of WT THP-1 cells with an empty lentiviral control vector

Viral stocks of unmodified puromycin-resistant pCW57 empty lentiviral vector were made by Dr Stuart Fawke. WT THP-1 cells were transduced with this vector as described in section 2.8.8. 72 hours after transduction, transduced cells were selected with 2 µg/mL puromycin for 72 hours. Cells were allowed to recover in THP-1 media without antibiotic for 7 days.

2.9 Proteomics

All chemicals were of AR grade or better and purchased from Merck/Sigma unless otherwise stated.

2.9.1 Cell culture

Proteomics analysis was performed on three independent biological replicates. EV, KO1 and KO2 cells were seeded at $0.25x10^6$ cells/mL in 30 mL THP-1 media in 150 mm dishes to harvest $11.5x10^7$ cells at

a cell density of $0.5x10^6$ cells/mL the next day. Cells were pelleted (300*g*, 5 min), washed in 15 mL PBS and transferred to a clean 15 mL falcon, pelleted and washed once more, then resuspended in 1.15 mL PBS. 1 mL of resuspended cells ($1x10^7$ cells) was transferred to a low bind Eppendorf tube for plasma membrane profiling, and 150 µL ($1.5x10^6$ cells) were transferred to a Bioruptor microtube (Diagenode) for whole cell proteomics and stored at -80 °C until processing (section 2.9.3). Cells were pelleted and the supernatant was discarded.

2.9.2 Sample preparation for plasma membrane profiling

2.9.2.1 Biotinylation of plasma membrane proteins

Surface sialic acid residues were oxidized with sodium meta-periodate and biotinylated using aminooxy-biotin, a reaction catalyzed by aniline. The oxidation/biotinylation mix (1 mM sodium meta-periodate (Pierce, 20504), 100 mM aminooxy-biotin (Insight biotech, 90113), 10 mM aniline (Sigma, 242284) in ice-cold PBS pH 6.7) was freshly prepared and mixed by vortexing. Aniline was added immediately prior to use. Cells were thoroughly resuspended in 1 mL oxidation/biotinylation mix and incubated at 4 °C for 30 min with end-over-end rotation in the dark. The oxidation reaction was quenched by addition of glycerol to a final concentration of 1 mM. Cells were washed three times with PBS and pelleted by centrifugation (300*g*, 5 min). The cell pellet was stored at -80 °C until samples from all three replicates were harvested. Downstream processing was performed at the same time for all samples.

2.9.2.2 Cell lysis

Biotinylated cell pellets were thawed on ice. Cells were lysed in 1 mL lysis buffer (10 mM Tris pH 7.4, 1% TX-100, 150 mM NaCl, 1× complete protease inhibitor without EDTA (Roche)) on an end-over-end rotator at 4 °C for 30 min. Lysates were clarified by centrifugation (13 000*g*, 4 °C, 10 mins) and the supernatant was transferred to a fresh tube. The clarification step was repeated a second time. Biotinylated proteins were enriched by incubation with 50 µL of washed high capacity neutravidin agarose bead slurry (Thermo Fisher Scientific, 29202) at 4 °C for 2 hours with end-over-end rotation.

2.9.2.3 Preparation of peptides for mass spectrometry

Extensive washing of beads was performed using a vacuum manifold and SnapCap filter columns (Pierce, 69725). Beads were washed 20 times with 400 µL Lysis buffer, followed by 20 washes with 0.5 % SDS in PBS (pH 7.4), then 10 washes with freshly made Urea buffer (50 mM triethylammonium bicarbonate buffer (TEAB), 6 M urea, pH 8.5). Columns were removed from the vacuum manifold and

capped, and beads were incubated with 400 µL reduction/alkylation solution (10mM Tris(2-carboxyethyl)phosphine (Sigma 646547**)**, 20mM iodoacetamide in Urea buffer) in the dark with mixing in a thermomixer (1300 rpm, 30 min). The reduction/alkylation solution was removed using the vacuum manifold, and beads were washed 10 times with Urea buffer, followed by 5 washes with 50 mM TEAB. Beads were resuspended in 400 µL 50 mM TEAB and transferred to a low bind Eppendorf tube. Beads were pelleted (500*g*, 2 min) and supernatant was removed using a fine gel loading tip with a constricted end. Beads were resuspended in 50 µL of 50 mM TEAB with 0.5 µg trypsin and incubated at 37 °C for 6 hours in a thermomixer (1500 rpm). Beads were pelleted (500*g*, 3 min) and supernatant was transferred to a new tube. Beads were resuspended with a further 50 µL of 50 mM TEAB, pelleted by centrifugation and the supernatant was added to the same tube for full recovery of peptides. Samples were dried in a vacuum centrifuge before proceeding with Tandem mass tag (TMT) labelling (section 2.9.4).

2.9.3 Sample preparation for whole cell proteomics

The methods described in section 2.9.3 were carried out by Dr James Williamson as part of an established collaboration with the lab of Prof Paul Lehner (Cambridge Institute for Therapeutic Immunology & Infectious Disease, University of Cambridge) and are included here for completeness.

2.9.3.1 Lysis and protein quantification

Cell pellets obtained in section 2.9.1 were resuspended in 50 µL resuspension buffer (76 mM HEPES pH 7.55, 6 mM $MgCl_2$, Benzonase (1400 U/mL) and 15 mM TCEP) by pipetting. 18.75 µL 20% lithium dodecyl sulfate was then immediately added to the cell suspension using a low retention pipette tip (RPT, StarLab) and pipetted to mix. Nucleic acids were fragmented by 30 s on/30 s off sonication in a Bioruptor Pico sonicator (Diagenode) for 10 minutes at 4°C. Samples were then incubated for 15 minutes at 37 °C to ensure complete reduction. Samples were alkylated by adding 6 µL of 187.5 mM methyl methanethiosulfonate (final concentration 15 mM) and incubating at room temperature for 15 minutes. 5 µL aliquots of each sample were diluted 2x in water and compared to a standard curve of BSA in the same buffer using a reducing agent-compatible BCA assay (Thermo Fisher). 25 µg of each sample was taken and the volumes of each lysate were equalised using resuspension buffer with 5% lithium dodecyl sulfate.

2.9.3.2 S-trap digestion

To each sample a 10% volume of 12% phosphoric acid was added to acidify samples to ~pH 2, completing denaturation. 6x volumes of wash buffer (100 mM HEPES pH 7.1, 90% methanol) was then

added and the resulting solution is loaded onto a S-trap (Protifi) using a positive pressure manifold (Tecan M10), adding not more than 150 µL of sample at a time (~80 PSI). In-house fabricated adaptors were used to permit the use of S-traps with the manifold. Samples were then washed 4x with 150 µL wash buffer. To remove any remaining wash buffer S-traps were centrifuged at 4000*g* for 2 minutes. To each S-trap, 30 uL of digestion solution (50 mM HEPES pH 8, 0.1% sodium deoxycholate) containing 1 µg Trypsin/lysC mix (Promega) was added. S-Traps were then loosely capped and placed in low adhesion 1.5 mL microfuge tubes in a ThermoMixer C (Eppendorf) with a heated lid and incubated at 37 °C for 6 hours. Where digestion was carried out overnight the Thermomixer was set to 4 °C after 6 hours. Peptides were recovered by adding 40 µL digestion buffer to each trap and incubating at room temperature for 15 minutes before slowly eluting with positive pressure (2-3 PSI). Traps were subsequently eluted with 40 µL 0.2% formic acid and 40 µL 0.2% formic acid, 50% acetonitrile in the same manner. Eluted samples were then dried in a vacuum centrifuge equipped with a cold trap prior to TMT labelling (section 2.9.4).

2.9.4 TMT labelling and clean-up

The methods described in section 2.9.4 to 2.9.7 were carried out by Dr James Williamson as part of an established collaboration with the lab of Prof Paul Lehner (Cambridge Institute for Therapeutic Immunology & Infectious Disease, University of Cambridge) and are included here for completeness.

Dried samples were resuspended in 21 µL 100 mM TEAB pH 8.5. After warming to room temperature, 0.5 µg TMTpro/0.2 µg TMT reagents (Thermo Fisher) were resuspended in 9 µL anhydrous acetonitrile which was added to the respective samples and incubated at room temperature for 1 h. A 3 µL aliquot of each sample was taken and pooled to check TMT labelling efficiency and equality of loading by liquid chromatography-mass spectrometry (LC-MS). Samples were stored at -80 °C in the interim. After checking each sample was at least 98% TMT labelled, total reporter ion intensities were used to normalise the pooling of the remaining samples such that the final pool should be as close to a 1:1 ratio of total peptide content between samples as possible. This final pool was then dried in a vacuum centrifuge. The sample was acidified to a final 0.1% trifluoracetic acid (~200 µL volume) and formic acid was added until the sodium deoxycholate visibly precipitated. 4 volumes of ethyl acetate were then added and the sample vortexed vigorously for 10 s. The sample was then centrifuged at 15 000*g* for 5 min at room temperature to effect phase separation. A gel loading pipette tip was used to withdraw the lower (aqueous) phase to a fresh low adhesion microfuge tube. If any obvious sodium deoxycholate contamination remained, the two-phase extraction with ethyl acetate was repeated. The sample was then partially dried in a vacuum centrifuge and brought up to a final volume of 1 mL

with 0.1% trifluoracetic acid. Formic acid was added until the pH was < 2, confirmed by spotting onto pH paper. The sample was then cleaned up by solid-phase extraction using a 50 mg tC18 SepPak cartridge (Waters) and a positive pressure manifold. The cartridge was wetted with 1 mL 100% Methanol followed by 1 mL acetonitrile, equilibrated with 1 mL 0.1% trifluoracetic acid and the sample was loaded slowly. The sample was passed twice over the cartridge. The cartridge was washed 3x with 1 mL 0.1% trifluoracetic acid before eluting sequentially with 250 µL 40% acetonitrile, 70% acetonitrile and 80% acetonitrile and dried in a vacuum centrifuge.

2.9.5 Basic pH reversed phase fractionation

Samples were resuspended in 40 µL 200 mM ammonium formate pH 10 and transferred to a glass HPLC vial. Basic pH reversed-phase fractionation was conducted on an Ultimate 3000 UHPLC system (Thermo Scientific) equipped with a 2.1 mm × 15 cm, 1.7 µm Kinetex EVO column (Phenomenex). Solvent A was 3% acetonitrile, Solvent B was 100% acetonitrile, solvent C was 200 mM ammonium formate (pH 10). Throughout the analysis solvent C was kept at a constant 10%. The flow rate was 500 µL/min and UV was monitored at 280 nm. Samples were loaded in 90% A for 10 min before a gradient elution of 0–10% B over 10 min, 10-34% B over 21 min, 34-50% B over 5 mins followed by a 10 min wash with 90% B. 15 s (100 µL) fractions were collected throughout the run. Fractions containing peptide (as determined by absorbance at 280 nm) were recombined across the gradient to preserve orthogonality with on-line low pH reversed-phase separation. For example, fractions 1, 25, 49, 73, 97 are combined and dried in a vacuum centrifuge and stored at -20 °C until LC-MS analysis. For whole cell proteome samples or plasma membrane enriched samples 24 / 12 fractions respectively are generated in this way.

2.9.6 Mass spectrometry

Samples were analysed on an Orbitrap Fusion instrument on-line with an Ultimate 3000 RSLC nano UHPLC system (Thermo Fisher). Samples were resuspended in 10 µL 5% DMSO/1% trifluoracetic acid. 5 µL of each fraction was injected. Trapping solvent was 0.1% trifluoracetic acid, analytical solvent A was 0.1% formic acid, solvent B was acetonitrile with 0.1% formic acid. Samples were loaded onto a trapping column (300 µm x 5 mm PepMap cartridge trap; Thermo Fisher) at 10 µL/min for 5 minutes. Samples were then separated on a 75 cm x 75 µm, 2 µm particle size PepMap C_{18} column (Thermo Fisher). The gradient was 3-10% B over 10 min, 10-35% B over 155 min, 35-45% B over 9 min followed by a wash at 95% B for 5 min and re-equilibration at 3% B. Eluted peptides were introduced by electrospray to the mass spectrometer by applying 2.1 kV to a stainless steel emitter (5 cm x 30 µm;

PepSep). During the gradient elution, mass spectra were acquired using Tune v3.3 and Xcalibur v4.3 (Thermo Fisher).

2.9.7 Data processing

Data were processed with PeaksX+, v10.5 (Bioinfor). Briefly, .raw files were searched iteratively in three rounds, with unmatched DeNovo spectra (at 0.1% PSM FDR) from the previous search used as the input for the next. The three iterations were as follows: 1) Swissprot Human (27/03/2020) + common contaminants; 2) the same databases as search 1 but permitting semi-specific cleavage; 3) trEMBL Human (27/03/2020), with specific cleavage rules. Identified proteins and their abundances were output to .csv, imported to R and submitted to statistical analysis using LIMMA, a moderated t-test available through the Bioconductor package[207-209]. LIMMA p-values were corrected for multiple hypothesis testing using the Benjamini-Hochberg method to generate a false discovery rate-adjusted p-value (q-value) for each comparison.

2.9.8 Mass spectrometry analysis for the GFP co-immunoprecipitation experiment

GFP co-immunoprecipitation samples were prepared as detailed in section 2.6.6. Mass spectrometry was performed by Dr Robin Antrobus and Dr Jack Houghton in the CIMR proteomics facility. Samples were resolved approximately 1 cm into a pre-cast 4-12% Bis-Tris polyacrylamide gel. The lanes were excised and cut in 4 equal chunks and the proteins reduced, alkylated and digested in-gel. The resulting tryptic peptides were analysed by LC-MS/MS using a Q Exactive coupled to an RSLCnano3000 (Thermo Scientific). Peptides were resolved on a 50 cm EASY-spray column (Thermo Scientific) with MS/MS data acquired in data-dependent acquisition mode.

Label-free quantitation values were obtained by processing raw files with MaxQuant (version 1.6.6.0) with the following parameters: specific trypsin digestion; Human database (UniProt reference proteome downloaded 03 Dec 2018 containing 21066 proteins); oxidation (M) and N terminal acetylation as variable modifications; carbamidomethylation (C) as a fixed modification; label-free quantification enabled; Match between runs enabled. Post-processing was carried out using R with bespoke scripts incorporating the “limma” statistical package.

2.9.9 Gene Ontology and pathway enrichment analysis

Gene Ontology and pathway enrichment was performed with the tool DAVID 6.8[210]. DAVID analysis was performed on lists of genes significantly up- and down-regulated (q-value < 0.05) against a background list of all genes detected in the experiment.

2.10 RNA sequencing

2.10.1 Cell culture

RNA sequencing (RNAseq) was performed on four technical replicates for each cell line (EV, KO1, KO2). For each cell line, cells were seeded at $0.25x10^6$ cells/mL in in 20 mL THP-1 media in four separate T75 flasks to harvest $1x10^7$ cells at a density of $0.5x10^6$ cells/mL the next day.

2.10.2 Cell lysis

Cells were pelleted (1000*g*, 5 min) and washed with 1 mL PBS. The cell pellet was thoroughly resuspended in 400 µL freshly prepared lysis buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 1% TX-100, 25 U/mL Turbo DNAse I (Invitrogen)) and incubated on ice for 20 minutes. Lysates were cleared by centrifugation (13 000*g*, 20 min, 4˚C) and the supernatant transferred to a fresh Eppendorf tube (low bind Eppendorf tubes were used throughout this experiment). Clarified lysates were aliquoted and stored at -70 ˚C.

2.10.3 RNA extraction

100 µL lysate was thawed and incubated with 200 µL freshly prepared proteinase K buffer (10 mM Tris-Cl pH 7.5, 1.5% SDS, 0.5 µg/mL Proteinase K) at 42 ˚C for 30 min. Samples and acidic phenol-chloroform were pre-warmed to 65 ˚C. 300 µL acidic phenol-chloroform was then added to each sample. Samples were vortexed thoroughly at high speed (10 s) and heated at 65 ˚C for 5 mins with occasional vortexing. Samples were then placed on ice for 5 minutes and the aqueous phase was separated by centrifugation (13 000*g*, 5 min, 4 ˚C). The upper aqueous phase was carefully collected into a fresh tube and 300 µL TE-buffered chloroform was added to each sample. Samples were thoroughly vortexed (30 s) and centrifuged (13 000*g*, 2 min, room temperature). The upper aqueous phase was carefully collected into a fresh tube.

2.10.4 Ethanol precipitation

RNA was mixed with 3 M sodium acetate pH 4.8 and 100% ethanol (1:0.1:3 v/v/v) and 30 µg GlycoBlue co-precipitant (Thermo Fisher Scientific) by vortexing thoroughly for 10 s. RNA was precipitated at -70 ˚C for 30 min. After centrifugation (13 000*g*, 15 min, 4 ˚C), the supernatant was removed using a fine gel loading tip. The RNA pellet was washed with 150 uL 95% ethanol, vortexed thoroughly, and centrifuged (13 000*g*, 2 min, room temperature). The supernatant was discarded, and the pellet was dried in a vacuum chamber for 5 minutes. RNA was resuspended in 20 µL water by vortexing.

2.10.5 Ribosomal RNA depletion

Ribosomal RNA was depleted from 5 µg of total RNA using a Ribo-Zero Gold rRNA Removal Kit (Illumina)[211] following the manufacturer's instructions. Depleted RNA was diluted in water for a volume of 180 µL, purified by ethanol precipitation and resuspended in 12 µL 10 mM Tris pH 7.5. Half of each depleted RNA sample was stored at -70 ˚C as a back-up and the second half carried forward for RNA fragmentation.

2.10.6 RNA fragmentation by alkaline hydrolysis

RNA was fragmented by alkaline hydrolysis in a 20 µL reaction containing depleted RNA, 5 mM Na_2CO_3, 45 mM $NaHCO_3$ and 1 mM EDTA and incubated at 95˚C for 15 min. The reaction was stopped by addition of 280 µL stop/precipitation solution (300 mM NaOAc pH 5.5, 50 µg/mL GlycoBlue) and fragmented RNA was recovered by ethanol precipitation and resuspended in 4 µL water.

2.10.6.1 RNA gel extraction

RNA was incubated with 6 µL denaturing gel loading buffer (Ambion) at 80 ˚C for 3 mins then cooled on ice prior to loading on a 15% (w/v) denaturing acrylamide gel (46% (w/v) urea, 14.25% acrylamide/0.75% bisacrylamide (Accugel), TBE) alongside a 10 bp DNA ladder and 28 and 34 nucleotide primers. The gel was run at 15 W for 40 min, then stained with SYBR Gold nucleic acid stain (Thermo Fisher Scientific) in TBE for 4 min. RNA fragments migrating between 28 and 34 nucleotides were excised from the gel using separate clean scalpels and eluted by addition of 600 µL RNA gel extraction buffer (300 mM NaOAc pH 5.5, 1 mM EDTA, 0.25% SDS) and overnight incubation at 4˚C on a rotating wheel. Eluted RNA was transferred to low bind Eppendorf tubes, washed by ethanol precipitation and resuspended in 15 µL 10 mM Tris pH 7.5.

2.10.7 Generation of RNA libraries

2.10.7.1 Removal of 3'-phosphates

RNA fragments were denatured at 80˚C for 2 min and cooled on ice. 3'-phosphates were removed in a 20 µL reaction with 20 U T4 Polynucleotide Kinase (NEB) and 20 U SUPERase-In RNase inhibitor (Thermo Fisher Scientific) in T4 Polynucleotide Kinase buffer without ATP. The reaction was incubated at 37˚C for 2 hours and heat-inactivated at 65˚C for 10 min. RNA was diluted in water for a volume of 100 µL, washed by ethanol precipitation and resuspended in 7 µL 10 mM Tris pH 7.5.

2.10.7.2 Ligation with 3'-adaptor

The RNA was ligated to a 3'-adaptor primer (5'-pNNNNNNNTGGAATTCTCGGGTGCCAAGG-3') pre-adenylated using a 5' DNA adenylation kit (NEB) following the manufacturer's instructions. Dephosphorylated RNA was mixed with 50 pmoles adenylated 3'-adaptor, denatured at 80˚C for 2 mins and cooled to room temperature. The ligation was performed with 400 U T4 RNA Ligase 2 truncated K227Q (NEB) in T4 RNA Ligase Reaction Buffer (without ATP) with 20 U SUPERase-In and 15% (w/v) PEG 8000 in a 20 µL reaction incubated at 14˚C overnight. RNA was washed by ethanol precipitation and resuspended in 4 µL 10 mM Tris pH 7.5. Ligated RNA fragments (56-62 nucleotides) were purified from a 15% denaturing gel as described above and dissolved in 10 µL 10 mM Tris pH 7.5.

2.10.7.3 3' adaptor-ligated product phosphorylation

3' adaptor-ligated products were phosphorylated with 20 U T4 Polynucleotide Kinase (NEB) in T4 Polynucleotide Kinase buffer with 1 mM ATP and 20 U SUPERase-In in a 20 µL reaction incubated at 37 ˚C for 2 hours and heat-inactivated at 65˚C for 10 min. Phosphorylated RNA was washed by ethanol precipitation and resuspended in 10 µL 10 mM Tris pH 7.5.

2.10.7.4 Ligation with 5'-RNA adaptor

RNA was ligated to the 5'-RNA adaptor (5'-GUUCAGAGUUCUACAGUCCGACGAUCNNNNNNN-3'; 100 pmoles) with 20 U T4 RNA ligase in T4 RNA ligase buffer with 20 U SUPERase-In in a 20 µL incubated at 14˚C overnight. The fully adapted RNA was washed by ethanol precipitation and resuspended in 6 µL 10 mM Tris pH 7.5.

2.10.7.5 Reverse transcription

3 µL fully adapted RNA was annealed to 50 pmoles reverse transcription primer (5'-GCCTTGGCACCCGAGAATTCCA-3') in a 4.5 µL reaction with 12.5 nmoles dNTP incubated at 65˚C for 5 min. RNA was subsequently reverse transcribed with 200 U Superscript III (Invitrogen) in first strand buffer (Invitrogen) with 2.5 mM $MgCl_2$, 10 mM DTT, 20 U SUPERase-In in a 20 µL reaction incubated at 55 °C for 50 min, followed by heat inactivation at 85 °C for 5 min.

2.10.8 PCR amplification and barcode addition

Standard PCR reactions were used to prepare amplicons using forward primer RP1 (5'-AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA-3') and reverse primer 5'-CAAGCAGAAGACGGCATACGAGATN_6GTGACTGGAGTTCCTTGGCACCGAGAATTCCA-3' (RPIX), where X is

primer number (1-12) and N_6 the reverse complement of the respective hexanucleotide index sequence detected during Illumina sequencing. (TruSeq Small RNA Library Prep Kit, indexes 1-12). Phusion High-Fidelity DNA polymerase mastermix (NEB) was used for amplification of 1 µL cDNA with 4 pmoles of each primer in a 20 µL reaction with the following cycling parameters: initial denaturation at 98˚C for 30 s, followed by 16 cycles of denaturation (98˚C, 10 sec), annealing (60˚C, 30 sec) and extension (72˚C, 15 sec), with a final extension step (72˚C, 10 min). The PCR product was purified with a QIAquick PCR purification kit (Quiagen) following the manufacturer's instructions and eluted with 30 µL water. The PCR was mixed with 10 µL loading dye (bromophenol blue in 30% sterile glycerol) and loaded onto a 10% w/v non-denaturing polyacrylamide gel (9.5% acrylamide/0.5% bisacrylamide (Accugel) in TBE) alongside a 10 bp ladder. The gel was run at 200 V for 35 min. Products with an approximate length of 163 bp were excised from the gel and eluted by addition of 600 µL DNA gel extraction buffer (300 mM NaCl, 10 mM Tris–HCl pH 8 and 1 mM EDTA) and incubation at 4 °C overnight on an end-over-end rotator. The amplicon libraries were purified by ethanol precipitation and resuspended in 15 µL 10 mM Tris pH 8. DNA concentration in each sample was determined using a Qubit fluorometer (Thermo Fisher Scientific). A pooled 2.11 nM library (20 µL, 0.41 ng per sample) was sequenced on a NextSeq 500 system (Illumina) at Cambridge Genomic Services (Department of Pathology, University of Cambridge).

2.10.9 RNAseq data analysis

Data analysis was performed by Dr Katy Brown (Department of Pathology, University of Cambridge). Log_2 fold changes were analysed with DESeq2 and corrected with lfcShrink[212]. False discovery rate adjusted q-values were calculated using fdrtool in R[213] for every combination of conditions.

2.11 Quantitative PCR

2.11.1 RNA extraction and reverse transcription

RNA was extracted from $5x10^5$ cells using a PureLink RNA micro kit (Invitrogen) following the manufacturer's instructions and stored at -80 °C. Reverse transcription was performed from 2 µg RNA using a High yield RNA-to-cDNA kit (Applied Biosystems, 4387406) following the manufacturer's instructions. A "no reverse transcriptase" control was produced by omitting the addition of reverse transcriptase enzyme and buffer to one sample from each batch to monitor genomic DNA contamination. cDNA was diluted to 10 ng/µL prior to use in quantitative PCR (qPCR) experiments.

2.11.2 qPCR

qPCR experiments were performed for three technical replicates per sample using the TaqMan FAM dye-labeled MGB probes (Thermo Fisher Scientific) listed in Table 2.4 and TaqMan Fast Advanced Master Mix following the manufacturer's instructions. 10-100 ng cDNA was added to each 20 µL reaction. Lack of genomic DNA contamination was confirmed by performing qPCR on the "no reverse transcriptase" control and ensuring that no amplification was detected or that the control Ct value was over 10 cycles higher than sample Ct values. Reactions were set up in white 96 well-plates (Bio-Rad) and executed on a CFX96 real-time thermal cycler (Bio-Rad) with the following cycle parameters: polymerase activation at 95 °C for 10 minutes, then 40 cycles of denaturation (95 °C, 15s) and annealing/extension (60 °C, 1 min).

Table 2.4. TaqMan gene expression assays used in qPCR experiments

Target gene	TaqMan assay ID
ACTB	Hs01060665_g1
CASC3	Hs00201226_m1
HMBS	Hs00609297_m1
GALC	Hs00164660_m1
UGT8	Hs00409958_m1
SLC7A11	Hs00204928_m1
FAS	Hs00236330_m1
MSRB3	Hs00827017_m1

Target gene	TaqMan assay ID
AXL	Hs01064444_m1
LILRB4	Hs00359756_g1
TNF	Hs00174128_m1
SLC4A4	Hs00186798_m1
MDFIC	Hs04177766_g1
PLXNC1	Hs00194968_m1
TRHDE	Hs01072025_m1
SDK1	Hs01010140_m1

2.11.3 Housekeeping gene validation

The stable expression of several housekeeping control genes (*ACTB*, *CASC3* and *HMBS*) was assessed by quantifying the expression of these genes in WT, EV1, KO1 and KO2 THP-1 cells and calculating the average pairwise variation of each control gene with the remaining control genes using the GeNorm algorithm[209] in qBase+ software (Biogazelle, Zwijnaarde, Belgium - www.qbaseplus.com).

2.11.4 qPCR data analysis

qPCR experiments were performed independently with at least three biological replicates. The expression of each gene relative to control was determined using the ΔΔCt method. Ct values for each biological replicate were calculated as the mean Ct of three technical replicates. *ACTB* was used as the housekeeping control gene in all qPCR experiments. ΔCt was calculated as:

$$\Delta Ct = Ct_{\text{housekeeping gene}} - Ct_{\text{gene of interest}}$$

Log_2 fold change in gene expression relative to control (ΔΔCt) was calculated by subtracting the mean $\Delta Ct_{control}$ value of the control sample across all biological replicates from the sample ΔCt_{sample} value for each biological replicate:

$$\Delta\Delta Ct_{sample} = \Delta Ct_{sample} - mean(\Delta Ct_{control})$$

The statistical significance of gene expression changes was determined by ordinary one-way ANOVA on ΔΔCt values, followed by Tukey's multiple comparison test in GraphPad Prism 8. The assumptions of independence of observations, normality of distributions and equality of variance between groups required for these statistical tests were met.

2.12 Microscopy

2.12.1 Brightfield microscopy on undifferentiated THP-1 cells

$5x10^4$ undifferentiated THP-1 cells were pelleted (125*g*, 5 min) and washed twice with PBS. Cells were incubated in 5 µg/mL Hoechst 33342 (H1399, Thermo Fisher Scientific) in PBS away from light at room temperature for 5 minutes. Cells were then imaged on an EVOS FL microscope (Thermo Fisher Scientific) with 40x magnification.

2.12.2 Immunofluorescence microscopy

2.12.2.1 Cell seeding

13 mm round glass coverslips were sterilised in 70% ethanol then washed in PBS in 24-well tissue culture plates. Adherent cells were seeded onto the glass coverslips as required.

2.12.2.2 Cell fixing

The culture medium was removed from the cells and cells were washed once in ice-cold PBS. Cells were fixed with 250 µL ice-cold, freshly prepared 4% formaldehyde in 250 mM HEPES, pH 7.4 for 10 minutes on ice. A second fixation step was performed with 250 µL 8% formaldehyde in 250 mM HEPES, pH 7.5 for 20 min at room temperature. Cells were washed twice with 500 µL PBS. Residual formaldehyde was quenched with 250 µL 50 mM NH_4Cl for 5 min. Cells were washed twice with 500 µL PBS and stored in PBS at 4 °C until the next step (for long term storage, 0.02% azide was added).

2.12.2.3 Cell permeabilisation and staining

Cells were permeabilised with 250 µL 0.1% saponin in PBS on a rocking platform (15-30 min, room temperature). Cells were then incubated with 250 µL blocking buffer (0.1% saponin, 5% FBS in PBS) with rocking (15-30 min, room temperature). Primary antibodies (Table 2.5) were diluted in blocking

buffer. Using tweezers and a bent needle, coverslips were placed cell-side down onto 30 µL drops of antibody solution on parafilm and incubated for 2 hours in a humidified chamber. Coverslips were then placed in a 24 well-plate and washed 5 times with 250 µL blocking buffer on a rocking platform (5 min per wash, room temperature). Coverslips were then placed onto 30 µL drops of secondary antibodies (Table 2.5) in blocking buffer and incubated for 1 hour in a humidified chamber. Cells were washed 5 times with 250 µL blocking buffer (5 min, room temperature), then briefly rinsed 3 times with 250 µL 0.1% saponin in PBS, followed by 3 rinses with 250 µL PBS. Coverslips were briefly rinsed in MQW, excess water was blotted off and coverslips were mounted cell-side down onto a 15 µL drop of mounting medium (10% w/v Mowiol 4-88, 0.66 µg/mL DAPI, 25% v/v glycerol, 0.1 M Tris pH 8.5) on a microscopy slide. Mounted samples were dried at room temperature overnight and visualised on an Axioimager Z2 widefield microscope (Carl Zeiss).

Table 2.5. Antibodies used for immunofluorescence experiments

Target	Source	Dilution	Provider and ID
Primary Antibodies			
LAMP-1	Rabbit polyclonal	1 in 1000	Abcam ab24170
EEA-1	Mouse monoclonal	1 in 500	BD Biosciences 610456
Secondary Antibodies			
Rabbit IgG AlexaFluor555	Donkey polyclonal	1:200	Life Technologies
Mouse IgG AlexaFluor488	Donkey polyclonal	1:200	Life Technologies

2.12.3 Retention Using Selective Hooks (RUSH) assay

HeLa cells stably expressing the Strep-KDEL sequence and the target protein expression vector pDCGx15 encoding a signal peptide, Streptavidin-binding peptide (SBP) at the N-terminus and a GFP-SPOT double tag at the C-terminus were generously provided by Dr David Gershlick (Cambridge Institute for Medical Research). Full-length PTPRM (residues 21-1451), the Fn3-4 domain (residues 478-768), and the extracellular domain (residues 21-768) of PTPRM were cloned into pDCGx15 by Eve Caroe.

Cells were seeded on sterilised coverslips in 24 well plates at 5 000 cells per well (Strep-KDEL HeLa cell line) or 10 000 cells per well (MO3.13 cell line) in DMEM with 10% FBS. Individual plates were seeded for each time point to prevent changes in temperature while processing samples for separate time points.

Cells were transfected on the next day with 250 ng pDCGx15 vector expressing PTPRM truncations pre-mixed in 25 µL OptiMEM with 0.75 µL FuGene 6 (as detailed in section 2.6.6.1). 4 hours after transfection, the transfection reagent was removed and 500 µL fresh media added to the cells.

A 585 mM biotin (B4501, Sigma) stock solution in DMSO was prepared with mixing at 37 °C, then aliquoted and stored at -20 °C. The day after transfection, a 2X biotin working solution (2.4 mM) was prepared by diluting the stock solution 1:250 in culture media and mixing thoroughly to dissolve any precipitate, then pre-warmed to 37 °C. At t=0, 500 µL of 2X biotin working solution was added onto the cells in 500 µL culture media for a final biotin concentration of 1.2 mM. At each time point, cells were placed on ice, fixed and mounted onto microscopy slides with Mowiol with DAPI following the protocol described in section 2.12.2. Cells were then imaged on an Axioimager Z2 widefield microscope (Carl Zeiss).

2.13 Cell-based assays

2.13.1 Cytokine array

$0.45x10^6$ cells were pelleted by centrifugation (200*g*, 5 min), washed in PBS, resuspended in 1.5 mL serum-free medium (RPMI-1640 medium with L-glutamine and sodium bicarbonate) and transferred to a 12-well plate. Conditioned media was harvested after 24 hours, cells and debris were pelleted by centrifugation (1000*g*, 5 min) and the supernatant was incubated with the antibody-based Human Cytokine Array C5 (RayBiotech) following the manufacturer's instructions. The array was imaged using a G:BOX imaging system (Syngene). Spot intensity calculation and background subtraction were performed in Image Studio Lite (LI-COR). Intensity fold change for each cytokine was calculated from raw intensity data.

2.13.2 Reactive oxygen species assay

Reactive oxygen species (ROS) were detected using the membrane permeable fluorogenic dye 2',7'-dichlorodihydrofluorescein diacetate (H_2-DCF-DA; D399, Invitrogen). The dye was stored under dry N_2 and a 5000x working stock (5 mM) was made in anhydrous DMSO before use. EV1, KO1 and KO2 cells were pelleted (125*g*, 5 min) and washed in PBS twice. Cells were resuspended in Dulbecco's phosphate-buffered saline (DPBS) with 1 µM H_2-DCF-DA at a cell density of $0.5x10^6$ cells/mL. 100 µL were transferred to a black lidded 96-well plate (655086, Greiner) in triplicate and the plate was incubated at 37 °C, 5% CO2 for 15 minutes. Fluorescence was measured using a Spark CM (Tecan) or Clariostar Plus (BMG) plate reader (Excitation: 492–495 nm, Emission: 517–527 nm).

2.13.3 Apoptosis assays

2.13.3.1 TUNEL assay

The APO-BrdU TUNEL Assay Kit (ThermoFisher) was used to detect apoptosis. $0.6x10^6$ WT and GALC KO oligodendrocytes were seeded in T75 flasks and differentiated on the next day. The differentiation media was changed after 4 days. The media of the staurosporine-treated control cells was replaced by differentiation media containing 2 µM staurosporine 24 hours before sample processing. 7 days after differentiation, floating cells were pooled with adherent cells detached by trypsinisation. $1x10^6$ cells were then labelled and stained with Alexa Fluor 488 following the kit protocol. Samples were analysed on the LSRFortessa cell analyser (BD Biosciences). The BrdU-conjugated fluorophore AlexaFluor 488 was excited with a 488 nm laser? with the 515/20 emission filter. $5x10^3$, $5x10^3$ and $8x10^3$ events were recorded for WT, staurosporine-treated and GALC KO samples, respectively, with a detector voltage of 291 V. $1x10^3$ events were analysed for the unstained WT sample, with a higher detector voltage of 421 V. Data was analysed with FlowJo.

2.13.3.2 DNA laddering assay

DNA fragmentation was assessed following a published protocol[214]. $140x10^3$ WT and GALC KO MO3.13 were seeded into 6 cm dishes. On the next day, cells were differentiated. One dish was kept as an undifferentiated sample. Dishes with WT and GALC oligodendrocytes were then processed 1, 2, 3, 6 and 7 days after differentiation as described below. The media of the dishes processed on day 6 and 7 was changed with fresh differentiation media 3 days after differentiation was initiated. One WT sample was treated with 2 µM staurosporine for 24 hours before harvesting at day 7 post-differentiation. Adherent cells were detached with trypsin and pooled with detached cells. Cells were washed once with PBS and spun down (200*g*, 5 min). The pellet was resuspended in 50 µL hypotonic buffer and kept on ice for 10 min. The lysate was centrifuged (4500*g*, 4 °C, 5 min). In a fresh tube, the supernatant was supplemented with SDS (1% final concentration). RNase A was added (2.5 mg/mL) and the sample was incubated for 2 hours at 55 °C. Proteinase K was added (2.5 mg/mL) and the sample was incubated for 2 hours at 37 °C. 100 µL 10 M ammonium acetate and 1 mL chilled absolute ethanol were added to the sample, and the sample was stored at -20 °C for 1-8 days. All samples were then centrifuged (13000*g*, 10 min). The pellet was washed with 500 µL 70% ethanol and spun down (13000*g*, 10 min). The pellet was air-dried and resuspended in 20 µL TE buffer. Samples were mixed with 10 µL of 6X DNA loading dye (NEB) and analysed by agarose gel electrophoresis (1.5% agarose gel, 35 V, 3.5 h).

3 CD1d-Saposin B interaction studies

This chapter has been expanded and reworked from a published article "The lipid transfer protein Saposin B does not directly bind CD1d for lipid antigen loading"[215] (Appendix I).

Chapter summary

As described in section 1.7.4, the lysosomal lipid transfer protein SapB has been shown to facilitate lipid exchange on the lipid antigen presenting molecule CD1d[165,166,169]. However, the mechanism of lipid transfer has not yet been elucidated. SapB is thought to act as a lipid solubiliser, extracting the lipid from a lipid bilayer, encapsulating it in its hydrophobic core and presenting it to lipid-interacting proteins in solution. Due to the hydrophobic property of lipids, their exposure to aqueous solutions is thermodynamically unfavourable. Therefore, lipid transfer between SapB and the hydrophobic lipid-binding groove of CD1d must require direct interaction between these proteins, as well as structural rearrangements for continuous shielding of the lipid from the aqueous phase. To gain an insight into the mechanism of CD1d lipid loading by SapB, I attempted structural studies of CD1d in complex with SapB. CD1d did not co-crystallise with SapB, calling into question whether SapB directly interacts with CD1d. This was addressed using biochemical assays to probe the interaction between CD1d and SapB.

Statement of acknowledgement

Dr Stephen Graham (Department of Virology, University of Cambridge) significantly aided crystal manipulation and diffraction data collection described in this chapter. Tomasz Benedyk (Graham lab, Department of Virology, University of Cambridge) helped with the design of the equilibrium assay experiment described in section 3.4.1. Dr Robin Antrobus (CIMR Proteomics facility) performed the mass spectrometry experiment described in section 3.5, and Dr Jack Houghton (CIMR Proteomics facility) helped with data analysis.

3.1 Purification of CD1d-β_2m complex from a human cell line

Several strategies have been developed previously for the purification of CD1d including refolding from *E. coli* inclusion bodies[216], baculovirus-insect cell expression[184] and mammalian expression systems[217]. An important consideration for CD1d protein expression and purification is the inclusion

of the essential partner protein β_2m. In a manner similar to the heavy chain of MHC class I, CD1d non-covalently associates with β_2m and this association is required for maturation of its four glycan chains[218]. In order to obtain high quality, near-native CD1d protein for *in vitro* studies, I co-expressed human CD1d with β_2m in the mammalian suspension cell line HEK293F as this would confer correct folding, native post-translational modification and full glycosylation of CD1d. Initial trials using polyethylenimine (PEI)-based transient co-transfection of CD1d and β_2m resulted in very low yields (0.1 mg protein per 100 mL media). Thus I used a piggyBac expression system[173] (detailed in section 2.4.1) to establish stable and inducible co-expression of full-length, untagged β_2m with the ectodomain of CD1d possessing a C-terminal hexahistidine (H_6) tag (Fig. 3.1A). Although the piggyBac expression system had been previously used in our lab, this was the first attempt at co-expression of two proteins using this system. Both proteins were expressed with an N-terminal secretion signal sequence targeting synthesis to the endoplasmic reticulum, resulting in protein secretion into the media. This expression strategy results in CD1d molecules that are bound to endogenous human lipids acquired during folding in the endoplasmic reticulum and trafficking through the Golgi apparatus[157]. CD1d-H_6 was purified from conditioned media by nickel affinity chromatography (Fig. 3.1B) followed by size exclusion chromatography (Fig. 3.1C-D). Untagged β_2m co-purified with CD1d, indicating that both proteins are correctly folded and form a stable complex. Although some CD1d alone was purified, I was able to separate this from CD1d-β_2m following size exclusion chromatography (Fig. 3.1C-D). Intriguingly, CD1d alone elutes earlier from the size exclusion column than the CD1d-β_2m complex, which is unexpected given that CD1d alone is expected to be smaller and therefore elute later. There are two possible explanations for this elution profile. The glycans on β_2m-associated CD1d may non-specifically interact with the column, retarding its elution. However, I would expect tailing of the elution peak in this case. Alternatively, CD1d may adopt a conformation with a larger hydrodynamic radius in the absence of β_2m, accelerating its elution.

Using this expression system, I obtained a high yield (1 mg protein per 100 mL media) of >95% pure CD1d-β_2m protein. The yield was 10 times higher than that obtained with transient co-transfection in the same cell line. In the rest of this chapter, purified CD1d-H_6-β_2m complex is referred to as CD1d.

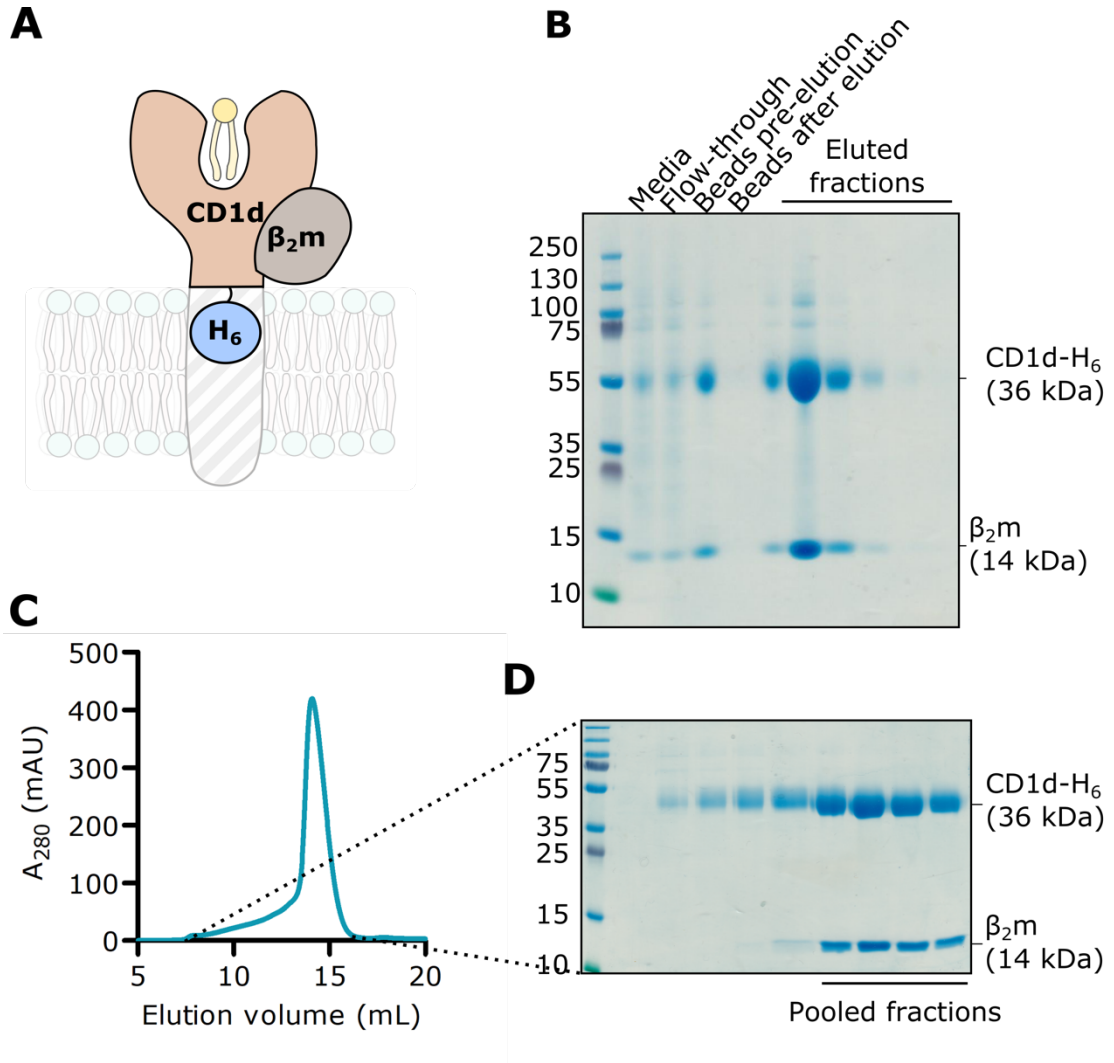

Figure 3.1. Human CD1d-β_2m expression and purification from mammalian cells
A. A hexahistidine (H_6)-tagged construct encoding the extracellular domain of CD1d (CD1d-H_6) was co-expressed with untagged full-length β_2m in mammalian cells. Both proteins were secreted into the media. B. CD1d-H_6 was captured from the media with Ni-NTA agarose and eluted with 300 mM imidazole in PBS. Fractions were analysed on a Coomassie-stained 4-12% Bis-tris SDS-PAGE gel, showing co-purification of untagged β_2m with CD1d-H_6. C. Eluted fractions were pooled, concentrated, and CD1d-H_6-β_2m was further purified by size exclusion chromatography (SEC) using a Superdex 200 10/300 column. D. SEC elution fractions were analysed on a Coomassie-stained 4-12% Bis-tris SDS-PAGE gel. The indicated fractions were pooled, concentrated, and used in downstream experiments.

3.2 Purification of SapB

The four saposin proteins (SapA-D) are produced from the proteolytic cleavage of the precursor protein PSAP which normally occurs upon delivery to the lysosome. To produce recombinant SapB alone, the segment of human PSAP corresponding to SapB (Fig. 3.2A) was expressed in *E. coli*. Untagged SapB was purified following a well-established protocol[219] that exploits the remarkable heat stability of saposins conferred by the presence of three disulphide bonds stabilising each saposin

monomer. A large fraction of contaminants was precipitated and removed by boiling of *E. coli* lysates. SapB was further purified by anion exchange chromatography (Fig. 3.2B-C) followed by size exclusion chromatography (Fig. 3.2D-E), which allows the separation of SapB from contaminating *E. coli* proteins and DNA. This protocol yields 50 mg pure SapB per litre of culture. Previous structural studies with *E. coli*-expressed SapB identified that it exists as a tight dimer that co-purifies with *E. coli* phosphatidylethanolamine[82].

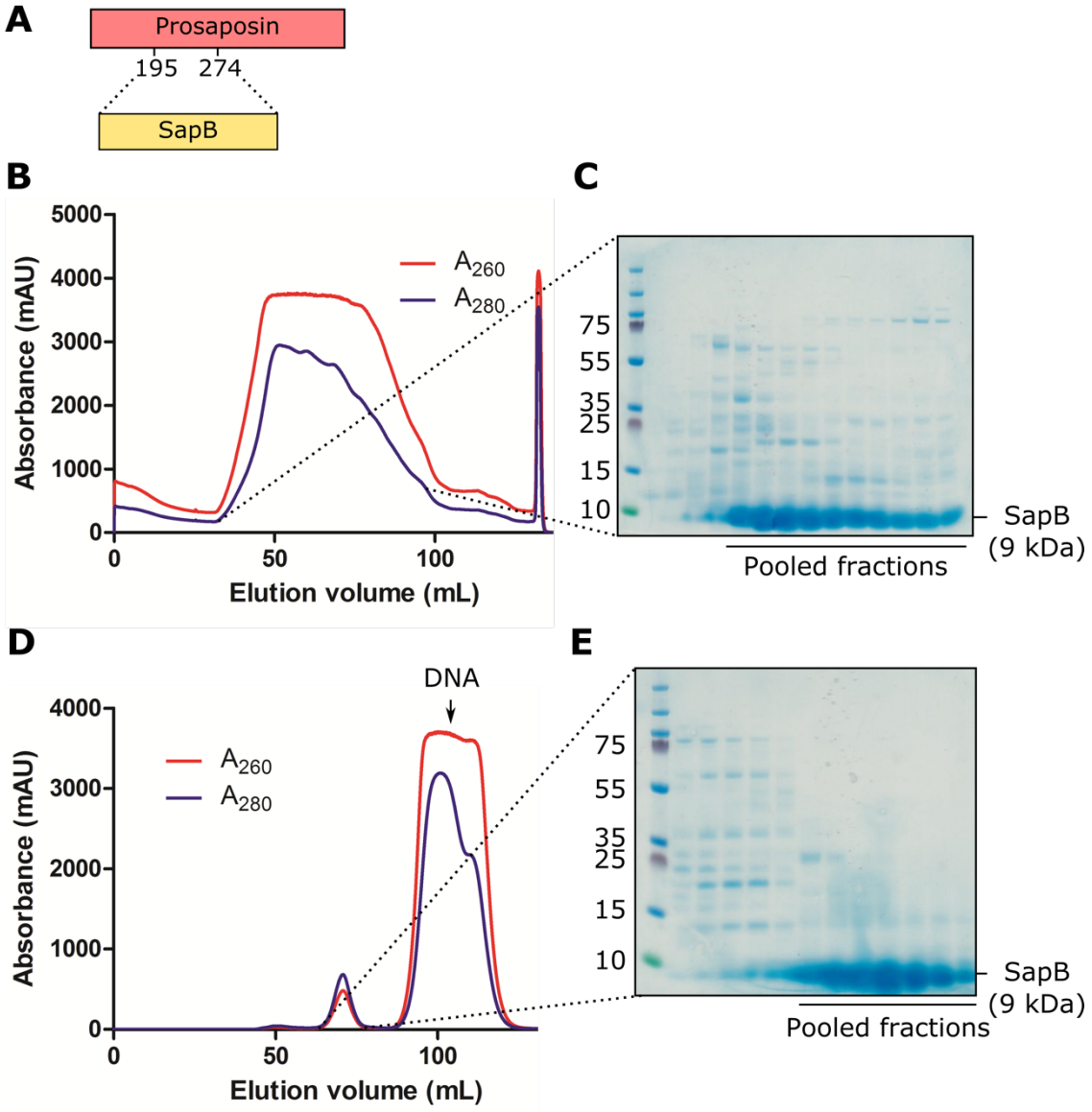

Figure 3.2. Human SapB expression and purification from *E. coli*
A. Untagged human SapB (residues 195-274 of the precursor protein PSAP) was recombinantly expressed in *E. coli*. B. SapB was purified by anion exchange chromatography using a 5 mL HiTrap QSepharose column. C. Eluted fractions were analysed on a Coomassie-stained 4-12% Bis-Tris SDS-PAGE gel. The indicated fractions were pooled and concentrated for further purification. D. SapB was further purified by size exclusion chromatography using a Superdex 75 16/600 column. Large amounts of contaminating DNA characterised by high absorbance at 260 nm elute around 90 mL. The first elution peak contains purified SapB. E. Eluted fractions corresponding to the first peak were analysed on a Coomassie-stained 4-12% Bis-Tris SDS-PAGE gel. The indicated fractions were pooled, concentrated and used in downstream experiments.

3.3 Co-crystallisation trials of CD1d and SapB

Previously published work strongly supported that SapB directly loads lipids onto CD1d, a process that would require a direct interaction between these two proteins in order to facilitate lipid exchange. Therefore, I attempted to co-crystallise CD1d with SapB to get structural insight into their interaction and the mechanism of CD1d lipid loading by SapB.

3.3.1 Initial co-crystallisation experiments

Using highly pure protein described in sections 3.1 and 3.2, I set up sitting-drop vapour diffusion crystallisation trials with CD1d and SapB in a 1:2 molar ratio, taking into account the fact that SapB functions as a dimer. Both proteins were purified in the same buffer (150 mM NaCl, 50 mM Tris pH 7.4). Crystallisation trials were set up with CD1d at 12.7 mg/mL and SapB at 4.5 mg/mL (Fig. 3.3A). Low-resolution diffraction quality crystals grew against a reservoir containing 20% v/v PEG monomethylether 550, 10% w/v PEG 20 000, 20 mM 1,6-hexanediol, 20 mM 1-butanol, 20 mM 1,2-propanediol, 20 mM 2-propanol, 20 mM 1,4-butanediol, 20 mM 1,3-propanediol, 0.1 M bicine/Trizma base pH 8.5 (Fig. 3.3B). Crystallisation conditions were optimized by screening around the pH, alcohol and precipitant concentration. To test whether these crystals contained both CD1d and SapB, I harvested several crystals, washed them in reservoir solution to remove residual uncrystallised components, then separated proteins contained in the crystal by SDS-PAGE followed by detection using Coomassie staining (Fig. 3.3C). CD1d and β_2m are visible on the gel, but SapB is not. These crystals composed of CD1d-β_2m alone were therefore not pursued further.

Crystals also grew as bunches of needle crystals against several related reservoir solutions containing 20% w/v PEG 6000 and divalent cations at pH 6 and 7 (Fig. 3.3D). These crystals did not diffract. Therefore, I washed several of these crystals and analysed them by SDS-PAGE, showing that these crystals were composed of SapB only (Fig. 3.3E).

I set up additional trials with a 1:3 molar ratio of CD1d to SapB, with SapB in excess in order to drive complex formation. Some non-crystalline protein aggregates were formed, but these could not be harvested for SDS-PAGE analysis, even after optimisation of crystallisation conditions.

The formation of crystals composed of CD1d-β_2m alone or SapB alone shows that the proteins used in this study are pure and crystallisable. However, no crystal containing both CD1d and SapB was grown. This suggests that the conditions in which these protein constructs crystallize are not conducive to

their interaction, or that a key component necessary for their interaction may be missing. In the next experiments, I have attempted to address these issues.

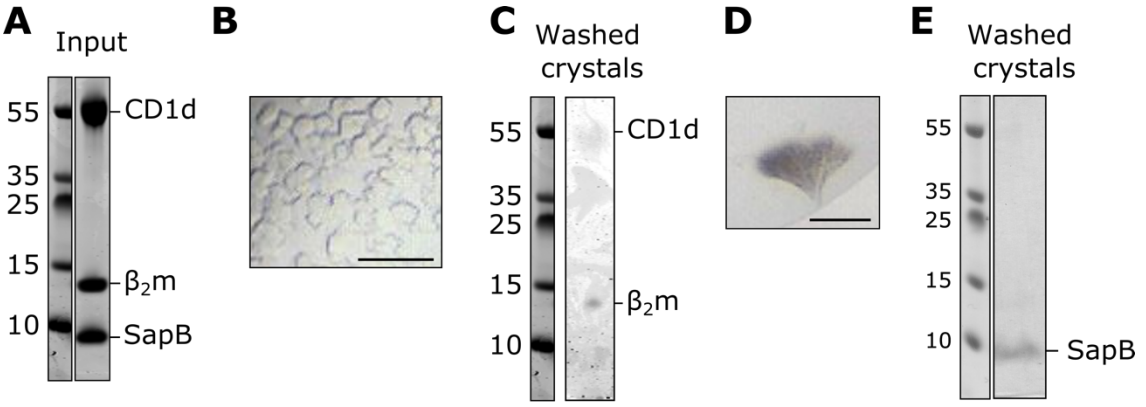

Figure 3.3. Crystallisation trials of CD1d-β_2m and SapB.
A. Coomassie-stained SDS-PAGE of the purified protein sample containing CD1d-β_2m and a 2-fold molar excess of SapB used in crystallisation trials. B. Crystals grown by sitting drop vapour diffusion against a reservoir containing 20% v/v PEG MME 550, 10% w/v PEG 20 000, 20 mM 1,6-hexanediol, 20 mM 1-butanol, 20 mM 1,2-propanediol, 20 mM 2-propanol, 20 mM 1,4-butanediol, 20 mM 1,3-propanediol, 0.1 M bicine/Trizma base pH 8.5. A condenser lens was used to image these thin crystals. C. Coomassie-stained SDS-PAGE of harvested and washed crystals shown in panel B. D. Crystals grown by sitting drop vapour diffusion against a reservoir of 20% w/v PEG 6000, 0.2 M calcium chloride and 0.1 M HEPES pH 7. E. Coomassie-stained SDS-PAGE of harvested and washed crystals shown in panel D. The scale bar represents 100 µm.

3.3.2 CD1d deglycosylation

CD1d possesses four *N*-glycosylation sites. Glycans have a high chemical and conformational heterogeneity and can therefore limit protein crystallisation. To facilitate CD1d crystallisation in different crystallisation conditions and improve crystal quality, I produced deglycosylated CD1d protein for crystallisation trials. Deglycosylated protein can be produced from mammalian cells by inhibiting α-mannosidase I with kifunensine[220] during protein expression. This permits normal protein folding and initial glycosylation, but glycan processing is limited so that glycans are sensitive to cleavage by Endoglycosidase H (Endo H) after protein purification[221]. Endo H treatment leaves a single *N*-Acetylglucosamine residue at each glycosylation site.

The expression of CD1d-β_2m from the cell line produced in section 3.1 is inducible. From the time of induction of protein expression, cells were treated with kifunensine throughout protein expression. CD1d was purified from the media of kifunensine-treated cells by nickel affinity purification and eluted with 100 mM citrate pH 4 (Fig. 3.4A). A low pH elution buffer was chosen because Endo H has a low pH optimum (pH 5.5). CD1d is likely to remain correctly folded and stable at low pH since it normally

localizes to late endosomal compartments. Indeed, using this purification protocol, the yield remained similar to that obtained with imidazole elution (section 3.1).

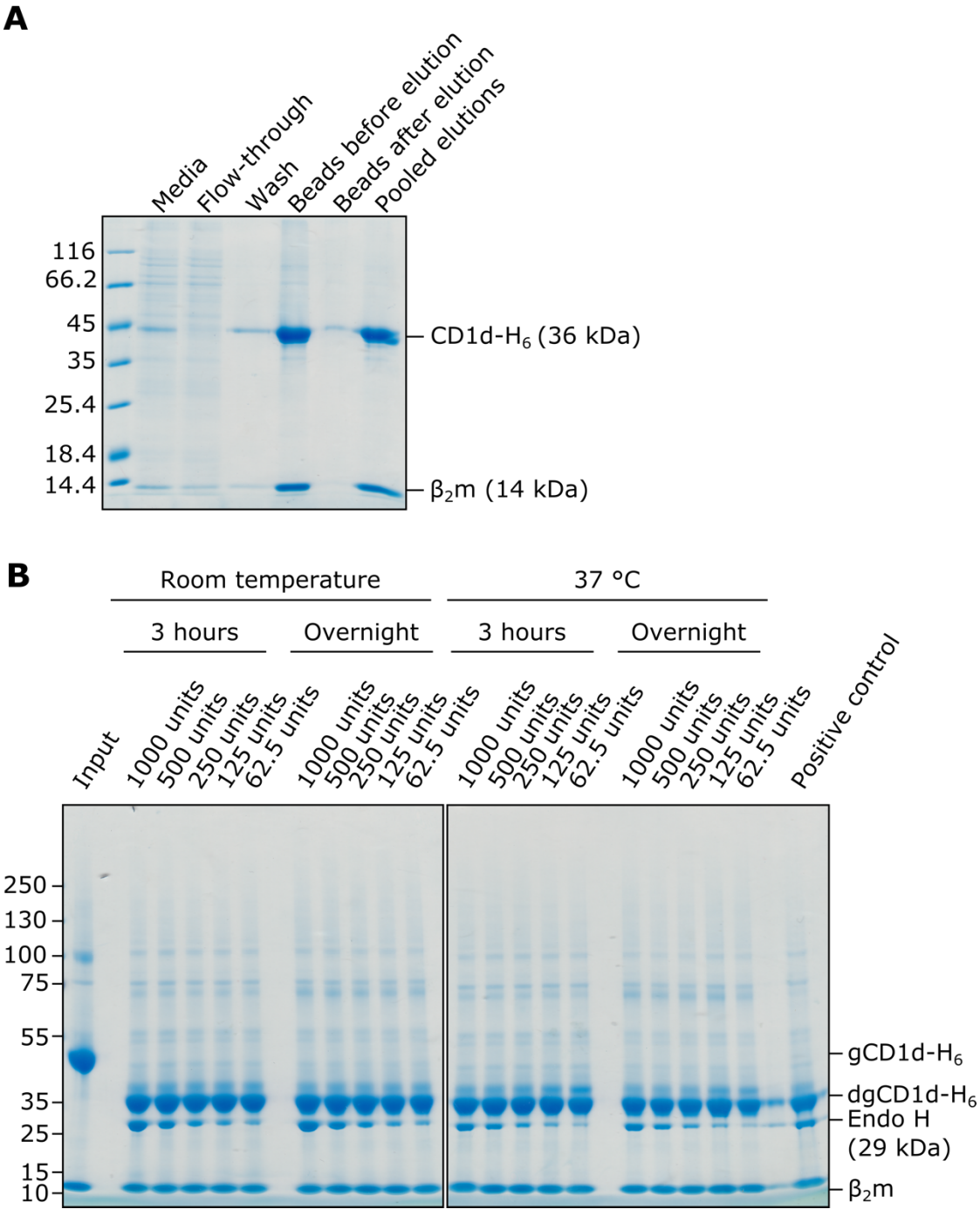

Figure 3.4. Optimisation of CD1d-β_2m deglycosylation
CD1d-β_2m was expressed in HEK293F cells in the presence of kifunensine to produce protein with high-mannose glycans susceptible to Endo H cleavage. A. Nickel affinity purification of CD1d-β_2m from the media of kifunensine-treated cells. B. Small-scale optimisation of conditions for Endo H treatment of CD1d-β_2m, analysed by Coomassie-stained 4-12% SDS-PAGE Bis-Tris gel. As a positive control, CD1d-β_2m was denatured prior to treatment with Endo H. Abbreviations: g: glycosylated; CD1dg: deglycosylated.

I first optimised the experimental conditions for Endo H treatment of CD1d in a small-scale experiment by varying the following factors: amount of Endo H added (a range of 1000 to 62.5 units was tested by serial dilution), time (3 hours or overnight) and temperature (room temperature or 37 °C). As a positive control, CD1d was denatured prior to treatment with Endo H. All conditions resulted in complete deglycosylation of CD1d as assessed by comparison with the positive control (Fig. 3.4B). Consequently, for large scale CD1d deglycosylation, CD1d was incubated with the smallest concentration of Endo H tested (3125 units per mg of CD1d) for 3 hours at room temperature. I observed the formation of precipitate in the sample during incubation with Endo H. After Endo H treatment, I pelleted the precipitate by centrifugation and analysed it by SDS-PAGE (Fig. 3.5A). This revealed that the precipitate is composed of deglycosylated CD1d without β_2m. This instability of deglycosylated CD1d without β_2m is advantageous since its loss by precipitation increases the homogeneity of the sample. Deglycosylated CD1d-β_2m was further purified by size exclusion chromatography to remove Endo H (Fig. 3.5B). Unfortunately, it was not possible to fully separate Endo H from deglycosylated CD1d by size exclusion chromatography due to their similar size, therefore only fractions containing CD1d were pooled and used in crystallisation experiments (Fig. 3.5C). A second peak with high absorbance at 280 nm was visible on the size exclusion trace. This did not correspond to protein as assessed by Coomassie staining. Although the content of this peak is unclear, this unknown component was well separated from the CD1d sample.

3.3.3 Co-crystallisation trials of deglycosylated CD1d and SapB

Co-crystallisation trials were set up with deglycosylated CD1d (6.5 mg/mL) and SapB (2.3 mg/mL) at a molar ratio of 1:2 in 100 mM citrate pH 6.0, 150 mM NaCl (Fig. 3.6A). Crystals grew in many crystallisation conditions, suggesting that deglycosylated CD1d is more readily crystallisable than glycosylated CD1d. A crystal was grown against a reservoir containing 20% w/v PEG 3350, 0.1 M bis-Tris propane pH 7.5 and 0.2 M di-sodium malonate (Fig. 3.6B) and a 2.52 Å diffraction dataset was collected from this crystal (Table 3.1, crystal (i)). The structure was solved by molecular replacement using a published structure of CD1d-β_2m (PDB ID: 3U0P) as a search model. Rigid-body refinement of this solution showed unbiased F_o-F_c difference density correctly indicating the presence of *N*-Acetylglucosamine (remaining after Endo H cleavage) on each molecule on residues Asn20 and Asn42 or Asn163 (Fig. 3.6C), absent in the search model, confirming the validity of the molecular replacement solution. Additional unmodelled density near the lipid-binding groove was evident that likely corresponds to the lipid head group (Fig. 3.6C). No additional positive difference density corresponding to SapB molecules was observed.

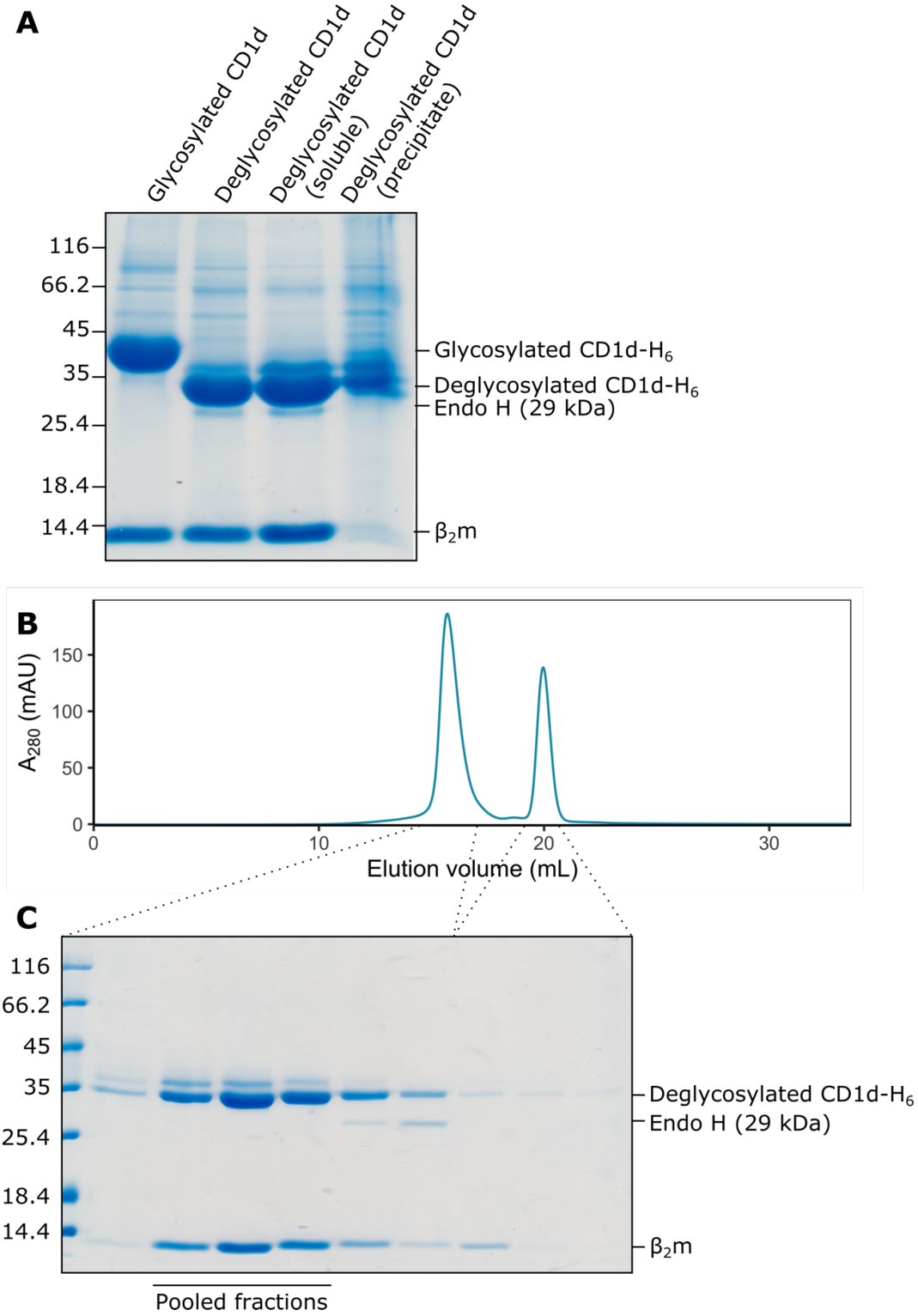

Figure 3.5. Large scale CD1d-β_2m deglycosylation
A. Large scale deglycosylation of CD1d-β_2m by Endo H treatment. After Endo H treatment, precipitate was formed and the contents of the mixture, soluble fraction and precipitate were analysed by SDS-PAGE on a 12% acrylamide gel. B. Soluble deglycosylated CD1d-β_2m was further purified by size exclusion chromatography on a Superdex 200 10/300 column. C. Size exclusion chromatography fractions were analysed by SDS-PAGE on a 12% acrylamide gel, showing incomplete separation between CD1d-β_2m and Endo H. The indicated fractions containing CD1d-β_2m only were pooled and used for experiments.

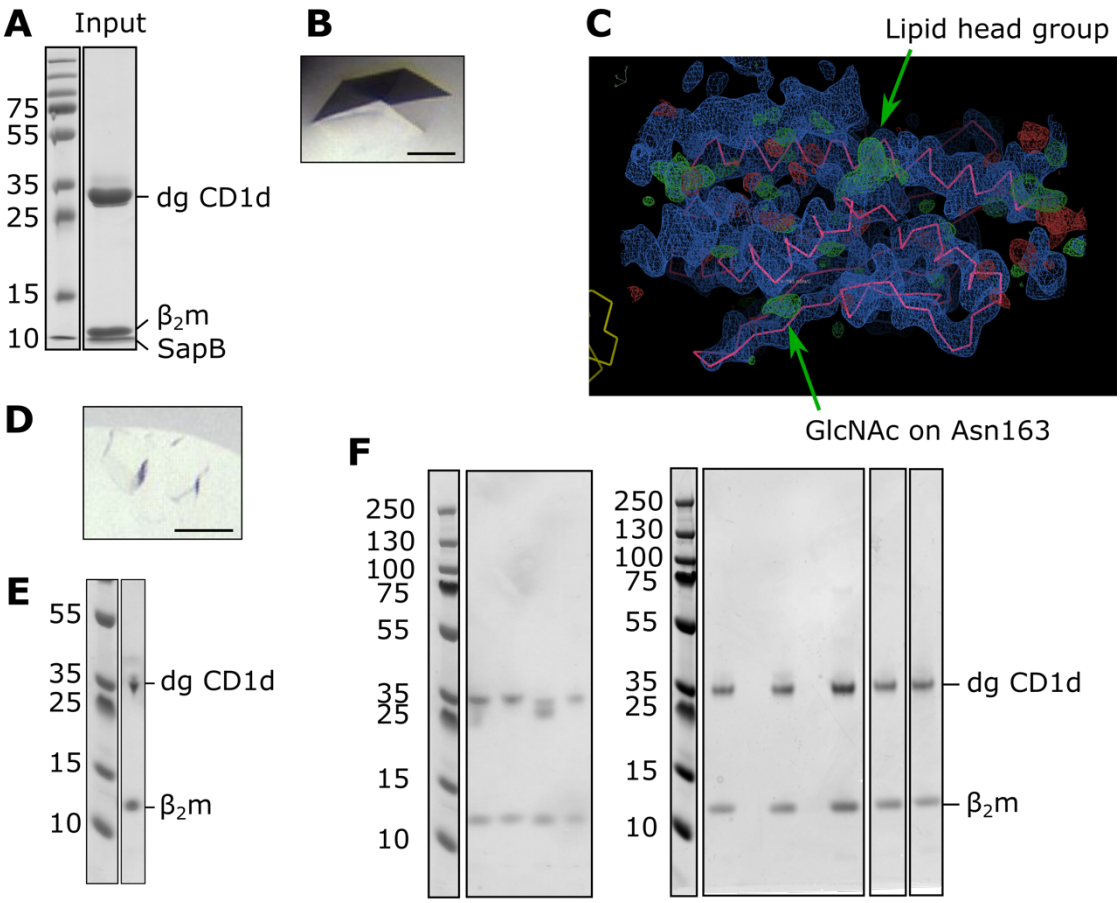

Figure 3.6. Crystallisation trials of deglycosylated CD1d-β_2m with SapB.
A. Coomassie-stained SDS-PAGE of the purified protein sample containing deglycosylated CD1d-β_2m and a 2-fold molar excess of SapB used in crystallisation trials. B. Crystal grown by sitting drop vapour diffusion against a reservoir containing 20% w/v PEG 3350, 0.1 M bis-Tris propane pH 7.5 and 0.2 M di-sodium malonate. A condenser lens was used to image this thin crystal. C. Electron density (2*Fo-Fc*, in blue) and difference density (*Fo-Fc*, in green for positive and red for negative) maps of the structure solved from the crystal shown in panel B. The lipid-binding groove of CD1d is seen from above (CD1d peptide backbone in pink). Unmodelled density corresponding to the lipid headgroup protruding from the CD1d lipid-binding groove, and residual *N*-Acetylglucosamine (GlcNAc) on Asparagine 163 is indicated by green arrows. D. Crystals grown by sitting drop vapour diffusion against a reservoir containing 16% v/v PEG 500 monomethylether, 8% w/v PEG 20000, 0.1 M bicine/Trizma base pH 8.5, 30 mM sodium nitrate, 30 mM sodium phosphate dibasic and 30 mM ammonium sulfate. E. Coomassie-stained SDS-PAGE of harvested and washed crystals shown in panel D. F. Coomassie-stained SDS-PAGE of crystals grown in various crystallisation conditions containing deglycosylated CD1d-β_2m and a 2-fold molar excess of SapB. Scale bar represents 100 µm. Abbreviations: dg: deglycosylated; GlcNAc: *N*-Acetylglucosamine.

Screening of many other crystals grown in conditions containing deglycosylated CD1d and SapB showed that they had the same space group and cell dimensions as the crystal described above, signifying that they have the same content. However, a crystal belonging to a different space group grew against a reservoir containing 20% v/v PEG 550 monomethylether, 10% w/v PEG 20000, 0.1 M

bicine/Trizma base pH 8.5, 30 mM sodium nitrate, 30 mM sodium phosphate dibasic and 30 mM ammonium sulfate and diffracted to 7 Å. This crystal was optimised by screening around the pH and precipitant concentration and increasing protein concentration (to 13.3 mg/mL CD1d-β_2m and 4.7 mg/mL SapB). This improved the resolution of a crystal grown in optimised crystallisation conditions (16% v/v PEG 500 monomethylether, 8% w/v PEG 20000, 0.1 M bicine/Trizma base pH 8.5, 30 mM sodium nitrate, 30 mM sodium phosphate dibasic and 30 mM ammonium sulfate; Fig. 3.6D) to 3.49 Å (Table 3.1, crystal (ii)). The structure was solved by molecular replacement and refined by rigid body refinement. The validity of the solution was confirmed by the presence of *N*-Acetylglucosamine at expected positions (Asn40, 42 and 163) on each of the six molecules in the asymmetric unit. No unmodelled density corresponding to SapB was observed, indicating that the crystal contained only CD1d. This was confirmed by SDS-PAGE analysis of optimised crystals (Fig. 3.6E). Similarly, crystals from a number of additional crystallisation conditions containing deglycosylated CD1d and SapB were analysed by SDS-PAGE and found to contain deglycosylated CD1d alone (Fig. 3.6F).

Table 3.1. Data collection statistics for CD1d crystals (i) and (ii). Values in parentheses are for the highest-resolution shell.

Data collection	**Crystal (i)**	**Crystal (ii)**
Beamline	I03	I04-1
Wavelength (Å)	0.9793	0.9159
Processing package	XIA2 DIALS	XIA2 3dii
Space group	*C* 2	*P* 2_1
Cell dimensions		
a,*b*,*c* (Å)	186.2, 57.1, 129.2	123.8, 122.9, 159.9
α,β,γ (°)	90, 100, 90	90, 95, 90
Resolution (Å)	54.53 – 2.52 (2.56 – 2.52)	123.27 – 3.49 (3.55 – 3.49)
R_{merge}	0.072 (2.438)	0.161 (1.572)
R_{pim}	0.031 (1.028)	0.067 (0.663)
$CC_{1/2}$	0.999 (0.813)	0.992 (0.714)
$I/\sigma I$	10.9 (0.8)	6.9 (1.0)
Completeness (%)	99.8 (98.0)	99.8 (99.9)
Redundancy	6.7 (6.6)	6.9 (6.6)
Molecular replacement		
CD1d-β_2m molecules per asymmetric unit	2	6

In summary, deglycosylated CD1d crystallises readily in a wide range of crystallisation conditions. The use of deglycosylated protein improved the quality of the crystals, allowing determination of the structure of CD1d to 2.52 Å. However, none of the crystals grown with these reagents contained both CD1d and SapB. This suggests that a key component necessary for their interaction may be missing.

3.3.1 Co-crystallisation trials of CD1d and SapB loaded with the lipid α-GalCer

Previous studies have shown that the ectodomain of CD1d co-purifies with a broad range of cellular lipids following synthesis and trafficking through the secretory pathway, but is likely to be primarily loaded with sphingomyelin[157-159]. SapB co-purifies with *E. coli* phosphatidylethanolamine[82], a lipid species that has also been shown to bind CD1d[222]. Although it seems likely that this lipid composition for SapB and CD1d should be compatible with protein-protein interactions required for lipid exchange, lipid loading onto CD1d has only been formally demonstrated *in vitro* for a few lipids, and in particular α-GalCer[165,166]. Therefore, both proteins were pre-loaded with α-GalCer for co-crystallisation trials. Lipid loading was performed based on a protocol used to obtain crystal structures of α-GalCer-loaded CD1d[177,222]. CD1d and SapB were incubated with a 3-fold excess of lipid in PBS overnight. α-GalCer-loaded proteins were then separated from free α-GalCer by size exclusion chromatography. Lipid exchange did not cause changes in the elution profile of CD1d (Fig. 3.7A-B) and SapB (Fig. 3.7C-D), suggesting that both proteins remained correctly folded and did not degrade over the course of lipid exchange.

I set up co-crystallisation trials with α-GalCer-loaded CD1d and SapB at a molar ratio of 1:2 and 1:3 in 50 mM Tris pH 7.4, 150 mM NaCl (CD1d at 11.9 mg/mL and SapB at 4.2 or 6.3 mg/mL; Fig. 3.8A). Crystals grew against a reservoir containing 0.1 M sodium acetate pH 5.0, 20% w/v PEG 6000 and 0.2 M ammonium chloride (Fig. 3.8B) and a diffraction dataset was collected (Table 3.2, crystal (iii)). The structure was solved by molecular replacement and refined by rigid body refinement as in section 3.3.3. For each of the four CD1d molecules in the asymmetric unit, unbiased F_o-F_c difference density indicated the presence of the expected *N*-Acetylglucosamine on residues Asp2, Asp42 and/or Asp 163, confirming the validity of the molecular replacement solution. Unmodelled density near the lipid binding groove corresponding to α-GalCer was also observed (Fig. 3.8C). No additional positive difference density corresponding to SapB molecules was observed. The absence of SapB was confirmed by SDS-PAGE analysis of the crystal (Fig. 3.8D). The harvested crystal was directly dissolved in SDS-PAGE buffer after data collection.

Other crystals grew against a reservoir containing 12.5% w/v PEG 1000, 12.5% w/v PEG 3350, 12.5% v/v 2-methyl-2,4-pentanediol, 0.03 M sodium fluoride, 0.03 M sodium bromide, 0.03 M sodium

iodide, 0.1 M bicine/Trizma base pH 8.5 (Fig. 3.8E). These crystals did not diffract, so I analysed the contents of remaining crystals by SDS-PAGE analysis. This showed that this crystal was also composed of CD1d only (Fig. 3.8F).

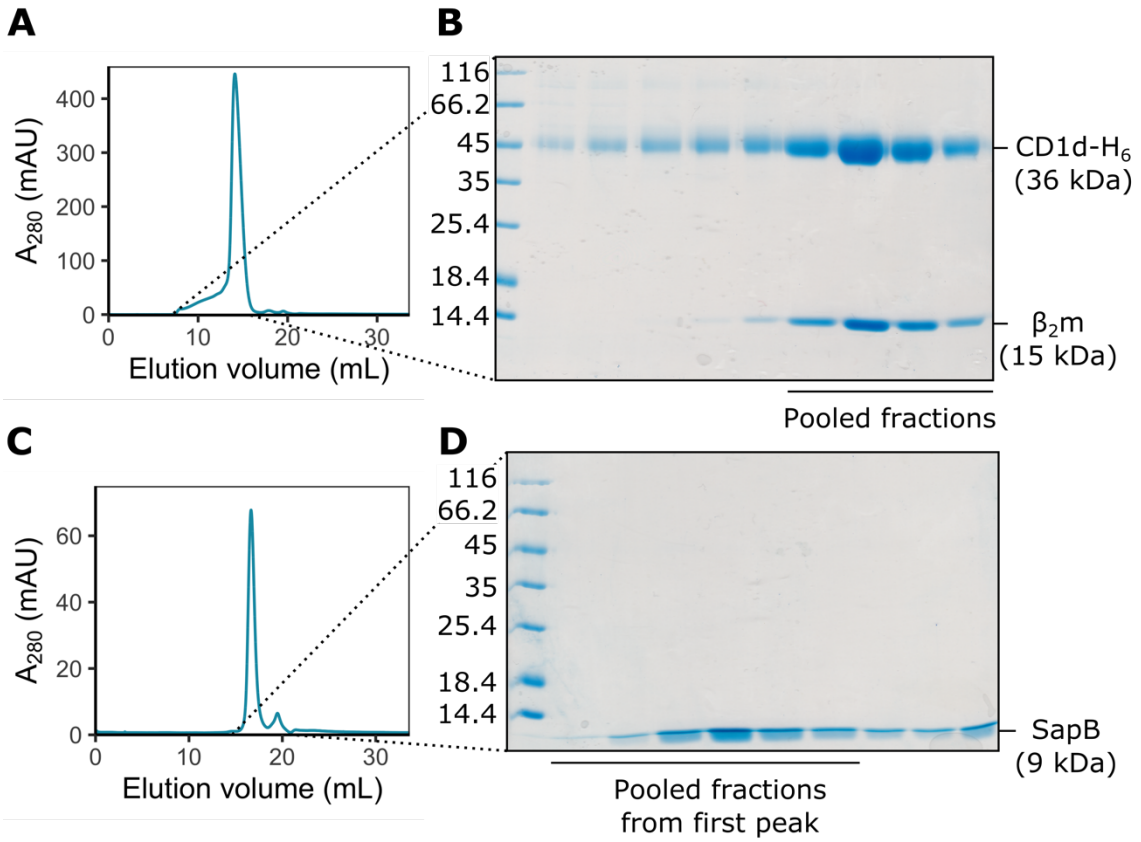

Figure 3.7. Re-purification of CD1d and SapB after loading with the lipid α-GalCer
Purified CD1d and SapB protein were loaded with the lipid α-GalCer by overnight incubation with a 3-fold molar excess of the lipid. Size exclusion chromatography of CD1d and SapB with a Superdex 200 10/300 column was used to separate lipid-loaded proteins from lipids in solution. A. Size exclusion chromatogram of α-GalCer-loaded CD1d. B. Eluted fractions were analysed on a Coomassie-stained 15% SDS-PAGE gel. C. Size exclusion chromatogram of α-GalCer-loaded SapB. D. Eluted fractions were analysed on a Coomassie-stained 15% SDS-PAGE gel.

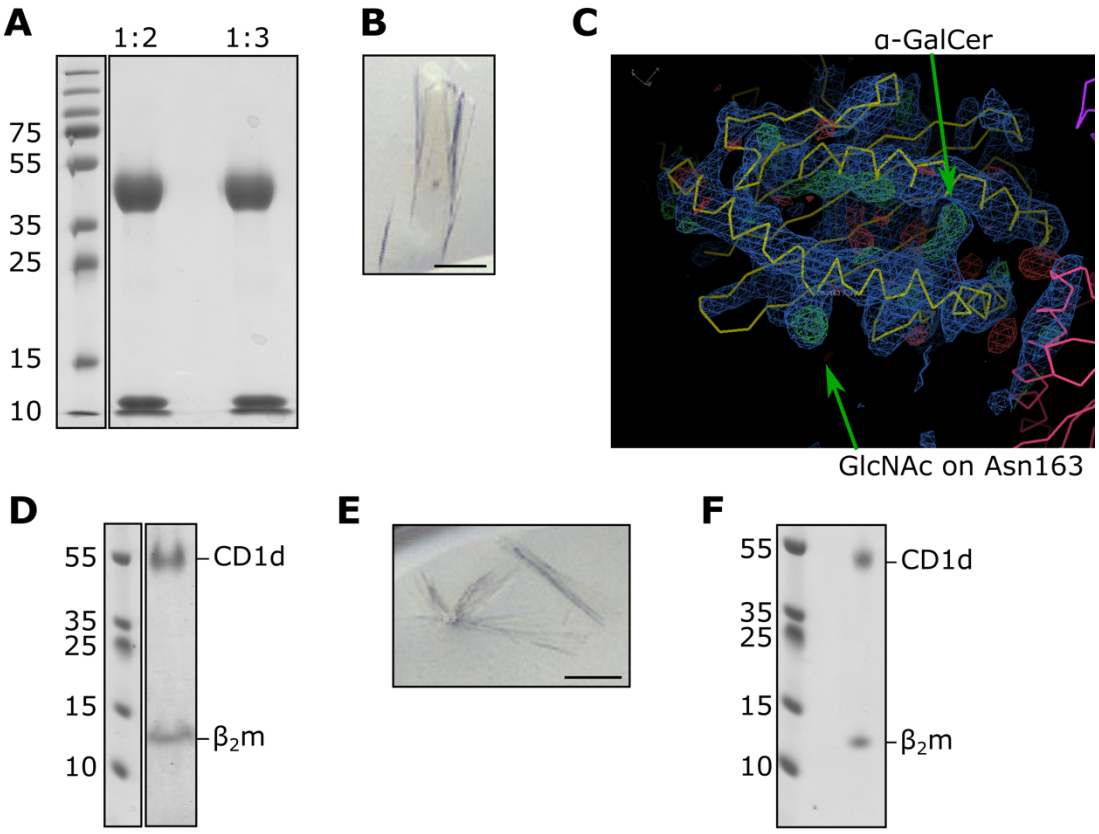

Figure 3.8. Crystallisation trials of α-GalCer-loaded CD1d and SapB
A. Coomassie-stained SDS-PAGE of the purified protein samples containing α-GalCer-loaded CD1d-β_2m and a 2- or 3-fold molar excess of α-GalCer-loaded SapB used in crystallisation trials. B. Crystal grown by sitting drop vapour diffusion against a reservoir containing 0.1 M sodium acetate pH 5.0, 20% w/v PEG 6000 and 0.2 M ammonium chloride. C. Electron density (2*Fo-Fc*, in blue) and difference density (*Fo-Fc*, in green for positive and red for negative) maps of the structure solved from the crystal shown in panel B. The lipid-binding groove of CD1d is seen from above (CD1d peptide backbone in yellow). Unmodelled density corresponding to α-GalCer in the CD1d lipid-binding groove, and residual *N*-Acetylglucosamine (GlcNAc) on Asparagine 163 is indicated by green arrows. D. Coomassie-stained SDS-PAGE of harvested crystal shown in panel B. E. Crystals grown by sitting drop vapour diffusion against a reservoir containing 12.5% w/v PEG 1000, 12.5% w/v PEG 3350, 12.5% v/v 2-methyl-2,4-pentanediol, 0.03 M sodium fluoride, 0.03 M sodium bromide, 0.03 M sodium iodide, 0.1 M bicine/Trizma base pH 8.5. F. Coomassie-stained SDS-PAGE of harvested crystal shown in panel E. Scale bar represents 100 µm. Abbreviations: GlcNAc: *N*-Acetylglucosamine.

Table 3.2. Data collection statistics for CD1d crystal (iii). Values in parentheses are for the highest-resolution shell.

Data collection	**Crystal (iii)**	
Beamline	I04	
Wavelength (Å)	0.9999	
Processing package	STARANISO	
Space group	*C* 2	
Cell dimensions		
a,*b*,*c* (Å)	273.3, 88.0, 127.9	
α,*β*,*γ* (°)	90, 114, 90	
	Isotropic	*Anisotropic*
Resolution (Å)	124.71 – 4.97 (5.05 – 4.97)	124.71 – 4.55 (4.99 – 4.55)
R_{merge}	0.438 (2.187)	0.198 (0.507)
R_{pim}	0.279 (1.366)	0.127 (0.323)
$CC_{1/2}$	0.967 (0.322)	0.987 (0.704)
$I/\sigma I$	1.9 (0.5)	3.4 (1.8)
Completeness (%)	99.8 (99.8)	e* 80.0 (56.7) s* 39.2 (8.1)
Redundancy	3.4 (3.5)	3.4 (3.4)
Molecular replacement		
CD1d-β_2m molecules per asymmetric unit	4	

3.3.2 Co-crystallisation trials of glycosylated CD1d with glycosylated SapB

SapB possesses one N-linked glycosylation site on residue Asn21, which remains unglycosylated when expressed in *E. coli*. Although unglycosylated SapB loads lipids onto CD1d[165,166], it is possible that this glycan on SapB may increase its affinity for CD1d. Therefore, I expressed H_6-tagged SapB (SapB-H_6) in mammalian cells, enabling the purification of glycosylated protein. SapB-H_6 was expressed using the pHLsec vector transfected into HEK293F cells using transient PEI-based transfection. SapB-H_6 was purified by nickel affinity chromatography from the media (Fig. 3.9A), followed by size exclusion chromatography (Fig. 3.9B-C). Following purification, two SapB-H_6 species were present, with apparent molecular weights of 12 and 15 kDa (Fig. 3.9B). This tagged SapB construct has a predicted molecular weight of 10 kDa, therefore these two species are likely to be SapB-H_6 protein modified with glycans of different sizes. The lack of separation of these two species by size exclusion chromatography suggests that they form heterodimers.

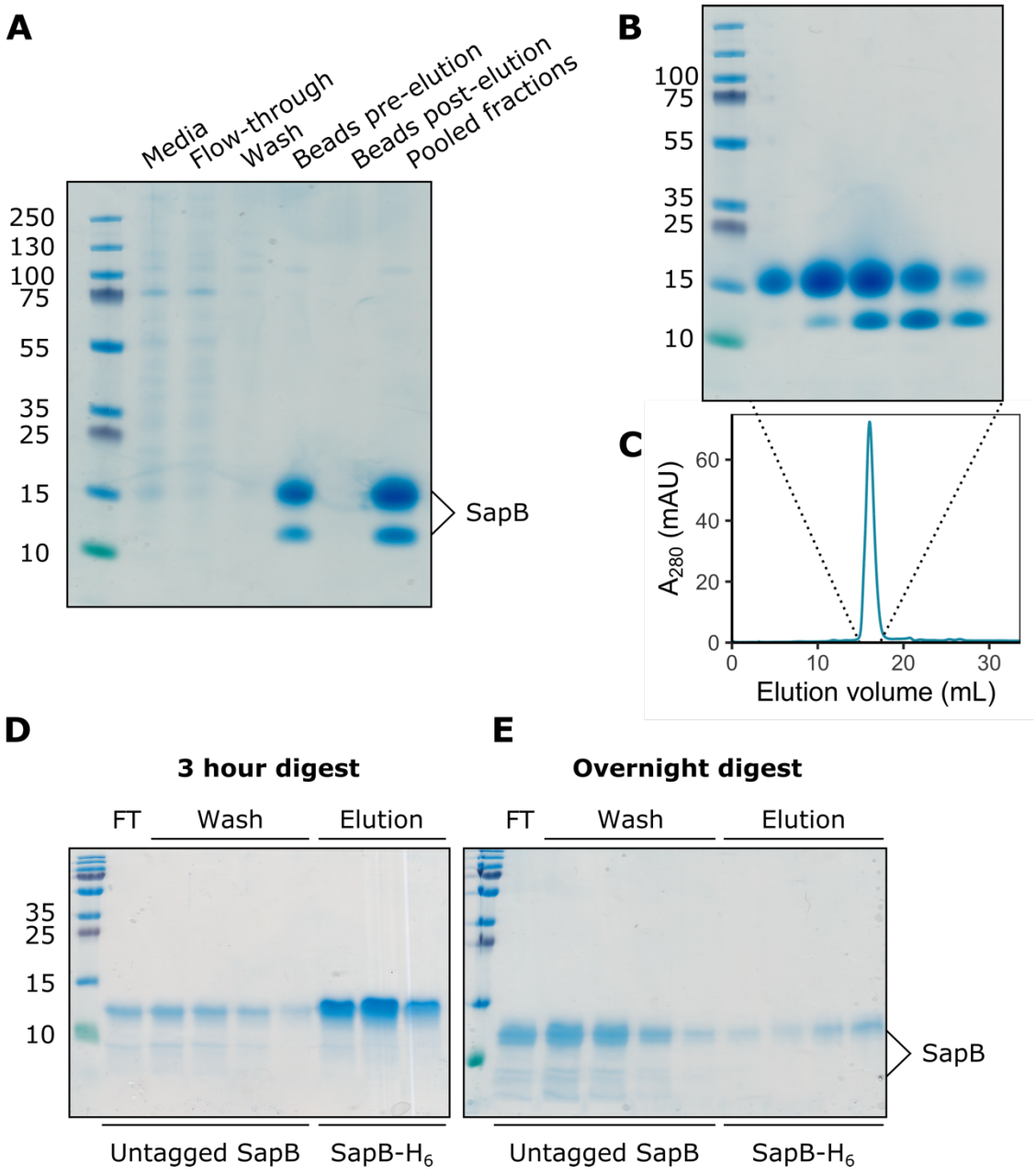

Figure 3.9. Purification of glycosylated SapB from mammalian cells
Secreted, hexahistidine-tagged full-length SapB (SapB-H_6) was expressed in mammalian HEK293F cells to enable purification of glycosylated SapB. A. SapB-H_6 was captured from the media with Ni-NTA agarose and eluted with 300 mM imidazole in PBS. Fractions were analysed on a Coomassie-stained 4-12% Bis-Tris SDS-PAGE gel. Two SapB species are present, likely reflecting the heterogeneity of glycan modifications. B. Eluted fractions were pooled, concentrated, and CD1d-H_6-β_2m was further purified by size exclusion chromatography (SEC) using a Superdex 200 10/300 column. SEC elution fractions were analysed on a Coomassie-stained 4-12% Bis-tris SDS-PAGE gel. All fractions were pooled. C. SEC chromatogram. D-E. The C-terminal His-tag of SapB was removed by incubation with carboxypeptidase A-agarose beads for 3 hours (D) or overnight (E). After this treatment, untagged SapB was separated from SapB-H_6 and free histidine by nickel affinity chromatography with Ni-NTA beads. Flow-through (FT) and wash fractions containing untagged SapB, and elution fractions containing SapB-H_6 were visualised on a Coomassie-stained 15% SDS-PAGE gel.

Because of the small size and compact conformation of SapB dimers, the addition of a His_6 tag may affect SapB folding and function. Therefore, I removed the tag by treatment of SapB-H_6 with carboxypeptidase A, which hydrolyses peptide bonds of C-terminal residues up to the lysine residue preceding the His tag. SapB-H_6 was incubated with carboxypeptidase A-agarose beads for 3 hours or overnight, then untagged SapB was separated from cleaved His and His-tagged SapB by affinity purification with Ni-NTA agarose. Tag removal efficiency was assessed by SDS-PAGE analysis of flow-through and wash fractions containing untagged SapB, and eluted fractions containing His-tagged SapB (Fig. 3.9D-E). Carboxypeptidase A digestion for 3 hours was found to be insufficient (Fig. 3.9D). After overnight digestion (Fig. 3.9E), a large fraction of SapB appeared untagged and this protocol was used to generate untagged glycosylated SapB for the next experiments. Approximately 0.5-1 mg untagged SapB was obtained from a 300 mL HEK293F culture. I noted that SapB bands appeared smeary, indicating some heterogeneity that may have been caused by excessive digestion of SapB by Carboxypeptidase A. However, due to the relatively small quantity of degraded SapB, this sample was carried through for crystallisation experiments.

I set up co-crystallisation trials with glycosylated CD1d (12.4 mg/mL) and glycosylated, untagged SapB (4.3 mg/mL) at a molar ratio of 1:2 in 20 mM HEPES pH 7.0, 150 mM NaCl (Fig. 3.10A). Crystals grew against a reservoir containing 12.5% w/v PEG 1000, 12.5% w/v PEG 3350, 12.5% v/v 2-methyl-2,4-pentanediol, 20 mM 1,6-hexanediol, 20 mM 1-butanol, 20 mM 1,2-propanediol, 20 mM 2-propanol, 20 mM 1,4-butanediol, 20 mM 1,3-propanediol and 0.1 M bicine/Trizma base pH 8.5 (Fig. 3.10B). However, SDS-PAGE analysis of these crystals showed that they were composed of CD1d only (Fig. 3.10C).

Crystals also grew against a reservoir containing 12.5% w/v PEG 1000, 12.5% w/v PEG 3350, 12.5% v/v 2-methyl-2,4-pentanediol, 20 mM D-glucose, 20 mM D-mannose, 20 mM D-galactose, 20 mM L-fucose, 20 mM D-xylose, 20 mM N-acetyl-D-glucosamine, 0.1 M MOPS/HEPES-Na pH 7.5 (Fig. 3.10D). These crystals were analysed by SDS-PAGE and were once again found to contain CD1d only (Fig. 3.10E). (Note that although glycosylated SapB runs at around the same size as β_2m, separation of SapB and β_2m is clearly visible by SDS-PAGE analysis using 4-12% Bis-Tris gradient gels, as seen in Fig. 3.14).

No crystal containing glycosylated SapB was found. This may be because the sample heterogeneity discussed above prevented crystal formation.

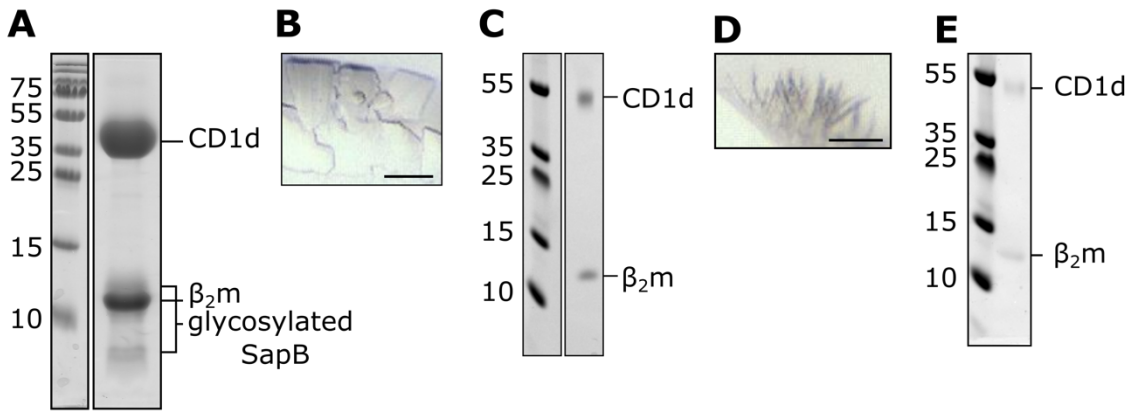

Figure 3.10. Crystallisation trials of CD1d and glycosylated SapB
A. Coomassie-stained SDS-PAGE of the purified protein samples containing CD1d-β_2m and a 2-fold molar excess of glycosylated SapB used in crystallisation trials. B. Crystal grown by sitting drop vapour diffusion against a reservoir containing 12.5% w/v PEG 1000, 12.5% w/v PEG 3350, 12.5% v/v 2-methyl-2,4-pentanediol, 20 mM 1,6-hexanediol, 20 mM 1-butanol, 20 mM 1,2-propanediol, 20 mM 2-propanol, 20 mM 1,4-butanediol, 20 mM 1,3-propanediol and 0.1 M bicine/Trizma base pH 8.5. C Coomassie-stained SDS-PAGE of harvested crystals shown in panel B. D. Crystals grown by sitting drop vapour diffusion against a reservoir containing 12.5% w/v PEG 1000, 12.5% w/v PEG 3350, 12.5% v/v 2-methyl-2,4-pentanediol, 20 mM D-glucose, 20 mM D-mannose, 20 mM D-galactose, 20 mM L-fucose, 20 mM D-xylose, 20 mM N-acetyl-D-glucosamine, 0.1 M MOPS/HEPES-Na pH 7.5. E. Coomassie-stained SDS-PAGE of harvested crystal shown in panel D. Scale bar represents 100 µm.

3.3.3 Summary of co-crystallisation experiments

I have performed extensive co-crystallisation trials of CD1d and SapB. I have identified several factors that may affect this interaction as well as protein crystallisation. I have used different molar ratios of CD1d to SapB, fully glycosylated or deglycosylated protein, and protein loaded with endogenous lipids or with α-GalCer (Table 3.3). Several crystals were obtained in each of these experiments, confirming the protein samples were of high quality and crystallisable, with the exception of glycosylated SapB. However, each of these crystals was found to contain CD1d or SapB only, indicating that they do not form a stable, crystallisable complex.

In a crystallisation experiment, the drop containing protein and precipitant gradually equilibrates with the reservoir, generally resulting in reduced drop volume and the progressive increase in concentration of the drop components. In this way, the protein components are steadily concentrated together. High concentrations of individual components can drive the formation of a protein complex even for weakly interacting proteins[223]. The absence of a co-crystal containing both CD1d and SapB suggests one of two things. (1) CD1d and SapB may interact in solution, but their complex is not crystallisable. Crystallisation is the process of formation of a regular array of macromolecules; if CD1d:SapB complexes cannot self-organise into such an array, no crystal can be formed. (2) CD1d and

SapB may not directly interact together. Although there is strong evidence that SapB directly loads lipids onto CD1d, a direct interaction has not yet been demonstrated. In fact, several publications report a failure to detect a direct interaction between saposins and human CD1d[160,165,166]. Unfortunately, these reports do not show any data or detail the approaches used to monitor binding, and the implication is that these interactions exist but are weak or transitory.

Table 3.3. Summary of co-crystallisation trials

CD1d	SapB	Molar ratio of CD1d to SapB
Glycosylated CD1d	Unglycosylated SapB	1:2
		1:3
α-GalCer loaded glycosylated CD1d	α-GalCer loaded unglycosylated SapB	1:2
		1:3
Deglycosylated CD1d	Unglycosylated SapB	1:2
Glycosylated CD1d	Glycosylated SapB	1:2

3.4 *In vitro* CD1d-SapB interaction assays

The fact that CD1d and SapB do not form a crystallisation complex cast doubt on the established idea that SapB directly interacts with CD1d. Therefore, I aimed to determine whether CD1d interacts with SapB using biochemical techniques adapted for weak and transient interactions.

3.4.1 Equilibrium binding assays of CD1d and SapB

Classical pull-down experiments involve the capture and immobilisation onto resin of a tagged bait protein followed by incubation with prey protein. Subsequent wash steps remove unbound protein, leaving behind protein complexes for detection. However, as the concentration of free prey protein is reduced at each wash, the equilibrium is pushed towards complex dissociation, meaning this method is not well suited for the capture of weak or transient interactions. Using this approach, no interaction was detectable between immobilised CD1d-β_2m and SapB. Therefore, an alternative approach was adopted to overcome the limitations of prey depletion by using an *in vitro* equilibrium binding assay (Fig. 3.11A). In this assay, the bait is bound to beads at a range of concentrations and the prey is added at a low, limiting concentration. Following incubation, and no wash steps, the supernatant is separated from the beads and unbound prey is detected using the sensitive protein dye SYPRO Ruby. Monitoring

the depletion of prey from the supernatant by the large excess of bait protein, rather than monitoring complex formation directly, represents a more sensitive measure of binding and thus is optimised for weak or transient interactions[224]. Equilibrium binding assays were carried out using CD1d-β_2m as bait at a concentration range of 0-40 µM and with SapB as the prey at 680 nM. Previous *in vitro* studies have identified that SapB can load α-GalCer onto CD1d at a pH range between 5 and 8[165]. Although this exchange appears to be most efficient at pH 6, this pH causes partial dissociation of the CD1d H_6-tag from the Ni-NTA resin. Therefore, the equilibrium binding assay was performed at pH 7. Previous work and preliminary experiments showed that β_2m dissociates from CD1d in the presence of detergent[225]. However, the gentle detergent digitonin was found to cause minimal disruption of the CD1d-β_2m complex[225] and was used in this assay. Despite the use of this optimised approach, no binding of SapB to CD1d could be detected (Fig. 3.11B). In these conditions, at high CD1d-β_2m concentrations, a small amount of β_2m is detectable in the supernatant due to dissociation from CD1d (Fig. 3.11B). By comparing how little appears to be lost based on the Coomassie-stained bead samples this highlights the sensitivity of this technique and of the SYPRO Ruby stain for protein detection.

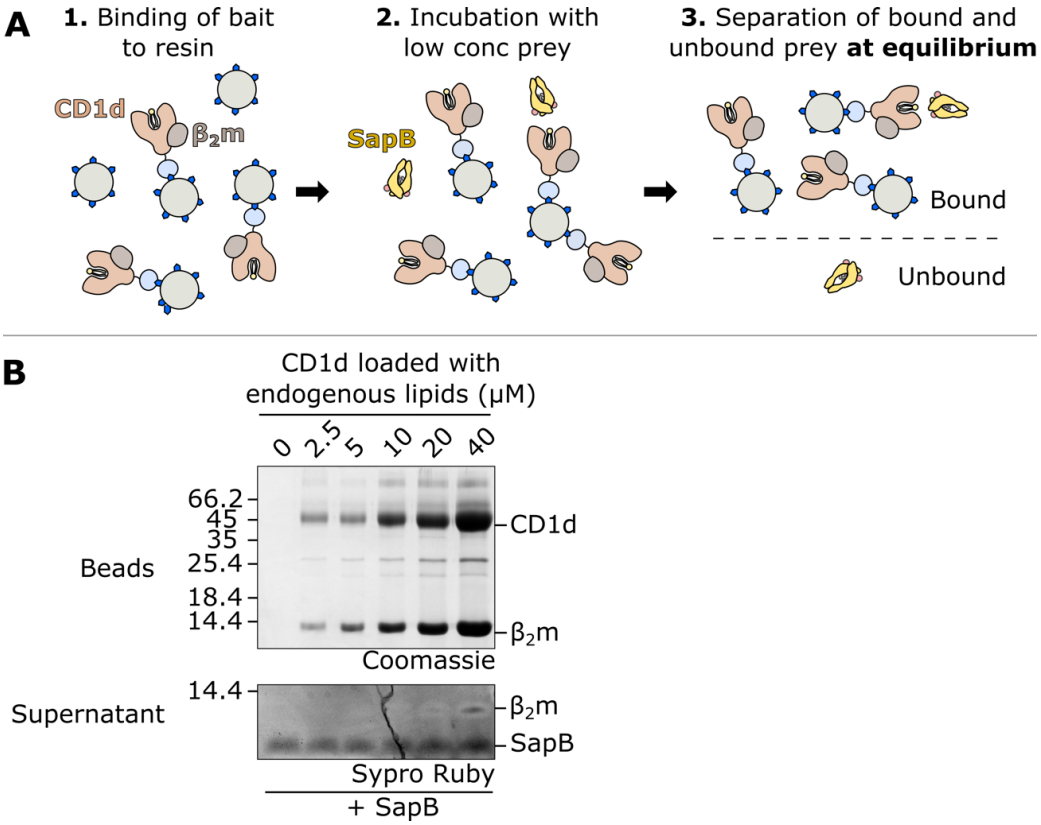

Figure 3.11. Equilibrium binding assay of CD1d-β_2m and SapB
A. Schematic diagram illustrating the equilibrium binding assay protocol. B. Increasing concentrations of CD1d-β_2m were bound to beads and detected by Coomassie-stained SDS-PAGE. After incubation with SapB, unbound SapB was detected in supernatants by SYPRO Ruby-stained SDS-PAGE. The lipids bound to CD1d-β_2m and SapB in this experiment are the endogenous lipids they co-purify with following expression in human cells and *E. coli.*

As discussed in section 3.3.4, it is possible that the lipids associated with CD1d and SapB may contribute to protein-protein interactions. Therefore, I performed the equilibrium binding assay with both proteins pre-loaded with α-GalCer, a lipid that has been shown to be loaded onto CD1d by SapB. α-GalCer lipid loading was done as described in section 3.3.4 prior to the experiment. However, this variation also did not result in a detectable interaction between SapB and CD1d (Fig. 3.12A). To exclude the possibility that a strong interaction between CD1d and SapB only occurs when SapB exchanges a lower affinity antigen for α-GalCer, I repeated the equilibrium binding assay where SapB was pre-loaded with α-GalCer while CD1d retained the co-purified endogenous lipids (Fig. 3.12B). Again, there was no substantial depletion of SapB from the supernatant in this assay. Taken together, these results strongly suggest that CD1d and SapB do not directly interact *in vitro*, or that their interaction is extremely weak with a K_d > 40 μM.

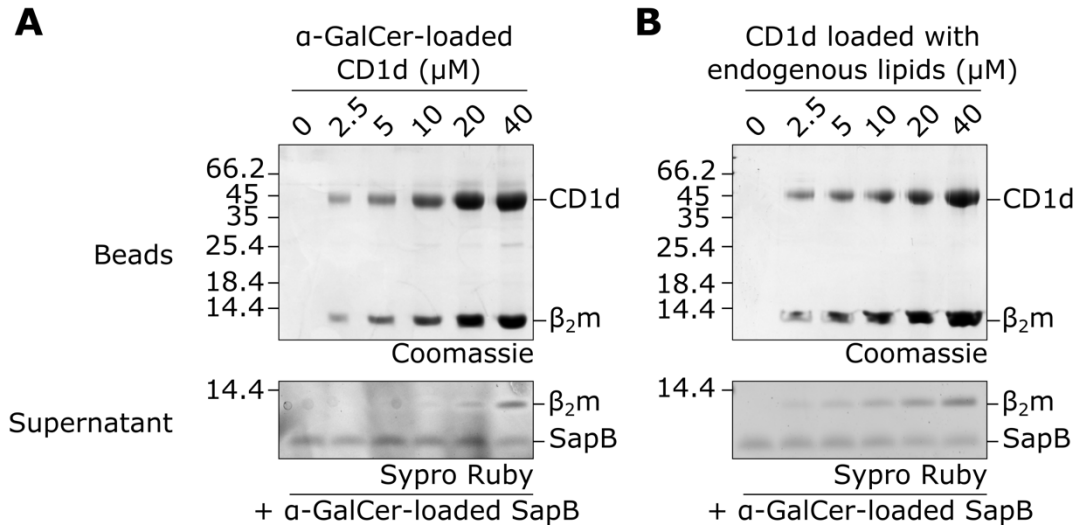

Figure 3.12. Equilibrium binding assay of α-GalCer-loaded CD1d-β_2m and SapB
An equilibrium binding assay was performed as using α-GalCer-loaded proteins. A. Increasing concentrations of α-GalCer-loaded CD1d-β_2m were bound to beads and detected by Coomassie-stained SDS-PAGE. After incubation with α-GalCer-loaded SapB, unbound SapB was detected in supernatants by SYPRO Ruby-stained SDS-PAGE. B. As for panel A but with SapB alone pre-loaded with the lipid α-GalCer. CD1d-β_2m is bound to endogenous lipids co-purified following expression in human cells.

3.4.2 *In vitro* cross-linking assays of CD1d and SapB

Cross-linking of proteins using short chemical tethers that react with surface-exposed lysine side-chains can be used to capture transient or weak interactions. Both CD1d and SapB have several surface-exposed lysines, and of particular importance, lysine residues are present near the lipid-binding groove of CD1d where SapB is predicted to bind (Fig. 3.13A-B). Therefore, it is possible that CD1d and SapB interact in a way that is amenable to intermolecular cross-linking of surface lysines by

amine-reactive cross-linkers. I used cross-linkers of different lengths to maximise the chances of capturing an interaction: DSSO, with a 10.3 Å spacer arm, and BS(PEG)5, with a 21.7 Å spacer arm. CD1d-β_2m and SapB were incubated with these cross-linkers either alone or following mixing (Fig. 3.13C). For the CD1d-β_2m samples alone, cross-linking between CD1d and β_2m is clearly evident, as expected. There is also some evidence of additional higher order oligomers, probably due to the large number of surface lysines on both CD1d and β_2m. However, no additional cross-linking products were observed when CD1d-β_2m and SapB were incubated together with the cross-linkers.

Although unglycosylated SapB loads lipids onto CD1d[165,166], it is possible that the N-linked glycan on Asn21 on SapB (Fig. 3.13B) may increase its affinity for CD1d. Therefore, I purified glycosylated SapB-H_6 (section 3.3.5) for use in a cross-linking experiment. This time, I did not remove the His_6 tag, as to do so would result in smearing of the bands corresponding to SapB on a gel, as discussed in section 3.3.5, and make SDS-PAGE analysis difficult.

I carried out a cross-linking experiment using glycosylated SapB-H_6 and glycosylated CD1d (Fig. 3.14). Once again, I did not observe formation of any additional cross-linked products when CD1d-β_2m was incubated with glycosylated SapB-H_6. I noted that although the same amount of unglycosylated SapB (Fig. 3.13C) and glycosylated SapB-H_6 (Fig. 3.14) were used in cross-linking assays, the Coomassie-stained bands corresponding to glycosylated SapB appear much stronger. This increased staining may be due to the additional histidine residues in the affinity tag increasing the amount of Coomassie reagent binding to this small protein. These results indicate that CD1d-β_2m and SapB do not interact together in a conformation that is amenable to cross-linking.

In summary, no direct interaction between CD1d and SapB has been detected using several sensitive *in vitro* assays using purified components.

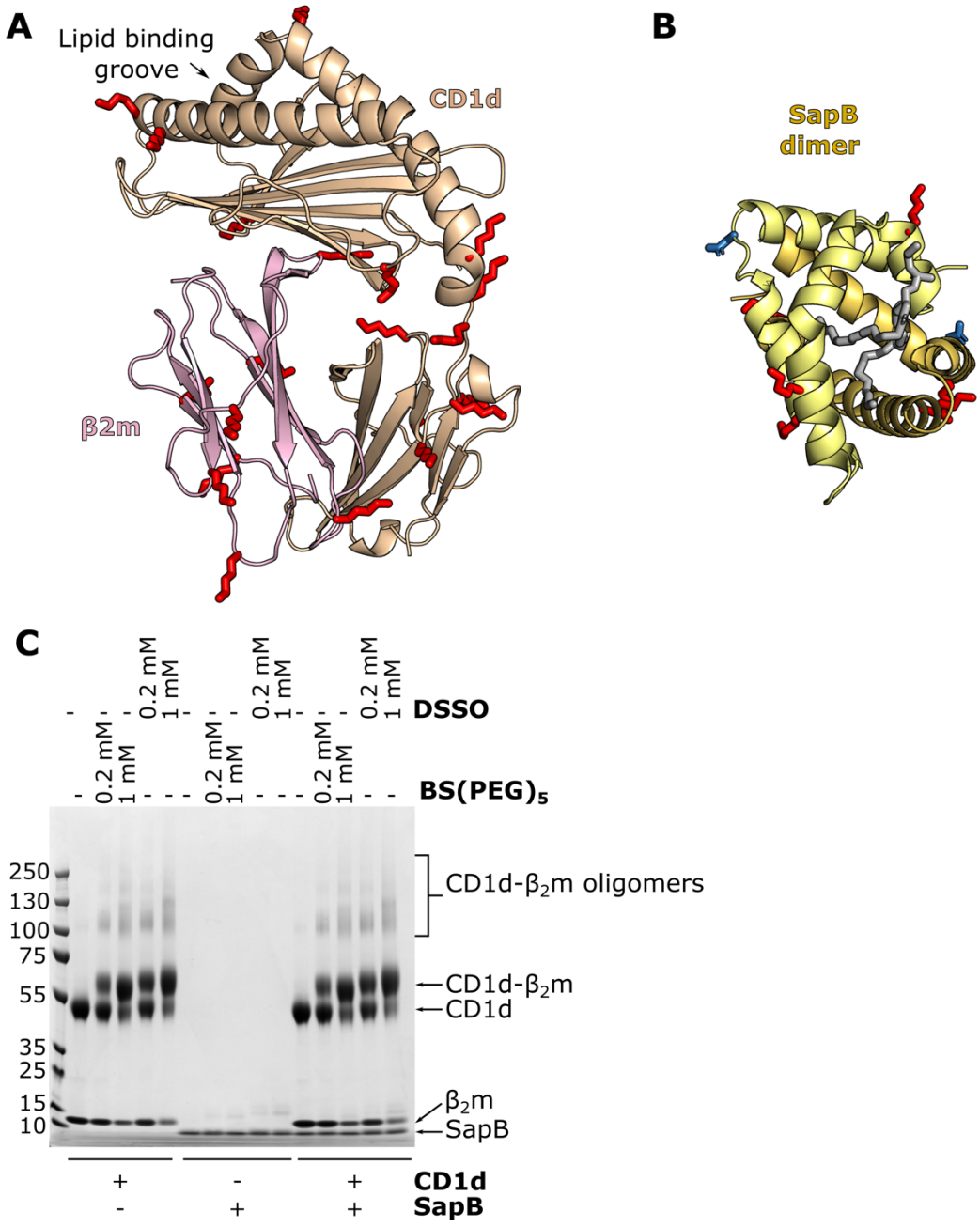

Figure 3.13. *In vitro* cross-linking of CD1d-β_2m and SapB
A. Ribbon diagram of the structure of CD1d (beige) and β_2m (pink) (PDB ID 3U0P) with the lipid antigen groove labelled (arrow). Surface-accessible lysine residues available for cross-linking are highlighted (red sticks). B. Ribbon diagram of the structure of a SapB dimer (yellow and orange, PDB ID 1N69) with bound lipid (in gray). Lysine residues are highlighted as in panel A. The N-linked glycosylation site of SapB is highlighted (blue sticks). C. Coomassie-stained SDS-PAGE of CD1d-β_2m and unglycosylated SapB samples following incubation with cross-linkers of different lengths: DSSO (10.3 Å) or BS(PEG)5 (21.7 Å).

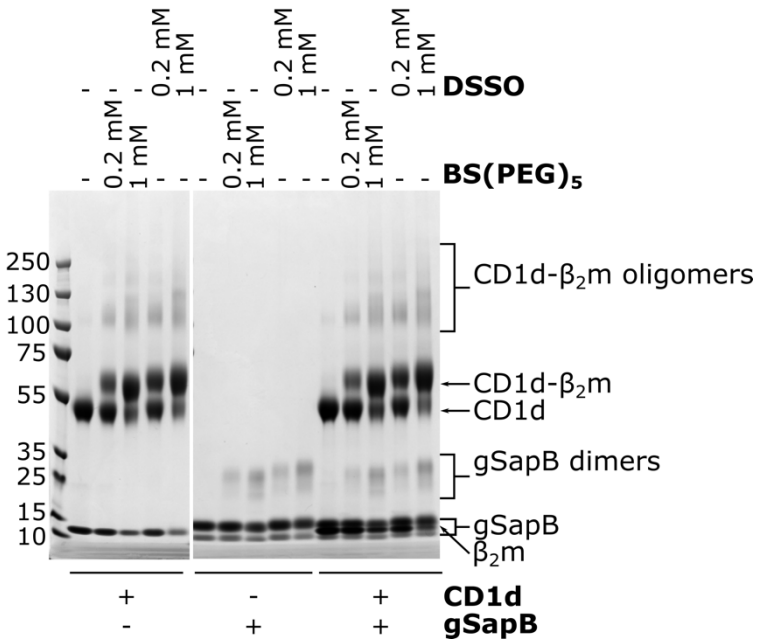

Figure 3.14. *In vitro* cross-linking of CD1d-β_2m and glycosylated SapB
Coomassie-stained SDS-PAGE of CD1d-β_2m and glycosylated SapB (gSapB) samples following incubation with cross-linkers of different lengths: DSSO (10.3 Å) or BS(PEG)5 (21.7 Å).

3.5 CD1d co-immunoprecipitation/mass spectrometry

The lack of a detectable interaction between CD1d and SapB *in vitro* suggests that SapB may require a cellular co-factor to interact with CD1d. To test whether CD1d interacts indirectly with SapB inside cells and identify potential protein partners that may facilitate this interaction, I performed a co-immunoprecipitation experiment followed by mass spectrometry, using CD1d as bait. This experiment was performed in HeLa cells due to the ease of transfection of this cell line and the presence of key proteins. Saposins are ubiquitously expressed and present in HeLa cells[226], and β_2m expression can be induced in HeLa cells by treatment with interferon-γ (IFN-γ)(Louise Boyle, personal communication). I transiently expressed recombinant, full-length CD1d with a C-terminal green fluorescent protein (GFP) tag (CD1d-GFP) (Fig. 3.15A) in HeLa cells. Alternatively, as negative controls, I expressed the transmembrane MHC class I antigen HLA-A2 with a C-terminal GFP tag (HLA-A2-GFP) or soluble, cytosolic GFP alone. In the same manner as CD1d, HLA-A2 forms a stable interaction with β_2m necessary for its correct folding. β_2m expression was induced in all transfected cells by treatment with IFN-γ. This was done 15 hours after transfection to avoid reducing transfection efficiency. 48 hours after IFN-γ treatment, GFP-tagged bait was immunoprecipitated from the lysate at endososomal pH (pH 6).

As a preliminary experiment, I performed GFP immunoprecipitation followed by Western blotting for GFP to ensure that all constructs are expressed and immunoprecipitated at pH 6 (Fig. 3.15B). All three constructs were successfully expressed and captured. However, expression levels were unequal. GFP was expressed at a much higher level than CD1d, and HLA-A2 expression was lower than CD1d. This may affect pull-down results, since a larger amount of bait protein can pull-down larger amounts of interaction partners. I envisaged optimising transfection protocols for each construct to equalize expression levels, however, this would result in different levels of cell stress and may cause changes in protein expression, post-translational modification and interactions across samples. Therefore, I chose to maintain this protocol for the co-immunoprecipitation/mass spectrometry experiment, keeping in mind this caveat. I performed a Western blot for β_2m to verify that CD1d and HLA-A2 are correctly folded and able to form an interaction with this protein partner (Fig. 3.15B). β_2m does not co-immunoprecipitate with GFP, but co-immunoprecipitates with CD1d and to a lesser extent with HLA-A2, which is expected due to lower expression of HLA-A2. This indicates that CD1d and HLA-A2 are correctly folded and an interaction with the known partner β_2m can be detected.

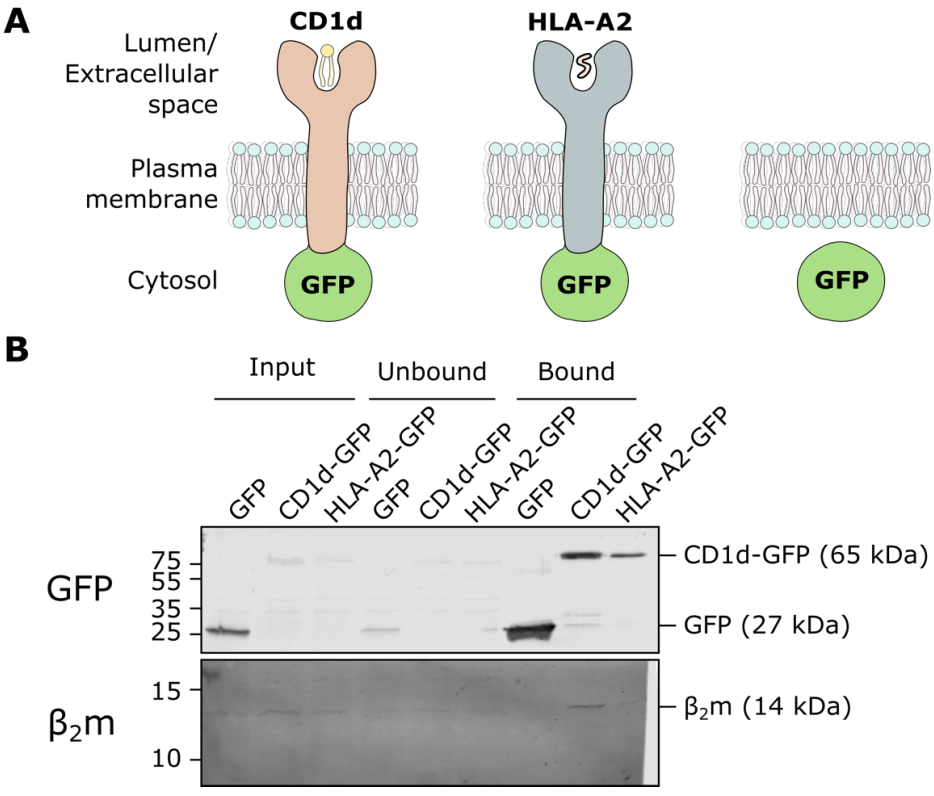

Figure 3.15. CD1d co-immunoprecipitation in HeLa cells
A. HeLa cells were transiently transfected with full-length CD1d with a C-terminal GFP tag, or, as negative controls, full-length HLA-A2 with a C-terminal GFP tag or GFP alone. Both CD1d and HLA-A2 are known to form a stable interaction with β_2m. B. GFP was immunoprecipitated from transfected lysates at endolysosomal pH (pH 6). Successful expression and capture of all protein constructs and co-immunoprecipitation of the CD1d and HLA-A2 partner protein β_2m was verified by Western blotting for GFP and β_2m.

I then performed the co-immunoprecipitation experiment followed by mass spectrometry in three biological replicates. Proteins interacting with CD1d were compared to proteins interacting with GFP (Fig. 3.16A) and HLA-A2 (Fig. 3.16B). The only protein significantly enriched by co-immunoprecipitation with CD1d was CD1d itself. No peptide corresponding to SapB co-immunoprecipitating with CD1d was detected by mass spectrometry. Two overlapping peptides mapped to the saposin SapC (NVIPALELVEPIK and NVIPALELVEPIKK) were detected after CD1d co-immunoprecipitation in two replicates, but this enrichment was not found to be significant.

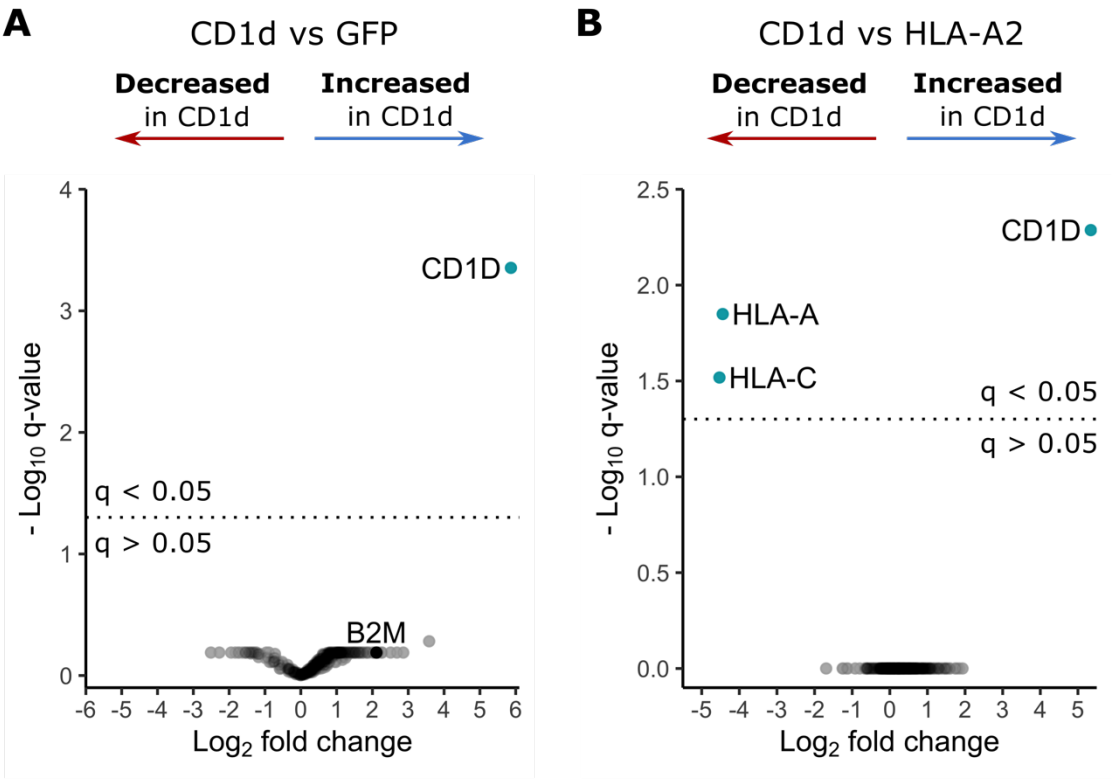

Figure 3.16. CD1d co-immunoprecipitation-mass spectrometry
CD1d-GFP or the negative controls GFP or HLA-A2-GFP were transiently expressed in HeLa cells and immunoprecipitated. Co-immunoprecipitated interacting proteins were quantified by mass spectrometry. A. Volcano plot showing pairwise comparison between the proteins pulled down by CD1d-GFP and GFP control. Each dot represents a single protein. Proteins significantly changed ($q > 0.05$) are coloured in blue. B. Volcano plot showing pairwise comparison between the proteins pulled down by CD1d-GFP and HLA-A2 control.

There is no significant enrichment of any CD1d-interacting proteins, including β_2m. HLA-A2 binds β_2m, therefore it is not expected that it should be significantly enriched by CD1d co-immunoprecipitation when compared with HLA-A2. However, β_2m does not interact with GFP, therefore it should be enriched by co-immunoprecipitation with CD1d. In this experiment, β_2m enrichment by CD1d co-immunoprecipitation is not significant and lesser than expected (2-fold change; Fig. 3.16A). The fact

that the known stable interaction partner of CD1d was not identified in this experiment suggests that these experimental conditions were suboptimal, for example due to misfolding of a large fraction of overexpressed CD1d causing non-specific interactions and therefore high variability between samples. Therefore, this assay is not suitable for the identification of transient interactions. Technical difficulties caused by protein over-expression may be overcome by endogenous co-immunoprecipitation experiments using an antibody recognising CD1d. However, the interaction between CD1d, SapB and unidentified proteins potentially facilitating their interaction may be hindered by antibody binding. Although alternative strategies could be used to pursue the aim of identifying CD1d interaction partners, it was decided at this point to cease work on this project.

3.6 Discussion

I attempted to obtain a snapshot of the lipid transfer mechanism between CD1d and SapB using X-ray crystallography. Despite extensive testing of various combinations of highly pure and crystallisable protein components, CD1d and SapB did not co-crystallise. Furthermore, I have been unable to identify a direct interaction between CD1d-β_2m and SapB *in vitro*. In addition to co-crystallisation trials, several biochemical approaches that have the capacity to capture weak and transient interactions were implemented, including equilibrium binding assays and protein cross-linking. Various factors that may influence the strength or specificity of these interactions were tested, including different pH and buffer systems, exchange of the lipids loaded onto both protein components, and changing the glycosylation state of each protein. None of these variants resulted in the detection of a direct interaction. Previous work in our lab studying saposin interactions with lysosomal hydrolases showed that although these interactions are sensitive to pH and salt concentrations, they were always detectable, and while detergent/lipid was required, neither the nature of the lipid/detergent nor the glycosylation state mattered[80].

Direct protein-protein interactions are crucial for several different lipid transfer pathways. The lipopolysaccharide (LPS) transfer cascade resulting in Toll-like receptor 4 (TLR4) activation involves several lipid transfer steps. LPS binding protein (LBP) extracts LPS and transfers it onto the receptor CD14. CD14 then transfers LPS to the TLR4 co-receptor MD-2. A direct, high affinity (K_d = 5.52 nM) interaction between LBP and CD14 is necessary for lipid transfer, and lipid transfer promotes rapid dissociation of CD14:LPS from LBP. Similarly, a direct interaction between CD14 and the TLR4:MD-2 complex is required for lipid transfer[171]. The transfer of lipids between proteins is also seen in the formation of Apolipoprotein B-containing lipoproteins. Microsomal Triglyceride Transfer protein assists this process by binding to Apolipoprotein B with high affinity (K_d = 10-30 nM)[227] and transferring

lipids onto Apolipoprotein B through an unknown mechanism. Both functions are required for lipoprotein assembly[172]. As another example, Niemann–Pick C1 (NPC1) and Niemann–Pick C2 (NPC2) proteins collaborate to export cholesterol from the lysosome. Soluble NPC2 binds cholesterol in the lysosomal lumen, interacts with the middle lumenal domain of the multipass transmembrane protein NPC1 and transfers cholesterol to NPC1 via a transfer tunnel linking the cholesterol-binding pockets of both proteins. The interaction between NPC2 and NPC1 (K_d = 0.4-2 μM) is required for cholesterol export[77]. In all cases, a high-to-medium affinity direct interaction between proteins engaged in lipid transfer is detectable. In this study, we were unable to detect a direct interaction between CD1d and SapB despite the assays being capable of detecting very weak interactions. Specifically, the equilibrium assay used in this study can detect protein interactions where the dissociation constant is as weak as 40 μM. In the absence of avidity effects, interactions weaker than 40 μM are unlikely to be physiologically relevant. Furthermore, the cross-linking assay can detect short-lived, transient interactions and the co-crystallography technique uses very high concentrations of protein components allowing for the detection of very weak interactions. It therefore seems unlikely that an interaction that is undetectable by these methods could permit direct lipid exchange between SapB and CD1d.

The study of lipid loading of CD1 molecules is technically challenging as biochemical methods often require the inclusion of detergents, which may interfere with lipid solubilisation and binding. Also, the detection of changes to CD1d-lipid complexes is limited by available reagents including specific antibodies and iNKT cell receptors. However, the inability to identify a direct interaction between CD1d and SapB raises a number of important questions regarding the role of saposins in lipid antigen presentation.

It is well established that knockout of the saposin precursor protein PSAP results in impaired presentation of lipid antigens by CD1d. Due to the dual role of saposins in lipid loading and lipid processing, the absence of PSAP will not only alter CD1d lipid loading, but will also significantly alter the composition of the lysosomal lipidome. Indeed, the loss of a single saposin results in the cellular accumulation of multiple, different GSLs. For example, loss of SapB alone, in mice or in humans, results in the accumulation of sulfatide, globotriaosylceramide, and lactosylceramide[228,229]. Therefore, in saposin knockout and rescue models, the effects on antigen processing and the changes to local lipid concentrations may have indirect and unpredictable effects on CD1d loading in the lysosome. Similarly, cell-based assays that involve recognition of CD1d-lipid complexes by NKT cells are made

complicated by the requirement of some glycolipid headgroups to be processed by lysosomal hydrolases[161], a process that is itself dependent upon saposins.

In vitro assays testing the role of different saposins in facilitating lysosomal hydrolase activity identify some of the difficulties in determining specificity of saposin-lipid binding. For example, *in vitro* assays have identified that SapC and SapA are both able to enhance the activity of β-galactosylceramidase (GALC) to promote degradation of galactosphingolipids[230,231]. However, this does not reflect the specificity of saposins in whole organisms as loss of SapA causes Krabbe disease, similar to that of loss of GALC, while mutations in SapC do not[64,65,67-69]. Why these *in vitro* assays do not recapitulate the specificity of saposins in a whole organism remains unclear but may suggest a non-specific detergent-like effect of saposins when present at relatively high concentrations. It is unclear whether a similar effect may be partially confounding the interpretation of SapB's influence on CD1d lipid loading. While *in vitro* studies demonstrating the ability of SapB to enhance CD1d stimulation of iNKT cells or binding to iNKT cell receptors are very compelling regarding the importance of SapB in this process, these do represent an indirect way of monitoring how SapB may be involved in lipid exchange[165,166]. In light of our data showing no detectable interaction between CD1d and SapB, the question remains: how can SapB facilitate lipid editing without directly binding CD1d? Modelling of how SapB might facilitate lipid exchange identified a role for SapB in enhancing the kinetics of lipid exchange in a manner similar to that of Tapasin in peptide exchange on MHC class-I molecules[166]. In order for Tapasin to carry out this role, it does bind directly to MHC class-I but as part of a larger complex including additional proteins[232,233]. Although the *in vitro* assays demonstrating the importance of SapB for CD1d lipid presentation did not require other proteins, it is possible that there may be additional factors involved in enhancing the affinity of SapB for CD1d in cells. A co-immunoprecipitation-mass spectrometry approach to identify CD1d binding partners was not technically successful, most likely due to issues with recombinant CD1d expression. The detection of protein interactions within the lumen of endolysosomal compartments presents many challenges. The low pH of endolysosomal compartments is not conducive to proximity labelling approaches such as Bio-ID and APEX, which rely on a biotinylation reaction requiring pH > 7. The presence of proteases can result in cleavage of recombinant protein tags, posing difficulties for tag-based approaches including Förster resonant energy transfer as well as proximity labelling. Overcoming these challenges for the identification of CD1d binding partners will be essential to clarify the mechanism of CD1d lipid loading.

Additional factors may influence the efficiency of lipid loading onto CD1d in cells. For example, the relative abundance of different lipid transfer proteins in the lysosome may influence the lysosomal

lipid composition, altering exchange kinetics. Therefore, it may be important that SapA-D exist in equimolar concentrations in the endolysosome, as ensured by their production from a common precursor. If this is the case, studies involving individual saposins or partial PSAP rescues will be difficult to interpret, potentially explaining some experimental inconsistencies. The lysosomal lipidome is not only influenced by the activity of saposins and sphingolipid processing hydrolases but also by the careful balance of other membrane components such as cholesterol and bis(monoacylglycero)phosphate. Lipid loading of CD1 in endolysosomal compartments is highly complex, requiring the delicate interplay of lipid binding, lipid processing and lipid exchange. For now, the mechanism of transfer of lipids onto CD1 molecules, as well as the molecular determinants of lipid transfer protein binding specificity, remains unclear.

4 Investigating the role of GALC in phagocytes

Chapter summary

Krabbe disease is accompanied by immune dysfunction. The hallmark of Krabbe disease which gives this disorder its alternative name, globoid cell leukodystrophy, is the presence of globoid cells in the brain. These are large multinucleated globular cells of unelucidated phagocytic origin[113]. Moreover, elevated levels of pro-inflammatory IL-6 and TNF-α have been observed in the brain of a mouse disease model[117]. The causes of these immune pathologies are unknown. The phagocytic origin of globoid cells and the central role of phagocytes in secreting pro-inflammatory cytokines have led me to hypothesise that immune dysfunction is driven by phagocytic cell dysfunction due to loss of GALC. Loss of the lysosomal enzyme GALC leads to the accumulation of its substrates GalCer and psychosine. Perturbation of glycosphingolipid degradation leads to substrate accumulation in the lysosome, but also throughout the cell, including at the plasma membrane[58,59]. As discussed in section 1.1, it is becoming increasingly appreciated that the correct lipid environment is crucial for the function, signalling and localisation of plasma membrane proteins, and in particular of immune signalling complexes. Notably, defects in sphingolipid metabolism result in dysregulation of Toll-like receptor (TLR) signalling and trafficking[16-19]. This suggests that GALC may play a role in normal phagocyte function. In this study, I explored how loss of GALC affects phagocytic cells and the proteins expressed on their surface. The phagocytic monocyte-like THP-1 cell line[234] was chosen as a model to study how sphingolipid imbalances affect phagocyte phenotype and function. Krabbe disease patients have extremely low GALC activity levels, and even small amounts of residual enzyme activity can be sufficient to prevent substrate accumulation. This is demonstrated by the fact that loss of up to 90% of enzyme activity can be asymptomatic[235]. A knockdown approach may not sufficiently reduce GALC levels. Therefore, I used CRISPR/Cas9 to generate GALC knockout (KO) THP-1 cell lines. Unfortunately, the results of this study were compromised by off-target effects of genetic manipulation in parental and control cell lines. The implications of these findings for the design of gene knockout strategies are discussed.

Statement of collaboration

Several collaborations have been essential to the work described in this chapter. The Lehner lab (Cambridge Institute of Therapeutic Immunology & Infectious Disease, University of Cambridge) kindly

provided the THP-1 Cas9 cell line and Dr Lidia Duncan provided the protocols for CRISPR/Cas9 knockout in these cells. Dr James Williamson generously trained me in sample preparation for plasma membrane profiling and performed mass spectrometry analysis of these samples. He also performed the sample processing and mass spectrometry analysis of whole cell lysates. The datasets generated through this collaboration are presented in sections 4.4.1, 4.4.5 and 4.4.7. Prof Ian Brierley (Department of Pathology, University of Cambridge) supervised and significantly aided sample preparation for RNA sequencing. RNA sequencing was done by Cambridge Genomic Services (Department of Pathology, University of Cambridge) and Dr Katy Brown (Department of Pathology, University of Cambridge) performed the data analysis, generating the data presented in section 4.4.3.

4.1 THP-1 cells possess high levels of GALC activity

GALC is known to be expressed in oligodendrocytes, where its importance is well established. However, GALC expression and function in other cell types has been scarcely investigated. Transcription profiling studies have shown that GALC is expressed in primary human phagocytic cells, including monocytes[236,237]. I tested whether there is detectable GALC activity in the monocyte-like THP-1 cell line using a GALC activity assay previously developed by our group. In this assay, endogenous GALC is enriched from cell lysates by immunoprecipitation using a specific monoclonal antibody. GALC-bound beads are then washed and incubated with the fluorogenic GALC substrate 4-methylumbelliferyl β-D-galactopyranoside (4MβDG) and the generation of fluorescent product is monitored over time[238]. The activity of GALC was higher in THP-1 cells than in the MO3.13 oligodendrocyte-like cell line (Fig. 4.1). This suggests that GALC has an important role in regulating galactosphingolipid turnover in THP-1 cells, indicating that THP-1 cells are an appropriate phagocytic cell line to monitor the effects of GALC loss. Endogenous GALC protein levels are too low to be detectable by Western blot in MO3.13 cells, and GALC detection by Western blot was similarly unsuccessful in THP-1 cells.

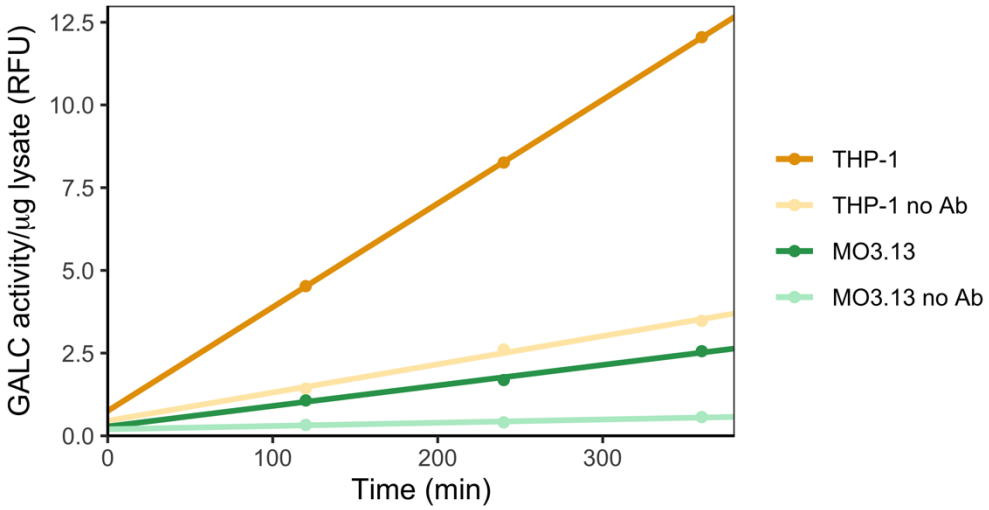

Figure 4.1. GALC activity in THP-1 cells
GALC activity assay on THP-1 and MO3.13 cell lysates. GALC was immunoprecipitated from cell lysates and incubated with the fluorogenic substrate 4MβDG. GALC activity is reported as the change of fluorescence in relative fluorescence units (RFU) over time. Background activity was measured in a control sample where no GALC antibody was added for the immunoprecipitation step ("No Ab"). This data is representative of three independent experiments.

THP-1 monocytes can be differentiated by phorbol 12-myristate 13-acetate (PMA) into a macrophage-like cell more responsive to lipopolysaccharide (LPS) stimulation[175]. THP-1 cells grow in suspension when undifferentiated and become adherent upon differentiation. I observed that GALC activity is tuned down by differentiation, and further reduced by stimulation of differentiated cells with the TLR2 agonist lipopolysaccharide (LPS) (Fig. 4.2A). Quantitative PCR (qPCR) indicates that *GALC* gene expression is significantly up-regulated by differentiation of THP-1 cells, and LPS stimulation of differentiated cells significantly down-regulates *GALC* gene expression (Fig. 4.2B). This is in agreement with previous transcription profiling studies showing differential GALC expression in primary human monocytes and macrophages[236,237]. Changes in total GALC activity are not correlated with changes in *GALC* gene expression, suggesting that GALC activity is also regulated by protein degradation or changes in RNA stability. GalCer levels are regulated by the complex interplay of GALC-mediated GalCer degradation, GalCer formation through the catabolism of complex galactosphingolipids and GalCer synthesis by UDP-galactose:ceramide galactosyltransferase 8 (UGT8). qPCR analysis showed that *UGT8* gene expression was significantly down-regulated in LPS-stimulated differentiated THP-1 cells compared to undifferentiated cells (Fig. 4.2C). Unfortunately, no reliable antibodies against UGT8 were available, therefore the presence of UGT8 protein could not be confirmed and quantified. However, down-regulation of UGT8 at the mRNA and protein level upon stimulation of macrophages with LPS and a TLR7/8 agonist has been previously reported[239]. Altogether, this data supports the hypothesis that galactosphingolipid levels are dynamically regulated in THP-1 cells to facilitate

processes such as immune signalling, phagocytosis and adhesion specific to each state. In the rest of this study, I focused on the role of GALC in undifferentiated monocyte-like THP-1 cells.

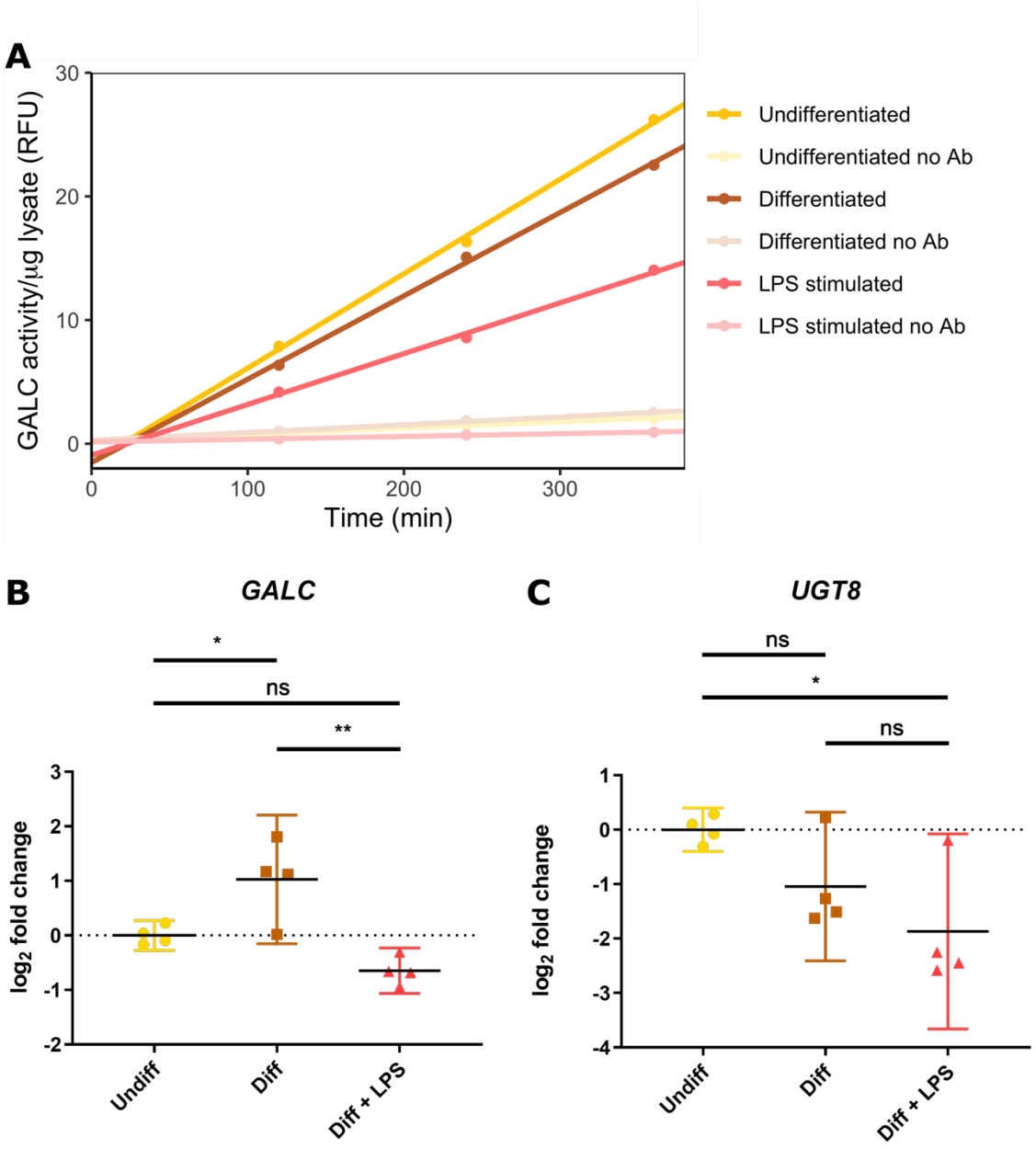

Figure 4.2. GALC expression is modulated by THP-1 differentiation and LPS stimulation
A. GALC activity assay in undifferentiated, differentiated and LPS-stimulated THP-1 cells. GALC was immunoprecipitated from cell lysates and incubated with the fluorogenic substrate 4MBDG. GALC activity is reported as the change of fluorescence in relative fluorescence units (RFU) over time. Background activity was measured in a control sample where no GALC antibody was added for the immunoprecipitation step ("No Ab"). This data is representative of two independent experiments. B-C. qPCR quantification of *GALC* (B) and *UGT8* (C) gene expression relative to *ACTB* in undifferentiated (undiff), differentiated (diff) and differentiated WT THP-1 cells stimulated with LPS (Diff + LPS). Gene expression was normalised to undifferentiated cells. p-values were calculated by one-way ANOVA followed by Tukey's multiple comparison test. Error bars represent the 95% confidence interval. ns: not significant (p-value > 0.05); *: $p \leq 0.05$.; **: $p \leq 0.01$.

4.2 CRISPR/Cas9-mediated knockout of GALC in THP-1 cells

To investigate the effect of GALC loss in phagocytes, I knocked out GALC in THP-1 cells using CRISPR/Cas9[240]. Expressing Cas9 and guide RNA (gRNA) targeting GALC in this cell line was challenging, because phagocytes are inherently resistant to transfection. A first attempt to transiently express Cas9 and gRNAs by electroporation of a co-expression plasmid into THP-1 cells was unsuccessful. No viable cells expressing this plasmid were obtained due to the low cell viability and low transfection efficiency using this technique. Subsequently, I used a THP-1 cell line stably expressing Cas9 (THP-1 Cas9) provided by a collaborator and delivered gRNAs targeting GALC or an empty vector (EV) gRNA by lentiviral transduction of this cell line (Fig. 4.3A). I used two different gRNAs (gRNA1 and gRNA2) targeting different parts of exon 3 (residues 73-93) of GALC (Fig. 4.3B). These gRNAs were previously successfully used in our lab to knock out GALC in MO3.13 oligodendrocytes. Generating two different knockout cell lines using distinct gRNAs allows to discriminate between on-target effects caused by loss of GALC, and potential off-target effects caused by non-specific binding of each gRNA. Cas9 THP-1 cells were infected with lentivirus vectors at a concentration such that 30% of cells were infected, as determined by flow cytometry using the lentiviral BFP marker. This ensures that the majority of transduced cells are infected by a single virus particle and only one insertion event takes place in each cell, minimising the risk of insertional mutagenesis. Loss of GALC may cause changes in the expression of adhesion molecules on the surface of THP-1 cells, which may result in adhesion of THP-1 cells normally growing in suspension. Therefore, after lentiviral transduction, I carried forward both suspension cells and any adherent cells detached with trypsin for the next steps. I selected transduced cells with puromycin and isolated single clones by limiting dilution, generating monoclonal populations labelled empty vector control clone 1 (EV1), knockout clone 1 (KO1) and knockout clone 2 (KO2) (Fig. 4.3A). These clones remained in suspension. I expanded and froze aliquots of these cell lines. Cas9 and gRNAs are stably expressed in these cell lines and pose a risk of spurious gene editing over time. Therefore, I ensured that all experiments were performed within 12 passages after thawing these aliquots to prevent the potential generation and expansion of cell populations with off-target mutations.

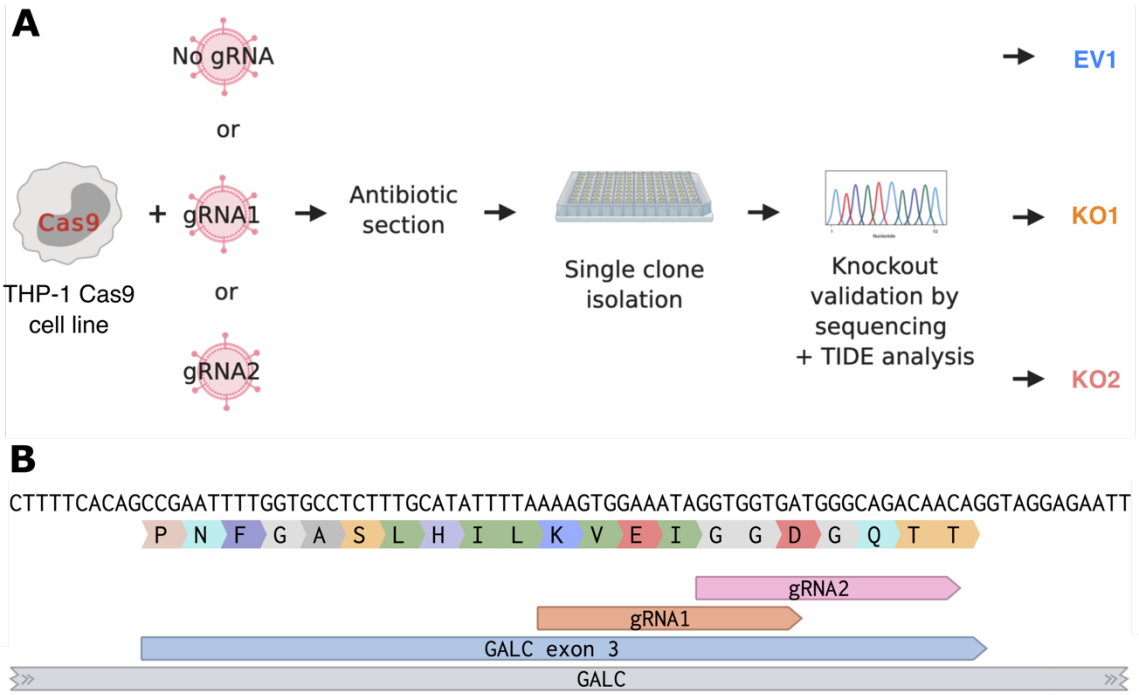

Figure 4.3. CRISPR/Cas9-mediated knockout of GALC in THP-1 cells
A. Graphical summary of the CRISPR/Cas9 knock-out protocol. B. DNA sequence of GALC exon 3. The arrows labelled gRNA1 and gRNA2 indicate the target DNA sequence for these gRNAs. Exon 3 encodes residues 73 to 93 of GALC.

4.2.1 GALC KO validation

To determine whether GALC gene editing by CRISPR/Cas9 successfully introduced insertions or deletions (indels), I isolated genomic DNA from individual clones and amplified the region of DNA encompassing the genetic lesion for Sanger sequencing (Fig. 4.4A-B). Indels were identified by TIDE analysis[206](Fig. 4.4C-D). For KO1, there is clear GALC gene editing across the gRNA1 site, however, TIDE analysis identifies three potential sequence traces instead of the expected two corresponding to each GALC allele (Fig. 4.4A). In accord with this, TIDE analysis identifies three deletions (-4, -10 and -21; Fig. 4.4C). The same result was obtained after sequencing a fresh batch of cells from this clone. This puzzling result may be explained by the presence of two clones in this cell population. Both clones may have a shared mutation on one allele and a divergent mutation on the second allele. Alternatively, the DNA region normally corresponding to GALC may have failed to amplify due to a large deletion in one allele. Although this suggests that KO1 may not be a monoclonal population, it is clear that the GALC gene is edited in this population with two missense deletions and one in-frame deletion. For KO2, there are two deletions identified in the sequence trace (Fig. 4.4B) and confirmed by TIDE analysis (-7 and -12), corresponding to a frameshifting deletion and an in-frame deletion (Fig. 4.4D).

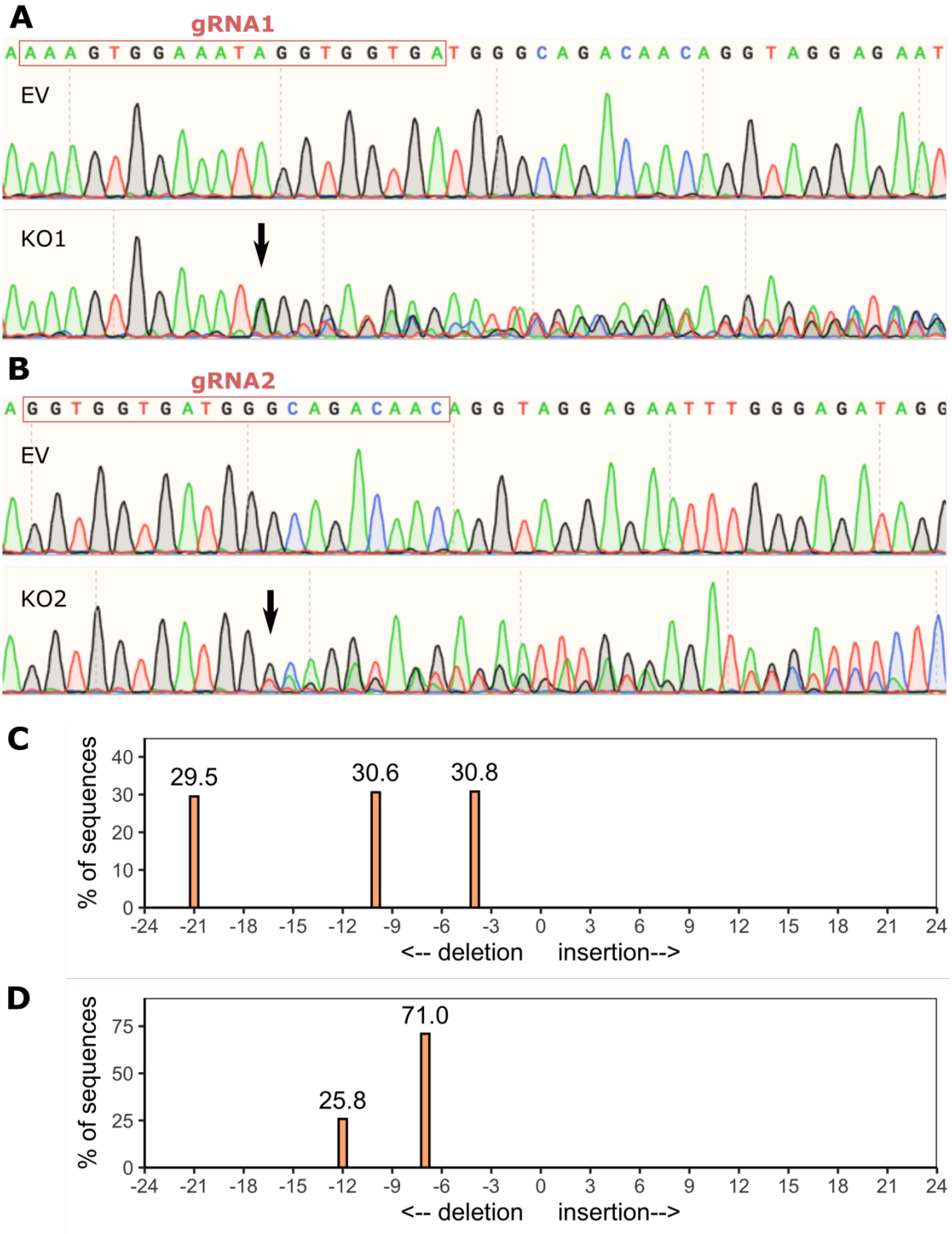

Figure 4.4. Analysis of GALC gene editing

A. Sanger sequencing chromatogram of GALC KO1 compared with EV1 control. Genomic DNA was extracted from cells and the DNA region corresponding to GALC was amplified by PCR prior to Sanger sequencing. The DNA sequence targeted by gRNA1 is highlighted. The arrow indicates the site where the sequence begins to differ from WT. B. Sanger sequencing chromatogram of GALC KO2 compared with EV1 control. The DNA sequence targeted by gRNA2 is highlighted. The arrow indicates the site where the sequence begins to differ from WT. C. TIDE analysis of gene editing across the GALC guide site 1 in GALC KO1 cells. D. TIDE analysis of gene editing across the GALC guide site 2 in GALC KO2 cells.

It is possible that in-frame deletions in one GALC allele in each clone still permit synthesis of small amounts of a deletion mutant GALC. However, these deletions affect a buried loop containing the residue Thr93 contributing to the substrate-binding pocket of GALC[241]. Therefore, these mutations are likely to result in GALC instability and misfolding, resulting in its degradation, and crucially, dysfunction of GALC due to substantial changes to the substrate binding site. I verified that GALC activity was lost in GALC KO cell lines using the highly sensitive GALC activity assays described above. GALC activity was virtually abolished in both GALC KO clones (Fig. 4.5). Residual GALC activity below 10% is a good mimic of the disease state and thus is likely to result in accumulation of GALC substrates[235]. I noted that GALC activity was 2 to 3-fold lower in the Cas9 THP-1 cell line and in the EV1 control cell line than in WT THP-1 cells.

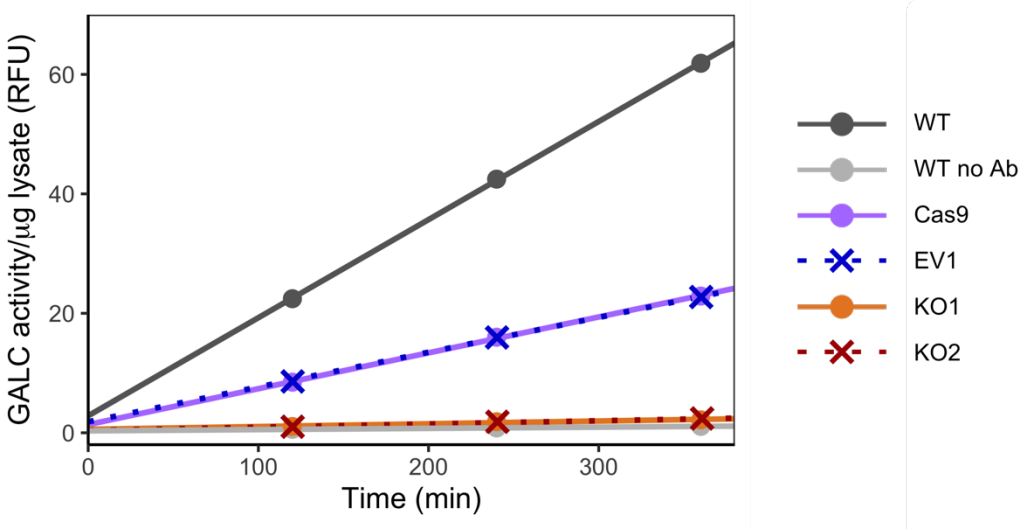

Figure 4.5. GALC KO validation with a GALC activity assay
A. GALC activity assay in WT, Cas9, EV1 and GALC KO1 and KO2 cell lysates. GALC was immunoprecipitated from cell lysates and incubated with the fluorogenic substrate 4MβDG. GALC activity is reported as the change of fluorescence in relative fluorescence units (RFU) over time. Background activity was measured in a control sample where no GALC antibody was added for the immunoprecipitation step ("No Ab"). This data is representative of three independent experiments.

4.3 Investigating the phenotype of GALC KO THP-1 cells

Due to the role of glycosphingolipids in plasma membrane integrity, I next investigated whether loss of GALC leads to notable changes in cell viability and morphology in THP-1 cells. Regular monitoring of cell density and viability by trypan blue staining followed by cell counting indicated that loss of GALC did not alter the growth rate nor viability of THP-1 cells.

Previous studies have found that mimicking Krabbe disease by treating cells with the GALC substrate psychosine induces multinucleation in several cell types[242,243]. I monitored changes to nuclear morphology by Hoechst 33342 staining of GALC KO THP-1 cells, but no multinucleation was observed in undifferentiated GALC KO cells (Fig. 4.6). This suggests that loss of GALC in THP-1 cells is not sufficient to trigger this phenotype. This effect may be cell type dependent, or triggered by additional factors, such as an inflammatory environment or the phagocytosis of apoptotic psychosine-ridden oligodendrocytes.

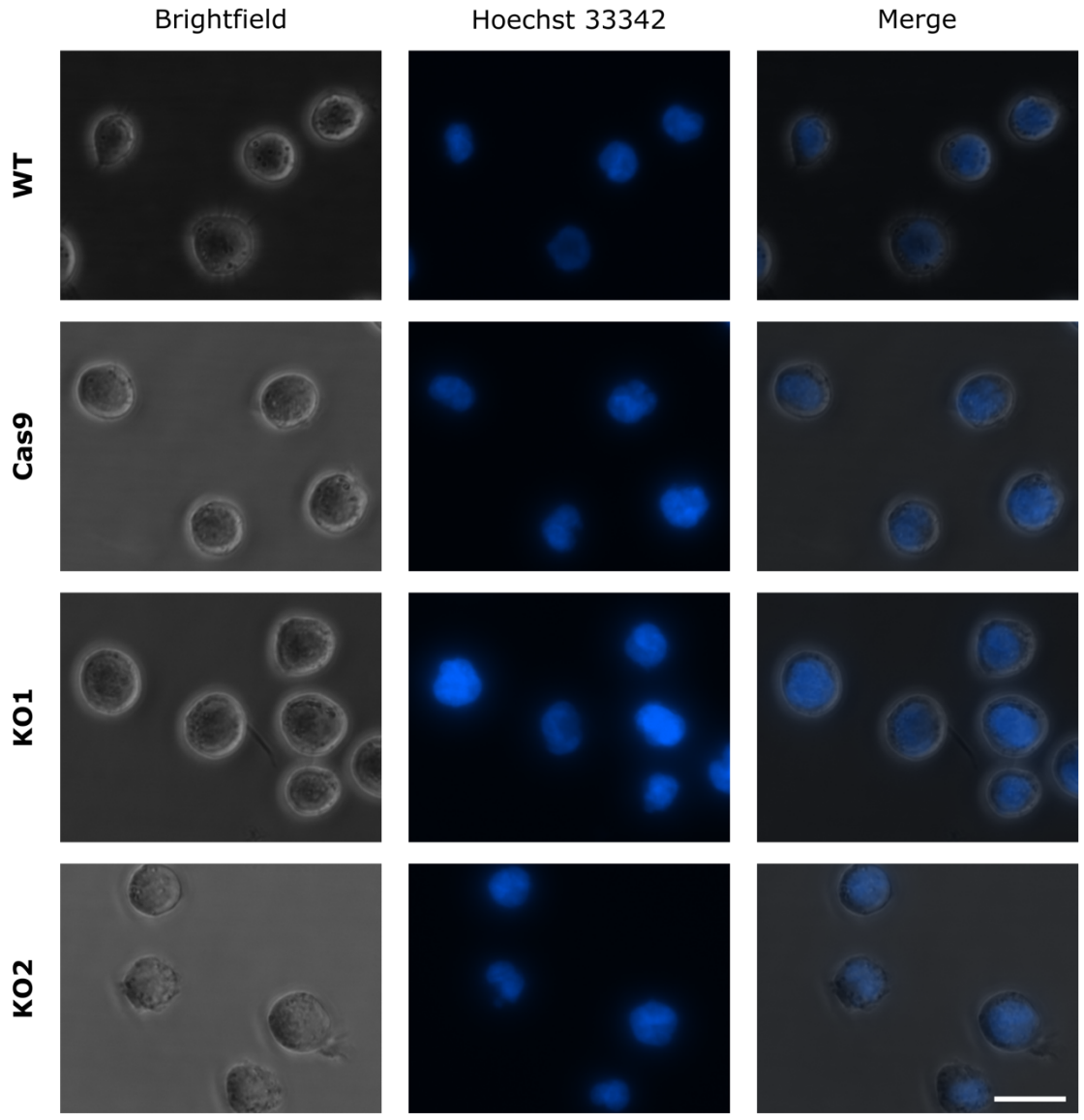

Figure 4.6. Loss of GALC does not trigger multinucleation of THP-1 cells.
A. Undifferentiated WT, Cas9, KO1 and KO2 THP-1 cells in suspension were stained with the nuclear stain Hoechst 33342 and visualised by brightfield and fluorescence microscopy. Scale bar: 20 µm.

A key function of phagocytic cells is the orchestration of inflammation and immune responses through the secretion of cytokines. To explore whether loss of GALC causes significant changes to the cytokines secreted by THP-1 cells, I used an antibody-based cytokine array to analyse the levels of 80 cytokines in the media of EV1, KO1 and KO2 cells (Fig. 4.7A-C). No cytokine had an abundance altered more than two-fold in both knockout clones. This indicates that loss of GALC does not lead to substantial loss or overexpression of cytokines in unstimulated THP-1 cells.

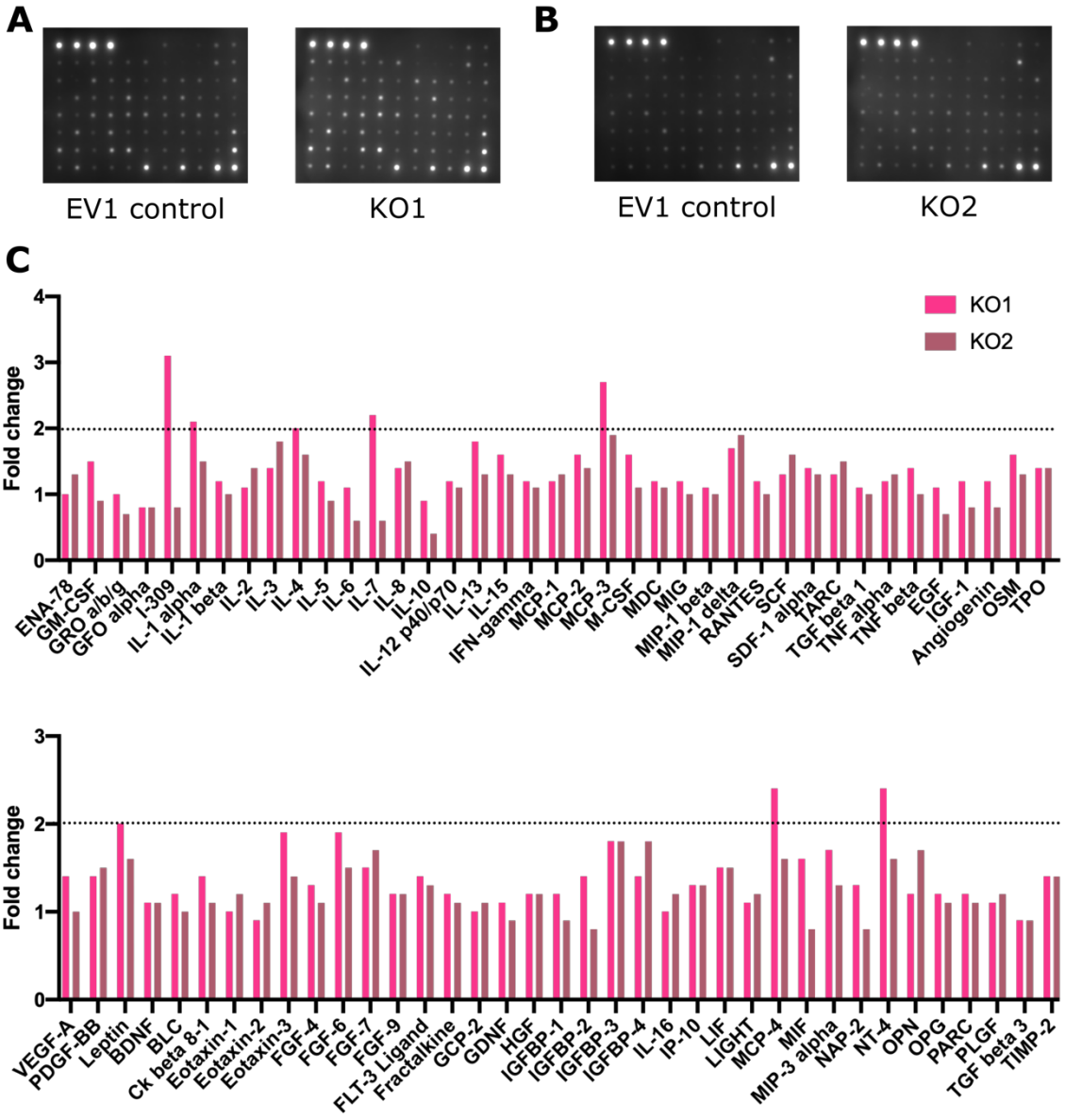

Figure 4.7. Quantification of cytokines secreted by GALC KO THP-1 cells
A. Antibody-based cytokine arrays incubated with media from EV1 and KO1 cells. B. Antibody-based cytokine arrays incubated with media from EV1 and KO2 cells. C. Relative quantification of cytokines secreted by GALC KO cells normalized to EV1. Fold change in KO1 cells is represented in pink, and fold-change in KO2 cells in purple. The raw intensity data from the cytokine array was used for this quantification. The dashed lines correspond to the chosen threshold of a fold change > 2.

4.4 Investigating protein changes at the plasma membrane

4.4.1 Plasma membrane profiling and whole cell proteomics

The dependence of immune receptors on their lipid environment, combined with the immune dysregulation seen in Krabbe disease patients, prompted the hypothesis that loss of GALC and subsequent accumulation of galactosphingolipids alters the proteins present at the plasma membrane through changes in protein stability, trafficking and/or gene expression. I used complementary unbiased approaches to detect such changes in protein abundance at the plasma membrane.

I performed plasma membrane profiling (PMP) in collaboration with Dr James Williamson. PMP allows the relative quantification of plasma membrane proteins by mass spectrometry. Proteins on the surface of intact cells are enriched by biotinylation of surface sialic acid residues followed by cell lysis and binding to streptavidin beads. After thorough washing of the beads, plasma membrane proteins are subjected to tryptic digestion, and peptides from individual samples are labelled with a unique Tandem mass tag (TMT) that allows relative quantification of each protein by mass spectrometry (Fig. 4.8). The proteome on the surface of GALC KO1 and KO2 cells was compared to EV1 control cells. Mass spectrometry was simultaneously performed on 3 biological replicates for each cell line. Each sample was harvested at the same cell density of $0.5x10^6$ cells/mL to avoid confounding effects caused by differences in cell density. Indeed, THP-1 cells have been shown to differentiate into distinct populations with different properties depending on cell density[244]. All the experiments that follow have been performed at the same cell density unless otherwise indicated.

PMP showed large and significant changes in the abundance of many proteins on the cell surface of GALC KO cells compared to EV1 control (Fig. 4.9, Appendix II). Changes in protein abundance at the plasma membrane may be caused by several mechanisms differentially affecting each protein. Changes in the lipid environment may drive changes in transmembrane protein folding and stability. Lipid imbalances may also lead to changes in gene expression. These two classes of mechanisms would affect protein levels in the whole cell. Alternatively, overall protein levels may be unchanged in the cell, but hindered trafficking to the plasma membrane or aberrant retention at the cell surface may lead to localised changes in protein abundance.

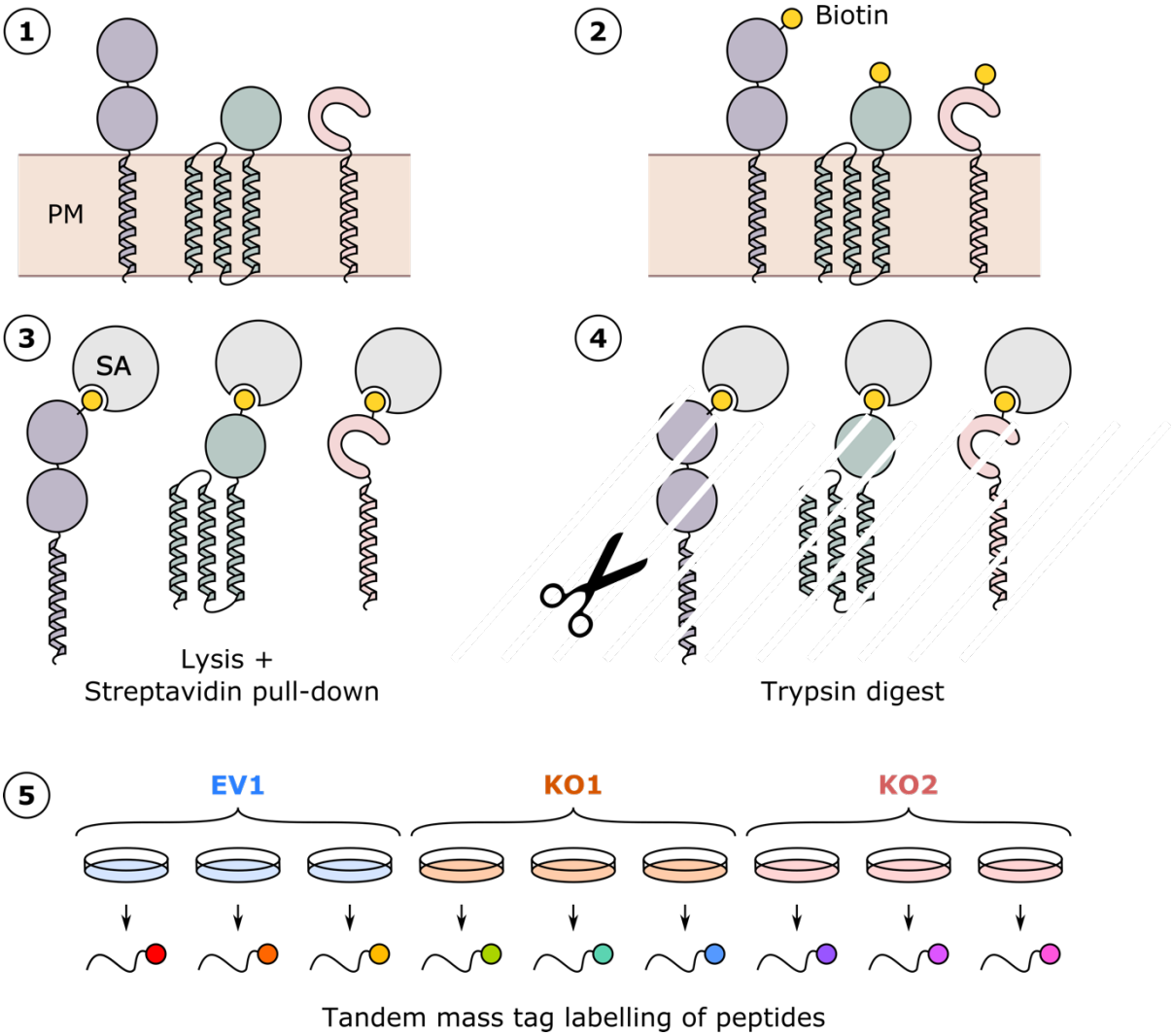

Figure 4.8. Graphical summary of the plasma membrane profiling protocol
Plasma membrane (PM) proteins were labelled with biotin on the surface of intact cells and pulled down by streptavidin (SA) beads after cell lysis. Plasma membrane proteins were then washed and digested with trypsin. Peptides from each sample (3 biological replicates per cell line) were then labelled with a unique Tandem mass tag for quantitative mass spectrometry.

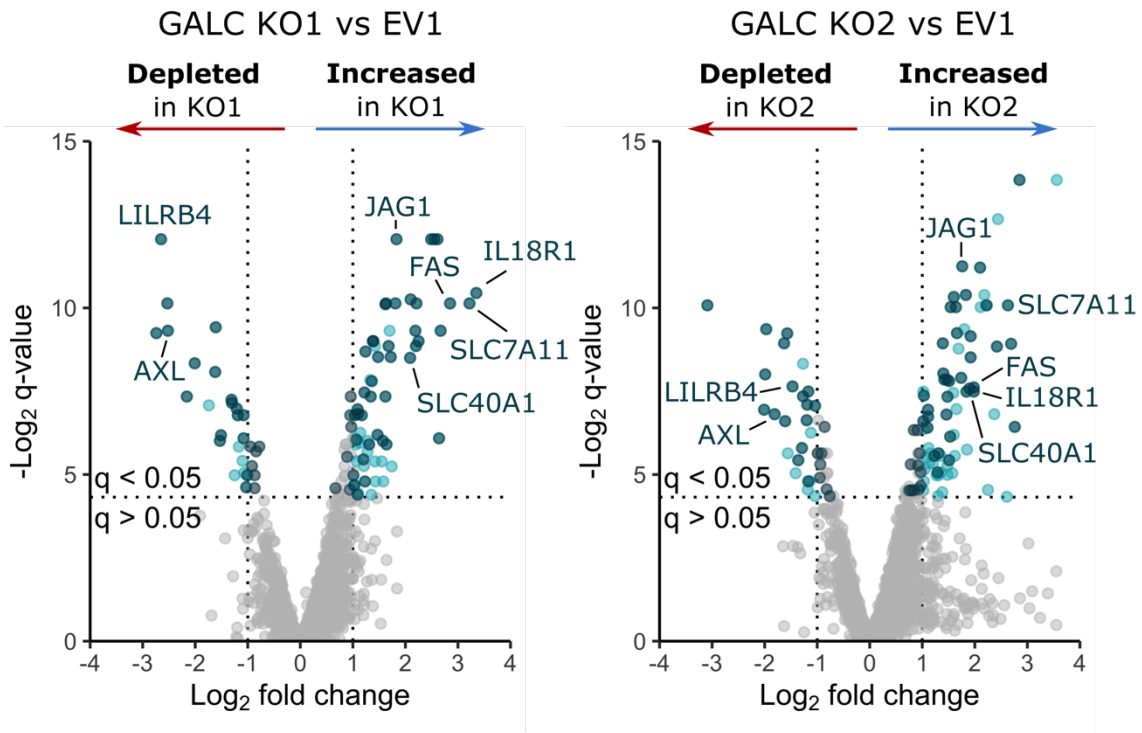

Figure 4.9. Plasma membrane profiling of GALC KO cells
A. Volcano plots showing pairwise comparison between the proteome at the surface of GALC KO cells and EV1 control THP-1 cells. Each dot represents a single protein. Proteins significantly changed ($q < 0.05$) more than two-fold are coloured in blue. Proteins significantly changed in the same direction in both GALC KO1 and KO2 are coloured in dark blue.

To discriminate between the potential mechanisms driving changes in protein abundance at the plasma membrane, I investigated cellular changes in protein levels by proteomics on whole cell lysates (WCP) in collaboration with Dr James Williamson in parallel with PMP. Both experiments were performed on the same cell samples. This permits direct comparison between the plasma membrane and whole cell proteomes, to distinguish between global changes caused by altered protein stability or gene expression and changes caused by protein mistrafficking and mislocalisation. A large number of significant changes in protein abundance in whole cell lysates of GALC KO cells compared to EV1 control were observed (Fig. 4.10A and Appendix III). These changes are well-correlated with changes observed at the plasma membrane. The change in abundance of proteins observed in both datasets (averaged over both knockout clones) tends to be in the same direction at the plasma membrane and in the whole cell (Fig. 4.10B). These changes are particularly consistent for proteins significantly changed in both datasets. In particular, *SLC7A11* is highly upregulated both at the plasma membrane and in the whole cell. This suggests that for the majority of proteins, changes in abundance at the plasma membrane is caused by differences in protein stability or gene expression. Off-diagonal data points correspond to proteins that are enriched or depleted in one dataset, but unchanged or anticorrelated in the other dataset. Such points are of particular interest because they may correspond

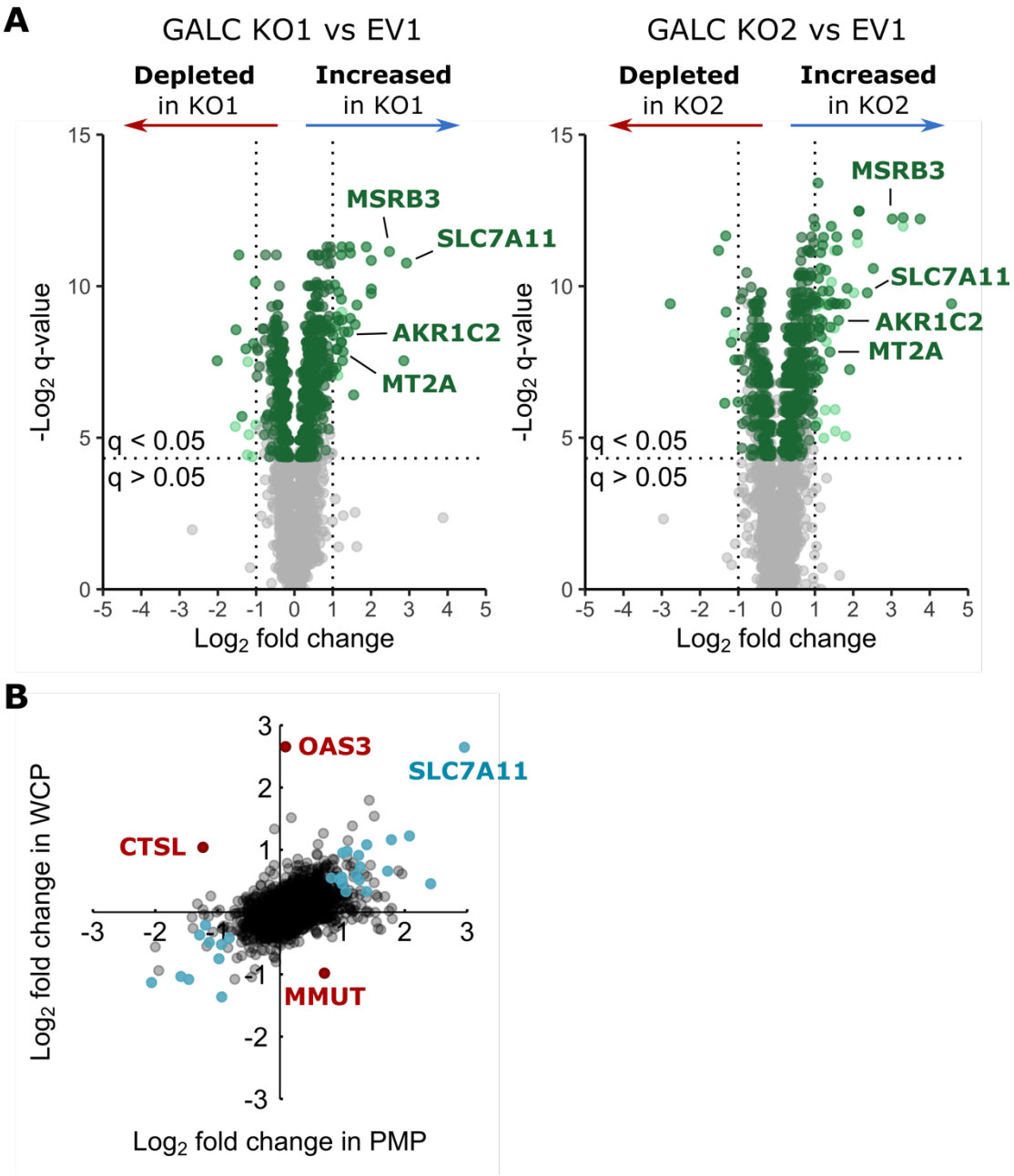

Figure 4.10. Whole cell proteomics of GALC KO cells

A. Volcano plots showing pairwise comparison between the cellular proteome of GALC KO cells and EV1 control THP-1 cells. Each dot represents a single protein. Proteins significantly changed ($q < 0.05$) more than two-fold are coloured in green. Proteins significantly changed in the same direction in both GALC KO1 and KO2 are coloured in dark green. B. Fold change in protein abundance at the plasma membrane vs. the whole cell. Fold change was calculated as the average fold change in KO1 and KO2. Proteins significantly changed ($q < 0.05$) in both clones at the plasma membrane and in the whole cell are coloured in blue.

to mistrafficked proteins. However, the most extreme off-diagonal points corresponding to OAS3, MMUT and CTSL (Fig. 4.10B) appear to represent unreliable measurements rather than mistrafficking. These three proteins are respectively nuclear, mitochondrial and lysosomal and are therefore not enriched in the PMP experiment. Only a single peptide has been detected for each of these proteins by plasma membrane profiling. Therefore, their quantification is unreliable. Thus, no protein mistrafficking or mislocalisation events were identified in GALC-deficient THP-1 cells.

Loss of GALC appears to cause changes in protein stability or gene expression rather than protein localisation in THP-1 monocytes. I investigated mechanisms relevant to Krabbe disease pathology potentially driving these changes.

4.4.2 Biologically relevant changes in protein abundance in GALC KO cells

Krabbe disease is accompanied by elevated TNF-α levels[117]. Interestingly, several TNF-α-responsive genes are up-regulated in GALC KO THP-1 cells: the interleukin-18 receptor IL18R1[245], the death receptor FAS[245], the Notch ligand JAG1[246] and the cystine/glutamate antiporter SLC7A11[247](Fig. 4.9). I hypothesised that loss of GALC leads to aberrant TNF-α signalling in monocytes. This may be caused by down-regulation of the receptors LILRB4 and AXL (Fig. 4.9), two negative regulators of TNF-α expression[248,249]. This hypothesis is intriguing because TNF-α can directly cause oligodendrocyte death[119,120]. Both the transmembrane precursor form of TNF-α and cleaved soluble TNF-α are active[250]. However, GALC KO cells did not secrete more TNF-α than EV1 cells (Fig. 4.11A and 4.7) and precursor TNF-α was not detected by Western blot in EV1 or GALC KO cells using two different antibodies. TNF-α was also not detected in proteomics experiments. Surprisingly, *TNF* gene expression appeared slightly down-regulated in GALC KO cells compared to EV control (Fig. 4.11B). Therefore, I did not find evidence that loss of GALC causes aberrant TNF-α signalling in monocytes.

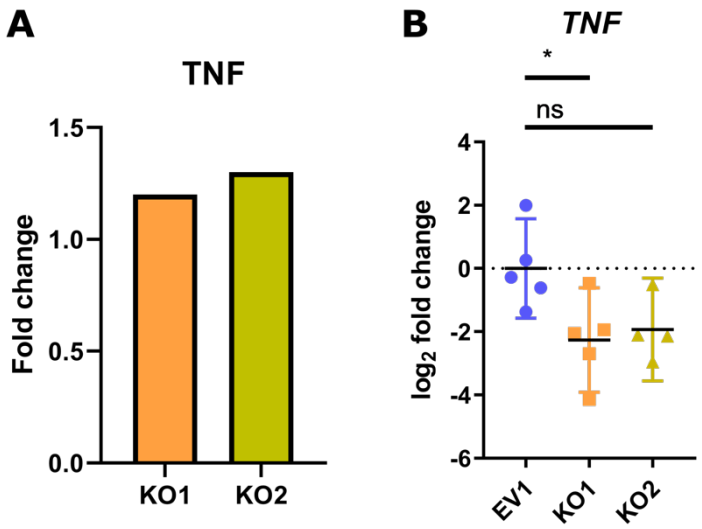

Figure 4.11. TNF is not up-regulated in GALC KO cells
A. Relative quantification of TNF-α secreted by GALC KO cells normalized to EV1, determined in the experiment shown in Figure 4.7. B. qPCR quantification of *TNF* gene expression relative to *ACTB* in GALC KO1 and KO2 cells normalised to EV1 control cells. p-values were calculated by one-way ANOVA followed by Tukey's multiple comparison test. Error bars represent the 95% confidence interval. ns: not significant (p-value > 0.05); *: p ≤ 0.05.

Krabbe disease is associated with oxidative stress and production of reactive oxygen species (ROS)[214,251-254]. Interestingly, several oxidative stress response genes are up-regulated in GALC KO THP-1 cells. At the plasma membrane, SLC7A11 imports cystine essential for synthesis of the antioxidant glutathione[255] and the iron exporter SLC40A1 protects cells from oxidative stress by maintaining iron homeostasis[256] (Fig. 4.9). In the cytoplasm, the antioxidant repair enzyme methionine-R-sulfoxide reductase B3 (MSRB3) reduces methionine sulfoxide to methionine[257], metallothionein-2A (MT2A) acts as a free radical scavenger[258], and the oxidoreductase AKR1C2 reduces reactive carbonyl species[259] (Fig. 4.10A). The expression of several of these genes can be regulated by the antioxidant transcriptional master regulator Nrf2[260-262]. Under non-stressed conditions, Nrf2 is repressed through sequestration in the cytoplasm and proteasomal degradation. In the context of oxidative stress, Nrf2 accumulates and translocates to the nucleus[263]. Proteomics on whole cell lysates showed that Nrf2 abundance is unchanged in GALC KO cells compared to EV1 control cells, indicating that upregulation of proteins involved in the oxidative stress response occurs through a different mechanism. I tested whether GALC KO THP-1 cells have increased levels of intracellular ROS with the fluorogenic dye 2',7'-dichlorodihydrofluorescein diacetate (H_2-DCF-DA)[251,252]. No differences in ROS levels were found between GALC KO and EV1 control cells using this assay (Fig. 4.12). It is possible that the observed upregulation of oxidative stress response proteins maintains redox homeostasis in GALC KO cells.

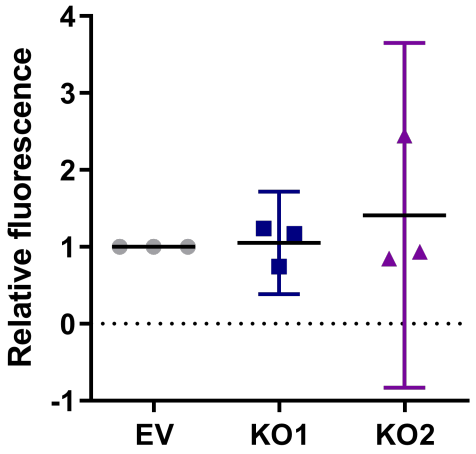

Figure 4.12. GALC KO cells do not have increased levels of reactive oxygen species
EV1 control, GALC KO1 and KO2 cells were incubated with the fluorogenic ROS reporter dye H_2-DCF-DA. Fluorescence was measured and normalised to EV1 in three independent experiments.

Although the upregulation of TNF-α-responsive proteins and oxidative stress response proteins in GALC-deficient THP-1 monocytes generated intriguing hypotheses linking the changes observed in these cells to Krabbe disease pathology, preliminary experiments did not support these hypotheses.

4.4.3 RNA sequencing

To identify whether changes in protein abundance in GALC KO THP-1 monocytes are caused by changes in gene expression or protein stability, I performed RNA sequencing (RNAseq) of GALC KO and EV1 control cells in collaboration with Prof Ian Brierley and Dr Katy Brown. RNA was extracted from four technical replicate samples for each cell line (EV1, KO1, KO2). Ribosomal RNA constitutes the majority of total RNA preparations and was depleted from RNA preparations to avoid wasting sequencing reads[211]. After enrichment of transcriptome RNA through depletion of ribosomal RNA, transcriptome RNA was fragmented and ligated to adaptors for deep sequencing. Unique indexing barcodes were added to each sample for relative RNA quantification. A large number of genes differentially expressed between GALC KO cells and EV1 control were identified (Fig. 4.13A and Appendix IV). There is strong overlap between differentially expressed proteins identified in PMP, WCP and RNAseq experiments (Fig. 4.13B), indicating that a large proportion of these changes are caused by differential gene expression. To identify the functional characteristics of differentially expressed genes, I performed Gene Ontology (GO) enrichment analysis with the DAVID functional annotation tool[210]. Biological Process GO terms[264,265] over-represented in the lists of significantly down-regulated and up-regulated genes in GALC KO cells versus EV1 control are shown in Table 4.1. Genes involved in viral response were up-regulated, and genes involved in transcription and

translation were down-regulated in GALC KO cells versus EV1 control. These processes are not known to be dysregulated in Krabbe disease. This led to the concern that their dysregulation may be an off-target effect of lentiviral transduction or off-target Cas9 activity in GALC KO or EV1 control cell lines.

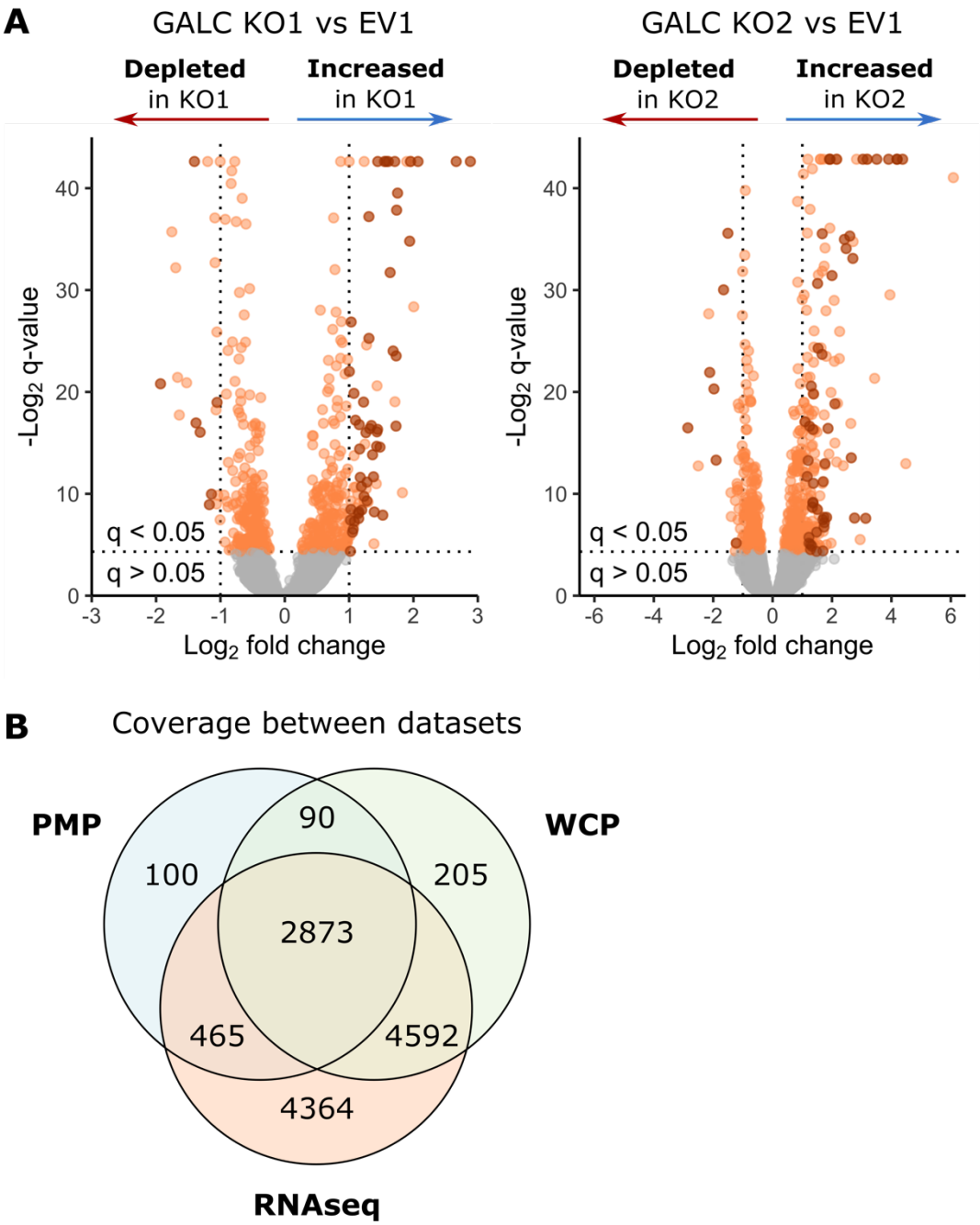

Figure 4.13. RNA sequencing of GALC KO THP-1 cells
A. Volcano plots showing pairwise comparison between the transcriptome of GALC KO cells and EV1 control THP-1 cells. Each dot represents a single gene. Significant gene expression changes ($q < 0.05$) are coloured in orange. Significant gene expression changes larger than 2-fold in both GALC KO1 and KO2 are coloured in brown. B. Venn diagram showing genes quantified in plasma membrane profiling (PMP), whole cell proteomics (WCP) and RNA sequencing (RNAseq) experiments.

Table 4.1. Enrichment analysis of genes differentially expressed in GALC KO cells. GO Biological Process terms enriched in differentially expressed genes in GALC KO1 and KO2 compared with EV1 control, against a background of all genes detected in the RNAseq experiment. Benjamini-Hochberg adjusted q-values are indicated. Significantly enriched Biological process terms (q-value < 0.05) are shown.

Biological processes enriched in up-regulated genes	**Count**	**Fold Enrichment**	**q-value**
defense response to virus	21	9.7	3.01E-11
type I interferon signaling pathway	14	15.1	2.30E-09
negative regulation of viral genome replication	10	15.8	3.04E-06
response to virus	13	8.7	8.13E-06
interferon-gamma-mediated signaling pathway	8	8.3	1.14E-02
Biological processes enriched in down-regulated genes			
translational initiation	44	22.7	6.21E-45
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	42	24.6	1.13E-44
SRP-dependent cotranslational protein targeting to membrane	39	28.6	2.40E-44
viral transcription	40	24.3	2.31E-42
translation	44	12.7	1.53E-33
rRNA processing	40	13.1	8.92E-31
cytoplasmic translation	6	16.8	2.60E-03
translational elongation	5	21.5	6.73E-03
formation of translation preinitiation complex	5	14.7	2.95E-02

4.4.4 Validation of gene expression changes by qPCR

Due to technical limitations on the number of samples that can be processed in parallel with proteomics and RNAseq approaches, a single control cell line was used in the experiments described above. The clonal cell line EV1 derived from the mixed population of EV gRNA-transduced Cas9-expressing THP-1 cells was chosen as a control for these experiments, because this cell line had been manipulated and grown in the same way as the GALC KO cell lines and therefore appeared to be the most appropriate control. However, reliance on a single clonal control cell line comes with the risk that differences in protein expression attributed to loss of GALC in the knockout cell line are in fact caused by abnormalities in the EV1 control cell line. Validation experiments testing the expression of shortlisted proteins in all parental control cell lines in addition to EV1, KO1 and KO2 cells were therefore required. Since a large proportion of changes observed in GALC KO cell lines were driven by gene expression changes, qPCR was used to validate these results. WT THP-1 cells, Cas9 THP-1 cells, and the mixed population of cells transduced with EV gRNA (mixed population EV) from which the EV1

clone was isolated were included in validation experiments. The choice of a housekeeping control gene that is stably expressed across samples is essential for the accuracy and reliability of qPCR experiments. The expression of several housekeeping genes was tested across WT, EV1, KO1 and KO2 lines using the GeNorm algorithm[266] (Fig. 4.14), and *ACTB* was found to be the most stable gene across these cell lines. *ACTB* was therefore used as a housekeeping gene for the experiments described in the rest of this chapter.

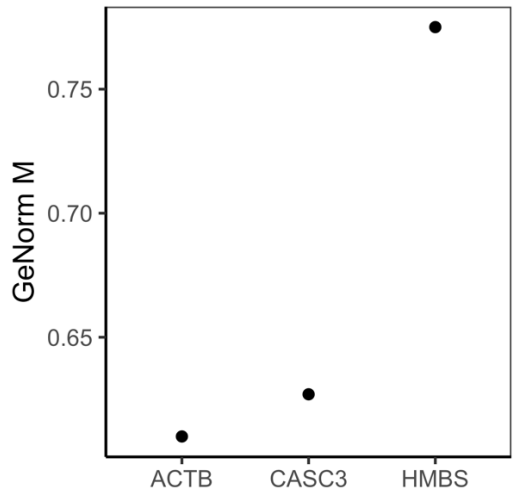

Figure 4.14. Average expression stability of housekeeping genes
The relative stability of several housekeeping genes (*ACTB*, *CASC3* and *HMBS*) was assessed by qPCR in WT, EV1, KO1 and KO2 cells and quantified by GeNorm analysis. Low GeNorm M values indicate higher stability of gene expression across cell lines.

The relative expression of several proteins shortlisted in section 4.4.2 (*SLC7A11*, *FAS*, *AXL*, *LILRB4* and *MSRB3*; Fig. 4.9A and 4.10A) and *TNF* was measured by qPCR across all cell lines. Strikingly, *SLC7A11* was strongly down-regulated in the clonal EV1 control cell line but expressed at similar levels in mixed population EV and Cas9 control cells and KO1 and KO2 cells (Fig. 4.15A). Thus, *SLC7A11* is not up-regulated in KO cells due to loss of GALC, but its expression is abnormally low in EV1 control cells. The same pattern is observed for *FAS* and *MSRB3* (Fig. 4.15B-C). *AXL* and *LILRB4* also appear dysregulated in EV1 control cells (Fig. 4.15D-E). *TNF* gene expression was remarkably low in Cas9 THP-1 cells (approximately 60-fold lower than WT) and all derived cell lines (Fig. 4.15F), explaining why TNF-α could not be detected by Western blot in EV1, KO1 and KO2 cells. The expression of several quantified genes differed significantly between Cas9 and WT THP-1 cells, although to a lesser extent than *TNF*. The mixed population of THP-1 cells transduced with Cas9 from which the clonal Cas9 THP-1 cell line was isolated by our collaborators was not available, so I could not identify which point in the process of making this cell line resulted in *TNF* dysregulation. However, lentiviral transduction of WT THP-1 cells with an empty lentiviral vector (pCW57) does not result in *TNF* down-regulation (Fig. 4.16),

indicating that *TNF* down-regulation in the Cas9 cell line is caused by insertional mutagenesis or off-target Cas9 activity in this clonal cell line rather than a process inherent to lentiviral transduction of THP-1 cells.

In summary, there are abnormalities in the EV1 clonal cell line used as a control in all experiments described so far in this chapter. Transduction with EV control gRNA may have introduced some off-target effects amplified by clonal selection. Consequently, many or all of the changes in protein expression detected by proteomics and RNAseq are artefacts. It is clear that mixed population EV or THP-1 Cas9 cells would be more suitable controls for these experiments, because the expression level of all genes quantified by qPCR is consistent across these cell lines. The parental Cas9 THP-1 clonal cell line appears different from WT THP-1 cells, and in particular shows strong down-regulation of *TNF* expression. This may affect the biological relevance of this cell line as a monocyte model because TNF-α production is one of the key roles of monocytes and TNF-α autocrine feedback loops regulate monocyte function[122]. With this caveat in mind, conclusions about fundamental cellular processes such as receptor trafficking and interactions with lipids may still be drawn from this model cell line.

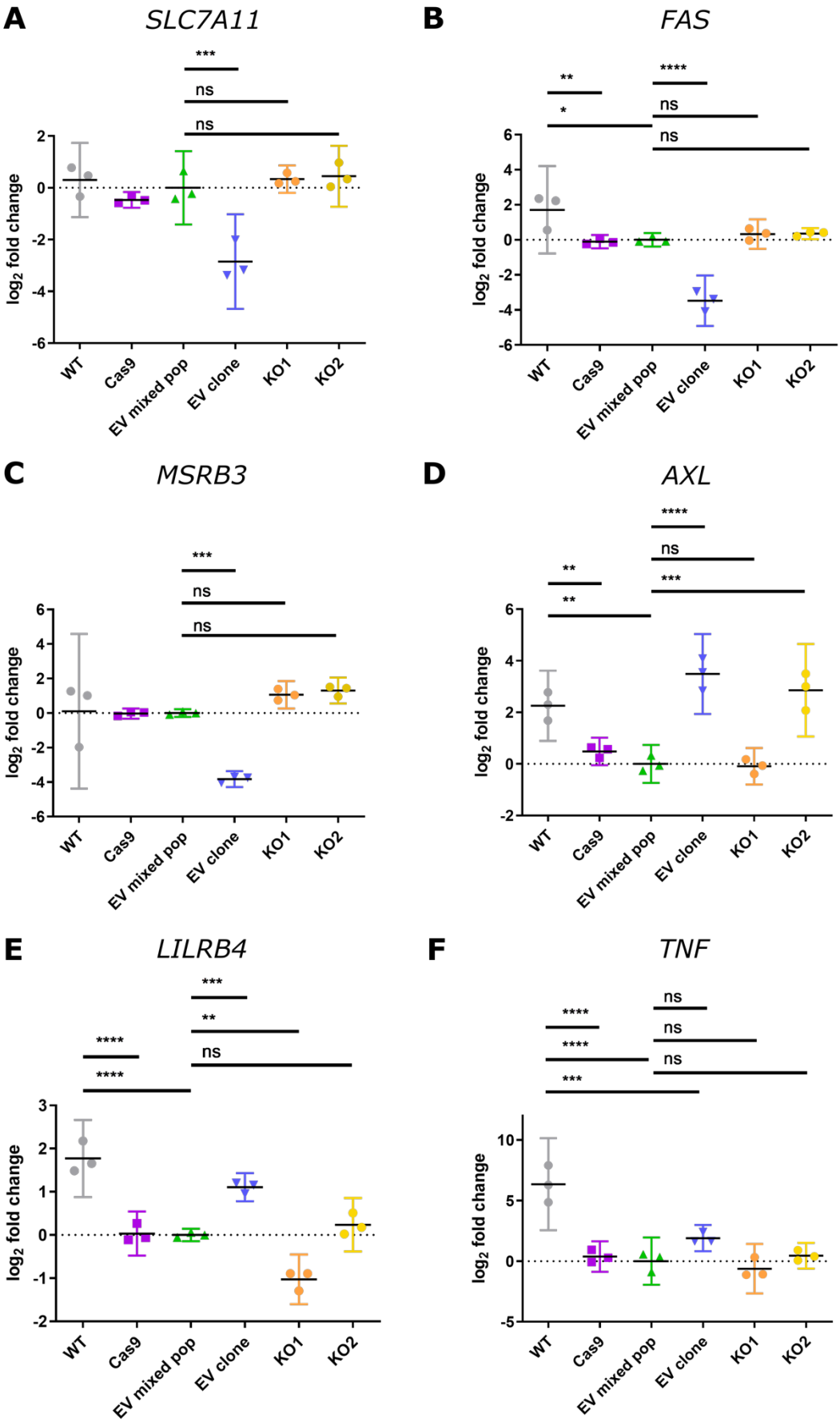

Figure 4.15. Validation of shortlisted genes by qPCR

qPCR quantification of *SLC7A11, FAS, MSRB3, AXL*, *LILRB4* and *TNF* gene expression relative to *ACTB* in WT, Cas9, EV mixed population and GALC KO1 and KO2 cell lines, normalised to EV mixed population control. p-values were calculated by one-way ANOVA followed by Tukey's multiple comparison test. Error bars represent the 95% confidence interval. ns: not significant (p-value > 0.05); *: p ≤ 0.05; **: p ≤ 0.01; ***: p ≤ 0.001; ****: p ≤ 0.0001.

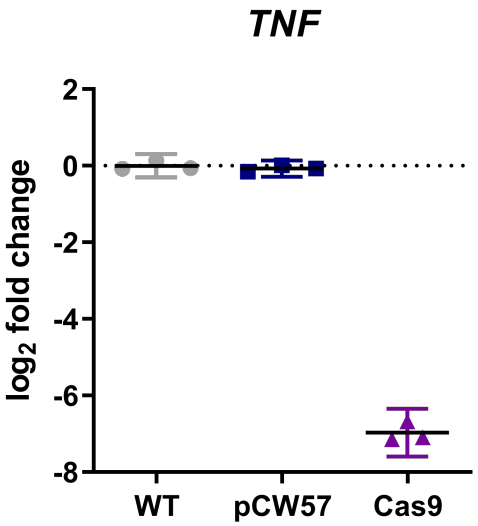

Figure 4.16. *TNF* gene expression is not down-regulated by lentiviral transduction
qPCR quantification of *TNF* gene expression relative to *ACTB* in WT THP-1 cells, a mixed population of THP-1 cells transduced with an empty lentiviral vector (pCW57) and the clonal THP-1 cell line stably expressing Cas9. The results of a single experiment performed with technical triplicates for each sample are shown. Error bars represent the standard deviation.

4.4.5 Repeat of plasma membrane profiling with additional controls

Newly available 16-plex TMT labelling probes for proteomics allowed for a repeat of the PMP and WCP experiments with additional controls. The relative abundance of proteins in WT, Cas9, EV mixed population, GALC KO1 and KO2 cell lines was quantified. The experiment was performed in collaboration with Dr James Williamson as described in section 4.4.1, with independent triplicate samples of each cell line. An additional single replicate of EV1 was also included for reference to the previous proteomics experiments. PMP results were obtained prior to WCP results. First, the plasma membrane proteome of control cell lines was compared to assess the best normalisation control for this experiment. Cas9 THP-1 cells appeared very different from WT cells (Fig. 4.17A). The abundance of 1347 plasma membrane proteins is significantly changed in Cas9 cells. This shows that the introduction of Cas9 by lentiviral transduction of WT THP-1 cells has introduced large and wide-ranging changes to its expression profile, either through insertional mutagenesis or off-target Cas9 activity in this clonal cell line. Consequently, Cas9 or mixed population EV cells are better normalisation controls than WT or clonal EV lines for assessing changes caused specifically by loss of GALC. There was no significant difference in the proteome at the surface of mixed population EV cells

compared to the Cas9 control cell line (Fig. 4.17B), indicating that the mixed population EV cell line is indeed a much better control than the clonal EV1 cell line. The mixed population EV cell line was therefore chosen as the normalisation control. Five proteins were significantly changed ($q < 0.05$) at the plasma membrane of GALC KO1 and KO2 cells compared to mixed population EV control cells (Fig. 4.17C, Table 4.2). The relative abundance of each of these proteins is similar in Cas9 and mixed population EV control cell lines, but different in WT cells for all proteins except PLXNC1 (Fig. 4.18).

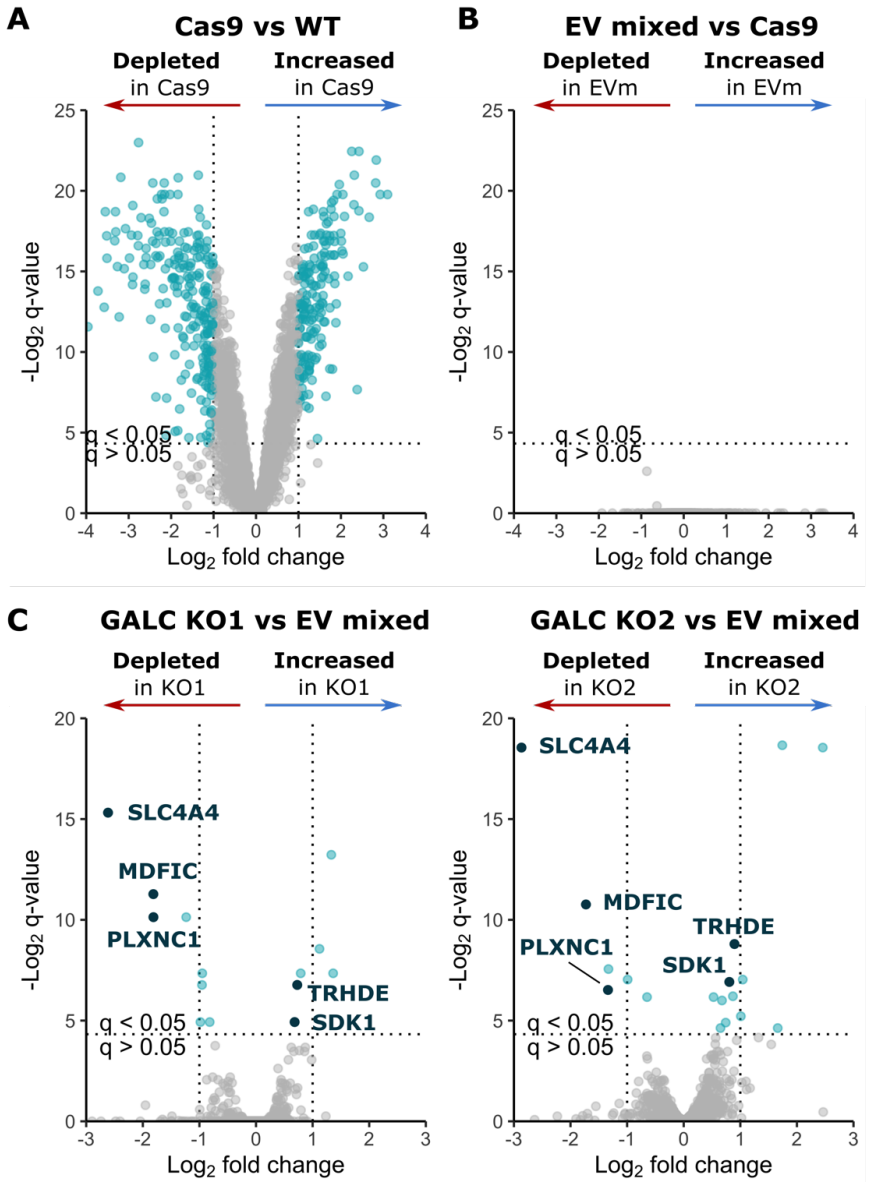

Figure 4.17. Plasma membrane profiling of control and GALC KO cell lines
A. Volcano plot showing pairwise comparison between the proteome at the surface of Cas9 and WT THP-1 cells. Proteins significantly changed ($q < 0.05$) more than 2-fold are coloured in cyan. B. Volcano plot showing pairwise comparison between the proteome at the surface of mixed population EV and Cas9 THP-1 cells. No significant differences were found. C. Volcano plots showing pairwise comparison between the proteome at the surface of GALC KO THP-1 cells and mixed population EV control cells. Each dot represents a single protein. Proteins significantly changed ($q < 0.05$) are coloured in cyan. Proteins significantly changed in both GALC KO1 and KO2 are labelled and coloured in dark blue.

Table 4.2. Proteins significantly changed at the plasma membrane of GALC KO1 and KO2 vs EV1 cells (Benjamini-Hochberg adjusted q-value < 0.05)

Gene ID	Log_2 (fold change KO1/EV1)	q-value	Log_2 (fold change KO2/EV1)	q-value
TRHDE	0.73	9.17E-03	0.90	2.25E-03
SDK1	0.68	3.29E-02	0.81	8.27E-03
PLXNC1	-1.81	8.91E-04	-1.34	1.09E-02
MDFIC	-1.81	4.03E-04	-1.73	5.76E-04
SLC4A4	-2.61	2.45E-05	-2.87	2.61E-06

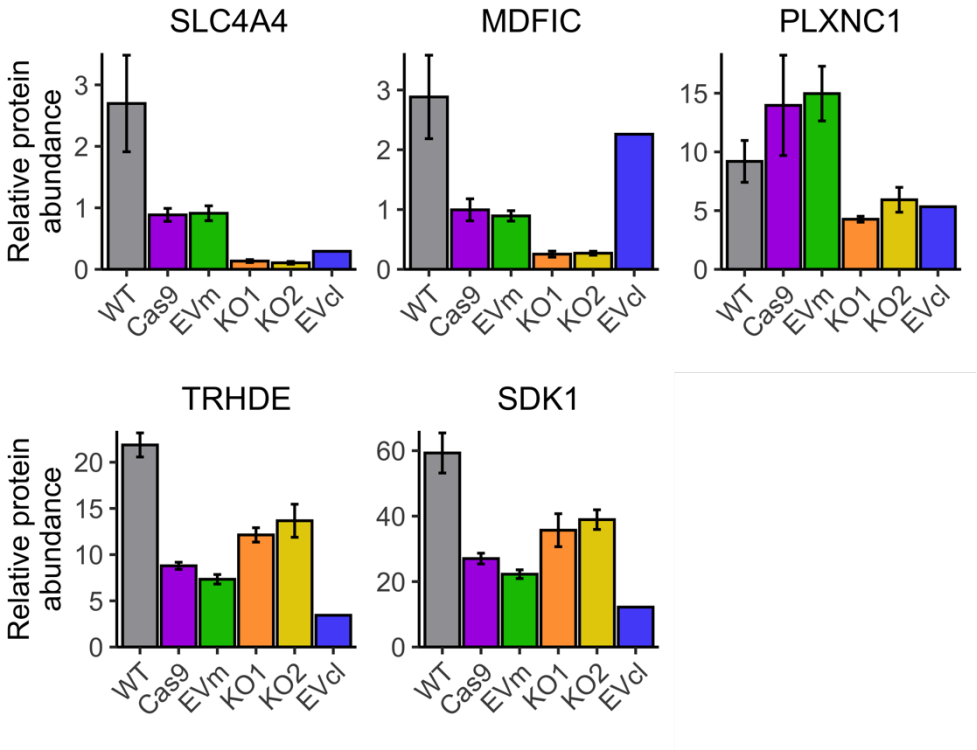

Figure 4.18. Relative abundance of shortlisted proteins in each sample
Relative abundance (shown on a linear scale) of shortlisted proteins quantified in WT, Cas9, EV mixed population (EVm), GALC KO1 and KO2 THP-1 cell lines by plasma membrane profiling-mass spectrometry. Error bars indicate the standard deviation across three independent biological replicates. A single replicate samples of the EV1 clonal cell line (EVcl) was analysed by mass spectrometry alongside these samples.

4.4.6 Validation of plasma membrane profiling results

To validate that the changes observed in GALC KO cells are specifically caused by loss of GALC, I created a GALC rescue cell line by transduction of GALC KO1 and KO2 THP-1 cells with a geneticin-resistant inducible lentiviral vector encoding GALC. Cas9 and gRNAs targeting GALC are stably expressed in the KO cell lines, therefore I introduced silent mutations into the GALC rescue construct sequence to

escape Cas9-mediated cleavage. The simplest way to achieve this is to abolish Cas9 binding by mutating the Protospacer adjacent motif (PAM) sequence (NGG) required for Cas9 binding to DNA, but it was not possible to introduce a silent mutation into the PAM site adjacent to the sequence targeted by gRNA1. After binding of the Cas9-gRNA complex to a PAM site, Cas9 only cleaves DNA if there is sufficient homology between the gRNA spacer sequence and the target DNA sequence. I therefore mutated the target DNA sequence of the GALC rescue construct. The sequence directly upstream of the PAM site has been found to be most sensitive to mismatches with the target sequence[267], therefore four mutations were introduced into this region (Fig. 4.19A). Some nucleotide pair mismatches have been reported to be more costly than others for Cas9 activity[268] and these nucleotide pair mismatches were introduced where possible. The PAM site adjacent to the gRNA2 target site is part of an intron. The recombinant DNA sequence encoding rescue GALC does not include introns and the PAM site adjacent to the gRNA2 target site is therefore lost (Fig. 4.19B). Although this is sufficient to abolish Cas9 activity at this site, silent mutations disrupting gRNA2 binding were also introduced (Fig. 4.19B). Site-directed mutagenesis of GALC rescue vectors proved challenging due to the large size of the template GALC vector. A QuikChange mutagenesis protocol modified for large plasmids[269] was attempted without success. Mutations were successfully introduced using a two-step protocol[174] using PCR to generate overlapping GALC gene segments containing the mutations of interest, which are then used as template for a second PCR to form the full-length mutant product[174] (section 2.2.7).

GALC KO1 and KO2 cells were transduced with the GALC gRNA1 escape mutant (Rescue 1) and gRNA2 escape mutant (Rescue 2) constructs, respectively. Cells were infected with a concentration of lentivirus vectors such that only 5-10% of cells were transduced in order to minimise the number of lentiviral insertion events per cell. Successfully transduced cells were selected by antibiotic selection. To avoid amplification of insertional mutations by clonal selection, both cell lines were maintained as a mixed population. GALC expression was induced by addition of various concentrations of doxycycline to cell cultures for 6 days prior to GALC activity measurement. GALC activity was rescued in the KO1 rescue cell line (KO1R; Fig. 4.20A) and the KO2 rescue cell line (KO2R; Fig. 4.20B). Cells induced with the lowest concentration of doxycycline tested (30 μg/L for KO1R and 25 μg/L for KO2R) showed partial rescue of GALC activity (Fig. 4.20C) and were chosen for subsequent experiments because partial rescue of GALC activity is likely to be sufficient to restore normal sphingolipid metabolism[235], whereas GALC overexpression may introduce lipid imbalances. Importantly, although average GALC activity in rescue cell populations appears to be similar to endogenous levels, GALC expression is expected to be heterogeneous across individual cells because transgene expression depends on

whether it integrates into a transcriptionally active or repressed region, a phenomenon known as position-effect variegation[270].

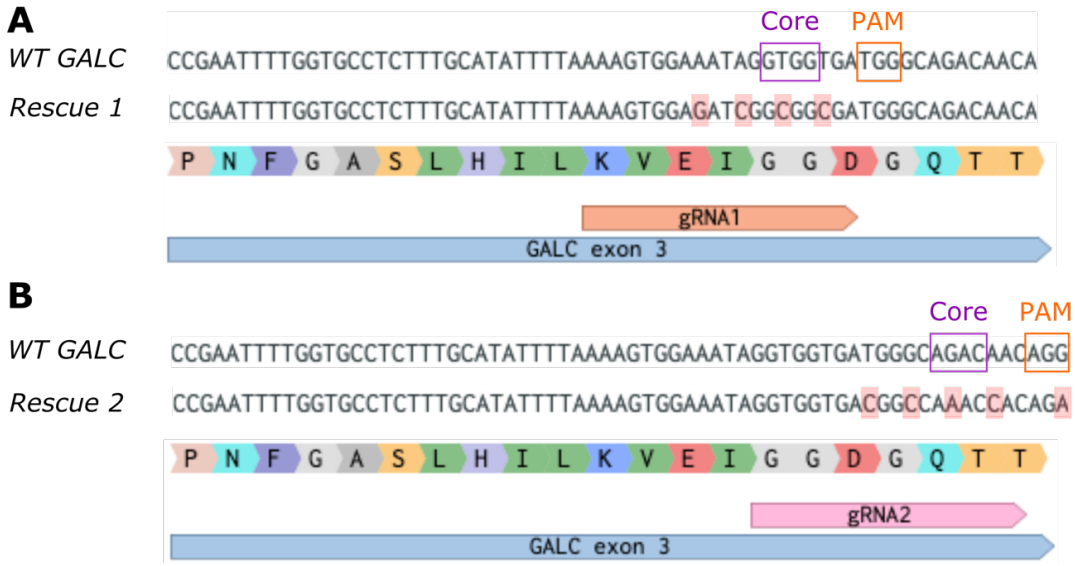

Figure 4.19. GALC rescue construct
A. Recombinant GALC rescue 1 construct transduced into the GALC KO1 cell line. Silent mutations (highlighted in red) were introduced in the gRNA1 target site to prevent genetic ablation of the rescue construct by Cas9 and gRNA1 stably expressed in this cell line. The PAM site recognised by Cas9 is highlighted in orange. The core gRNA sequence most sensitive to nucleotide pair mismatches is indicated in purple. B. Recombinant GALC rescue 2 construct transduced into the GALC KO2 cell line, annotated as A. In addition to the four silent mutations introduced in the gRNA2 target site, the adjacent PAM site (NGG) is lost in the recombinant GALC construct because this site is included in the intron of endogenous GALC.

Relative gene expression of proteins changed at the plasma membrane of GALC KO cells was assessed by qPCR in WT, Cas9, mixed population EV, GALC KO1, KO2, and doxycycline-induced rescue KO1R and KO2R cell lines and normalised to EV mixed population (Fig. 4.21). The changes observed at the plasma membrane by PMP (Fig. 4.18) are replicated in this experiment, providing independent validation of these results and showing that each of these changes is driven by differential gene expression. Rescue of GALC expression rescued *SLC4A4* (Fig. 4.21A) and *MDFIC* (Fig. 4.21B) gene expression to a limited extent but did not rescue *PLXNC1*, *TRHDE* and *SDK1* gene expression. These results suggest that these changes may not be directly caused by loss of GALC.

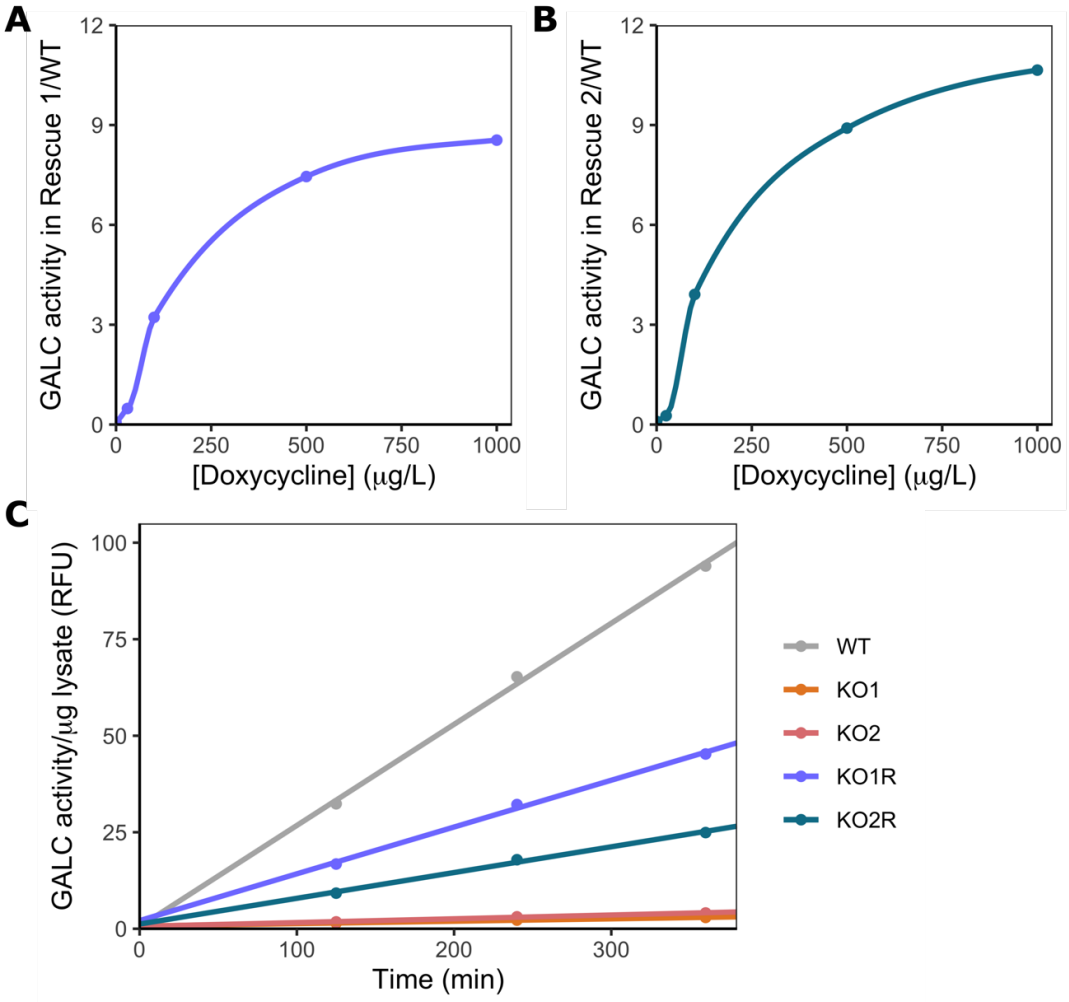

Figure 4.20. Rescue of GALC activity

GALC expression was induced in the KO1R and KO2R rescue cell lines by addition of various concentrations of doxycycline. GALC activity was assessed by activity assays after 6 days of doxycycline induction. A. Ratio of GALC activity in KO1R cells induced with various concentrations of doxycycline to GALC activity in WT THP-1 cells. B. Ratio of GALC activity in KO2R cells induced with various concentrations of doxycycline to GALC activity in WT cells. C. GALC activity in WT and GALC knockout THP-1 cells (KO1 and KO2) compared with KO1R cells induced with 30 μg/L doxycycline and KO2R cells induced with 25 μg/L doxycycline.

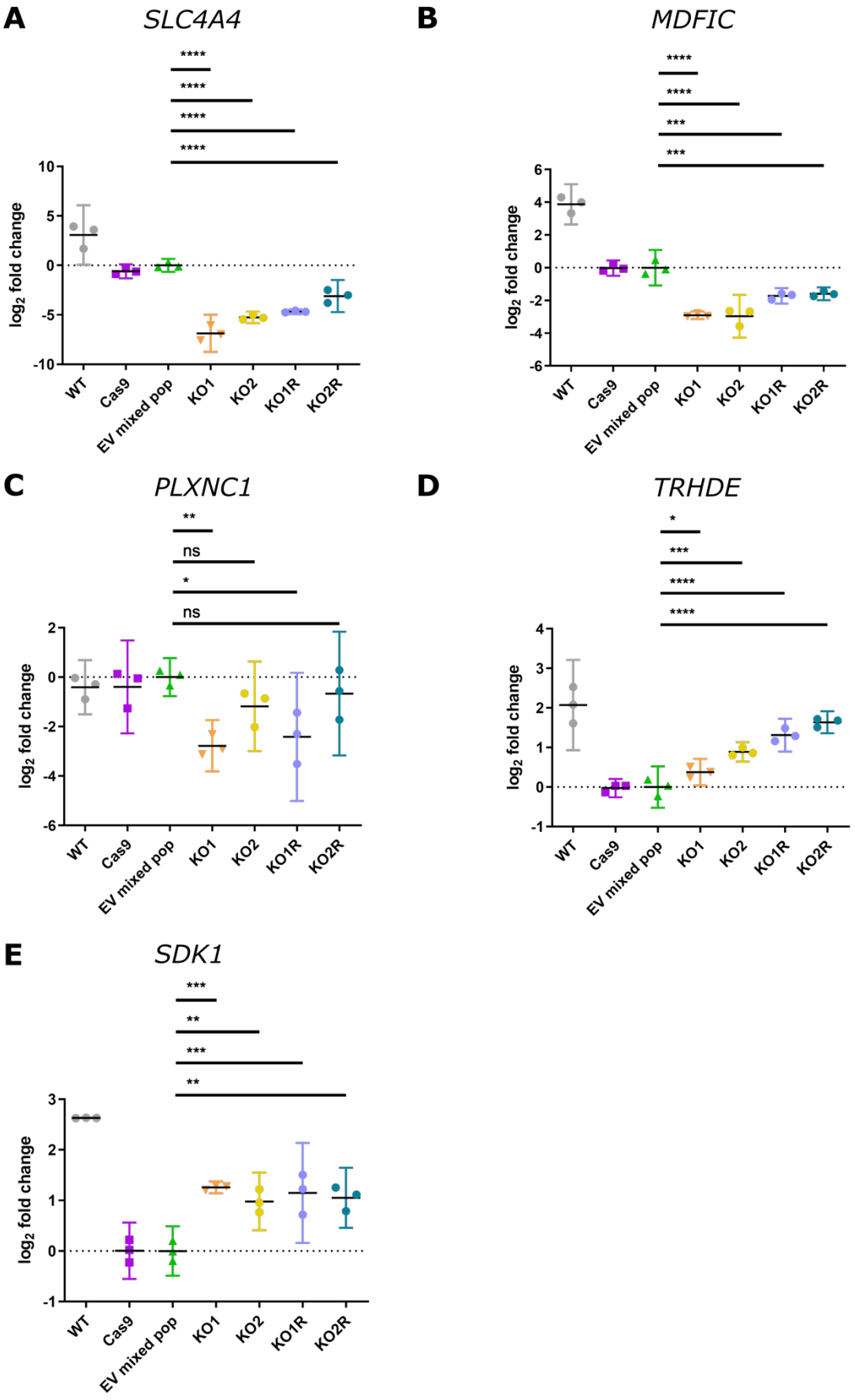

Figure 4.21. Validation of proteins differentially expressed in GALC KO THP-1 cells

qPCR quantification of *SLC4A4, MDFIC, PLXNC1, TRHDE* and *SDK1* gene expression relative to *ACTB* in WT, Cas9, EV mixed population, GALC KO1 and KO2, and GALC rescue KOR1 and KOR2 THP-1 cell lines, normalised to EV mixed population control. p-values were calculated by one-way ANOVA followed by Tukey's multiple comparison test. Error bars represent the 95% confidence interval. ns: not significant (p-value > 0.05); *: p ≤ 0.05; **: p ≤ 0.01; ***: p ≤ 0.001; ****: p ≤ 0.0001.

4.4.7 Repeat of whole cell proteomics with additional controls

WCP analysis of WT, Cas9, EV mixed population, GALC KO1 and KO2 cell lines showed that a striking total of 2680 proteins were significantly up-regulated and 2228 proteins were significantly down-regulated in the Cas9 THP-1 cell line compared with the WT THP-1 cell line (Fig. 4.22A, Appendix V). I identified pathways dysregulated in these cells by pathway enrichment analysis with DAVID[210], focusing on pathways from the KEGG database[271]. Pathways over-represented in the lists of significantly down-regulated and up-regulated genes in Cas9 THP-1 cells versus WT THP-1 cells are shown in Table 4.3. Notably, the DNA mismatch repair pathway is up-regulated, perhaps due to continuing off-target Cas9 activity in these cells. The phagosome and lysosome pathways are most significantly enriched in down-regulated genes. Importantly, many proteins involved in glycosphingolipid metabolism, including GALC (as observed in Figure 4.5), the saposin precursor PSAP and the enzymes immediately upstream and downstream of GALC in the galactosphingolipid degradation pathway (ARSA and AC, respectively; Fig. 1.3) are significantly down-regulated in the Cas9 cell line (Table 4.4). The dysregulation of glycosphingolipid metabolism strongly suggests that this cell line is inappropriate to study the effects of GALC deficiency. Given the variety of dysregulated pathways, it is not possible to identify a single cause for these changes. It is possible that continuous off-target Cas9 activity or a large chromosomal recombination event underlies these changes.

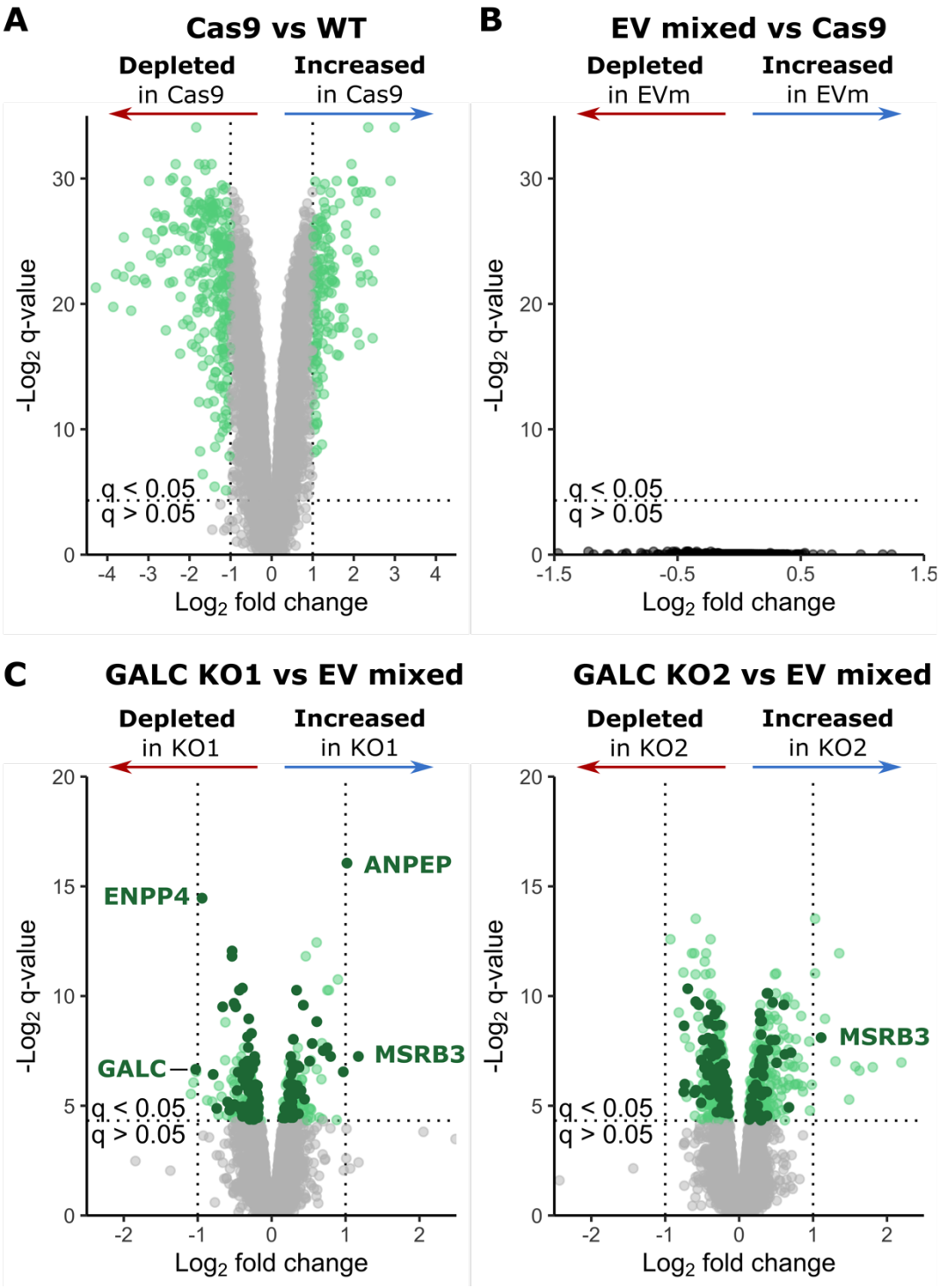

Figure 4.22. Whole cell proteomics of control and GALC KO cell lines

A. Volcano plot showing pairwise comparison between the cellular proteome of Cas9 and WT THP-1 cells. Proteins significantly changed ($q < 0.05$) more than 2-fold are coloured in green. B. Volcano plot showing pairwise comparison between the cellular proteome of mixed population EV and Cas9 THP-1 cells. No significant differences were found. C. Volcano plots showing pairwise comparison between the cellular proteome of GALC KO THP-1 cells and mixed population EV control cells. Each dot represents a single protein. Proteins significantly changed ($q < 0.05$) are coloured in green. Proteins significantly changed in the same direction in both GALC KO1 and KO2 are coloured in dark green.

Table 4.3. Pathways dysregulated in the Cas9 THP-1 cell line. KEGG pathways enriched in lists of significantly up-regulated and down-regulated proteins in Cas9 THP-1 cells compared to WT, against a background of all quantified proteins in the WCP experiment. Benjamini-Hochberg adjusted q-values are indicated. Significantly enriched pathways with a q-value < 0.005 are shown.

Pathways enriched in up-regulated proteins	**Count**	**Fold enrichment**	**q-value**
Leukocyte transendothelial migration	4	7.5	9.92E-01
Pathogenic Escherichia coli infection	3	9.1	9.92E-01
Focal adhesion	4	4.9	9.92E-01
T cell receptor signaling pathway	3	5.7	1.00E+00
Pathways enriched in down-regulated proteins			
Thyroid hormone synthesis	4	10.9	5.27E-01
Morphine addiction	3	11.4	5.37E-01
Bile secretion	3	10.9	5.37E-01
Citrate cycle (TCA cycle)	3	10.1	5.37E-01
Endocrine and other factor-regulated calcium reabsorption	3	9.4	5.37E-01
Retrograde endocannabinoid signaling	3	8.7	5.37E-01
Insulin secretion	3	8.5	5.37E-01
Gastric acid secretion	3	8.2	5.37E-01
Aldosterone synthesis and secretion	3	8.0	5.37E-01
Salivary secretion	3	8.0	5.37E-01
Glutamatergic synapse	3	6.2	7.49E-01

Despite the concerns highlighted above, available WCP data was used to identify proteins significantly enriched in whole cell lysates of GALC KO1 and KO2 cells compared with mixed population EV control cells (Fig. 4.22C and Appendix VI). Although the abundance of 125 proteins was significantly changed in GALC KO cells, the amplitude of these changes was small and only MSRB3 was up-regulated more than 2-fold in both GALC KO clones. Although *MSRB3* gene expression appears slightly higher in KO cells compared to mixed population EV (Fig. 4.15C), this increase is not significant, it is therefore unclear whether MRSB3 up-regulation is caused by changes in gene expression or increased MSRB3 stability in GALC KO cells. Pathway enrichment analysis on proteins significantly up- and down-regulated in GALC KO cells yielded non-significant results (Table 4.5). These pathways were not investigated further due to low confidence in these cell lines and in the pathway enrichment analysis results.

Table 4.4. Glycosphingolipid metabolism proteins down-regulated in Cas9 cells. Relative protein abundance of proteins described by the GO term "glycosphingolipid metabolic process" significantly down-regulated in whole cell lysates of Cas9 versus WT THP-1 cells.

Gene ID	Gene name	Log_2 (fold change Cas9/WT)	q-value
HEXB	Beta-hexosaminidase subunit beta	-1.47	1.04E-08
ARSB	Arylsulfatase B	-1.44	6.22E-09
ARSD	Arylsulfatase D	-1.25	4.63E-08
SUMF1	Formylglycine-generating enzyme	-0.93	1.08E-04
GM2A	Ganglioside GM2 activator	-0.81	8.63E-07
ARSA	Arylsulfatase A	-0.77	4.39E-05
ASAH1	Acid ceramidase	-0.75	1.66E-06
HEXA	Beta-hexosaminidase subunit alpha	-0.64	3.40E-06
GLB1	Beta-galactosidase	-0.62	1.42E-05
GALC	Galactocerebrosidase	-0.61	9.13E-03
PSAP	Prosaposin	-0.53	1.15E-03
GBA	Lysosomal acid glucosylceramidase	-0.53	1.54E-05
CTSA	Lysosomal protective protein	-0.46	4.35E-06
NEU1	Sialidase-1	-0.23	2.27E-02

Table 4.5. Pathway enrichment analysis of genes differentially expressed in GALC KO cells. KEGG pathways enriched in lists of significantly up-regulated and down-regulated proteins in GALC KO1 and KO2 compared to mixed population EV cells, against a background of all quantified proteins in the WCP experiment. Benjamini-Hochberg adjusted q-values are indicated.

Pathways enriched in up-regulated proteins	Count	Fold enrichment	q-value
Leukocyte transendothelial migration	4	7.5	9.92E-01
Pathogenic Escherichia coli infection	3	9.1	9.92E-01
Focal adhesion	4	4.9	9.92E-01
T cell receptor signaling pathway	3	5.7	1.00E+00
Pathways enriched in down-regulated proteins			
Thyroid hormone synthesis	4	10.9	5.27E-01
Morphine addiction	3	11.4	5.37E-01
Bile secretion	3	10.9	5.37E-01
Citrate cycle (TCA cycle)	3	10.1	5.37E-01
Endocrine and other factor-regulated calcium reabsorption	3	9.4	5.37E-01
Retrograde endocannabinoid signaling	3	8.7	5.37E-01
Insulin secretion	3	8.5	5.37E-01
Gastric acid secretion	3	8.2	5.37E-01
Aldosterone synthesis and secretion	3	8.0	5.37E-01

4.5 Immunofluorescence microscopy of Cas9 THP-1 cells

Endolysosomal proteins appear to be dysregulated or differentially regulated in the Cas9 cell line compared to WT THP-1 cells. I used immunofluorescence microscopy for the lysosomal marker LAMP-1 and the early endosomal marker EEA-1 to assess changes to these compartments in cells stably expressing Cas9. To facilitate adhesion to microscopy slides, THP-1 cells were differentiated with PMA to form macrophage-like adherent cells (Fig. 4.23). The majority of differentiated WT cells maintained a rounded morphology, with only a few cells spreading out and forming processes. In contrast, most differentiated Cas9 cells spread over the surface and a large proportion of these cells formed processes. Differentiated KO1 and KO2 cells appeared similar to Cas9 cells.

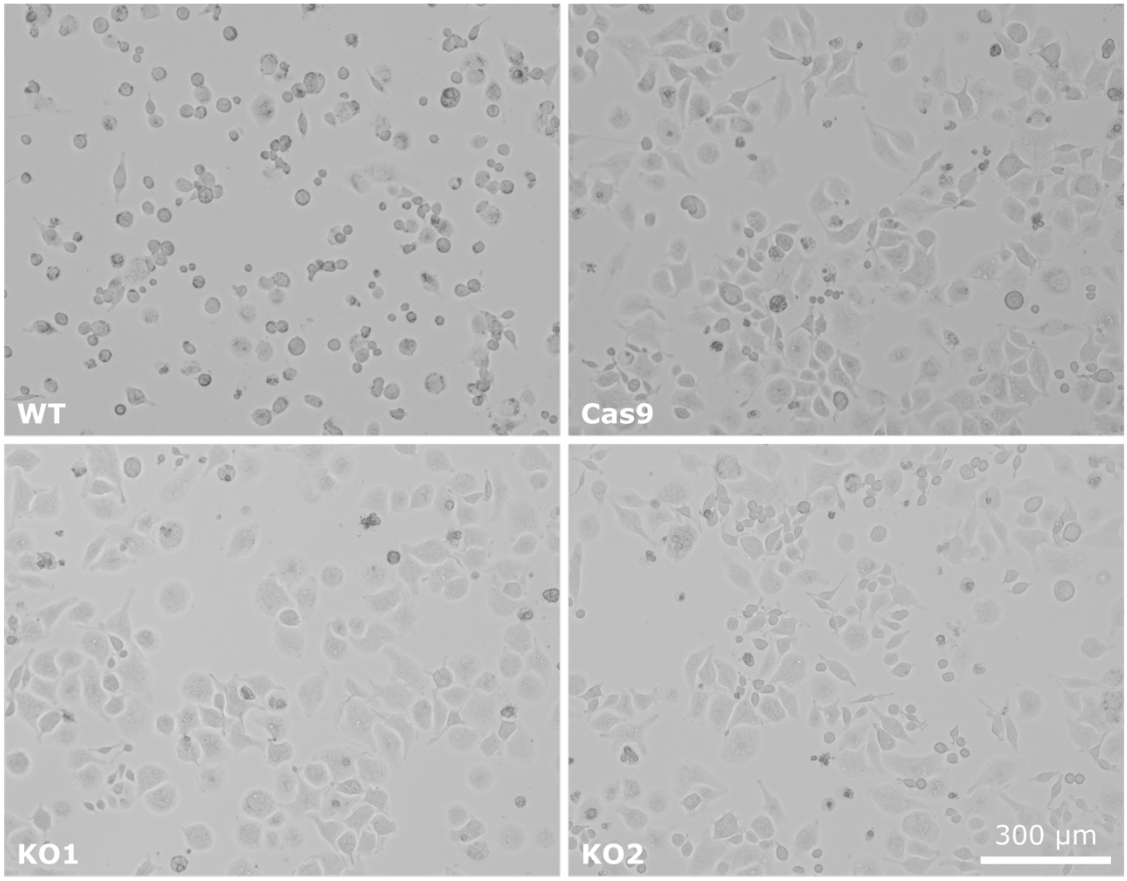

Figure 4.23. The morphology of differentiated Cas9-derived cell lines differs from WT
WT, Cas9, KO1 and KO2 THP-1 cells were differentiated and visualised with a brightfield microscope.

Nuclear staining showed a striking difference in the nucleus morphology of differentiated Cas9 cells compared to WT (Fig. 4.24). Monocytes and macrophages have the capacity to form multilobated nuclei, possibly to support cell passage through small gaps by providing flexibility to the nucleus[272]. In addition, macrophages can fuse to form giant multi-nucleated cells through a poorly understood process[273]. PMA-differentiated THP-1 cells have been previously observed to have heterogeneous cell

and nuclear morphology, with some evidence of multilobation and multinucleation[274]. Similarly, spontaneously activated THP-1 cells have been reported to form giant multinucleated cells with up to 20 nuclei[275]. In this experiment, differentiated WT cells show some multilobated nuclei and some cells appear to have two nuclei (Fig. 4.24). In Cas9 cells and Cas9-derived cell lines, this phenotype is more pronounced, with a large number of multinucleated cells possessing more than two nuclei in a flower-shaped arrangement (Fig. 4.24). Considering the marked propensity of the Cas9 THP-1 cell line to form multi-nucleated cells, this cell line is not a good model to study how loss of GALC triggers the formation of multinucleated globular cells characteristic of Krabbe disease. The lysosome and early endosome markers are concentrated at the centre of these flower-shaped multinuclei in Cas9-derived cell lines (Fig. 4.24). This suggests that the localisation and trafficking of endo-lysosomal compartments differs from WT in these cell lines. This poses additional concerns about endolysosomal function in these cells. No differences were observed between Cas9, mixed population EV control cell lines and GALC KO1 and KO2 cell lines. Loss of GALC and subsequent galactosphingolipid does not appear to cause substantial changes in the number or localisation of lysosomes in Cas9 THP-1 cells, although such effects may be masked by abnormal lysosome localisation in these cells.

In conclusion, the parental THP-1 Cas9 clonal cell line used to generate GALC KO THP-1 cells shows dramatic and wide-ranging abnormalities. Although Cas9-derived cell lines appeared normal when undifferentiated, these cells show a different cell morphology and a higher propensity to form multinucleated cells upon differentiation than WT THP-1 cells. The abundance of approximately half of all quantified proteins is significantly different from WT in this cell line. The endolysosomal compartments and sphingolipid metabolic enzymes are particularly affected by these changes. This cell line is therefore inappropriate for the study of changes caused by disruption of glycosphingolipid metabolism. This chapter illustrates that genetic manipulation of cells can cause unexpected and wide-reaching changes amplified by the selection of single cell clones. Future studies of lipid metabolism in phagocytic cells will require a carefully designed gene knockout strategy to overcome these challenges.

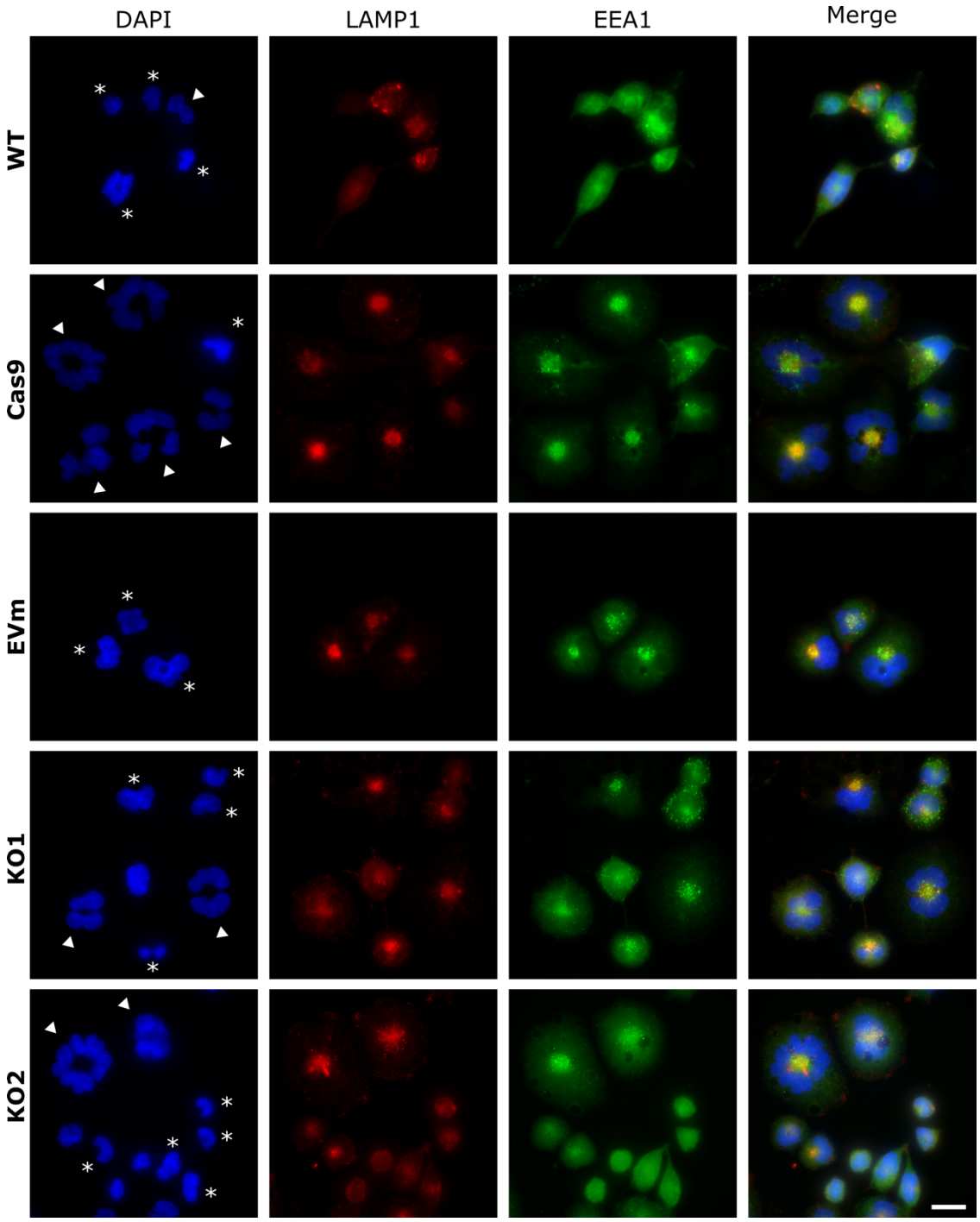

Figure 4.24. Immunofluorescence microscopy of THP-1 cell lines
WT, Cas9, mixed population EV (EVm), GALC KO1 and KO2 THP-1 cells were differentiated and stained with antibodies against the lysosomal marker LAMP-1 and the early endosome marker EEA-1. Nuclei were stained with DAPI. The symbol * indicates multilobated nuclei; arrows indicate multinucleated cells. Scale bar: 20 µm.

4.6 Discussion

Disentangling the mechanisms of inflammation, demyelination and oligodendrocyte cell death in Krabbe disease has been a major challenge for understanding the pathogenic mechanisms driving this

disease. I asked whether loss of GALC in phagocytic cells causes them to acquire an activated, globoid cell phenotype using CRISPR/Cas9 KO in the THP-1 monocytic cell line. I found that loss of GALC in monocytes alone is not enough to promote globoid cell formation and an inflammatory phenotype. A recently published study by Weinstock *et al*.[276] independently confirmed this observation. The authors performed a site-specific KO of *GALC* in peripheral macrophages of WT mice. They observed that these mice did not present globoid cell infiltration of peripheral nerves. Cultured bone marrow derived macrophages (BMDMs) appeared normal and expressed normal levels of the pro-inflammatory cytokines TNF-α and IL-1β. In contrast, in mice with a double knockout of GALC in macrophages and Schwann cells, the myelinating cells of the peripheral nervous system, they observed globoid cell infiltration of demyelinating peripheral nerves. Moreover, incubation of cultured GALC KO BMDMs with exogenous GalCer induced globoid cell formation and up-regulation of TNF-α, IL-1β and integrated stress response genes. Altogether, this study provides strong evidence that globoid cell formation is driven by phagocytosis of myelin debris and subsequent accumulation of undigested GALC substrates in GALC-deficient macrophages. This triggers a stressed, pro-inflammatory, multinucleated macrophage phenotype.

The glycosphingolipid repertoire at the surface of immune cells is diverse, dynamic and only partially understood. Regulation of various glycosphingolipid metabolic enzymes upon immune cell differentiation and stimulation has been observed for several immune cell types including monocytes and macrophages[123]. I have found that the expression levels of the galactosphingolipid anabolic and catabolic enzymes UGT8 and GALC is modulated upon THP-1 cell differentiation and LPS stimulation. This is in agreement with previous reports finding changes in GALC expression upon differentiation of primary monocytes[236] and down-regulation of UGT8 expression upon TLR stimulation of a mouse macrophage cell line[239]. The dynamic regulation of galactosphingolipid synthesis and degradation suggests that these lipid species have functional roles in monocytes and macrophages. Quantification of glycosphingolipids by lipidomics is very challenging and requires the implementation of specialised procedures to discriminate between GalCer- and glucosylceramide (GluCer)-based glycosphingolipids. Our lab is working on implementing such a technique through a collaboration, but we are currently unable to quantify galactosphingolipids in THP-1 cells. Weinstock *et al*. [276] did not detect any GalCer in WT or GALC KO mouse BMDM. However, very small amounts of psychosine were detected in WT macrophages and this was significantly increased in GALC KO macrophages[276]. Thus, the high expression levels of GALC in THP-1 cells may be primarily to aid with digestion of phagocytosed GalCer rather than recycling of large amounts of self-synthesised galactosphingolipids. It is possible that psychosine is functionally relevant for these cells, but endogenous psychosine accumulation does not

cause activation of macrophages. It should be noted that differences exist in the glycosphingolipid profile of mouse and human immune cells, and these differences have not yet been fully characterised[123].

I attempted to understand the potential roles of galactosphingolipids in monocytes by investigating the effect of galactosphingolipid accumulation on the proteome at the cell surface. Unfortunately, the dysregulation of sphingolipid metabolism and endolysosomal compartments in the parental Cas9 cell line may have masked any effects of GALC loss in the GALC KO cell line, limiting the amount of meaningful data that could be obtained from these cell lines. Loss of GALC in the Cas9 THP-1 cell line caused few changes in protein expression at the cell surface through changes in gene expression. Rescue of GALC expression in the KO cell lines only partially rescued the expression of the sodium/bicarbonate cotransporter *SLC4A4* and the transcriptional regulator *MDFIC*, making clear interpretation of these results elusive. I did not observe any changes in the cytokine profile of GALC KO cells compared to the EV1 control cell line, nor any changes in morphology compared to the Cas9 and mixed population EV cell lines. However, the issues encountered in these cell lines severely undermine confidence in these observations.

The generation of GALC KO and control cell lines posed significant problems. Due to the inherent resistance of phagocytic cells to transfection, a lentiviral transduction strategy was used to generate these cell lines. Two of the control clonal cell lines showed large and wide-ranging changes in protein expression compared to the parental cell line. These changes may have been caused by a mutation caused by random insertion of the lentiviral vector into the genome (insertional mutagenesis) or Cas9 off target effects[277] caused by unintended cleavage of DNA at sites showing mismatches with the gRNA causing mutations and potentially large deletions and genomic rearrangements. Although lentiviral transduction and CRISPR/Cas9 KO are commonly employed for biological research, the possibility of such off-target effects warrants caution while using these tools. As illustrated in this study, the use of all parental cell lines in control experiments is crucial for the interpretation of results obtained through the genetic manipulation of cell lines. Thanks to recent advances in proteomics labelling, it is now possible to analyse up to 16 samples by quantitative proteomics. Several precautions can be taken when generating KO cell lines. Where possible, a mixed population of transduced cells should be used instead of a single cell clone. This prevents the amplification of off-target effects, as I observed by direct comparison of the proteome of the Cas9 THP-1 cell line and mixed population EV cells, showing no significant difference in protein expression. However, a homogeneous cell population is sometimes required. Homologous recombination into a "safe harbour" locus can be employed to generate a

homogeneous cell line and circumvent the issue of insertional mutagenesis. Safe harbour loci are validated sites which can be disrupted without any known adverse effects on the cell, and which can support robust transcription of the inserted gene in most cell types[278]. THP-1 cells genetically engineered to facilitate Zinc finger nuclease-mediated genomic integration into the adeno-associated virus site 1 safe harbour locus are commercially available. Using this strategy to insert a DNA sequence encoding Cas9 and gRNAs can eliminate lentivirus insertion-mediated off-target effects. Moreover, several strategies have been recently developed to minimise off-target effects of CRISPR/Cas9[279]. For example, Cas9 variants have been designed to reduce the efficiency of cleavage and thus increase its specificity. These strategies can be combined to generate reliable KO cell lines in phagocytic cells.

To conclude, recent evidence suggests that immune dysfunction in Krabbe disease is caused by storage of undigested myelin debris in GALC-deficient phagocytic cells, triggering the formation of globoid cells and contributing to a pro-inflammatory environment[276]. Despite the secretion of pro-inflammatory factors by activated globoid cells, macrophages appear to have an overall protective role in Krabbe disease by removing myelin debris and promoting remyelination, as demonstrated in a macrophage-deficient Krabbe mouse model[280]. Loss of GALC does not intrinsically activate phagocytic cells and does not drive the pathogenesis of Krabbe disease. However, the dynamic regulation of galactosphingolipid anabolic and catabolic enzymes suggests that these lipid species could play active roles in cellular and immune processes. Since endogenous galactosphingolipid accumulation in GALC KO macrophages does not appear to trigger a distinctive phenotype in this study and a previous report[276], GALC KO may not be the most appropriate tool to study this role. Instead, abolishing the synthesis of galactosphingolipids by knocking out the anabolic enzyme UGT8 may be a better tool to uncover the role of these lipids in phagocytic cells.

5 Study of acid ceramidase function

Chapter summary

The lysosomal enzyme acid ceramidase (AC) hydrolyses ceramide to sphingosine, thus regulating the levels of these ubiquitous bioactive lipids involved in many aspects of cell function, including apoptosis and inflammation. Dysregulation of AC activity has dire consequences. Loss-of-function mutations in AC cause Farber disease, a rare lysosomal storage disorder that causes a range of severe symptoms resulting in early death. In contrast, up-regulation of AC expression is associated with some cancers and resistance to cancer treatment. In addition, AC may contribute to the pathogenesis of several lysosomal storage diseases by hydrolysing aberrantly accumulated glycosphingolipids to form harmful lysosphingolipids. The first aim of my PhD had been to determine the structure of AC as there was substantial preliminary work already performed within the Deane lab. However, within the first month of my PhD this structure was published by another lab[128] and so the focus and aims of this project were re-addressed. Specifically, the structure left several key aspects of AC function unknown. It is unclear how AC accesses its membrane-embedded substrate, and how it could process glycosphingolipids with bulky head groups that are not present on its known substrate ceramide. This chapter describes my attempts to investigate these aspects of AC function using structural, biochemical and cell-based methods.

Statement of acknowledgement

Dr Sam Spratley, a former PhD student in our group, performed preliminary work for this project, establishing a purification strategy for AC and solving a preliminary structure of AC with the help of Dr Stephen Graham (Department of Virology, University of Cambridge). Dr Stephen Graham also assisted with crystal manipulation and diffraction data collection described in this chapter. Prof Randy Read (CIMR, University of Cambridge) assisted with the molecular replacement solution for the SapA assembly presented in this chapter. One of the stable cell lines for the expression of AC with an N-terminal Protein A tag was established by Dr Janet Deane.

5.1 Understanding the interaction of AC with lipid membranes

AC is thought to interact with lysosomal membranes through a large hydrophobic patch participating in hydrophobic interactions with the lipid bilayer, as well as electrostatic interactions with negatively

charged lipid head groups[90,128] (Fig. 1.7), but the molecular details of these interactions are unknown. It is also unclear how ceramide substrates leave the lipid bilayer to access the deep active site of AC and how this process is facilitated by saposin D (Fig. 1.8). I aimed to understand how AC interacts with lipid membranes by obtaining a crystal structure of AC in complex with a lipid-containing nanodiscs. Small discs of lipid bilayer can be enclosed by saposin A (SapA) molecules to form stable and uniform assemblies suitable for structural studies. These discs have been used to provide a native lipid environment for nuclear magnetic resonance[93] and electron microscopy[94] studies of transmembrane proteins. SapA nanodiscs are also appropriate for the study of lipid processing enzymes, as demonstrated by the fact that the enzyme Sialidase-3 can degrade its ganglioside substrate embedded in SapA nanodiscs[281].

5.1.1 Stable expression of human AC in mammalian cells

I established a stable, inducible HEK293F cell line expressing hexahistidine (H_6)-tagged AC using the PiggyBac system[173] (Fig. 5.1A). AC was purified from the media following a protocol optimised by Dr Sam Spratley, consisting of a single nickel affinity chromatography step followed by elution with low pH buffer (100 mM citrate, pH 4; Fig. 5.1B). AC was subsequently stored in the elution buffer. The yield was high: 12 mg of pure protein were purified from 500 mL of media. SDS-PAGE analysis of the purified product reveals two forms of AC, as observed previously[126,128]. The immature, uncleaved form of AC can be detected as a single band at approximately 55 kDa while mature AC is cleaved into two subunits, α (16 kDa) and β (40 kDa), linked by a disulphide bond. Mature AC is the active form of AC formed by autoproteolysis of the peptide bond between Cys143 and Ile142, an event triggered by the low pH of the lysosome [127,128]. The molecular weight of AC is higher than predicted from its sequence (47 kDa) because it is glycosylated at five sites on the β subunit[126].

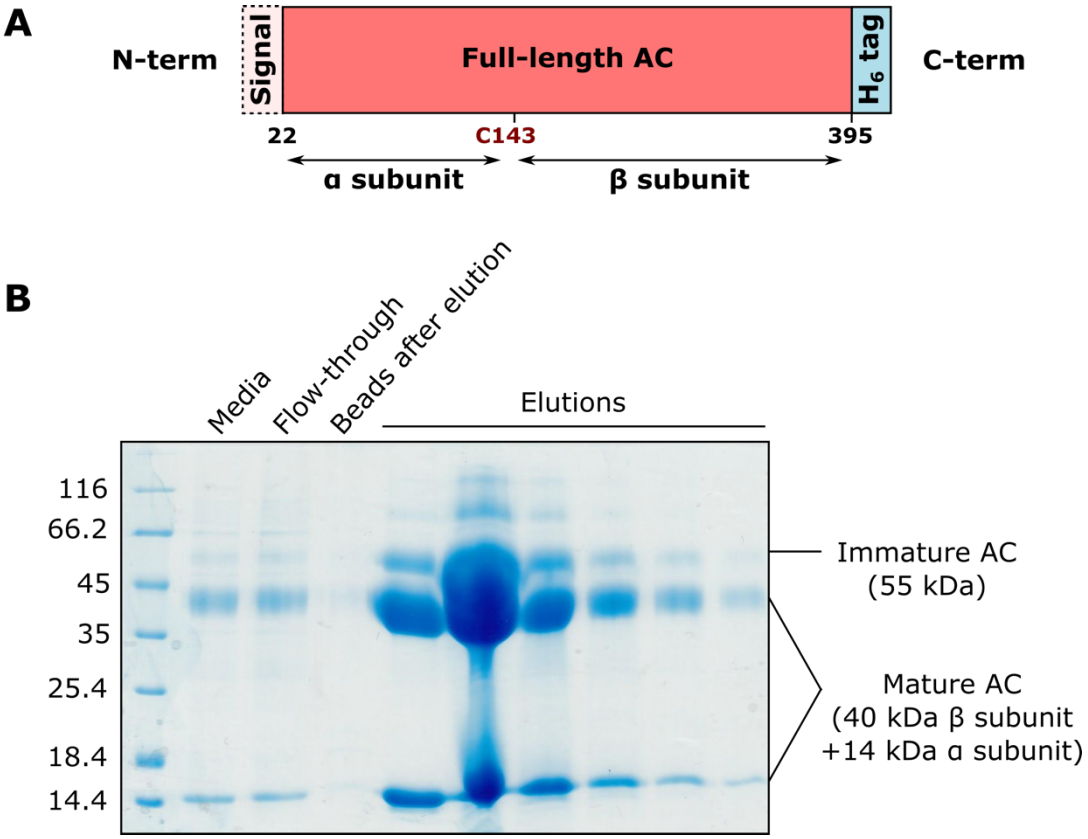

Figure 5.1. Nickel affinity purification of recombinant AC from mammalian cells
A. Recombinant AC was expressed in HEK293F cells. The construct encoded full-length AC possessing an N-terminal cleaved secretion signal peptide and a C-terminal hexahistidine (H_6) tag. B. AC was captured from the media with Ni-NTA beads and eluted in 1 mL fractions of 100 mM citrate pH 4. Fractions were analysed on a Coomassie-stained 12% SDS-PAGE gel.

5.1.2 Preparation and characterisation of SapA lipoprotein nanodiscs

I expressed untagged recombinant mouse SapA in *E. coli*. SapA was purified using the same purification protocol as for SapB (section 3.2). *E. coli* lysate was boiled to remove most proteins, except for heat-stable proteins including SapA. I then purified SapA by anion exchange chromatography (Fig. 5.2A-B) and size exclusion chromatography (Fig. 5.2C), yielding 8 mg of pure SapA from 1 L of culture (Fig. 5.2D).

As described in section 1.4, SapA can adopt a closed, globular monomeric conformation (Fig. 1.4C), or, in the presence of lipids or detergents, SapA can extend into an open conformation exposing its internal hydrophobic face and allowing SapA to assemble into oligomers enclosing lipid or detergent molecules (Fig. 1.4B, D, E). SapA-lipid nanodisc formation requires incubation of SapA and lipids at low pH (pH 4.8) or solubilisation of lipids with the gentle detergent n-Dodecyl-β-D-maltoside (DDM) prior to incubation with SapA at neutral pH[93]. SapA has been found to assemble with a very wide range of lipids to form nanodiscs[95] but most studies of this system have employed phosphatidylcholine

(PC)[81,93,95,281,282]. PC is one of the most abundant lipids in lysosomal membranes and AC has been shown to interact with liposomes primarily composed of PC (80 mole % PC and 20 mole % cholesterol (Chol))[90]. Therefore, SapA-PC nanodiscs were a good initial system for this study. The number of SapA and lipid molecules included in the nanodisc, and overall nanodisc size depend on several factors, including the lipid to SapA molar ratio[93] and the pH of the buffer[282].

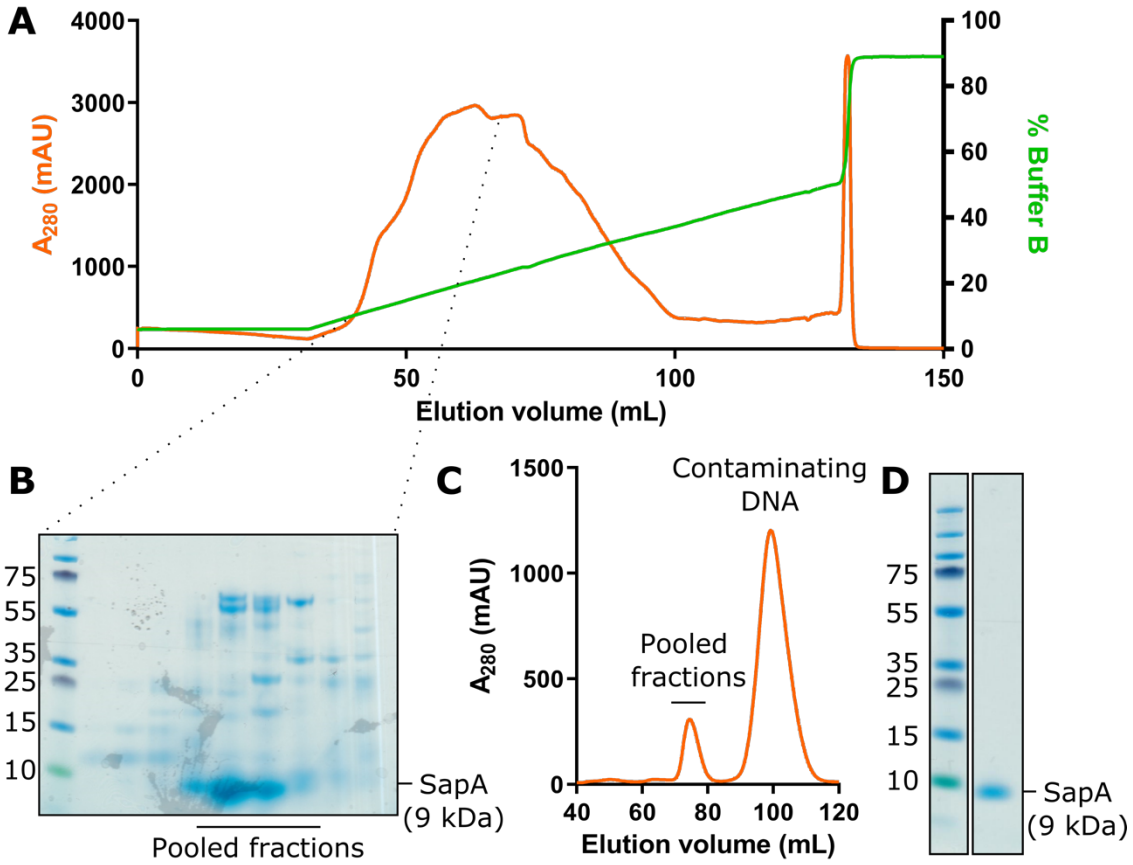

Figure 5.2. Purification of untagged recombinant mouse SapA from *E. coli*
Mouse SapA was purified by anion exchange chromatography followed by size exclusion chromatography. A. Anion exchange chromatogram. B. The first fractions of the peak were analysed on a Coomassie-stained 4-12% Bis-Tris SDS-PAGE gel. The indicated fractions were pooled and concentrated for size exclusion chromatography. C. SapA was further purified by size exclusion chromatography on a Superdex 75 16/600 column. The indicated fractions containing SapA were pooled and concentrated for future experiments. Large amounts of contaminating DNA (absorbing at 280 nm) elute around 100 mL. D. SDS-PAGE analysis of purified SapA.

When preparing SapA-PC nanodiscs I adapted published protocols[93,94] as follows. Although human SapA has been used to form nanodiscs in previous studies, here I used the mouse ortholog which possesses 79.5% sequence identity and 91.6% similarity to human SapA. The interaction of AC with lipid membranes has been shown to be pH-dependent and strongly enhanced at pH 4[90]. Nanodiscs

were therefore prepared at pH 4 as it is similar to the low pH of 4.8 used to promote nanodisc formation in Chien *et al*.[93]. Previous publications reported complete solubilisation of dried lipids in aqueous buffer upon addition of the detergent DDM[93,94]. However, in my initial experiments lipid solubilisation was incomplete, resulting in a very low yield with less than 10% of SapA forming nanodiscs. I improved lipid solubilisation by making liposomes in aqueous buffer by 10 cycles of sonication and vortexing of the lipid solution prior to the addition of DDM. This greatly improved the final yield to over 70%. Solubilised lipids were then incubated with SapA following the protocol by Frauenfeld *et al*.[94] with a molar ratio of PC to SapA of 30:1 as in Chien *et al.*[93]. The incorporation of lipids into SapA nanodiscs increases the molecule size, allowing separation from free SapA and detergent-solubilised lipids by size exclusion chromatography (Fig. 5.3A-B). SapA-PC nanodiscs eluted at 15.8 mL from a Superdex 200 10/300 column. This elution volume was consistent across several preparations. The elution volume is dependent on the hydrodynamic radius of a molecule, allowing comparison of eluted macromolecules across different experiments when the same size exclusion chromatography column is used. In order to evaluate the composition of the nanodiscs I had produced, I compared the size exclusion chromatography profile with that from Chien *et al*.[93] (Fig. 5.3C). While I used a lipid to SapA ratio of 30:1, which should have produced a large nanodisc assembly eluting at 13.3 mL, the elution volume of my nanodiscs more closely matched that of a smaller nanodisc assembly produced using a 1:5 ratio and eluting at 15.9 mL. Li *et al*. [282] showed that nanodiscs eluting at this volume contain two SapA and 23-29 PC molecules by mass spectrometry and size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS).

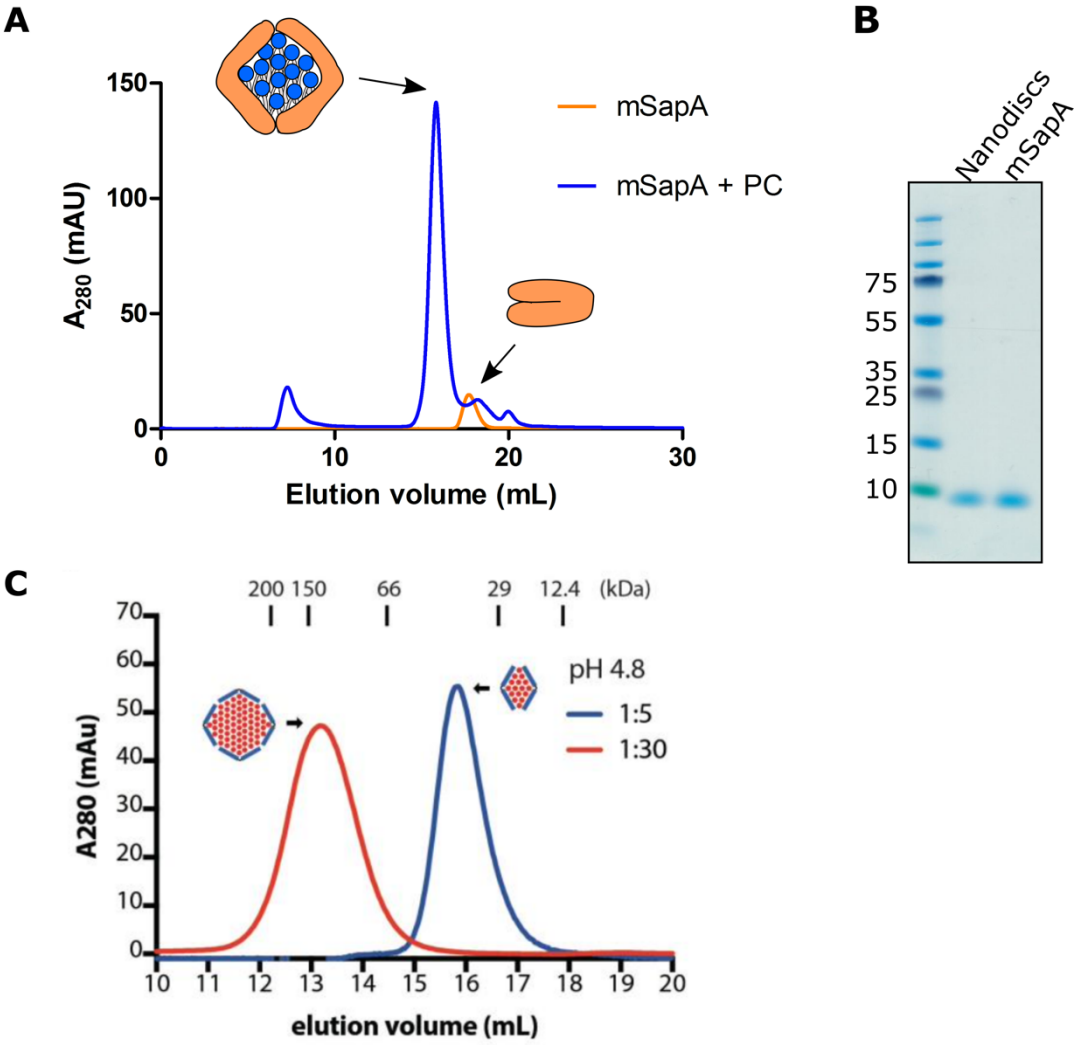

Figure 5.3. SapA-phosphatidylcholine nanodiscs preparation
Lipoprotein nanodiscs were made by incubating SapA with detergent-solubilised phosphatidylcholine (PC). A. Size exclusion chromatography using a Superdex 200 10/300 column allows separation of PC-SapA nanodiscs (elution peak at 15.8 mL) from SapA, which elutes at 17.7 mL. B. Coomassie-stained SDS-PAGE gel of PC-SapA nanodiscs. C. This panel is reprinted with permission from Chien *et al.*[93]. Size-exclusion chromatography on a Superdex 200 Increase 10/300 of SapA-PC nanodiscs made at pH 4.8 using the indicated SapA to lipid ratio (1:5, in blue and 1:30, in red) by Chien *et al*. The proposed SapA arrangement is indicated for each peak.

I characterised the molecular weight, size and shape of the SapA-PC nanodiscs I produced by SEC-MALS. This technique uses size exclusion chromatography to separate macromolecules based on their hydrodynamic radius, then measures the light scattering, differential refractive index and UV signal of the sample. This facilitates the calculation of the molecular mass, independent of shape, and therefore the determination of the oligomeric state of macromolecules in solution. The observed average molecular mass of SapA-PC assemblies at pH 4 was 37.1 ± 0.5 kDa (Fig. 5.4). However, the sample was heterogeneous, as indicated by a polydispersity value (Mw/Mn) of 1.010 and higher molecular masses observed at the beginning of the peak. This is likely due to variable amounts of PC (770 Da) and/or

DDM (511 Da) molecules contained within the SapA nanodiscs. For nanodiscs containing a SapA dimer with 23-29 PC molecules, Li *et al.*[282] measured molecular masses ranging from approximately 35 to 50 kDa, with an average MW of 40.9 ± 1.6 kDa by SEC-MALS. These values are similar to the molecular mass distribution observed in this experiment (Fig. 5.4). In conclusion, I made SapA-PC nanodiscs with a molecular mass closely resembling previously published data, suggesting that they are composed of a SapA dimer encapsulating approximately 20 to 30 PC molecules.

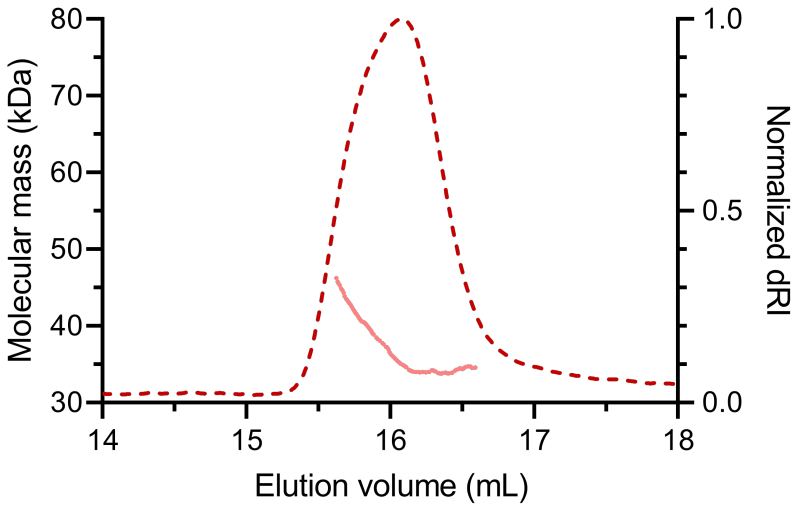

Figure 5.4. SEC-MALS analysis of SapA-PC assemblies at pH 4
The molecular mass distribution (pink line) is shown across the elution profile (normalized differential refractive index, dRI, red dotted line) when eluting from a Superdex 200 10/300 Increase column. The observed molecular mass is 37 kDa, with higher molecular masses observed in the start of the peak.

5.1.3 Crystallisation of AC with SapA-PC nanodiscs

To investigate how AC interacts with membranes, I attempted to co-crystallise AC with SapA-PC nanodiscs. Both AC and nanodiscs were purified in pH 4 buffer, the optimal pH for AC interaction with membranes. AC and nanodiscs were concentrated to their solubility limit (4.5 mg/mL and 3.5 mg/mL, respectively) and mixed in equimolar quantities. I set up a high-throughput sparse matrix crystallisation screen using this mixture. In parallel with this approach, concentrated AC and nanodiscs were mixed and incubated at room temperature for 75 minutes, then concentrated further to the solubility limit of the mixture and used to set up the same crystallisation screen for comparison. Good quality crystals grew in several conditions (Fig. 5.5A). There was no significant difference between the conditions obtained in the screens set up with fresh or pre-incubated mixture, so I used freshly mixed AC and nanodiscs for subsequent screens. I optimised crystallisation conditions by changing the pH, precipitant concentration, and presence/absence of the additives glycerol and $MgCl_2$.

Crystals diffracted to a maximum resolution of 3.2 Å (Table 5.1). All crystals belonged to the space groups *P* 6_1, *P* 6_1 2 2 or *P* 6_2 2 2, with the same unit cell dimensions, indicating that the same crystal form with the same composition grew in different conditions. The best quality datasets (based on low resolution merging statistics) were assigned the space group *P* 6_1, suggesting that this is the most likely space group. I could not solve the phase problem by molecular replacement using the unpublished low-resolution structure of AC previously obtained by our group and/or two copies of human SapA in its open conformation (PDB ID: 4DDJ).

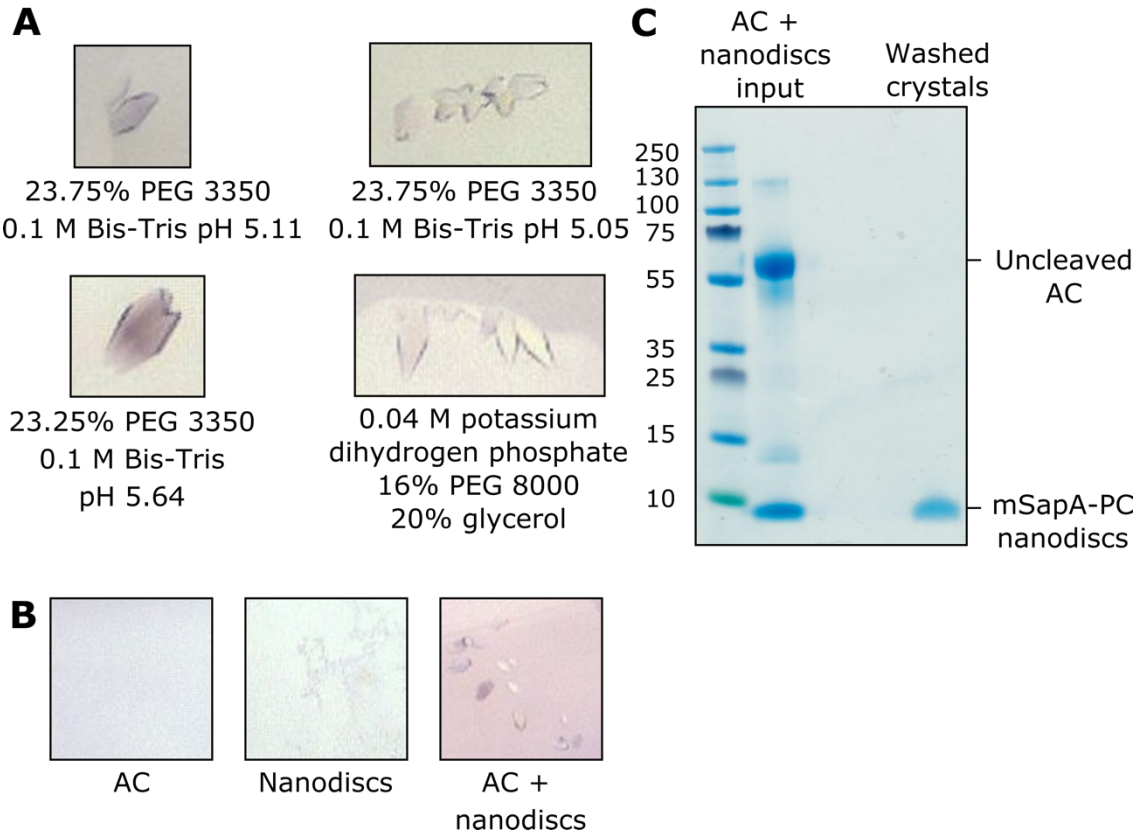

Figure 5.5. Crystals grown in conditions containing AC and PC-SapA nanodiscs
A. Crystals grew in several conditions containing AC and PC-SapA nanodiscs. B. All crystallisation conditions being equal (20.25% PEG 3.35K, 0.1 M Bis-Tris pH 5.11), crystals grew in drops containing both AC and PC-SapA nanodiscs, but not AC alone or PC-SapA nanodiscs alone. C. Coomassie-stained SDS-PAGE gel of washed crystals compared to the input mixture containing both AC and nanodiscs. Approximately 20 medium-sized crystals were washed in a drop containing reservoir solution.

To test whether the crystals grown were composed of AC and nanodisc complexes rather than AC or nanodiscs alone, I set up crystallisation plates with optimised conditions and added either the AC-nanodisc mix, AC alone supplemented with nanodisc buffer instead of nanodiscs, or nanodiscs alone supplemented with AC buffer instead of AC. Many crystals grew in drops containing both AC and

nanodiscs, while drops containing only one component contained precipitate but not crystals (Fig. 5.5B). This was a good indication that crystals were composed of AC and nanodisc complexes. However, SDS-PAGE of washed crystals clearly showed that the crystals were in fact composed of nanodiscs only, since AC present in the input mixture was completely depleted in the crystals (Fig. 5.5C).

Table 5.1. Representative data collection statistics. Values in parentheses are for the highest-resolution shell.

Beamline	i04
Wavelength (Å)	0.97950
Processing package	autoPROC
Space group	$P\ 6_1$
Cell dimensions	
a,*b*,*c* (Å)	48.9, 48.9, 317.9
α,β,γ (°)	90.0 90.0 120.0
Resolution (Å)	52.982 - 3.208 (3.418 - 3.208)
R_{merge}	0.166 (2.673)
R_{pim}	0.058 (1.027)
$CC_{1/2}$	0.996 (0.497)
$I/\sigma I$	10.2 (0.9)
Completeness (%)	s* 82.1 (24.0) e* 89.7 (39.1)
Redundancy	9.0 (7.8)

5.1.4 Pull-down experiments of SapA nanodiscs by AC

Since the crystals grown in presence of AC and nanodiscs contained only nanodiscs, I needed to establish whether AC could interact with a lipid bilayer enclosed by SapA by pull-down assays. The interaction of AC with membranes is strongly enhanced at pH 4, corresponding to lysosomal pH, and appears drastically reduced at pH 6[90]. Pull-down experiments that reflect the conditions of the lysosome are complicated by the loss of interaction between many protein tags and their ligand, such as polyhistidine and Ni-NTA at low pH. The interaction between biotin and streptavidin, however, is stable up to pH 4, making it a useful tool for pull-downs. I cloned and expressed a new construct of AC with a C-terminal biotin acceptor peptide (BAP) followed by a H_6 tag (AC-BAP-H_6) (Fig. 5.6A) using the same stable mammalian expression system as in section 5.1.1, yielding 6 mg of pure protein from 500 mL media (Fig. 5.6B). I biotinylated AC-BAP-H_6 *in vitro* with recombinant BirA purified from *E. coli*[178]. The C-terminal BAP tag is on the β-subunit, therefore biotinylation can be observed on the β-subunit

of mature AC and uncleaved AC. I assessed biotinylation efficiency by a gel-shift assay (Fig. 5.6C), showing a biotinylation efficiency up to 70%. There was a significant loss of protein by precipitation when the incubation time or the amount of BirA was increased, indicating that the protocol detailed in section 2.4.15 is optimal.

The binding of AC to membranes is enhanced by negatively charged lipids, and in particular the negatively charged phospholipid bis(monoacylglycero)phosphate (BMP) enriched in endolysosomal compartments[90]. AC was shown to interact strongly with liposomes containing 55 mole % PC, 25 mole % BMP and 20 mole % Chol. This composition was chosen to make negatively charged nanodiscs following the same protocol as for PC nanodiscs (Fig. 5.7A) since SapA nanodiscs can be made with a number of lipids mixtures, including a very similar mixture (50 mole % PC, 25 mole % Chol, 20 mole % BMP, 5 mole % GalCer)[81].

For the nanodisc pull-down experiment, biotinylated AC was immobilised on streptavidin agarose beads. Bead-bound AC was incubated with an 8-fold molar excess of SapA, SapA-PC nanodiscs or SapA-(PC+BMP+Chol) nanodiscs (Fig. 5.7B). After washing the beads to remove unbound protein, AC was eluted by boiling the beads in SDS-PAGE buffer and eluted protein was analysed by SDS-PAGE. No SapA-PC nanodiscs nor SapA-(PC+BMP+Chol) nanodiscs were bound to AC at pH 4 (Fig. 5.7C), indicating that AC does not form a stable interaction with lipid membranes enclosed within SapA nanodiscs, in contrast with the interaction between AC and liposomal membranes.

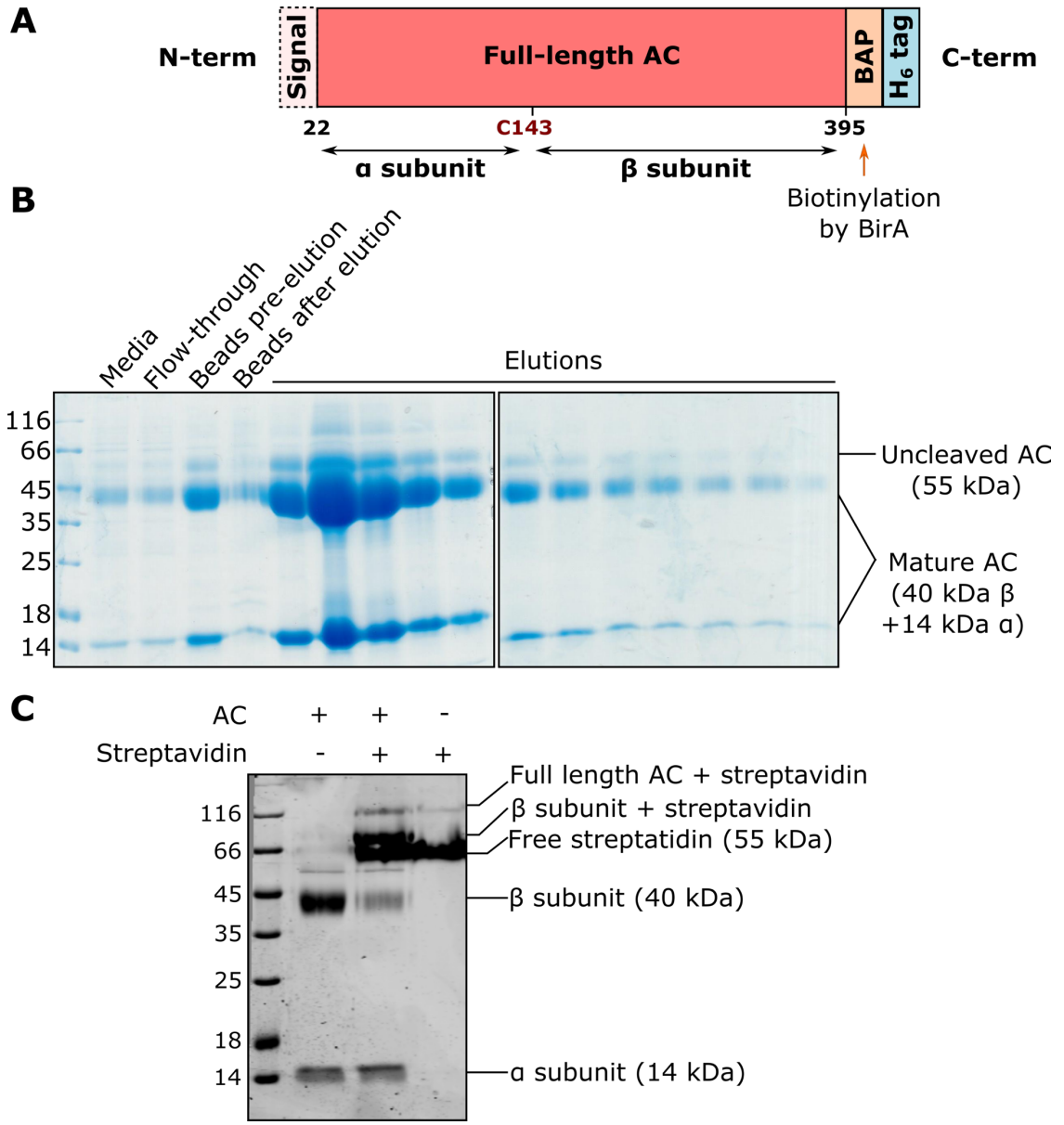

Figure 5.6. Preparation of biotinylated AC

A. Secreted AC with a C-terminal biotin acceptor peptide (BAP) followed by a hexahistidine tag was stably expressed in HEK293F cells. B. AC was captured from the media with Ni-NTA beads and eluted in 0.5 mL fractions of 100 mM citrate pH 4. Fractions were analysed by coomassie-stained 12% SDS-PAGE gel. C. AC was biotinylated *in vitro*. The proportion of biotinylated protein was determined by a gel-shift assay and calculated by densitometry. Streptavidin binds biotin with very high affinity, shifting the molecular weight of biotinylated AC.

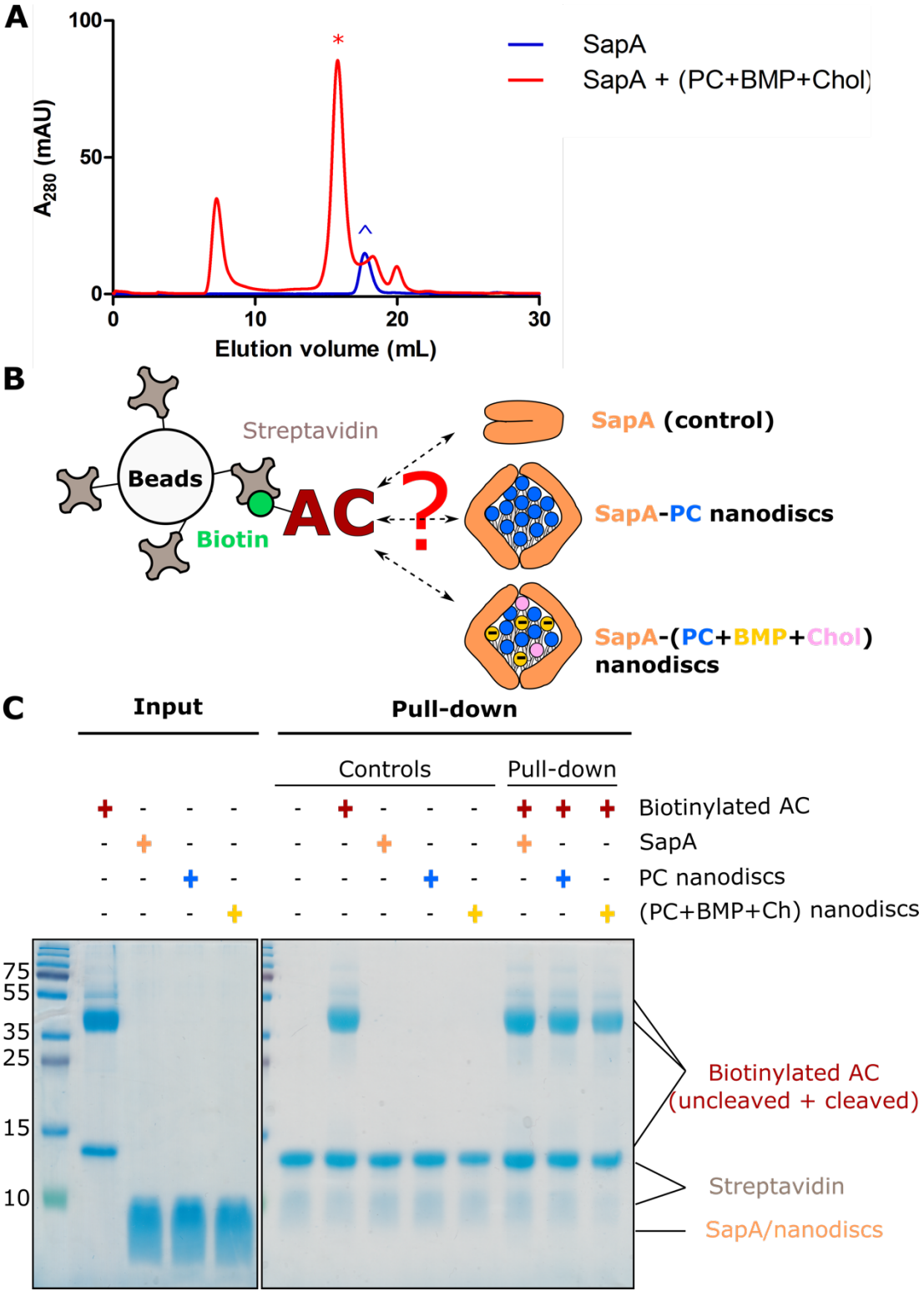

Figure 5.7. AC does not interact with SapA nanodiscs

A. Size exclusion chromatogram of SapA nanodiscs composed of PC, BMP and Chol (*) compared to SapB alone (^) B. Pull-down assay protocol. Biotinylated AC is immobilised on Streptavidin beads to test its interaction with SapA nanodiscs with different lipid compositions. C. Coomassie-stained SDS-PAGE gel of input and bound protein in a pull-down assay of PC-SapA and (PC, BMP, Chol)-SapA nanodiscs by biotinylated AC. The α subunit of AC runs at the same size as streptavidin detached from the resin during elution. The controls show that SapA and SapA nanodiscs do not interact non-specifically with the resin. No SapA or SapA nanodiscs were pulled down by AC.

5.1.5 SapA assembly structure solution and refinement

The structure of the SapA nanodisc crystal grown in 23.25% PEG 3350, 0.1 M Bis-Tris pH 5.64 (crystal shown in Fig. 5.5A, bottom left panel; diffraction dataset described in Table 5.1) was solved by molecular replacement using three copies of open SapA monomers (PDB ID: 4DDJ) as a search model. However, this solution was incomplete, as indicated by apparent but weak electron density likely corresponding to a fourth SapA monomer, as well as a Translation Function Z (TZF) score for this solution below 8 (TZF = 6.0) suggesting that the solution may be incomplete, overfilled, partly incorrect or completely incorrect. However, the fourth SapA chain could not be placed correctly by automated molecular replacement. With the help of Randy Read, this fourth monomer was successfully placed manually into the region of weak electron density using non-crystallographic symmetry. Although the electron density corresponding to the fourth SapA monomer (chain D) is very poor, it appears to have been placed correctly, as supported by lower R_{free} after rigid body refinement of this model (R_{free} = 0.406) versus the model with three SapA monomers (R_{free} = 0.465) or the incorrect automated molecular replacement solution with four SapA chains (R_{free} = 0.466). The statistic R_{free} describes how well the model describes a small subset of randomly selected reflections that are not used in the refinement of the structural model.

The electron density maps of this low-resolution dataset are poorly defined, with clear density visible only for the backbone of α-helical segments and the side chains of some bulky residues. This makes model building and improvement very challenging, because the placement of most amino-acid side chains and flexible loops is extremely ambiguous, error-prone and difficult to achieve. The model-building software package ISOLDE was recently developed to help guide model building into low resolution maps[190]. ISOLDE uses molecular dynamics to simulate interactions between all atoms of the model (or section of a large model) in real time while the model is manipulated by the user. This powerful approach allows refinement of atom positions based on the electron density map as well as interactions with neighbouring atoms. ISOLDE also provides real-time visual information on local deviations of the model from ideal protein geometry, informing model improvements. I ran a molecular dynamics simulation on the SapA tetramer model in ISOLDE, which eliminated clashes between atoms and corrected some Ramachandran outliers. I manually corrected obvious errors in the model such as Ramachandran, peptide bond and rotamer outliers and non-proline *cis* peptide bonds. These types of outliers can occur at very low frequencies; however, they are energetically unfavourable and require strong stabilisation by surrounding atoms. In the absence of strong density and stabilising interactions, these outliers are likely a consequence of erroneous modelling.

After model building in ISOLDE, the model was refined using the automated refinement software phenix.refine. Following two iterations of this cycle, R_{free} started increasing, indicating that the model was becoming over-fitted to the data and could not be improved any further. The final refinement statistics are presented in Table 5.2. Selected statistics can be compared to PDB entries of similar resolution[283] (Fig. 5.8). A typically well-refined structure should have a small and roughly equilateral polygon. Here, the Molprobity clashscore[191] is extremely low for a structure at this resolution because clashes between atoms have been minimised by molecular dynamics simulations in ISOLDE. Of note, the average protein B-factors are unusually high for this structure. B-factors quantitate the uncertainty about the position of each atom. High B-factors can be caused by flexibility of certain parts of the molecule (if localised to specific areas), mosaicity (disorder) within the crystal, or a problem with the protein refinement strategy. In this case, the Wilson B-factor of the raw crystallography dataset is also high (Wilson B-factor= 86.6), suggesting that the crystal had high mosaicity, possibly due to the presence of disordered lipids within the structure. B-factors are similar across all four chains, indicating that mosaicity is not caused by variability in the position of the fourth SapA chain (Table 5.2), despite the very poor electron density for this chain.

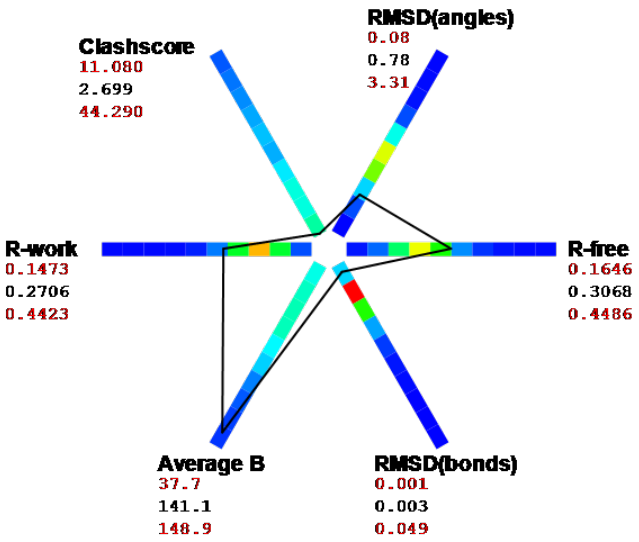

Figure 5.8. SapA tetramer refinement statistics
Histograms of the distribution of selected statistics across 1539 PDB entries of similar resolution, with the range specified by numbers printed in red. Statistics for the SapA tetramer structure are printed in black; the connecting polygon (in black) shows where these values fall in the distribution. Histogram bins are colored by the number of structures in each bin: 0 (blue), 385 (green), 770 (red).

Table 5.2. Final refinement statistics for the SapA tetramer structure.

Resolution (Å)	52.982 - 3.154
No. reflections	6460
R_{work}/R_{free}	0.2706/0.3068
No. atoms	
Protein	2432
Water	0
B-factors	
Protein	141.07
Chain A	139.08
Chain B	138.10
Chain C	144.95
Chain D	142.16
Ramachandran	
Favoured (%)	98.72
Outliers (%)	0.32
R.m.s deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.777
Rotamer outliers (%)	1.71
Cis peptides (Proline)	0
Cis peptides (General)	0
Twisted peptides (Proline)	0
Twisted peptides (General)	0
Clashscore	2.70

5.1.6 SapA forms a novel tetrameric structure

In this structure, SapA forms a tetrameric assembly distinct from previously described nanodiscs. Four SapA molecules in an open conformation form a diamond-shaped assembly with symmetrical contacts (Fig. 5.9A). This is very different from the SapA nanodisc structure (Fig. 1.4E) expected based on the size exclusion chromatography elution profile of these nanodiscs (as detailed at the end of section 5.1.2). It is also unlike the larger nanodiscs containing four SapA molecules for which a structure has been predicted (Fig. 5.9B)[282] and observed (Fig. 5.9C)[93] previously.

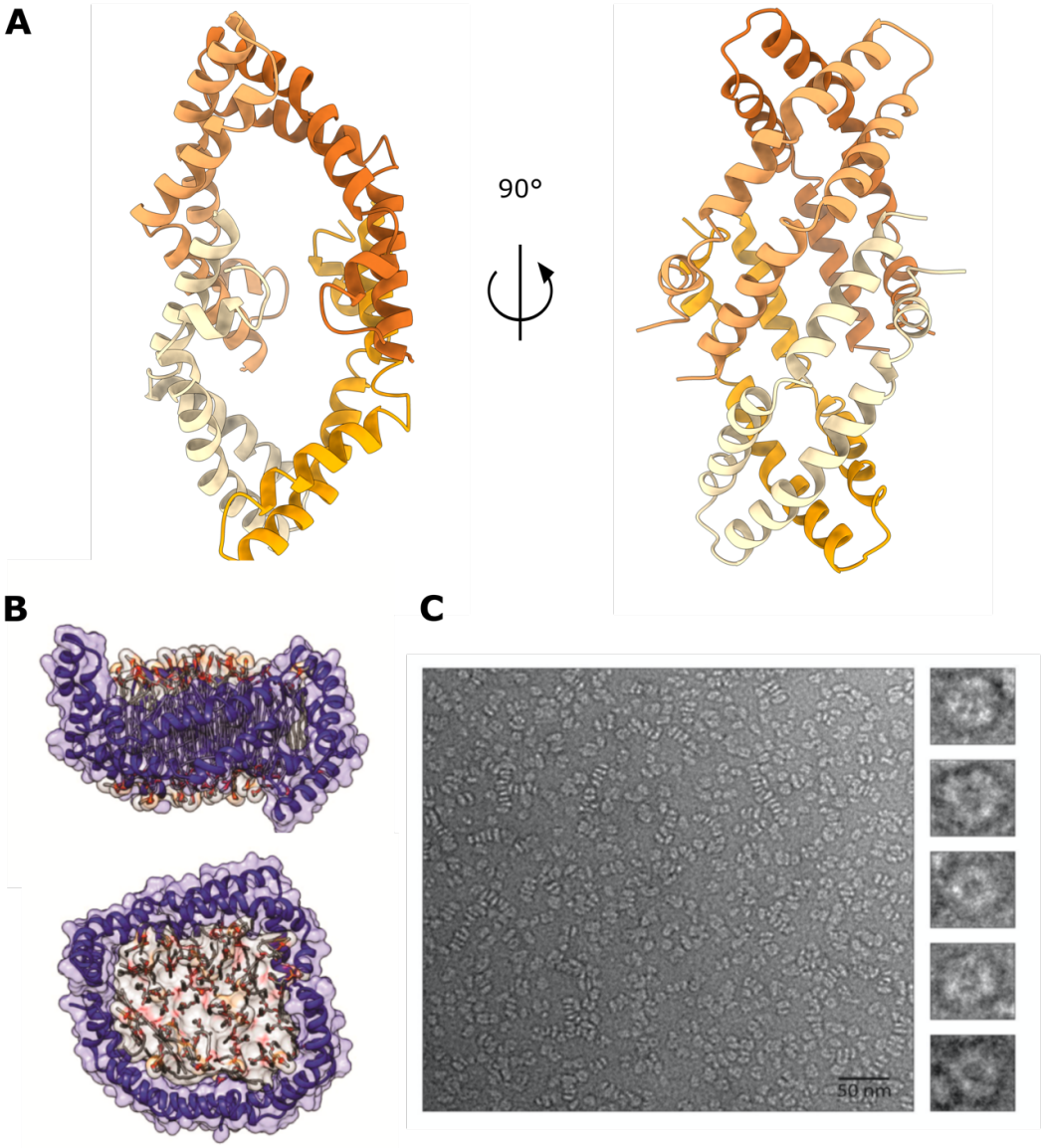

Figure 5.9. Novel tetrameric SapA assembly structure
A. Front and side view of the novel diamond-shaped tetrameric SapA assembly. SapA monomers are represented as ribbons and coloured by chain. B. This panel is reprinted with permission from Li *et al*.[282]. Top and side view of a predicted structure of a SapA nanodisc containing four SapA molecules (purple ribbon) forming a belt around a lipid core (lipids are represented as sticks) obtained by manual placement of SapA molecules around the lipid core followed by molecular dynamics simulations. C. This panel is reprinted with permission from Chien *et al*.[93]. Negative-stain electron microscopy images of SapA nanodiscs containing four SapA molecules. The scale bar represents 50 nm. The insets show top-down views of representative nanodisc images in a 20 nm box.

This tetrameric assembly is maintained by several protein-protein contacts between SapA monomers (Fig. 5.10A). These contacts are symmetrical and each SapA monomer participates in the same interactions. In particular, there are hydrophobic interactions between Trp37 on one monomer and Met61 and Ile58 on another monomer (Fig. 5.10A, left panel); between Tyr54 on one monomer and Ile38 and Leu43 on another; and between Leu2 and Ile6 on one monomer and Leu17 on another (Fig.

5.10A, right panel). Each SapA monomer also forms an intermolecular hydrogen bond between Tyr54 and the backbone of Cys35 on another monomer (Fig. 5.10A, lower panel). All of the residues participating in these interactions are either identical between mouse and human SapA or conserved in terms of their hydrophobic properties (for example, Leu to Met; Fig. 5.10B). This suggests that this novel tetrameric assembly is not driven by interactions specific to mouse SapA.

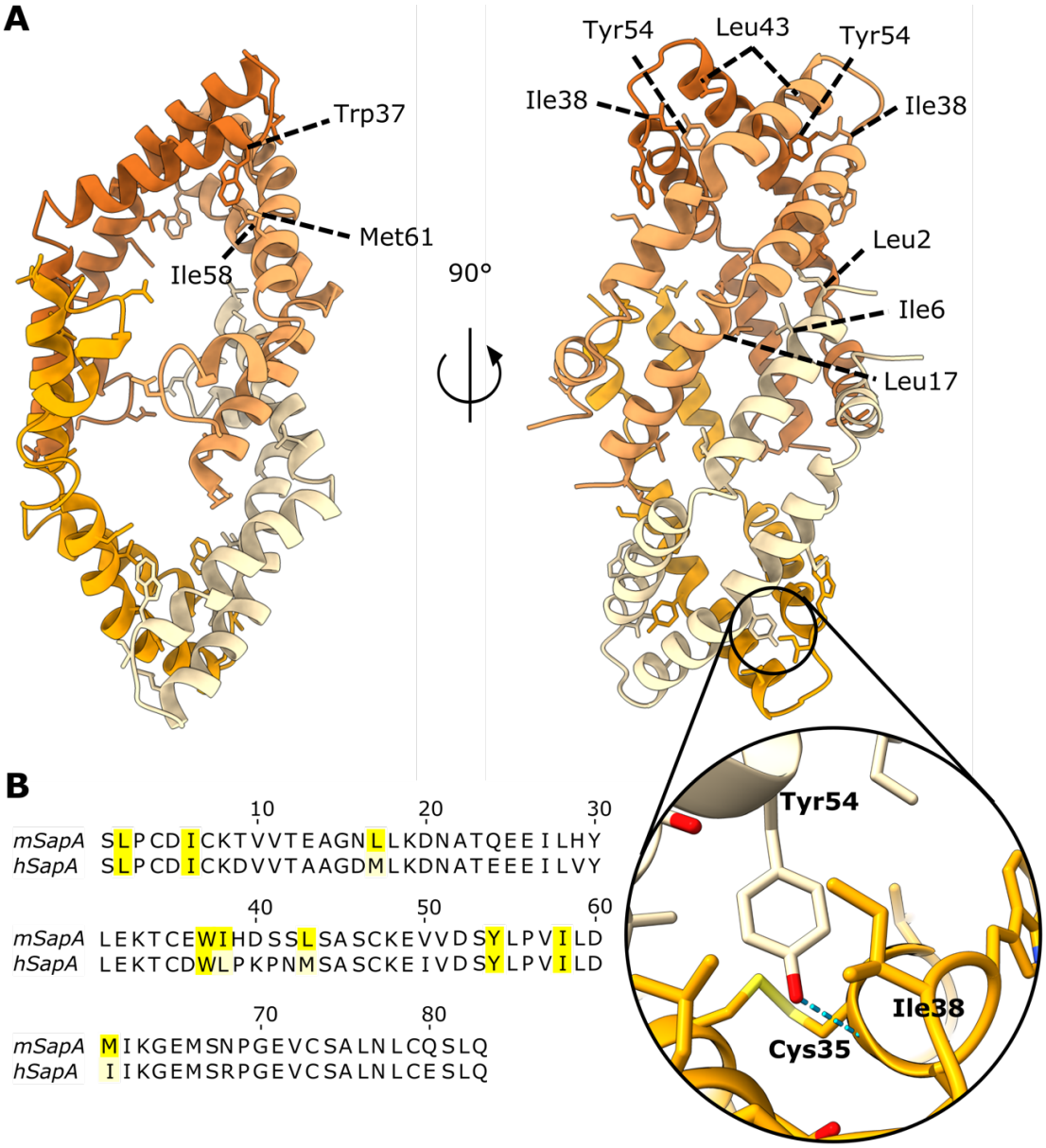

Figure 5.10. The tetrameric SapA assembly is maintained by protein-protein interactions
The SapA tetramer is maintained by intermolecular protein-protein contacts, including multiple hydrophobic interactions (Trp37, Met61 and Ile58; Ile38, Tyr54 and Leu43; and Leu2, Ile6 and Leu17). In addition, Tyr54 forms an intermolecular hydrogen bond with the backbone of Cys35 (lower panel). Each of the four symmetrical SapA monomers participates in the same interactions. B. Sequence alignment of mouse SapA (mSapA) and human SapA (hSapA) showing that the residues involved in these contacts (highlighted in yellow) are conserved (identical residues highlighted in yellow, conservative substitutions highlighted in pale yellow in the hSapA sequence).

Two structures of SapA in its open conformation have been previously published and can be compared to this structure. The first SapA nanodisc structure (PDB ID: 4DDJ; Fig. 1.4E) shows human SapA enclosing an ordered bilayer of the detergent lauryldimethylamine oxide (LDAO). Our group has also published a structure of mouse SapA in complex with the lipid processing enzyme GALC in the presence of LDAO (PDB ID: 5NXB; Fig. 1.4D). The open conformation of mouse SapA monomers within the novel tetrameric structure is almost identical to the open SapA conformations reported previously except for small loop movements (Fig. 5.11A). Human SapA within the nanodisc structure has a root mean square deviation (RMSD) from mouse SapA within the tetramer structure of 0.9 $Å^2$ over 80 Cα atoms, with 79.7% sequence identity; GALC-associated mouse SapA has an RMSD of 1.0 $Å^2$ over 80 Cα atoms, with 100% sequence identity. The dimer formed by chains A and B of the tetramer can be compared to the published SapA dimer structures (Fig. 5.11B). The orientation of SapA tetramer chains A and B relative to each other is similar to the head-to-head arrangement found in the GALC-associated SapA dimer. Both structures are maintained by intermolecular hydrophobic contacts near the loop formed by residues 38-42; however, some of the residues participating in these contacts and the position of the SapA monomers relative to each other are different. In contrast, SapA nanodiscs are formed by a head-to-tail arrangement of SapA dimers without contact between monomers, the entire assembly being held together by the hydrophobic acyl core (Fig. 1.4E)[81].

Unlike SapA-LDAO nanodiscs and LDAO-solubilised, GALC-bound SapA, the electron density provides no evidence that this SapA tetramer structure includes detergents or lipids. However, the core of this structure is hollow and potentially solvent-accessible if it is not filled with amphipathic molecules (Fig. 5.12, left panel). Considering the highly hydrophobic nature of the internal SapA surface (Fig. 5.12, right panel), exposure of this hollow core to solvent would be thermodynamically unfavourable. It is therefore likely that disordered PC and/or DDM molecules fill this internal cavity. Approximately 6 to 8 PC or DDM molecules may fit within this tetrameric SapA structure.

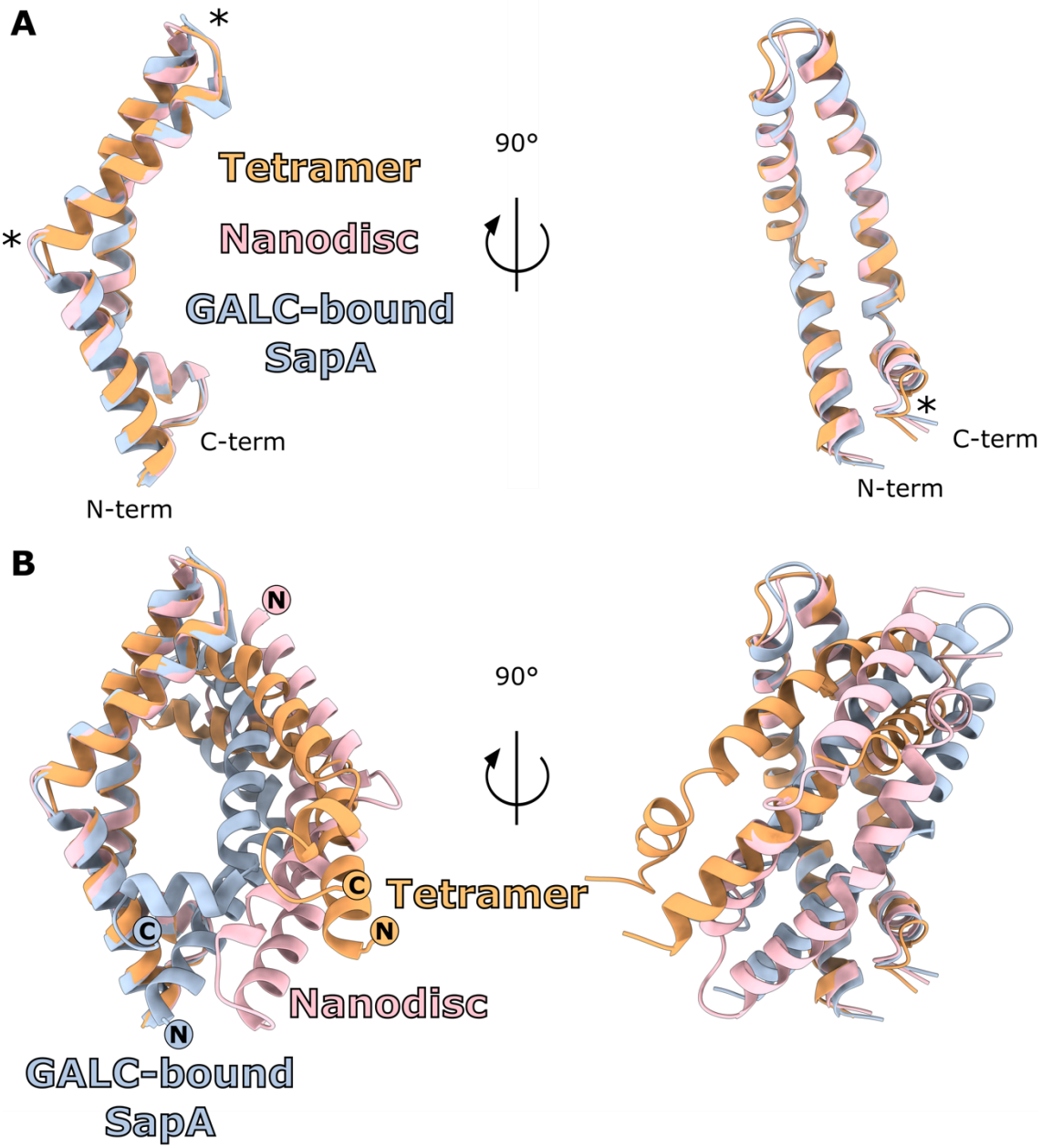

Figure 5.11. Comparison of SapA structures
A. The open conformation of mouse SapA monomers found in the novel tetrameric structure (in orange), human SapA-LDAO nanodiscs (in pink; PDB ID: 4DDJ) and GALC-associated, LDAO-solubilised dimeric mouse SapA (in blue; PDB ID: 5NXB) is almost identical, except for small movements indicated by (*). B. The dimer formed by chains A and B of the tetramer can be compared to the published dimeric SapA structures. In the tetramer, SapA monomers (orange) are arranged head-to-head in a manner similar to GALC-associated SapA dimer (blue); however, their position and the protein-protein contacts involved are different. In contrast, nanodiscs are made with a head-to-tail arrangement of SapA monomers (pink) and do not involve any protein-protein contacts. N- and C-termini are marked with circles.

The diamond-shaped tetrameric SapA assembly observed by crystallography is unsuitable for structural studies with AC as it does not provide a large lipid bilayer surface for interaction with AC.

An important question is whether this assembly exists in solution or is an artefact of crystallisation. SEC-MALS showed that the average molecular mass of SapA-PC assemblies in solution is 37.1 ± 0.5 kDa (Fig. 5.4; section 5.1.2). This is consistent with the theoretical mass of the SapA tetramer observed by X-ray crystallography (4 × 9.12 = 36.48 kDa). It is therefore possible that this tetrameric assembly was produced instead of the expected SapA nanodiscs. However, this molecular mass is also consistent with the mass of dimeric SapA nanodiscs with 23-29 PC molecules observed by Li *et al.*[282]. Therefore, it cannot be excluded that freshly prepared dimeric SapA nanodiscs underwent structural rearrangements in crystallisation conditions, forming the diamond-shaped tetrameric SapA assembly observed by crystallography.

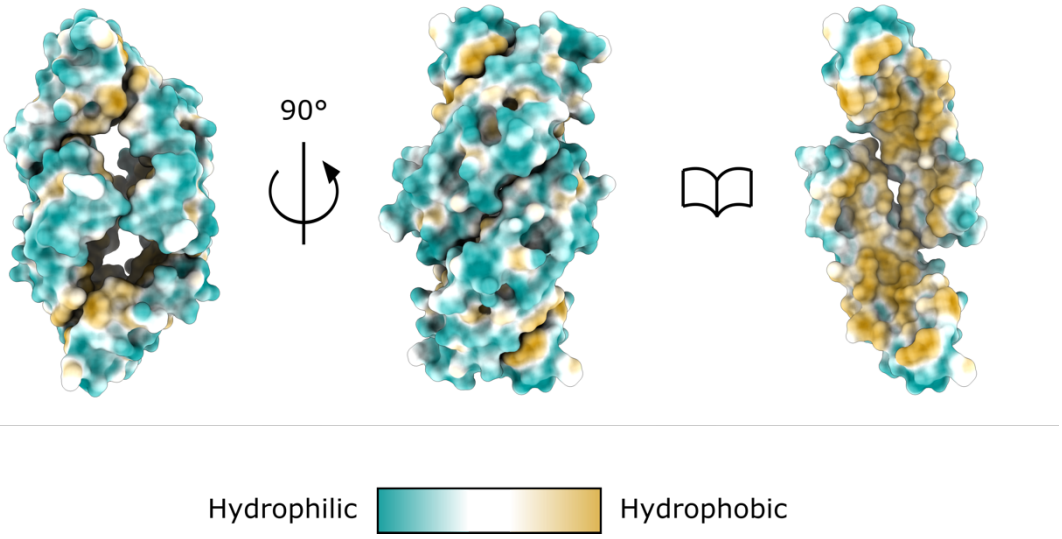

Figure 5.12. The core of the SapA tetramer is highly hydrophobic
A. Surface representation of the SapA tetramer coloured based on residue hydrophobicity. The left panel is in the same orientation as shown in Fig. 5.10A, left panel, while the central panel is rotated by 90° to be in the same orientation as Fig. 5.10A, right panel. The right panel maintains the orientation of the central panel, with two SapA chains removed to reveal the highly hydrophobic core of the SapA tetramer.

5.2 Can AC process glycosphingolipids?

The work described throughout the rest of this chapter was performed prior to the publication of a study demonstrating *in vitro* activity of AC towards GalCer and involving AC in Krabbe disease pathogenesis[284]. Prior to this report, several lines of evidence indicated that AC can deacylate GluCer[142-144], Gb3[142], and potentially GalCer when these substrates are accumulated in cells deficient in the hydrolytic enzymes that process these glycosphingolipids to ceramide, suggesting that AC can deacylate these bulkier substrates; however, no *in vitro* activity of AC against glycosphingolipid substrates had yet been reported.

The recently published structure of AC shows a hydrophobic channel leading to the active site which the authors proposed to be compatible with binding of ceramide, but not the bulkier head groups of glycosphingolipids[128]. However, glycosphingolipid processing enzymes can show plasticity around their active site to accommodate their substrate. Indeed, structural studies by our group showed that the active site pocket of GALC undergoes conformational changes upon binding of a substrate analogue[285]. Such plasticity around the active site could only be observed in a crystal structure of AC with a glycosphingolipid substrate mimic. Therefore, I aimed to test whether AC can hydrolyse glycosphingolipid substrates *in vitro* and obtain a crystal structure capturing this process.

5.2.1 AC C-terminal tag removal and autoproteolysis

The published structure of AC showed that the C-terminus of AC is positioned near the active site and that the C-terminal residue (Trp395) forms part of the hydrophobic area around the active site essential to AC activity[128]. Consequently, the C-terminal H_6 tag of the AC construct used in the work described so far (Fig. 5.1A) may obstruct the active site and hinder optimal AC function. Therefore, I removed the C-terminal H_6 tag of AC by treatment with Carboxypeptidase A. H_6-tagged AC was purified as documented in section 5.1.1 and incubated with carboxypeptidase A in 20 mM Tris pH 7.4, 300 mM NaCl for 4 or 15 hours. After 4 hours significant quantities of AC retained binding to Ni-NTA, indicating there was insufficient proteolysis (Fig. 5.13A), but after 15 hours, a large fraction of AC was no longer H_6-tagged and no longer interacted with Ni-NTA beads (Fig. 5.13B). Flow-through and wash fractions containing untagged AC were pooled and used for subsequent experiments. To ensure complete autoproteolysis of AC into its active, mature form, AC was incubated in pH 4 buffer at 37 °C for 27 hours. After this process, AC was almost fully cleaved (Fig. 5.13C).

5.2.2 Activity assay of untagged AC

I performed an AC activity assay using this untagged, mature AC sample. C12 ceramide (d18:1/12:0) was chosen as a substrate for this assay (Fig. 5.13D), as it has been shown to be the ceramide most efficiently hydrolysed by AC[203]. Although AC has a pH optimum of 4.5, it retains high activity at pH 4[129] therefore activity assays were performed at pH 4.

Several AC activity assay protocols have been reported in the literature (reviewed by Bedia *et al*.[286]). I opted for a protocol using unmodified ceramide as a substrate (rather than fluorescently labelled or radiolabelled lipids) coupled with detection by thin layer chromatography (TLC) followed by lipid visualisation by charring with sulfuric acid[287]. Labelled lipids can be detected with more sensitivity,

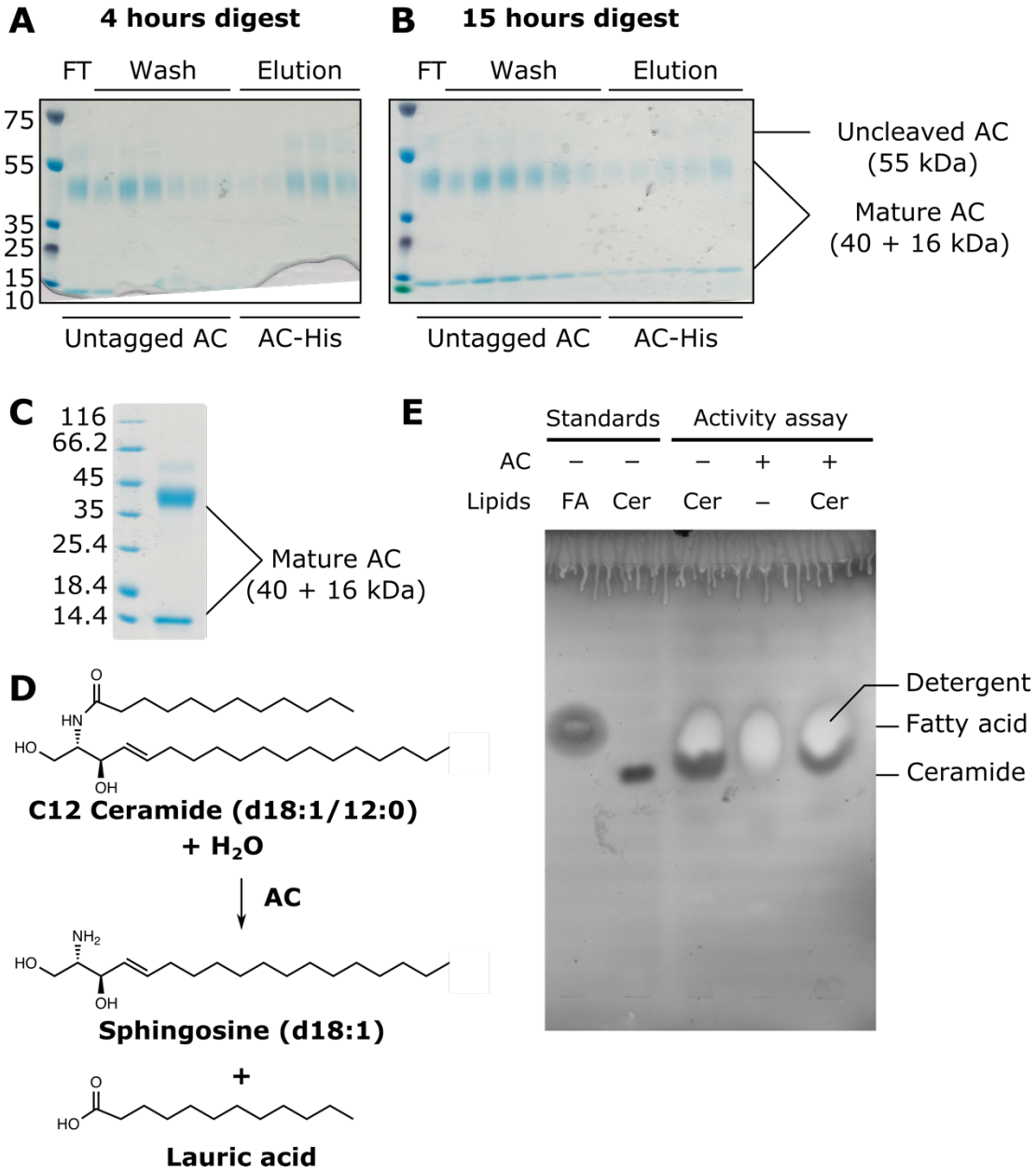

Figure 5.13. Activity assay of AC after H_6 tag removal

A-B. The C-terminal His_6 tag of AC was removed by incubation with carboxypeptidase A-agarose beads for 4 hours (A) or 15 hours (B). After this treatment, untagged AC was separated from His-tagged AC and free histidine by nickel affinity chromatography with Ni-NTA beads. Flow-through (FT) and wash fractions containing untagged AC, and elution fractions containing His-tagged AC were visualised on a Coomassie-stained 4-12% Bis-Tris gel. C. AC was incubated at pH 4 at 37 °C for 27 hours to allow for autoproteolysis, forming mature AC. Cleavage efficiency was assessed by visualising the sample on a Coomassie-stained 12% SDS-PAGE gel. D. Reaction catalysed by AC in this activity assay. C12 Ceramide (d18:1/12:0) is hydrolysed to sphingosine (d18:1) and lauric acid (C12:0 fatty acid). E. AC was incubated with detergent-solubilised ceramide. Lipid products were extracted and loaded onto the TLC plate. 10 µg of lipid standards were directly loaded onto the TLC plate for comparison. Sphingosine is expected to remain at the bottom of the plate with the solvent system used in this experiment. No formation of lauric acid or sphingosine was detected. Abbreviations: FA - fatty acid; Cer – ceramide.

facilitating enzyme kinetic studies; however, a qualitative output would be sufficient to answer whether AC degrades glycosphingolipids, so I chose a simpler and more economical detection method. I tested several TLC solvent systems for their ability to resolve ceramide. 80:10:1 (v/v/v) chloroform, methanol, water was found to be optimal[287].

Lipids must be solubilised for this activity assay, and a variety of detergents or detergent mixtures have been used previously (including Triton X-100 (TX-100), Tween 20, NP-40 and Sodium cholate[286]). Based on the literature, I tested several detergent mixtures for ceramide solubilisation. Visual inspection of the detergent-solubilised lipid solution showed that 1% TX-100 resulted in the clearest lipid solution, but it still remained somewhat cloudy even after sonication, indicating that ceramide was not fully solubilised. To promote lipid solubilisation prior to the activity assay, I followed the protocol I had used to solubilise lipids for SapA nanodiscs: lipids were vortexed and sonicated in activity assay buffer without detergent to promote liposome formation, then TX-100 was added to the solution, which was vortexed and sonicated further. Even with this protocol, ceramide did not fully solubilise.

Untagged, mature AC (4.8 µM) was incubated with TX-100-solubilised ceramide (at a concentration of 370 µM if ceramide was fully solubilised) at 37 °C for 4 hours. The reaction was stopped, and lipids were extracted and analysed by TLC as described above (Fig. 5.13E). Ceramide and the fatty acid reaction product lauric acid were used as standards. Although a broad detergent band overlaps with the ceramide and lauric acid positions and may disturb the separation of these species, I observed no formation of the hydrolysis products fatty acid and sphingosine, which is expected to remain at the bottom of the plate in this experiment. The lack of detectable AC activity may be the consequence of poor lipid solubilisation, making too little substrate available to AC, or blockage of the active site of the enzyme by additional residues at the C-terminal, even after H_6 tag cleavage. Indeed, this recombinant enzyme contains a leucine and a lysine residue between the end of the native human AC sequence and the H_6 tag.

5.2.3 N-terminally tagged AC expression, purification and tag removal

To overcome the potential problem of AC active site hindrance by a C-terminal tag, I used a different AC construct for subsequent structural studies and activity assays (Fig. 5.14A). N-terminally tagged AC is active and crystallisable[128], so I used a construct with a cleavable N-terminal Protein A (ProtA) tag. The tag is separated from AC by a Tobacco Etch Virus (TEV) protease site, allowing removal of the tag by digestion with TEV protease. The residues immediately following the signal sequence on the N-

terminus of AC (residues 22-26) were disordered in our group's preliminary low-resolution structure of AC (obtained by Dr Sam Spratley, Dr Janet Deane and Dr Stephen Graham), as well as the published structure[128]. These residues were not included in this AC construct (ProtA-AC27) intended for structural studies because truncation of disordered terminal disordered regions may improve protein crystallisation.

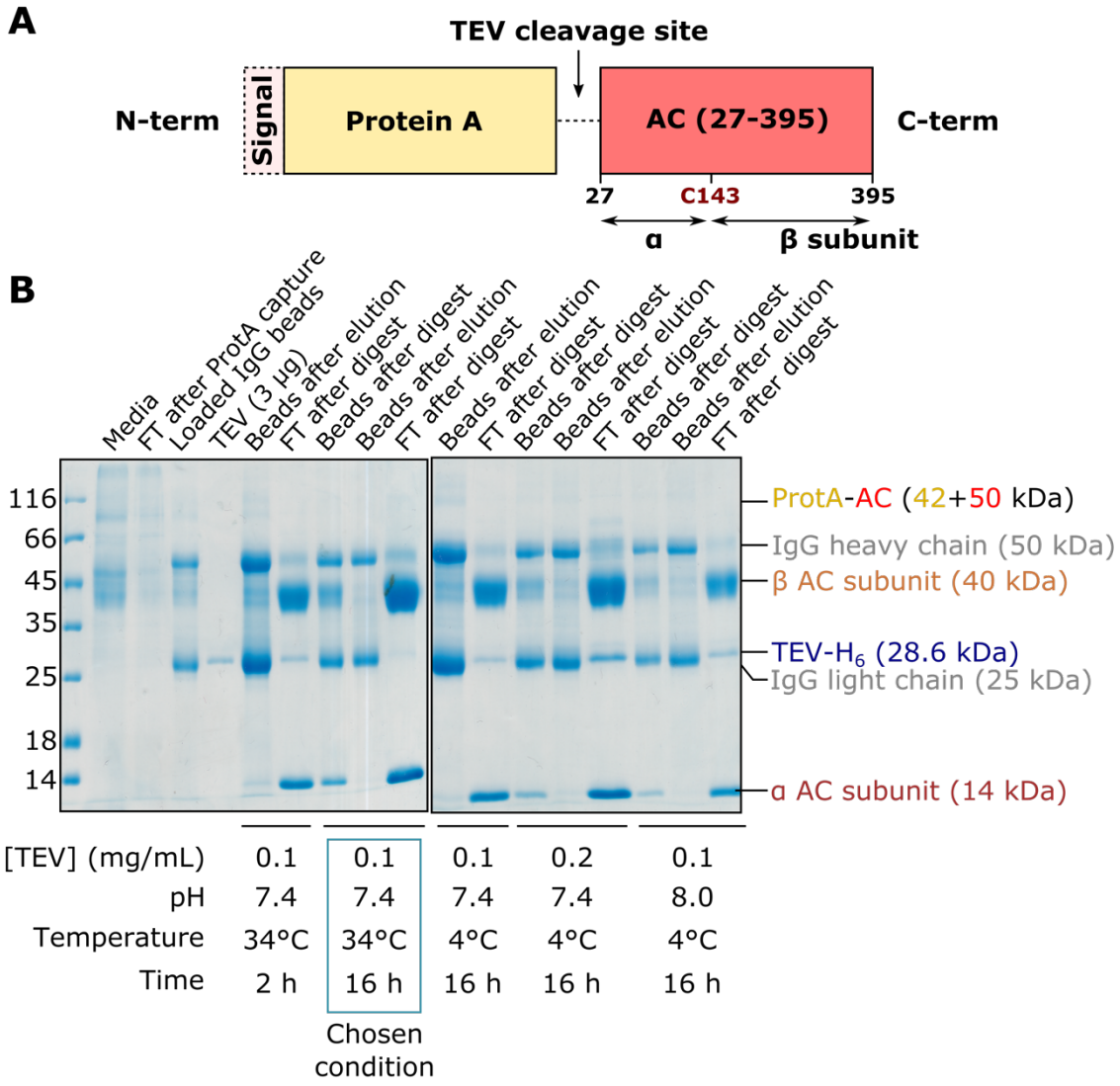

Figure 5.14. ProtA-AC27 tag removal optimisation
A. Recombinant AC was expressed in HEK293F cells. The construct encoded AC (residues 27-395) preceded by a cleaved secretion signal peptide, Protein A and a TEV cleavage sequence. B. ProtA-AC27 was captured from the media by IgG affinity purification. This was followed by on-bead TEV protease digest to cleave the tag. Small-scale test were performed to identify optimal conditions for TEV cleavage. Samples from each condition were analysed on a Coomassie-stained 12% SDS-PAGE gel. The indicated condition was chosen for subsequent experiments based on the high amount of untagged AC27 eluting after TEV cleavage in this condition.

A stable cell line expressing ProtA-AC27 was established by Dr Janet Deane using the PiggyBac expression system. I purified ProtA-AC27 from the media by IgG affinity purification. After protein capture from the media, ProtA-AC27 was subjected to on-beads digestion with TEV protease. I optimised the TEV cleavage step by changing the digestion time, temperature and pH of the reaction buffer. The best yield was obtained by overnight cleavage at 34 °C (Fig. 5.14B).

I performed a large-scale preparation of this construct using this cleavage protocol (Fig. 5.15A). After tag removal, AC27 was eluted from the beads by gravity flow. TEV protease is tagged with a H_6 tag, allowing its separation from AC27 by nickel affinity chromatography. After TEV removal, AC27 was further purified by size exclusion chromatography (Fig. 5.15B-C). The final yield of untagged AC27 was 3 mg from 500 mL conditioned media. AC was then allowed to self-cleave at pH 4 for a further 27 hours at 37 °C (Fig. 5.15D) prior to crytallisation trials (section 5.2.4) and activity assays (section 5.2.5).

5.2.4 Crystallisation trials of AC27 with psychosine

I hypothesised that AC may be able to accommodate the head group of GalCer in its active site pocket and deacylate this substrate to psychosine. I hoped to obtain a crystal structure of AC showing the conformational changes permitting this potential reaction. For these crystallisation trials, AC was mixed with the proposed reaction product psychosine because psychosine is soluble in aqueous buffer, whereas GalCer is not. Although enzymes have a stronger affinity for their substrate than their product, they do have some affinity for the product and can bind it at high enough product concentrations.

For crystallisation trials, AC27 was concentrated to its solubility limit in pH 4 buffer (4.1 mg/mL) and a psychosine stock was made in the same buffer with a concentration close to its solubility limit (2 mM). Equal volumes of the protein and psychosine solutions were mixed for a 20-fold molar excess of psychosine. This mixture was incubated for an hour to allow binding of psychosine to AC, then the mixture was concentrated until AC27 was at a concentration of 4.1 mg/mL based on protein absorbance at 280 nm. This mixture was then used in high-throughput crystallisation screens. Although approximately 400 crystallisation conditions were tested, no diffraction-quality crystals grew in these conditions.

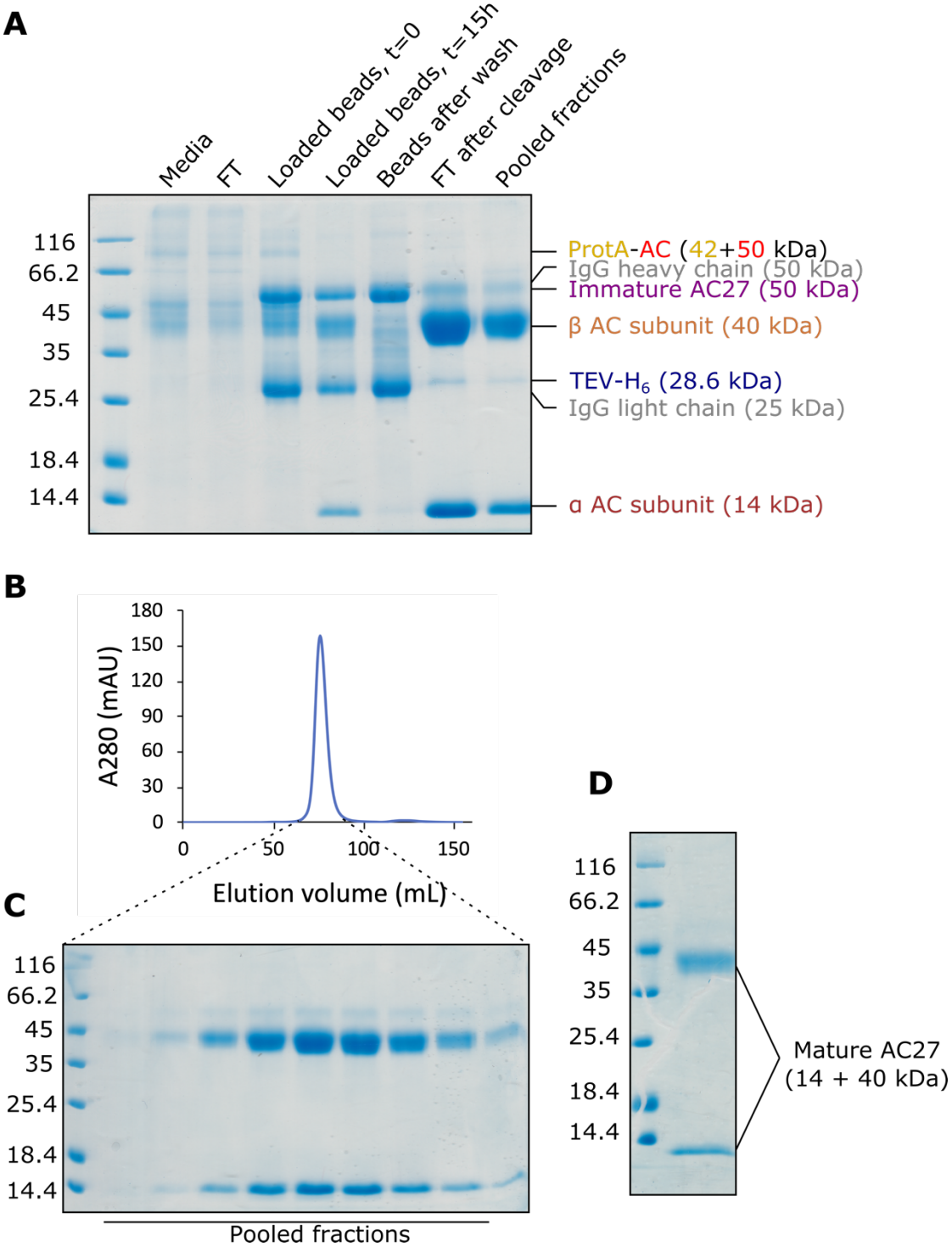

Figure 5.15. Large scale AC27 preparation

A. ProtA-AC27 was captured from the media by IgG affinity purification. The tag was then removed by on-bead TEV digestion. Samples from each step were analysed on a Coomassie-stained 12% SDS-PAGE gel. B. AC27 was further purified by size exclusion chromatography on a Superdex 200 16/600 column. C. SEC elution fractions were analysed on a Coomassie-stained 12% SDS-PAGE gel. The indicated fractions were pooled and concentrated. D. AC27 was incubated at pH 4 at 37 °C for 27 hours to allow for autoproteolysis, forming mature AC. Cleavage efficiency was assessed by visualising the sample on a Coomassie-stained 12% SDS-PAGE gel.

5.2.5 Activity assay of AC27

I repeated the AC activity assay with untagged AC27 after N-terminal ProtA tag removal with improvements to the lipid solubilisation protocol. Ceramide was resuspended in activity assay buffer with 1% TX-100. I heated the lipid solution at 85 °C for 30 seconds, followed by sonication for 30 seconds and vortexing for 10 seconds. This cycle was repeated 4 times. This resulted in clear lipid solution with very little particulate matter visible, indicating that the lipids were almost fully solubilised. After incubation of AC27 (3 µM) with ceramide (250 µM) at 37 °C for 4 hours, the lipids were extracted and analysed by TLC (Fig. 5.16). However, once again, no fatty acid and sphingosine reaction products were detected.

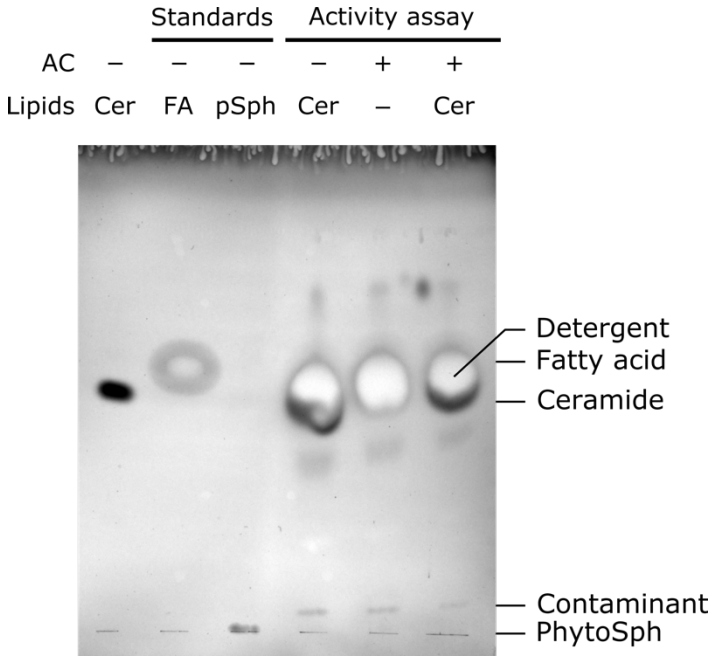

Figure 5.16. Activity assay of AC27
AC27 was incubated with detergent-solubilised ceramide. Lipid products were extracted and loaded onto the TLC plate. 10 µg of lipid standards were directly loaded onto the TLC plate for comparison. Sphingosine is expected to remain at the bottom of the plate with the solvent system used in this experiment, similarly to phytosphingosine (PhytoSph). No formation of lauric acid or sphingosine was detected in this assay. Abbreviations: FA - fatty acid; Cer – ceramide.

The difficulty in establishing a viable AC activity assay to test the hypothesis that AC deacylates glycosphingolipids combined with the absence of structural insights into the potential mechanism of glycosphingolipid deacylation led me to focus on a different, cell-based way to investigate whether AC contributes to Krabbe disease by processing GalCer to psychosine.

5.3 Role of AC in Krabbe disease

The pathogenesis of Krabbe disease is thought to be caused by accumulation of the cytotoxic sphingolipid psychosine in oligodendrocytes, causing their apoptosis and subsequent demyelination of neurons[108,109]. I hypothesised that in Krabbe disease, AC deacylates accumulated GalCer to psychosine, contributing to oligodendrocyte cell death. I aimed to test this hypothesis by looking at the effect of AC inhibition on apoptosis levels of GALC knock-out (KO) oligodendrocytes. Our lab has generated a GALC KO MO3.13 oligodendrocyte cell line by CRISPR/Cas9. Loss of GALC in this cell line was validated by sequencing and activity assays (Dr Stuart Fawke and Dr Janet Deane).

First, I sought to establish robust apoptosis assays in this cell line. Apoptosis of GALC knock-down MO3.13 oligodendrocytes[288] or MO3.13 oligodendrocytes fed with exogenous psychosine[214] has been observed by caspase-3 (Casp-3) activation assays, Terminal Deoxynucleotidyl Transferase dUTP Nick End Labelling (TUNEL) and DNA laddering assays 3 days after differentiation of these cells. I attempted to replicate these assays to assess the extent of oligodendrocyte apoptosis after differentiation of GALC KO MO3.13 oligodendrocytes with phorbol 12-myristate 13-acetate (PMA). I assayed Casp-3 activation, an early marker of apoptosis, by Western blotting with an antibody specifically recognising cleaved (active) Casp-3. No Casp-3 activation was detected in wild-type (WT) or GALC KO oligodendrocytes, 3 or 7 days after differentiation (Fig. 5.17). As a positive control, differentiated WT oligodendrocytes were treated with the non-selective kinase inhibitor staurosporine at 2 μM for 24 h, which has previously been shown to induce some Casp-3 activation in this cell line[289], but I did not detect any Casp-3 activation in this positive control, suggesting that antibody staining not be optimal.

TUNEL is an alternative method for the quantification of apoptotic cells. During apoptosis, genomic DNA is cleaved into 200 base pair (bp) fragments that can be labelled by TUNEL. I detected late apoptotic cells by TUNEL followed by flow cytometry in WT and GALC knockout oligodendrocytes 7 days after differentiation (Fig. 5.18). WT oligodendrocytes were treated with staurosporine as a positive control. Detached cells were included in the experiment because apoptosis often causes cell detachment. WT oligodendrocytes did not undergo significant apoptosis during this assay (1.05%). Following treatment with staurosporine, 13% of WT cells were apoptotic, indicating that the assay could detect apoptosis. However, levels of apoptosis similar to WT control were observed in GALC KO cells. DNA fragments can also be extracted from apoptotic cells and visualised by agarose electrophoresis. The DNA of apoptotic cells forms a ladder, whereas the DNA of necrotic cells presents as a smear due to non-specific cleavage. I followed a protocol previously used to detect apoptosis in differentiated MO3.13 oligodendrocytes after GALC knock-down[214]. I extracted DNA from WT and

GALC KO oligodendrocytes before differentiation and 1, 2, 3, 6 and 7 days after differentiation. No DNA laddering was detected in WT (data not shown) or GALC KO oligodendrocytes (Fig. 5.19), before or after differentiation. However, no DNA laddering was observed in the staurosporine-treated positive control, indicating that this assay is not sensitive enough to detect low levels of apoptosis or that staurosporine-treated MO3.13 cells undergo necroptosis rather than apoptosis, as suggested by the presence of a faint smear in this sample.

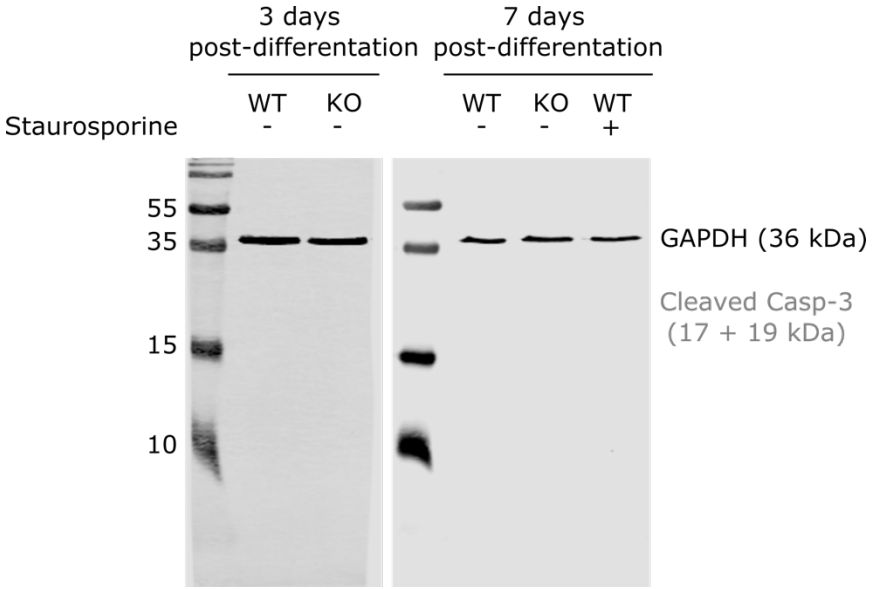

Figure 5.17. Casp-3 activation was not detected in WT or GALC KO oligodendrocytes
Casp-3 activation was assayed by Western blotting in WT and GALC KO oligodendrocytes, 3 and 7 days after differentiation. WT oligodendrocytes were treated with the kinase inhibitor staurosporine as a positive control.

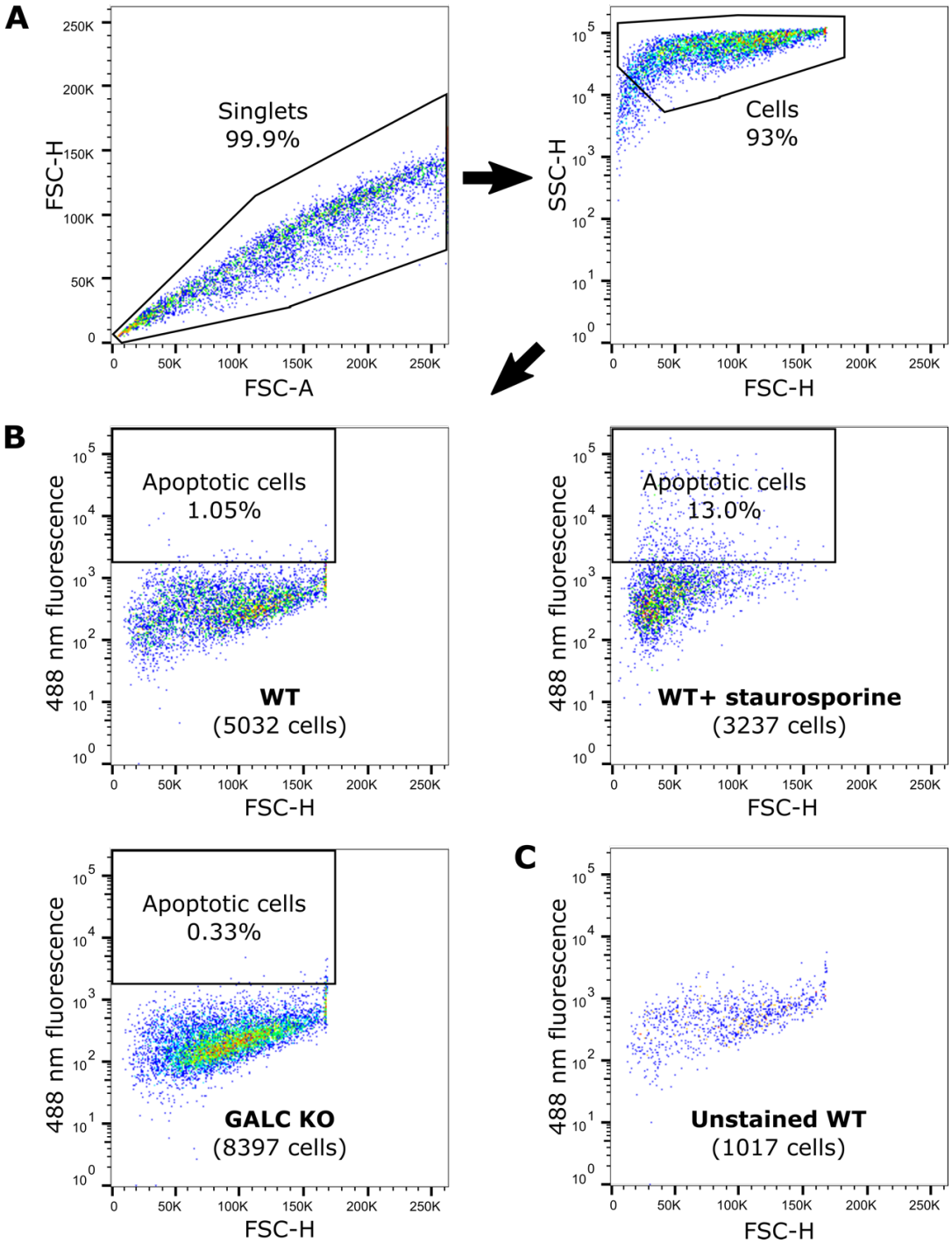

Figure 5.18. No apopotosis was detected in GALC KO oligodendrocytes by TUNEL assay
DNA fragmentation was assayed in WT and GALC KO oligodendrocytes 7 days after differentiation. WT oligodendrocytes were treated with the kinase inhibitor staurosporine as a positive control. Cells were labelled by TUNEL and stained with AlexaFluor488 for analysis by flow cytometry. A. Representative gating strategy (WT cells): singlets were gated through the forward scatter height (FSC-H) vs forward scatter area (FSC-A) plot, then cells were defined on side (SSC-H) vs forward scatter height (FSC-H) plots to exclude small debris. B. Apoptotic cells are detected by fluorescence at 488 nm. Only cells with very high fluorescence intensity were gated, because MO3.13 oligodendrocytes used in this assay are highly autofluorescent at 488 nm, as shown in C. The percentage of apoptotic cells is indicated. C.Unstained WT oligodendrocytes (gated as in A.) are autofluorescent at 488 nm. The voltage of the fluorescence detector was set higher than in B.

Interestingly, these results suggest that despite loss of GALC activity and accumulation of lipid substrates in GALC KO oligodendrocytes (verified by Dr Janet Deane and collaborators), there is no evidence of significant cell death contrary to expectations based on previously published work. Consequently, inhibiting psychosine production through AC inhibition in GALC KO cells would not result in detectable changes in cell death levels.

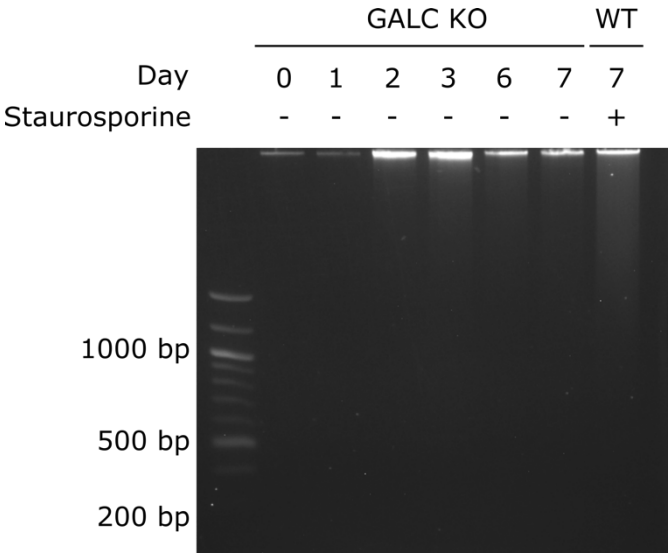

Figure 5.19. No DNA fragmentation was detected in GALC KO oligodendrocytes
DNA fragmentation was assayed by gel electrophoresis of DNA extracted from undifferentiated and differentiated GALC KO oligodendrocytes, 1 to 7 days after differentiation. WT oligodendrocytes were treated with the kinase inhibitor staurosporine as a positive control.

5.4 Discussion

How AC interacts with lipid bilayers to access its membrane-embedded substrate is a key unresolved question in our understanding of the mechanism of AC function. I probed the interaction of AC with membranes using a recently developed technique exploiting the lipid-binding properties of SapA to form lipoprotein nanodiscs[93,94]. Unfortunately, AC did not co-crystallise or interact with nanodisc-enclosed lipid bilayers. Several factors may have hindered this interaction. The recombinant AC protein used in these experiments included a C-terminal H_6 tag. When the coordinates of the first published AC structure became available, these showed that the C-terminus of AC is close to the active site and part of the hydrophobic area essential for AC activity[128]. It is therefore possible that the addition of histidine residues, charged and hydrophilic at pH 4, would disrupt the hydrophobic association of AC with membranes. An N-terminally tagged AC construct would be more suited to study the interaction of AC with membranes. Moreover, although I produced SapA nanodiscs in a manner similar to published protocols, I obtained an unexpected tetrameric SapA assembly very

different from the dimeric SapA-lipid nanodisc I was aiming to produce. This assembly does not present a membrane surface available for interactions with AC.

The production of a tetrameric SapA assembly instead of SapA nanodiscs could be explained by two differences between my protocol and the original protocols for nanodisc formation. (1) Previous protocols use human SapA for nanodisc formation, whereas mouse SapA was used in this study. However, these orthologs are highly similar, and the residues involved in the interactions maintaining this tetrameric assembly are identical or similar in their hydrophobic properties between human and mouse SapA. It is therefore unlikely that this deviation accounts for the difference in the final product. (2) I prepared SapA nanodiscs at pH 4, which promotes the interaction between AC and membranes, whereas SapA nanodiscs were formed at pH 4.8 to 7.5 in previous publications[93,94,282]. Li *et al*.[282] observed that the oligomeric state of SapA within nanodiscs is dependent on the final pH of the nanodisc solution, rather than the pH at which these were formed. This supports the idea that SapA assemblies are highly dynamic, adopting distinct oligomeric states depending on the buffer condition. In fact, I could not conclusively show that freshly prepared nanodiscs in solution adopted the tetrameric conformation observed by crystallography rather than the expected dimeric nanodisc conformation, since the predicted molecular mass of both assemblies is similar making them indistinguishable by SEC-MALS. It is therefore possible that freshly prepared nanodiscs initially adopted a dimeric SapA nanodisc conformation, then underwent a structural rearrangement in various crystallisation conditions to form the tetrameric SapA assembly observed by X-ray crystallography. Alternatively, the initial mixture may have contained both dimeric SapA nanodiscs and SapA tetramers, but only the latter self-assembled into crystals. The ability of SapA to form variable oligomeric assemblies may have relevance for its function as a lipid transfer protein in the lysosome.

Besides SapA nanodiscs, other lipoprotein and polymer systems have been developed for the study of membrane proteins, such as apolipoprotein A-based nanodiscs[290] and styrene maleic acid lipid particles (SMALPs)[291], but these methods are not adapted to the low pH stimulating the interaction of AC with membranes. However, recently developed technologies such as acid-compatible SMALPs[292] can function at low pH and may hold promise for the elucidation of the mechanism of AC function.

The metabolic pathway involved in the formation of psychosine in Krabbe disease oligodendrocytes has long remained a mystery. Based on recent reports showing that AC can deacylate the glycosphingolipids glucosylceramide (GluCer) and Gb3[142-144], I hypothesised that AC can also deacylate

accumulated GalCer. To demonstrate this, I attempted to implement an AC activity assay using a simpler, less expensive but less sensitive lipid detection method than previously described AC activity assays[286]. Unfortunately, this method was unsuccessful, and a more sensitive detection method may have been necessary to detect the lipid products, for example through the use of fluorescently labelled substrate. It is possible that the enzyme I used in these assays is not functional: the first recombinant AC protein used in these assays is likely to possess at least one additional hydrophilic residue at the C-terminal even after tag removal, potentially disrupting the essential hydrophobic surface near the active site of AC[128]. The second recombinant AC protein tested was intended for structural studies and did not include the first five residues of AC. Both of these changes could have disrupted AC activity. A recently published study did in fact show activity of AC against fluorescently labelled, detergent solubilised GalCer *in vitro*[284]. In support of the role of AC in the production of psychosine and its importance for the pathogenesis of Krabbe disease, AC was shown to be necessary for psychosine accumulation in Twitcher mice, a mouse model of Krabbe disease, and genetic ablation of AC eliminated Krabbe disease symptoms in this model[284]. Inhibition of AC activity is therefore a promising therapeutic target for Krabbe disease. The AC inhibitor carmofur, approved for clinical use as a chemotherapeutic agent, alleviates disease in Twitcher mice if one allele of the gene encoding AC (*ASAH1*) is knocked-out. However, carmofur is not sufficient to alleviate symptoms in Twitcher mice with normal AC expression, and carmofur is associated with significant drug toxicity[140]. Therefore, an improved AC inhibitor targeting the hydrolysis of GalCer but not ceramide could specifically reduce psychosine formation, alleviate Krabbe disease symptoms and present a lower toxicity profile. Published AC structures and the proposed catalytic mechanism of AC[128,141] cannot explain its activity on glycosphingolipid substrates with a bulky head group because of clashes between the proposed catalytic residues and the head groups of glycosphingolipids modelled into the active site by molecular docking[128]. It is therefore likely that plasticity of the active site allows AC to accommodate lipids with bulky head groups and/or that AC employs an alternative catalytic mechanism for the hydrolysis of glycosphingolipids. A structure of glycosphingolipid-bound AC would be valuable for structure-based design of inhibitors specifically targeting deacylation of glycosphingolipids.

Such a therapeutic approach may be beneficial for other diseases because AC has been involved in deacylation of several glycosphingolipid substrates. AC promotes formation of lyso-Gb3 in Fabry disease[142], a sphingolipidosis caused by loss of α-galactosidase A associated with neuropathy and organ failure. Lyso-Gb3 has been identified a risk factor for complications of Fabry disease[293] and shown to sensitise nociceptive neurons, possibly causing or contributing to Fabry-disease associated neuropathy[294]. Moreover, AC promotes formation of glucosylsphingosine in the sphingolipidosis

Gaucher disease caused by loss of glucocerebrosidase (GCase)[142-144]. Mutations in the gene encoding GCase are the highest risk factor for Parkinson's disease[295] and glucosylsphingosine is thought to promote α-synuclein aggregation[296], a hallmark of Parkinson's disease. AC inhibition has been shown to reduce the accumulation of α-synuclein in neurons derived from a patient with Gaucher disease[143]. Considering the ability of AC to process glycosphingolipids with distinct head groups, such as GalCer, GluCer and Gb3 (Fig. 1.9), AC may also contribute to the formation of other lysosphingolipid species, such as lyso-sulfatide and lyso-sphingomyelin. It is currently unknown whether AC processes glycosphingolipid substrates only when these are accumulated to abnormal levels, or if this is a process occurring at low levels under normal conditions, facilitating the formation of small quantities of bioactive lyso-glycosphingolipids playing a role in normal cell function.

The accumulation of psychosine in GALC-deficient oligodendrocytes is thought to cause oligodendrocyte cell death, driving the demyelination and neurodegeneration observed in Krabbe disease[108,109]. This hypothesis is strongly supported by studies showing that GALC knock-down or feeding of extracellular psychosine directly causes apoptosis in differentiated MO3.13 oligodendrocytes[214,288]. In contradiction with these results, I did not detect any apoptosis of differentiated GALC KO MO3.13 oligodendrocytes with the apoptosis assays used in previous studies. Exogenous psychosine feeding experiments may not reflect GALC loss-of-function phenotypes, because psychosine concentration and cellular localisation may differ. However, the phenotype of GALC KO cells should be similar to or more severe than GALC knock-down cells. It is conceivable that lentiviral knock-down of GALC in the reference study[288] resulted in off-target effects increasing susceptibility to cell death. It has certainly been our experience that the generation of cell lines by lentiviral transduction frequently results in severe off-target effects, as illustrated in Chapter 4. Alternatively, knocking out GALC by transient transfection of Cas9 and gRNA may have introduced off-target Cas9 effects reducing susceptibility to cell death in our cell line. However, a second GALC KO MO3.13 oligodendrocyte cell line generated using a distinct gRNA against GALC also did not show noticeably elevated cell death levels, based on regular trypan blue staining of this cell line. It remains unclear why our observations differ from published results.

The absence of spontaneous cell death in GALC KO oligodendrocytes lead us to ask how loss of GALC results in oligodendrocyte cell death and demyelination. This has been a key area of focus in our lab and some of this work will be described in Chapter 6.

6 Understanding the interaction of PTPRM with lipids

Previous work and summary

The causes of demyelination and oligodendrocyte cell death in Krabbe disease are still unelucidated. It has been hypothesised that psychosine accumulation due to loss of GALC activity is cytotoxic and directly causes oligodendrocyte cell death. However, as shown in Chapter 5, genetic ablation of GALC in the MO3.13 oligodendrocyte cell line does not trigger cell death. The myelin sheath is formed by wrapping of oligodendrocyte membranes around nerve axons, a process dependent on adhesion molecules and other plasma membrane proteins. Aberrant accumulation of galactosphingolipids in GALC-deficient oligodendrocytes may disrupt this process by changing the expression, stability or trafficking of plasma membrane proteins, leading to defective myelination and subsequent oligodendrocytes cell death. Our group has investigated the consequences of galactosphingolipid accumulation on the plasma membrane proteome of GALC KO oligodendrocytes following the approach described in Chapter 4. Plasma membrane profiling was combined with whole cell proteomics and RNA sequencing to identify proteins down-regulated at the plasma membrane of GALC KO oligodendrocytes. This approach revealed that the adhesion protein Protein Tyrosine Phosphatase Receptor Type μ (PTPRM) is specifically down-regulated at the plasma membrane of GALC KO oligodendrocytes while its total cellular levels and RNA levels are unchanged. This suggests that aberrant galactosphingolipid accumulation hinders normal trafficking of PTPRM to the plasma membrane. We hypothesised that PTPRM specifically interacts with galactosphingolipid headgroups, leading to its intracellular retention in GALC-deficient oligodendrocytes. In support of this hypothesis, recent work in the Deane lab has shown that purified PTPRM Fn3-4 (the membrane proximal fibronectin III domains 3 and 4 of the extracellular domain) interacts with nitrocellulose-immobilised sulfatide on a sphingolipid array (Fig. 6.1A). Sulfatide is a galactosphingolipid formed by sulfatation of galactosylceramide and is a major myelin component[103]. Accumulation of sulfatide occurs in metachromatic leukoystrophy, a sphingolipidosis with very similar clinical features to Krabbe disease[297]. As sulfatide lies upstream of GalCer in the degradative pathway (Fig. 1.3) it may also accumulate in GALC KO oligodendrocytes. Potential interactions with lipid headgroups are likely to be mediated by the membrane-proximal fibronectin III domain 4 (Fn4) of PTPRM. A structure of the extracellular domain of PTPRM is available[298] but the Fn4 domain was not observed in this structure (Fig. 6.1B). Our group has recently obtained a first structure of the membrane-proximal Fn4 domain (Fig. 6.1C). This structure revealed an atypical fibronectin III fold with large membrane-oriented loops possibly involved in interactions with lipid headgroups. To probe the details of this potential

interaction, I aimed to obtain a crystal structure of the membrane-proximal domains of PTPRM interacting with a galactosphingolipid analogue. Although I was not successful in obtaining a structure of ligand-bound PTPRM, I solved and refined a second structure of the Fn3-4 domains of PTPRM which provided intriguing preliminary insights into its function and regulation. Additionally, I tested a microscopy-based approach to monitor PTPRM trafficking in GALC KO oligodendrocytes.

Statement of acknowledgement

This project has received numerous contributions from other members of the lab. Work preceding this project was performed by Dr Stuart Fawke, Eve Caroe, Dr Stephen Graham and Dr Janet Deane. Eve Caroe has performed the preparative work for this project, including the cloning of all expression constructs described in this chapter and the first protein purifications. Her contributions are acknowledged throughout this chapter. X-ray diffraction data for the structure described in this chapter were collected by Dr Janet Deane and Dr Stephen Graham. Alex Nicholson and Dr Shannon McKie have performed subsequent PTPRM interaction experiments with lipids and lipid headgroups, informing the discussion of this chapter. Iain Hay has provided valuable knowledge about R2B family phosphatases. Finally, Dr David Gershlick (Cambridge Institute for Medical Research) provided the reagents and protocols for RUSH experiments.

6.1 Purification of PTPRM Fn3-4

PTPRM forms *trans* homophilic interactions between opposite membranes at adherens junctions. The rigid, ruler-like ectodomain of PTPRM allows it to act as a sensor of intercellular distances and stabilise the distance between membranes[298]. This locks the intracellular phosphatase activity at adherens junctions, where PTPRM can dephosphorylate substrates such as cadherins and catenins[299]. PTPRM possesses one MAM (meprin/A5/μ), one immunoglobulin-like and four fibronectin III domains combined with an uncharacterized intracellular juxtamembrane domain, an active phosphatase domain (D1) and an inactive phosphatase domain (D2) (Fig. 6.2A). The membrane-proximal Fn4 domain hypothesised to interact with lipid headgroups is unstable when expressed on its own (Radu Aricescu, MRC Laboratory of Molecular Biology, personal communication). We therefore expressed the Fn3 and Fn4 domains of PTPRM (residues 478-723) with an N-terminal secretion sequence and a C-terminal His_6 tag (Fig. 6.2B) by transient transfection in HEK293F cells. The cloning and first preparations of this construct were performed by Eve Caroe. Following the same purification protocol, I purified PTPRM Fn3-4 from expression media by nickel affinity purification (Fig. 6.2C) followed by size exclusion chromatography (Fig. 6.2D-E). This construct appears as a 35-45 kDa smear of higher

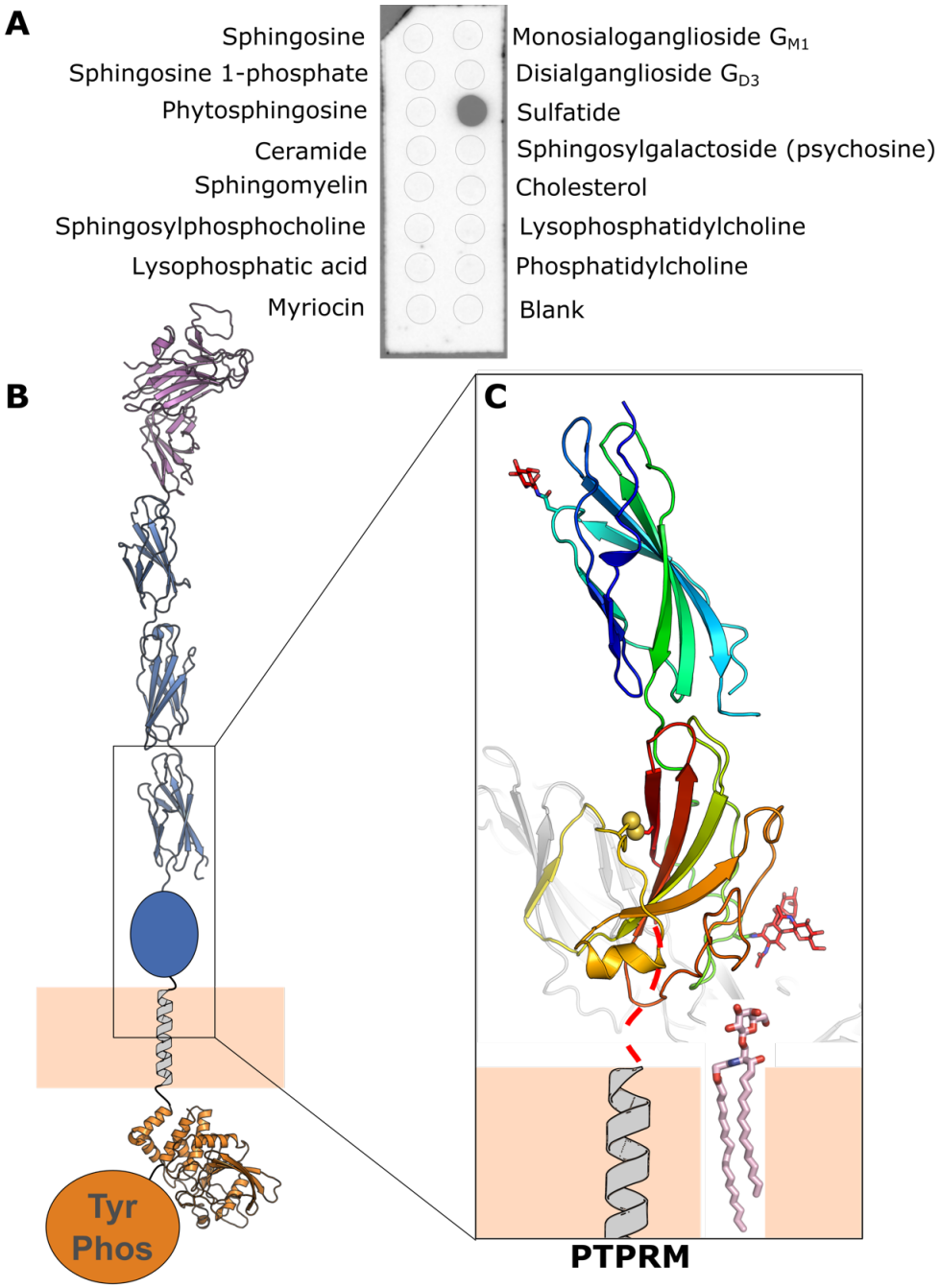

Figure 6.1. The membrane-proximal domain of PTPRM may interact with lipids
Figure courtesy of Dr Janet Deane, reused with permission. A. Purified PTPRM Fn3-4 was used to probe a sphingolipid array (work performed by Eve Caroe). PTPRM interacts with sulfatide in this assay. B. Schematic representation of the structure of PTPRM. The published structure of the extracellular domain of PTPRM without the membrane-proximal domain is shown as a purple and blue ribbon (PDB ID: 2V5Y). The structure of one of the two intracellular tyrosine phosphatase (Tyr Phos) domains is shown as an orange ribbon (PDB ID: 1RPM). C. The crystal structure of the membrane-proximal fibronectin III 3 and 4 domains obtained by our group is shown as a ribbon, rainbow-coloured from the N-terminus (blue) to the C-terminus (red). A 22 amino-acid linker from C-terminus of this construct to the transmembrane region of PTPRM, not included in this construct, is represented as a red dashed line. A second protein chain present in the crystal is shown in grey. Glycans are shown in stick representation (oxygen, red; nitrogen, blue). A membrane-embedded galactosylceramide molecule modelled in an interaction with PTPRM Fn3-4 is shown in stick representation.

molecular weight than predicted based on its sequence (29 kDa) on SDS-PAGE gels due to the presence of at least two potentially heterogeneous N-linked glycans.

During endogenous expression of PTPRM, a fraction of PTPRM molecules is normally cleaved by the proprotein convertases furin and PC5[300]. This processing event occurs within the Fn4 domain and both proteolytic fragments remain associated after cleavage[301]. Furin and PC5 are isozymes cleaving at the furin cleavage site Arg-X-[Arg/Lys/X]-Arg-↓-X, where X can be any amino-acid[302]. In PTPRM, this motif corresponds to the sequence Arg-Pro-Arg-Arg-Thr, with the cleavage occurring between Arg632 and Thr633. Although there are potential N-glycosylation sites on the C-terminal cleavage fragment[298], the preliminary structure of PTPRM Fn3-4 obtained in our lab suggests the absence of N-linked glycans on this fragment, which is therefore expected to have a molecular weight of 11 kDa calculated from its protein sequence. A faint band likely corresponding to the C-terminal fragment is indeed visible around 11 kDa (Fig. 6.2E). Based on the apparent molecular weight of uncleaved PTPRM (37 kDa), the glycosylated N-terminal fragment is expected to have a molecular weight of approximately 26 kDa and a faint band of this size is also apparent. An additional faint band is visible at 18 kDa, possibly corresponding to a glycosylated fraction of the C-terminal fragment (Fig. 6.2E). Altogether, this suggests that a very small fraction of PTPRM expressed in HEK293F cells is proteolytically processed by furin/PC5. This small proportion is expected since the Fn3-4 construct is expressed at much higher levels than endogenous levels and directly secreted by the cell instead of remaining embedded within the plasma membrane.

A significant amount of protein was lost when PTPRM-bound Ni-NTA beads were washed prior to protein elution (Fig. 6.2C, "Wash" lane), possibly because the His_6 tag is partially buried, reducing the affinity of the construct for Ni-NTA. The wash buffer contains 20 mM imidazole to outcompete binding of contaminant proteins weakly interacting with the resin through histidine residues. Unfortunately, the wash buffer appears to also outcompete binding of PTPRM to the beads, resulting in loss of some protein during washes. However, reducing the imidazole concentration significantly reduced the purity of the final product, so I chose to maintain a concentration of 20 mM imidazole in the wash buffer despite the reduced yield. Despite these losses, I obtained an excellent yield of 1 mg of highly pure protein from a 100 mL HEK293F culture with this protocol.

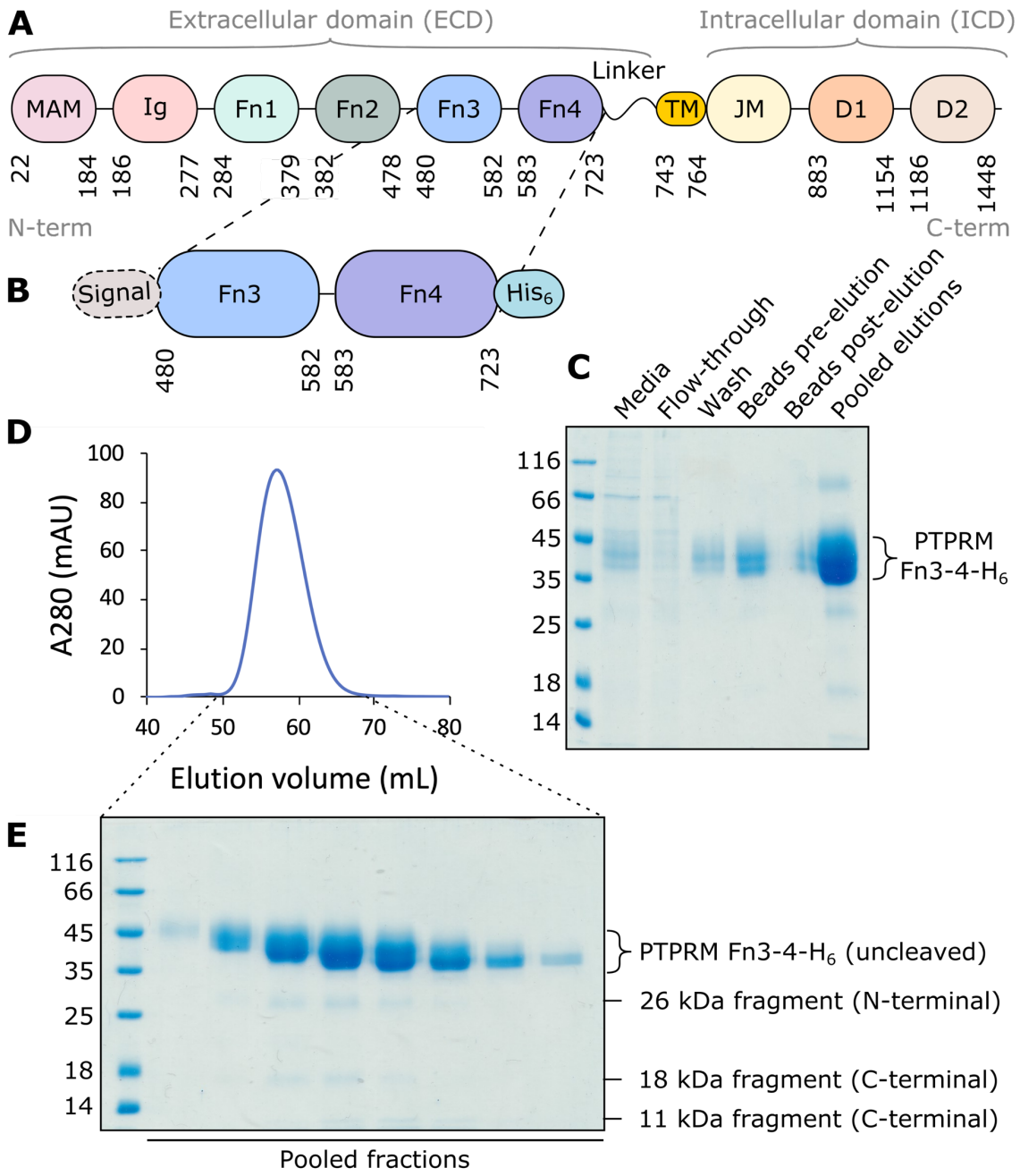

Figure 6.2. Human PTPRM Fn3-4 expression and purification

A. The ectodomain of PTPRM includes a MAM and Ig-like domain, followed by four fibronectin III (Fn) domains and a transmembrane (TM) region. The intracellular region of PTPRM includes a juxtamembrane domain, an active phosphatase D1 domain and an inactive D2 domain. B. The Fn3-4 domains (residues 480-723) of PTPRM were expressed with a C-terminal hexahistidine (H_6) tag in HEK293F cells. An N-terminal signal peptide targets this construct for secretion into the media and is cleaved off during expression. C. PTPRM Fn3-4 was captured from the media with Ni-NTA agarose and eluted with 300 mM imidazole in TBS. Fractions were analysed on a Coomassie-stained 12% SDS-PAGE gel. D. Eluted fractions were pooled, concentrated, and PTPRM Fn3-4 was further purified by size exclusion chromatography (SEC) using a Superdex 75 16/600 column. E. SEC elution fractions were analysed on a Coomassie-stained 4-12% Bis-tris SDS-PAGE gel. The indicated fractions were pooled, concentrated, and used in downstream experiments. Potential fragments formed by proprotease convertase cleavage (discussed in the text) are indicated.

6.2 X-ray crystallography

I aimed to obtain a crystal structure of PTPRM with a galactosylceramide or sulfatide headgroup analogue. Sulfatide is formed by addition of a sulfate group to the 3-position of the galactose moiety of galactosylceramide (Fig. 6.3A-B). Unfortunately, synthetic 3-*O*-sulfated galactose (Gal-3S) is prohibitively expensive for use in structural studies. Therefore, I used galactose and the Gal-3S structural isomer 4-*O*-sulfated galactose (Gal-4S; Fig. 6.3C) as galactosylceramide and sulfatide analogues. Crystal structures of ligand-bound protein can be obtained in two ways. (1) If diffraction-quality protein crystals have been grown and are reproducible, these can be soaked with ligand prior to crystal freezing and data collection. This simple method allows to screen for binding of several ligands but depends on the accessibility of the ligand-binding site through solvent channels within the crystal. (2) Co-crystallisation screens can be set up with protein and ligand mixed prior to plating. This approach is less efficient because new crystals are of unknown quality and may require optimisation, however it allows the formation of different crystal forms which may not be obtainable using the first method, especially if the ligand binding site is inaccessible or ligand binding causes large conformational changes. I pursued both strategies for structural studies of PTPRM with lipid analogues.

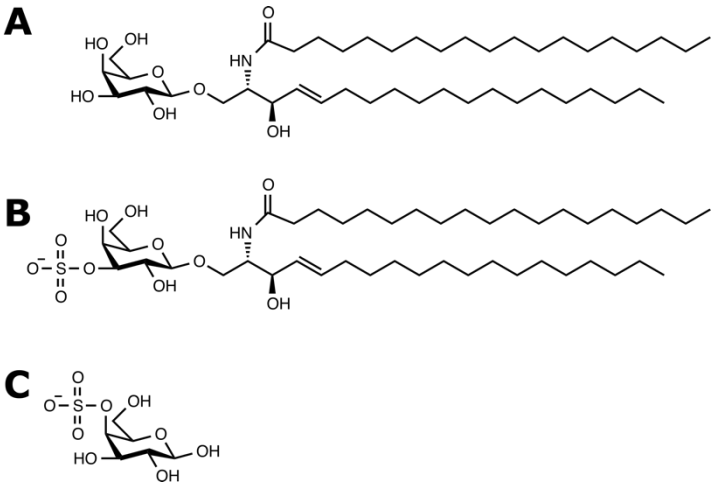

Figure 6.3. PTPRM may interact with galactosphingolipids including sulfatide
A. Skeletal formula of galactosylceramide. B. Skeletal formula of sulfatide. Sulfatide is formed by 3-*O*-sulfation of galactosylceramide. C. 4-O-sulfated galactose is a structural isomer of the sulfatide head group.

6.2.1 Optimisation of the previously obtained crystal

A crystal of PTPRM Fn3-4 diffracting to 2.89 Å was previously grown by Eve Caroe against a reservoir condition containing 20% (w/v) PEG 3350, 0.2 M ammonium nitrate (Fig. 6.4A). However, Eve Caroe's attempts to reproduce and optimise these crystals in optimisation screens, as well as in the original

crystallisation condition, were unsuccessful. Instead, these optimisation trials yielded drops with a multitude of microcrystals (Fig. 6.4B). This suggests that the latter crystallisation condition permitted too many crystal nucleation events, yielding many small crystals, whereas crystal nucleation in the original crystallisation condition was a fortuitous event possibly dependent on an unpredictable factor such as the presence of a contaminant in the crystallisation condition.

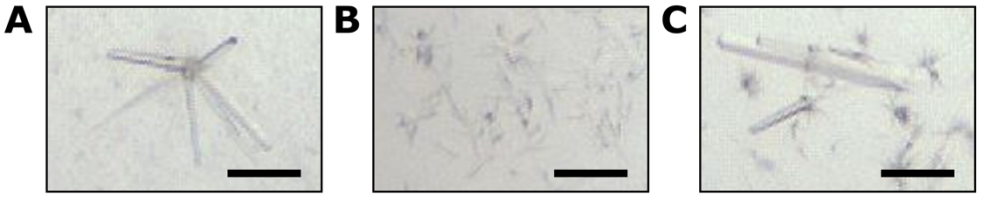

Figure 6.4. Attempts to reproduce crystals
A. Original crystal diffracting to 2.89 Å grown by Eve Caroe against a reservoir condition containing 20% PEG 3350, 0.2 M ammonium nitrate. B. Microcrystals grown by Eve Caroe while attempting to reproduce the crystal shown in B. These microcrystals were used in a subsequent microseeding experiment. C. The crystal form shown in A was reproduced against a reservoir condition containing 16% PEG 3350, 0.23 M ammonium nitrate. Scale bar: 100 µm.

Crystallisation depends on the formation of a nucleus followed by ordered growth of a crystal lattice around the nucleus. Crucially, the optimal conditions for both events may be different. De-coupling nucleation from crystal growth can help overcome crystallisation issues. This can be done by using previously obtained crystals as seeds in a new screen of crystallisation conditions[179]. I produced a seed stock by crushing the microcrystals obtained by Eve Caroe (Fig. 6.4B) with a Teflon bead and set up a high-throughput crystallisation screen with various dilutions of this seed stock added to several dilutions of the original crystallisation condition. Unfortunately, no crystals grew in this experiment.

In parallel, I attempted to reproduce and optimise the original crystal by screening around the precipitant and ammonium nitrate concentration of the original crystallisation condition. I succeeded in growing large crystals (Fig. 6.4C) which were used in subsequent soaking experiments (section 6.2.3).

6.2.2 Co-crystallisation screens of PTPRM Fn3-4 with galactose

As an alternative approach to obtain a structure of galactose-bound PTPRM, I set up co-crystallisation screens of PTPRM Fn3-4 with galactose. PTPRM Fn3-4 was incubated with a 5-fold molar excess of galactose (2 mM) and the mixture was used in high-throughput crystallisation screens. Encouragingly, crystals and microcrystals were grown in several conditions where crystal growth was not observed in previous crystallisation experiments without galactose (Fig. 6.5A and B), suggesting that galactose

might be contained within the crystal and contributing to crystal contacts. Intriguingly, some of these conditions also contained sulfate, which may mimic sulfatide. I optimised these microcrystals and obtained crystals of sufficient size to harvest for data collection (Fig. 6.5C and D).

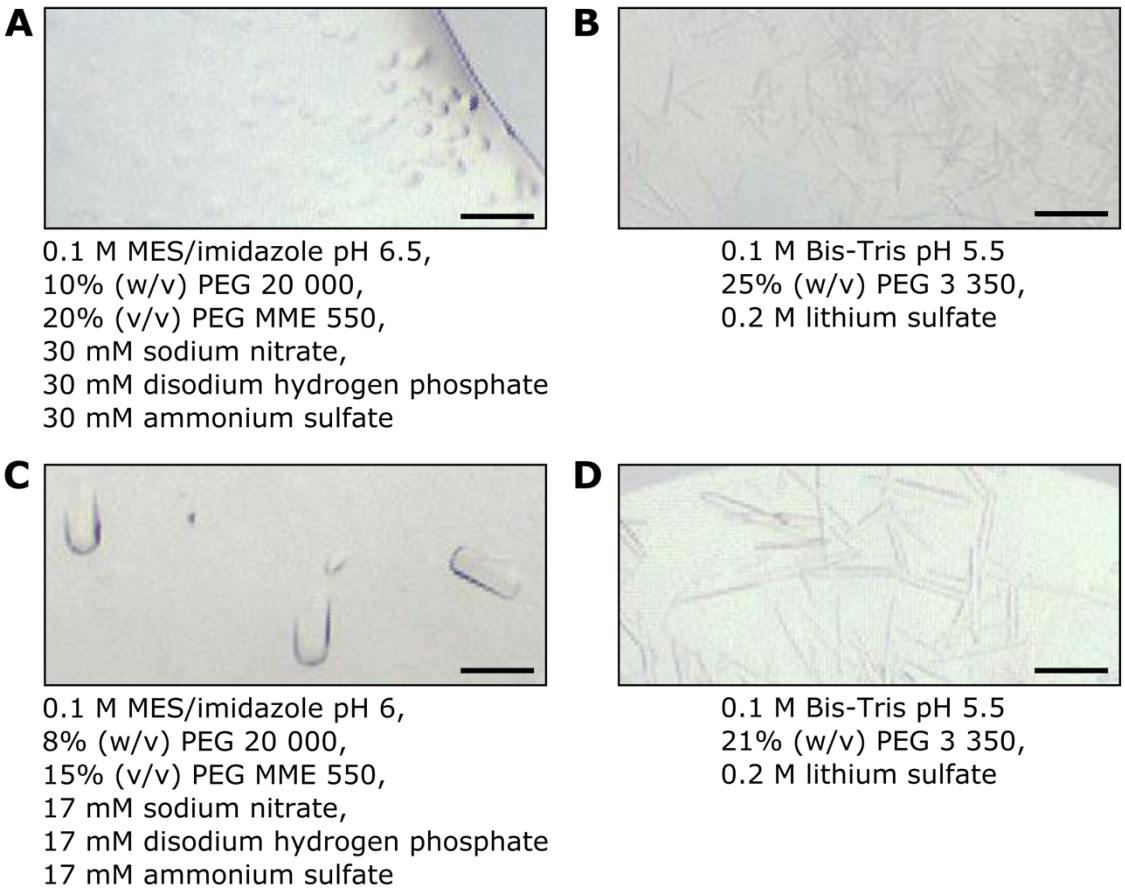

Figure 6.5. Crystals grown in the presence of galactose
A-B. Crystals obtained in high-throughput crystallisation screens of PTPRM Fn3-4 with a 5-fold molar excess of galactose. C-D. Crystals grown after optimisation of the conditions shown in A-B. The reservoir condition is detailed below each picture. Scale bar: 100 µm.

6.2.3 Crystal soaks

Several of the crystallisation condition described in sections 6.2.1 and 6.2.2 contained multiple crystals of similar appearance. Some crystals were harvested in their native state, whereas others were soaked in 10 mM galactose or Gal-4S for 10 to 30 min prior to harvesting. Soaked crystals belonging to the crystal form first obtained by Eve Caroe did not diffract sufficiently well for a dataset to be collected. I was able to collect a full dataset of diffraction images from a native crystal, but it was of lower resolution (3.4 Å) than the original dataset, and since these crystals did not differ from the original, this dataset was not of any use.

Data were collected to 2.83 Å from a galactose-soaked crystal grown in co-crystallisation experiments of PTPRM Fn3-4 with galactose (Table 6.1). This crystal was grown against a reservoir containing 0.1 M MES/imidazole pH 6, 8% (w/v) PEG 20 000, 15% (v/v) PEG MME 550, 17 mM sodium nitrate, 17 mM disodium hydrogen phosphate and 17 mM ammonium sulfate and soaked with 10 mM galactose for 10 min prior to harvesting (Fig. 6.5C). Processing of these data identified that this was a new crystal form. The presence of galactose and sulfate in the crystallisation condition combined with a galactose soak provided our best chance for observing an interaction with galactose or sulfate.

Table 6.1. Data collection statistics for PTPRM Fn3-4. Values in parentheses are for the highest-resolution shell.

Beamline	i03
Wavelength (Å)	0.976254
Processing package	xia2 dials
Space group	$P\,3_2\,2\,1$
Cell dimensions	
a,*b*,*c* (Å)	87.5, 87.5, 311.2
α,*β*,*γ* (°)	90, 90, 120
Resolution (Å)	77.81 – 2.83 (2.88 – 2.83)
R_{merge}	0.2019 (12.9024)
R_{pim}	0.069 (4.564)
$CC_{1/2}$	0.9928 (0.2628)
$I/\sigma I$	6.30 (0.08)
Completeness (%)	100 (100)
Redundancy	9.66 (8.96)
Wilson B-factor	80.1

6.2.4 PTPRM Fn3-4 model building and refinement

The crystal structure of PTPRM Fn3-4 was solved by molecular replacement using the model of PTPRM Fn3-4 previously obtained by Eve Caroe, Dr Janet Deane and Dr Stephen Graham as a search model. The asymmetric unit contained four PTPRM Fn3-4 molecules related by non-crystallographic symmetry (NCS). NCS restraints were therefore used during refinement. At this resolution, the backbone electron density was unambiguous except for flexible loops, and side chain conformations were interpretable for many residues. The entire model could be built into the electron density. Molecular dynamics simulations and model geometry validation in ISOLDE greatly improved the model. Ambiguous features of the model are discussed below.

Continuous $2F_o$-F_c electron density suggestive of a disulfide bond was visible between Cys642 and Cys716 in all four PTPRM Fn3-4 chains (Fig. 6.6A). Surprisingly, some negative difference density was also visible at the location of the bond, as well as positive difference density corresponding to a different conformation for the side chain of Cys642, suggesting that the disulfide bond was broken in a fraction of protein copies within the crystal or during data collection. These effects were most pronounced for chain C. Disulfide bonds are particularly sensitive to radiation damage during X-ray diffraction data collection[303]. It is therefore likely that this disulfide bond was present in the crystal and broken during data collection, so that the first part of the dataset describes an oxidised structure, whereas the second part describes a reduced disulfide bond resulting in electron density for both conformations. Accordingly, in a model of this structure with a reduced disulfide bond, positive difference density at the disulfide bond location suggests that this bond is missing (Fig. 6.6B). The oxidised and the reduced model both appear partially correct and have a similar R_{free}, suggesting that both models describe the data equally well. Since the reduced form appears to be the product of radiation damage, and this bond is likely to be oxidised in the native state of this protein, I chose to model this disulfide bond in its oxidised state in the final structure model.

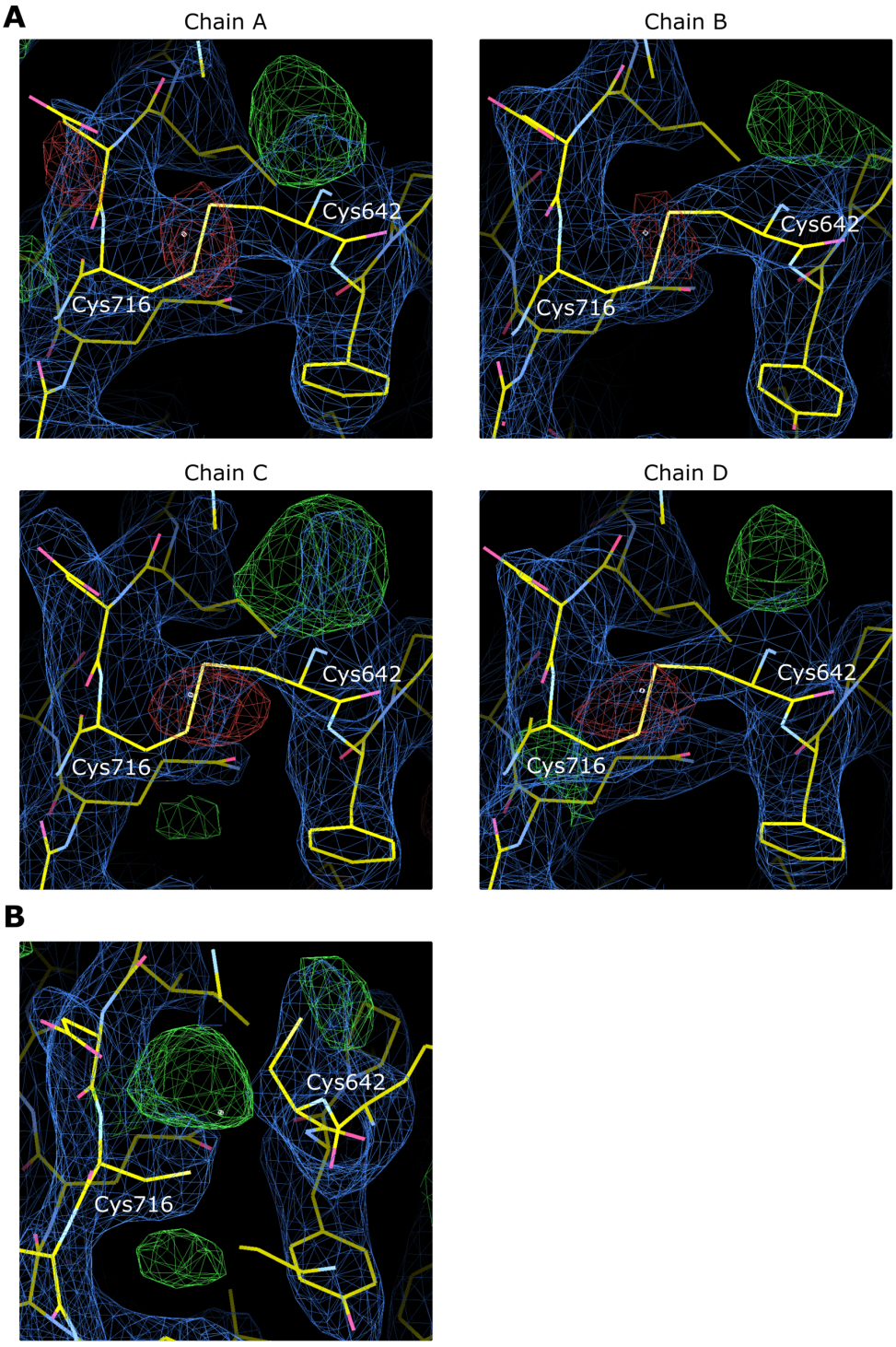

Figure 6.6. The disulfide bond between Cys642 and Cys716 is partially reduced
A. Electron density (2*Fo-Fc*, in blue) and difference density (*Fo-Fc*, in green for positive and red for negative) maps showing continuous electron density between Cys642 and Cys716 strongly suggesting the existence of a disulfide bond. The presence of negative difference density at the modelled disulfide bond location and positive difference density corresponding to an alternative conformation for the side chain of Cys642 in all PTPRM Fn3-4 chains, particularly in chain C, indicate that the modelling of a disulfide bond between Cys642 and Cys716 is partially incorrect. B. Modelling the side chains of Cys642 and Cys716 in the reduced, free dithiol form in chain C results in positive difference density which can be explained by the presence of a disulfide bond in a fraction of the molecules in the crystal.

Clear electron density corresponding to N-linked glycans could be seen on Asn544 (Fn3 domain) and Asn598 (Fn4 domain) on all chains, except at position 598 in chain A (Fig. 6.7A-F). All N-linked glycans have a common core pentasaccharide composed of two *N*-Acetylglucosamine residues and three mannose residues ($Man_3GlcNAc_2$; Fig. 6.7G)[304]. Several of these sugar residues could be built into the electron density (Fig. 6.7C and 6.7F). For both N-linked glycans, additional electron density was observed corresponding to fucosylation of the innermost *N*-Acetylglucosamine, a common post-translational modification[305] (Fig. 6.7H). The most extensive electron density for the glycan at position 544 was observed in chain B, and in chain D for the glycan at position 598. In other chains, there was less electron density. Therefore, for each chain, I built glycan residues into visible electron density, using the best-defined glycan as a template when the glycan conformation was ambiguous. This difference in the extent of visible electron density is most likely caused by differences in glycan flexibility within the crystal[304]. Indeed, the best electron density is visible for glycan residues positioned close to a symmetry molecule restricting their movement (Fig. 6.7C and F). Conversely, at position 598 in chain A, where no reliable electron density corresponding to a glycan is visible, a potential glycan would have space for flexible movement within the crystal, therefore not yielding consistent electron density. Importantly, the restriction of glycan movement due to crystal packing interactions means that the conformations observed in this structure represent one of potentially many conformations in solution, where glycans are likely to be very flexible. No electron density corresponding to glycans at the remaining predicted N-linked glycosylation sites in the Fn3 and Fn4 domains of PTPRM (Asn534, Asn651 and Asn681)[298] was observed.

The final refinement statistics are shown in Table 6.2. and compared to the refinement statistics of deposited structures at a similar resolution (Fig. 6.8A). The clashscore[191] and outlier statistics are excellent as a result of the molecular dynamics simulations performed in ISOLDE. Average B-factors are unusually high; however, these are high throughout the structure, with expected increases around flexible loops with poorly defined electron density (Fig. 6.8B). These high B-factor values throughout the structure suggest that there is disorder or mosaicity within the crystal. This may arise from glycan flexibility and heterogeneity within the crystal as well as radiation damage[303]. As discussed above, chain C appears to have suffered more X-ray radiation damage absorbed by the disulfide bond between Cys642 and Cys716, which may explain why this chain has higher B-factors than the other three chains.

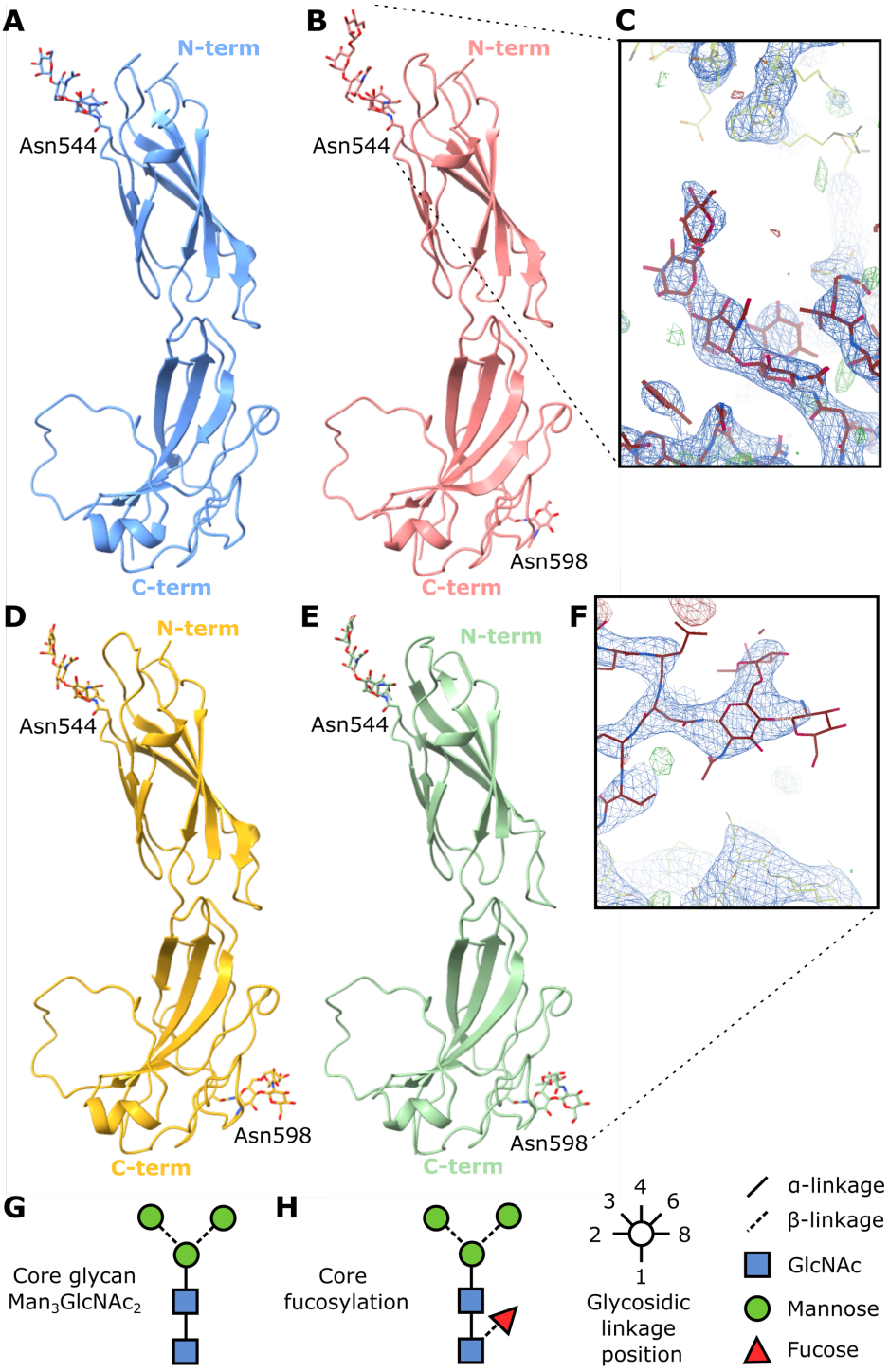

Figure 6.7. Building of N-linked glycans

Ribbon representation of each chain in the final PTPRM Fn3-4 structure showing N-linked glycans in stick representation (oxygen, red; nitrogen, blue). A. N-linked glycans on chain A (in blue). B. N-linked glycans on chain B (in red). C. Electron density (2*Fo-Fc*, in blue) and difference density (*Fo-Fc*, in green for positive and red for negative) maps of the glycan linked to Asn544 on chain B (stick representation, in red). The neighbouring symmetry molecule is shown (stick representation, in light green). D. N-linked glycans on chain C (in yellow). E. N-linked glycans on chain D (in green). F. Electron density and difference density maps of the glycan linked to Asn598 on chain D (stick representation, in red). The neighbouring symmetry molecule is shown (stick representation, in light green). G. Schematic representation of the common core pentasaccharide $Man_3GlcNAc_2$. H. Fucosylated core pentasaccharide.

Table 6.2. Refinement statistics for PTPRM Fn3-4.

Resolution (Å)	75.804 - 2.832
No. reflections	30183
R_{work}/R_{free}	0.2448/0.2952
No. atoms	
Protein	7692
Glycans	297
Water	0
B-factors	117.66
Protein	116.37
Glycans	150.95
Chain A	116.63
Chain B	116.44
Chain C	120.97
Chain D	116.56
Ramachandran	
Favoured (%)	98.14
Outliers (%)	0
R.m.s deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.023
Rotamer outliers (%)	0.72
Cis peptides (Proline)	0
Cis peptides (General)	0
Twisted peptides (Proline)	0
Twisted peptides (General)	0
Clashscore	2.35

Due to the presence of NCS and the use of NCS restraints during refinement, all four copies of PTPRM are very similar (RMSD ≤ 0.6 Å) but show small, localised differences. Chain A possessed the best-quality electron density and this model will be discussed throughout the rest of this chapter.

Unfortunately, no unexplained electron density corresponding to galactose was observed in the refined PTPRM Fn3-4 structure. Although I did not succeed in obtaining a structure elucidating how PTPRM binds to galactosphingolipids, I obtained a good quality model of the Fn3 and Fn4 domains of PTPRM.

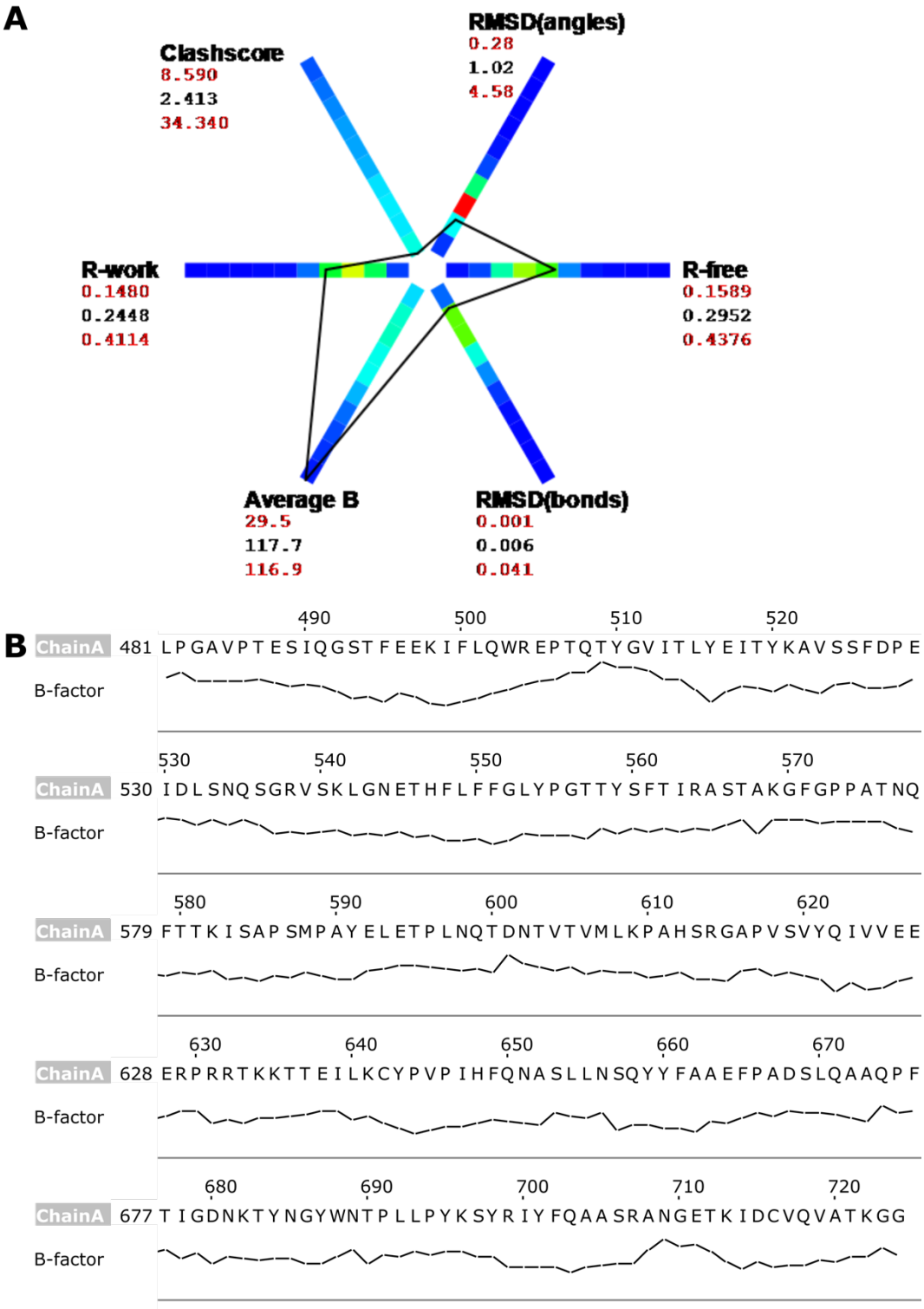

Figure 6.8. PTPRM Fn3-4 refinement statistics

A. Histograms of the distribution of selected statistics across 1854 PDB entries of similar resolution, with the range specified by numbers printed in red. Statistics for the PTPRM Fn3-4 structure are printed in black; the connecting polygon (in black) shows where these values fall in the distribution. Histogram bins are colored by the number of structures in each bin: 0 (blue), 519 (green), 1038 (red). B. B-factors of PTPRM Fn3-4 chain A residues. A similar pattern is observed in all chains.

6.2.5 Structural features of the Fn3 and Fn4 domains of PTPRM

The structure of the Fn3 domain has been reported previously within the context of the ectodomain of PTPRM[298]. In this structure, the Fn4 domain was not modelled. A key feature of the ectodomain of PTPRM is its rigid extended architecture posited to be essential for its function as a sensor of inter-cellular distances. This rigidity is provided by a network of interactions stabilising the interface between consecutive domains[298]. This holds true for the interface between the Fn3 and Fn4 domains, which is similarly stabilised by a network of hydrophobic and hydrogen bonding interactions (Fig. 6.9A). The structure of the Fn3 domain in the context of the Fn3-4 domains or the ectodomain of PTPRM are very similar (RMSD 1.1 Å over 80 Cα atoms) except for some small loop movements (Fig. 6.9B). The loop including Tyr510 is highly disordered in the Fn3-4 structure, as indicated by extremely poor electron density and high B-factors around this residue (Fig. 6.8B). This is because Tyr510 forms part of the interface between the Fn2 and Fn3 domains[298] and is therefore stabilised in a different conformation in the PTPRM ectodomain structure (Fig. 6.9B). Similarly, the loop including Asp527, Pro528 and Glu529 is destabilised in the ectodomain structure due to the absence of an ordered Fn4 domain in this structure, because these three residues are involved in interface contacts between the Fn3 and Fn4 domains (Fig. 6.9A). Finally, the loop bearing the glycan linked to Asn544 appears flexible. The presence of a symmetry molecule in close proximity to this loop in the Fn3-4 crystal suggests that crystal packing interactions drive this alternative loop conformation. Overall, these observations suggest that the Fn3 domain of PTPRM has a rigid, compact structure with limited flexibility and extensive contacts with adjacent domains, in accord with a rigid spacer role for this domain.

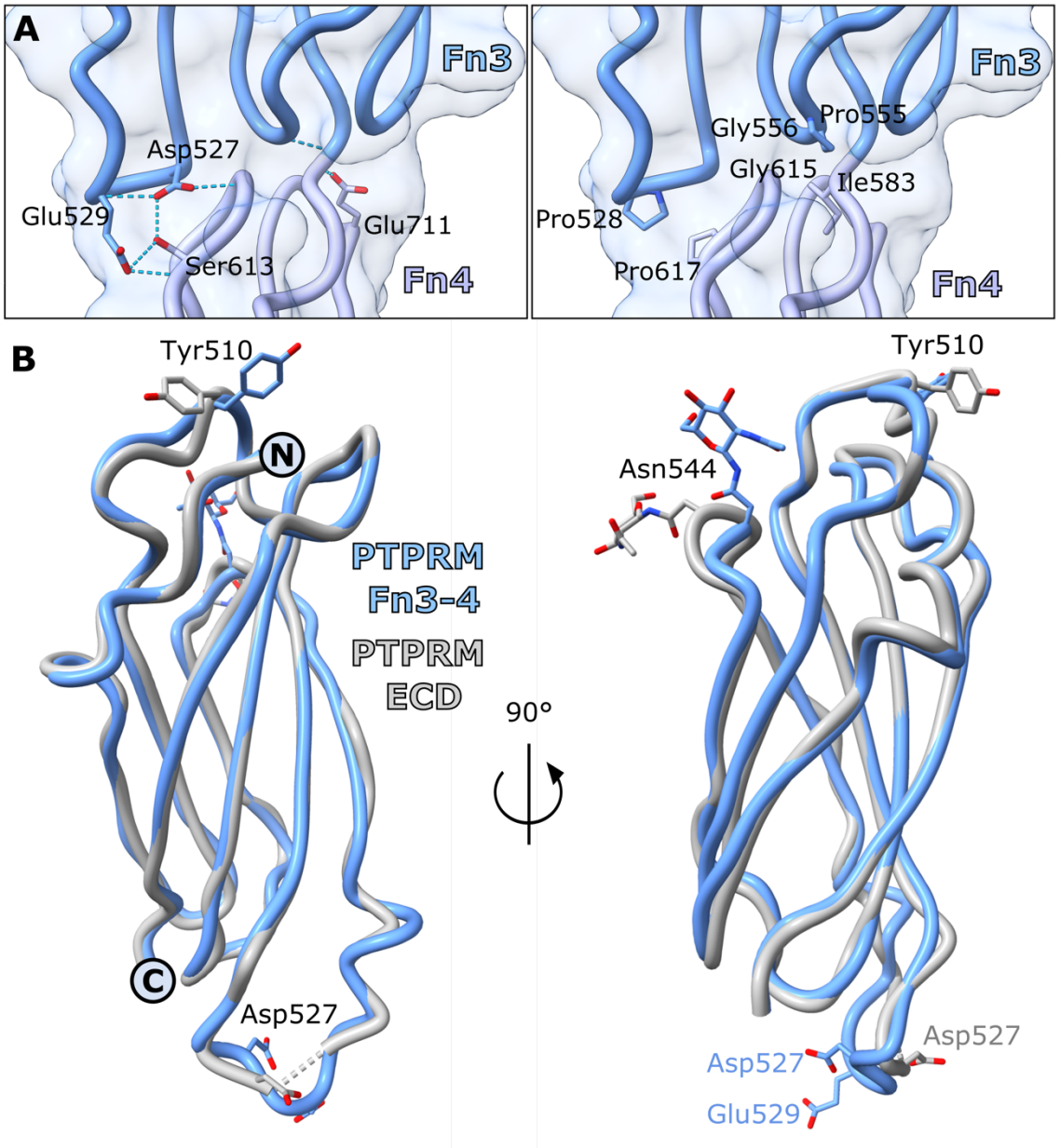

Figure 6.9. Interface between the Fn3 and Fn4 domains
A. The interface between the Fn3 (blue) and Fn4 (lilac) domains of PTPRM is stabilised by a network of interactions. The protein backbone is shown as liquorice, with key residues drawn in stick representation (oxygen, red; nitrogen, blue). The protein surface is shown as a transparent surface. The left panel shows potential hydrogen bond interactions, represented as blue dashes. The right panel shows residues potentially involved in hydrophobic interactions. B. Structural alignment of the Fn3 domain within the context of the Fn3-4 construct (liquorice backbone representation, in blue) and of the published structure of the extracellular domain (ECD) without the Fn4 domain (in grey; PDB ID: 2V5Y). An unmodelled disordered loop is represented as dashes. Residues discussed in the text are drawn in stick representation (oxygen, red; nitrogen, blue). Termini are marked as circles.

This is in contrast with the Fn4 domain of PTPRM, which has intriguing features suggestive of additional functions for this domain. The secondary structure of the Fn4 domain differs from the classical fibronectin III fold composed of two β-sheets with 3 and 4 β-strands forming a β-sandwich[306]

(Fig. 6.10A-B). In the Fn4 domain of PTPRM, the first β-sheet is replaced by three loops which are not organised as β-strands. The second β-sheet formed by the β-strands C, D, F and G is maintained in the Fn4 domain. Profile-profile alignment to distantly related fibronectin III domains using the Fold and Function Assignment System (FFAS) server[197] and structure based-alignment using the Dali server[198] show that the Fn4 domain of PTPRM has unusual insertions in its sequence (Fig. 6.10C-D). Structural alignment of the Fn4 domain of PTPRM with the closest structural homolog for which a structure is available, the A74 fibronectin III domain of Titin[307], demonstrates that in addition to the compact fibronectin III-like fold, a large loop is protruding out of the main fold of the Fn4 domain of PTPRM (residues 628-642; Fig. 6.10E). This loop, labelled the F-loop, appears to be flexible, adopting different conformations in each structure of the Fn3-4 domains (Fig. 6.11). Structural alignment of the two chains of the Fn3-4 structure previously obtained in our lab with the new Fn3-4 structure shows a similar overall conformation (chain A RMSD 0.8 Å over 220 Cα atoms; chain B RMSD 1.3 Å over 241 Cα atoms), with the exception of two membrane-proximal loops including the F-loop (Fig. 6.11). The F-loop was found in different conformations in two models and was disordered and unmodelled in the third model. The electron density for this loop is very clear in the new Fn3-4 structure and in chain B of the previous structure because in both cases, this loop forms contacts with another protein chain in the asymmetric unit. Likewise, a loop equivalent to the A strand in a classical fibronectin III fold (Fig. 6.10A-B), named the A-loop (residues 590-600), was found in different conformations in two structures, and disordered in the third (Fig. 6.11, right panel). The presence of flexible loops near the membrane-proximal end of the domain suggests they may facilitate interactions with lipids or other membrane proteins. The A-loop carries the glycan linked to Asn598 (Fig. 6.11). This glycan may protect the A-loop from proteolysis or contribute to interactions with lipids or proteins.

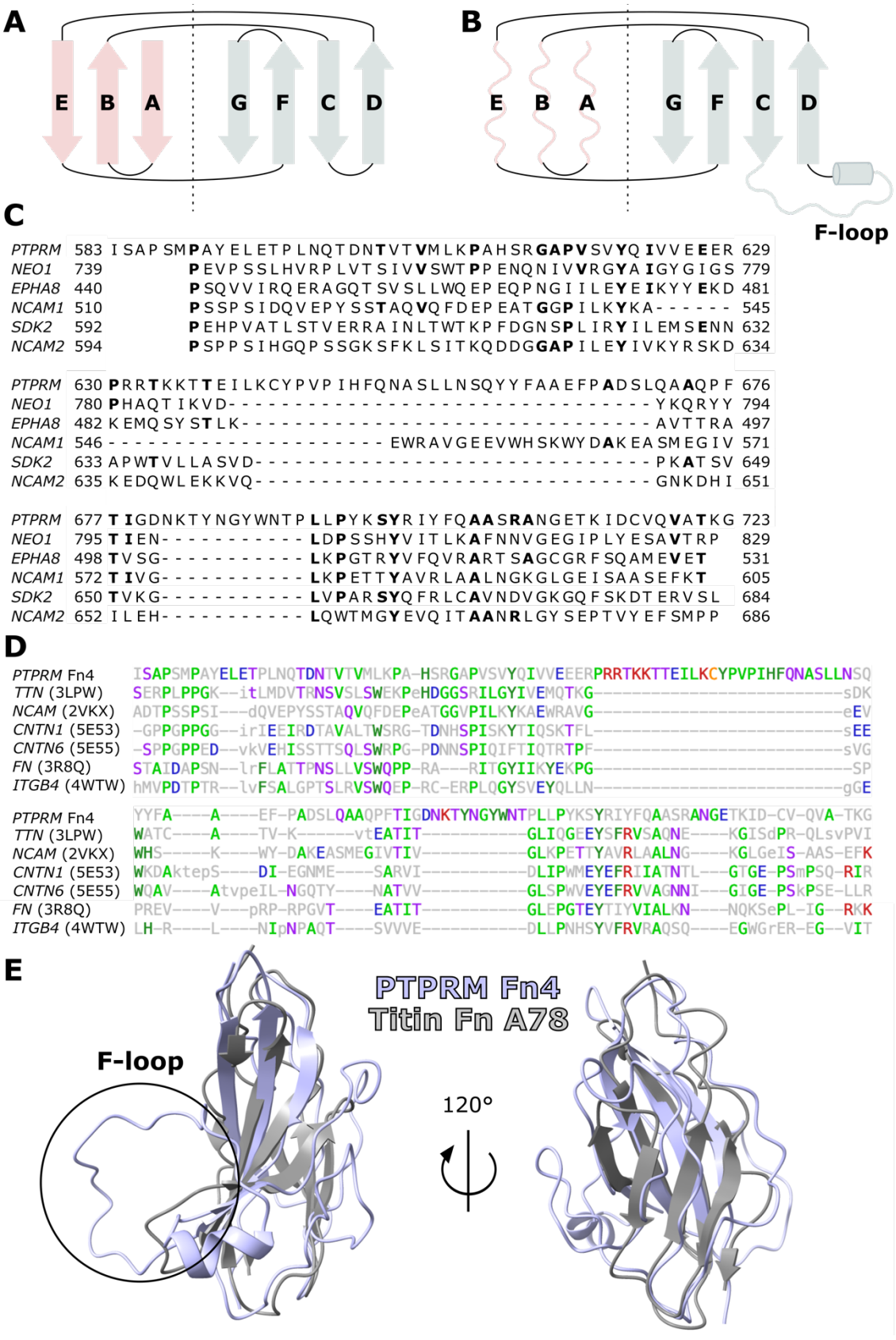

Figure 6.10. The Fn4 domain of PTPRM has an unusual structure

A. Schematic diagram showing the secondary structure for a typical fibronectin III fold. β-strands (shown as arrows) form two β-sheets arranged as a β-sandwich. B. Secondary structure of the Fn4 domain of PTPRM. Several β-strands are replaced by loops. A long loop labelled the F-loop and a

short α-helix (shown as a cylinder) are inserted between strands C and D. C. FFAS profile-profile alignment of the Fn4 domain of PTPRM with fibronectin III domains with distant homology. D. Sequence alignment of the Fn4 domain of PTPRM with the closest structural homologs deposited in the PDB (PDB IDs are indicated in parentheses) identified using the DALI server. Uppercase means structurally equivalent positions with PTPRM. Lowercase means insertions relative to PTPRM. The most frequent amino acid is coloured in each column. E. Structural alignment of Fn4 and its closest structural homolog, the fibronectin III domain A78 of human Titin (PDB ID: 3LPW). The F-loop is circled.

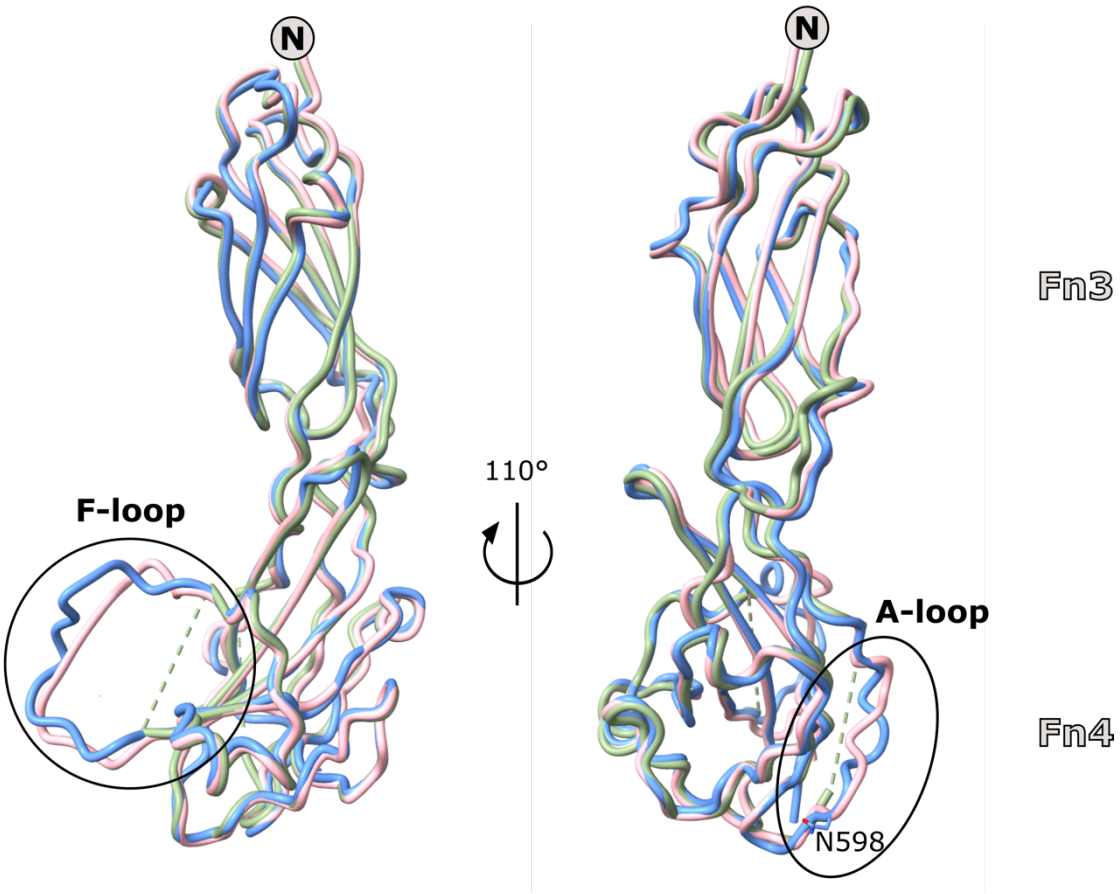

Figure 6.11. The Fn4 domain possesses flexible loops
Structural alignment of the PTPRM Fn3-4 structure described in this chapter (blue liquorice) with the unpublished Fn3-4 structures previously obtained in our lab (chain A: green liquorice, with unmodelled, disordered loops represented as dashes, chain B: pink liquorice). The F-loop and A-loop (circled) are flexible and adopt different conformations in each structure. Asn598 is drawn in stick representation. The linked glycan is not shown for clarity. The N-terminus is marked as a circle.

PTPRM is cleaved by the proprotein convertases furin and PC5 between Arg632 and Thr633. After cleavage, the ectodomain fragment remains associated with the transmembrane fragment[301]. The structure of the Fn4 domain reveals that this cleavage event occurs in the unusual F-loop extending away from the Fn4 domain (Fig. 6.12), hence the name F-loop for Furin-loop. Both proteolytic fragments form part of the main fibronectin III fold and the extensive interactions contributing to this

fold are expected to maintain this fold intact after furin/PC5 cleavage, explaining why both fragments remain together after cleavage. It has been hypothesised that cleavage by proprotein convertases introduces some flexibility in the membrane-proximal region of PTPRM[298] but this seems unlikely based on this observation. However, proteolytic fragments of the F-loop are likely to undergo large conformational changes after cleavage. These conformational changes may underlie the currently unknown effects of proprotein convertase cleavage on PTPRM function. Interestingly, the longest loop fragment is covalently bonded to the surface of the fibronectin III fold via a disulfide bond between Cys642 and Cys716, the sole disulfide bond within the Fn4 domain (Fig. 6.12). This bond does not tether the ectodomain fragment to the C-terminal transmembrane fragment after proprotein convertase cleavage. Instead, it appears to restrict the movement of this unusual loop before and after cleavage, stabilising this flexible region in a specific orientation likely to be important for cleaved and/or uncleaved PTPRM function.

The R2B family of receptor-type protein tyrosine phosphatases includes PTPRM, PTPRT, PTPRK and PTPRU. These proteins share a common domain architecture, including the membrane-proximal Fn4 domain. Like PTPRM, PTPRK is cleaved by proprotein convertases and is expressed on the cell surface as a mixture of cleaved and uncleaved protein[308]. PTPRU is also expressed as a mixture of cleaved and uncleaved forms (Iain Hay, personal communication), whereas proteolytic cleavage of PTPRT has not been reported. Multiple sequence alignment of the Fn4 domain across human R2B family proteins shows that the unusual insertion corresponding to the F-loop is conserved in R2B family members (Fig. 6.13A, in contrast with distantly related fibronectin III domains (Fig. 6.10C-D). Despite the presence of this loop in other R2B family members, the loop sequence is poorly conserved (Fig. 6.13A-B). The furin cleavage site is also found in the F-loop region of PTPRK and PTPRU, although the exact position and sequence is not conserved (Fig. 6.13A). PTPRU appears to have two overlapping furin cleavage sites, with the first potential site at the same position as the PTPRM cleavage site, and a second site three residues downstream, at the same position as the PTPRK cleavage site. PTPRT does not possess a canonical furin cleavage site. Interestingly, the disulfide bond between Cys642 and Cys716 described above is conserved across all four proteins. Altogether, this comparison suggests that all R2B family members possess a long membrane-proximal loop anchored to the main fold of the Fn4 domain by a conserved disulfide bond. This loop is cleaved by proprotein convertases in all R2B family members except PTPRT.

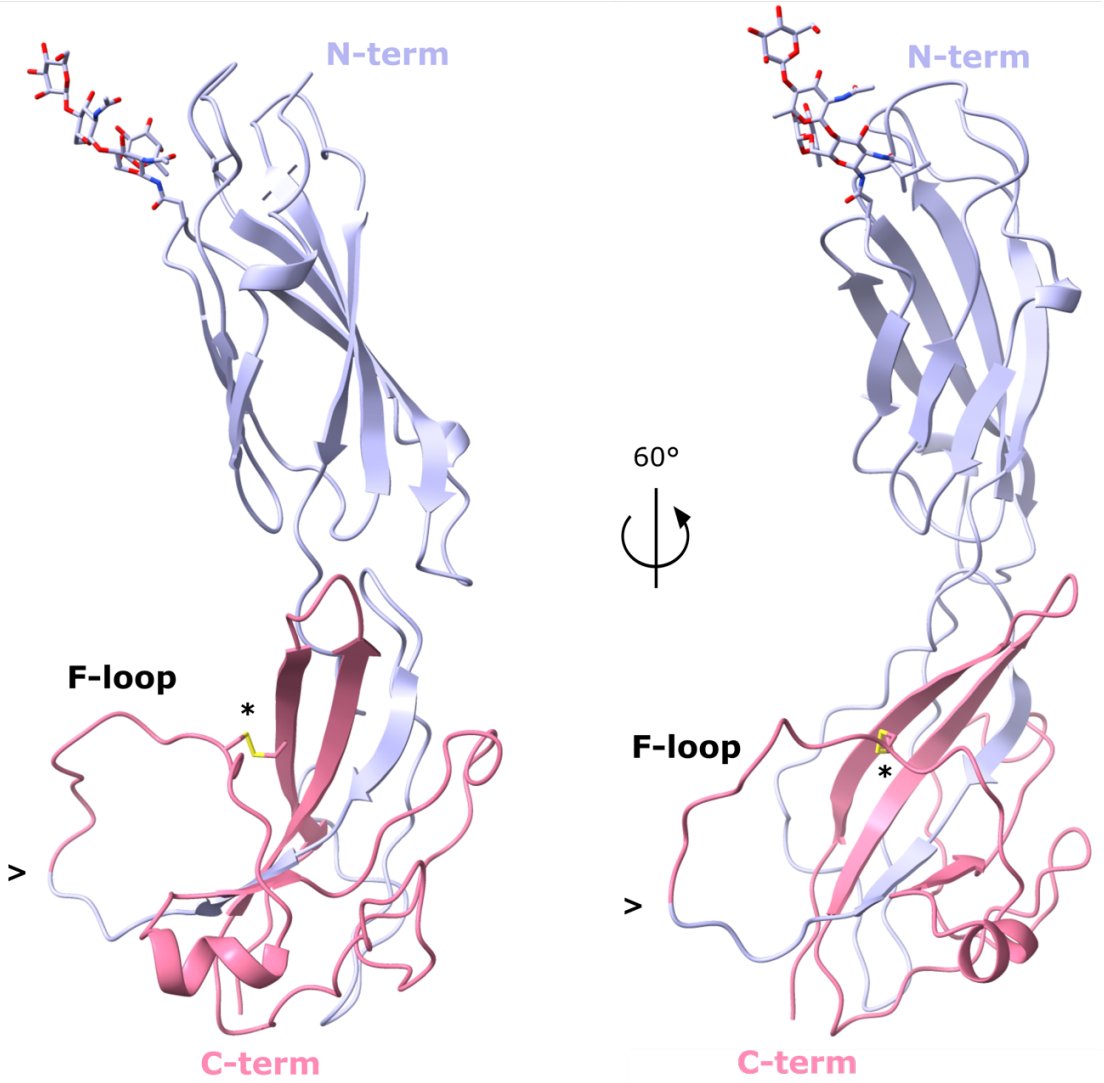

Figure 6.12. Furin cleavage occurs on a long membrane-proximal loop
Ribbon representation of the PTPRM Fn3-4 structure showing the furin cleavage site indicated by the symbol (>). After cleavage, the transmembrane fragment (pink) remains associated with the ectodomain fragment (lilac). The disulfide bond indicated by (*) drawn in stick representation (sulfur, yellow) anchors the loop to the main Fn4 fold. Glycans are shown in stick representation (oxygen, red; nitrogen, blue).

I have solved and refined a structure of the Fn3-4 domains of PTPRM structure, revealing intriguing aspects of PTPRM structure potentially relevant for its function or regulation by proprotein convertases. These findings may be applicable to other R2B family receptor phosphatases. Unfortunately, my attempts to elucidate how PTPRM interacts with galactosphingolipids through crystallisation experiments with the headgroup analogues galactose or Gal-4S did not yield a structure of ligand-bound PTPRM. Subsequent isothermal titration calorimetry experiments performed by Alex Nicholson showed that PTPRM Fn3-4 does not interact with galactose. Furthermore, Gal-4S is not an ideal sulfatide headgroup analogue because protein-ligand interactions are often isomer-specific. We

have since initiated a collaboration with Prof Henrik Jensen (Department of Chemistry, Aarhus University) to synthesise the sulfatide headgroup Gal-3S. Future work in our lab will use this sulfatide analogue for interaction assays and structural studies with PTPRM. The refined structure described in this chapter will facilitate structural studies by providing a good quality molecular replacement model for the Fn3-4 domains of PTPRM.

6.3 PTPRM trafficking in GALC KO oligodendrocytes

Previous work in our lab showed that PTPRM is specifically down-regulated at the plasma membrane of GALC KO oligodendrocytes, but total expression levels and mRNA levels remain unchanged. Therefore, we hypothesised that the interaction of PTPRM with accumulated galactosphingolipid headgroups leads to its mistrafficking or intracellular retention in GALC-deficient oligodendrocytes. I aimed to test this hypothesis by monitoring PTPRM trafficking in GALC KO cells.

PTPRM expression levels are extremely low in MO3.13 oligodendrocytes, rendering its detection by immunofluorescence challenging. Our lab has attempted endogenous GFP tagging of PTPRM without success. Subsequently, I set out to monitor the trafficking of recombinant GFP-labelled PTPRM constructs by microscopy with the “retention using selective hooks” (RUSH) technique[309,310]. This method allows one to monitor target protein trafficking through the secretory pathway after synchronous release from the ER. The target protein is fused to a streptavidin-binding peptide (SBP) and GFP and co-expressed with a "hook" protein, streptavidin fused to the ER retrieval signal peptide KDEL (Strep-KDEL). The target protein therefore remains in the ER until biotin is added to the cells, outcompeting the SBP-streptavidin interaction and allowing the synchronous release of the target protein from the ER and triggering trafficking towards its destination compartment.

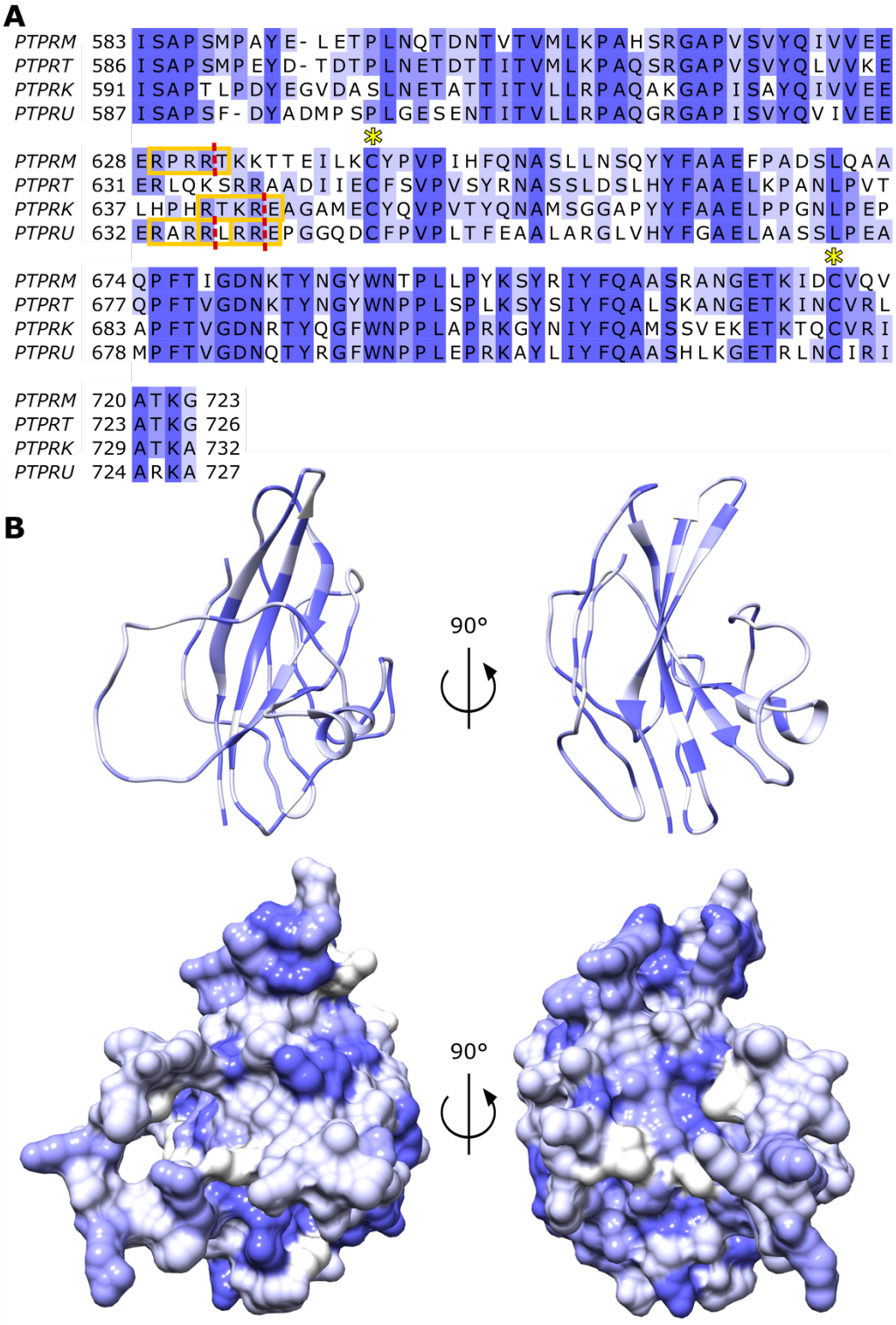

Figure 6.13. The F-loop is conserved among R2B family receptor phosphatases
A. Multiple sequence alignment of the membrane-proximal Fn4 domain across human R2B family members (PTPRM, PTPRT, PTPRK and PTPRU), coloured by percentage identity (gradient from dark purple for conserved residues to white for non-conserved residues). Potential furin cleavage sites are highlighted in orange, with red dashes indicating the peptide bond subject to hydrolysis. PTPRU has two overlapping furin cleavage sites. The conserved cysteines participating in a disulfide bond are indicated with the yellow symbol (*). B. Structure of the Fn4 domain of PTPRM (top panel: ribbon representation; bottom panel: surface representation) showing residue conservation. Residues are coloured using the same colour scheme as in A.

The trafficking of several PTPRM truncations including the transmembrane region was tested. Full-length PTPRM, the extracellular domain (ECD) or the membrane-proximal Fn3-4 domains were fused to SBP and GFP by Eve Caroe (Fig. 6.14A). I tested the expression and trafficking of these constructs in HeLa cells stably expressing the hook peptide Strep-KDEL. Full-length PTPRM accumulated in intracellular puncta and was not released to the plasma membrane upon biotin addition, suggesting that this recombinant construct forms aggregates of misfolded protein (Fig. 6.14B). In contrast, the ECD accumulates in the ER prior to the addition of biotin (Fig. 6.14C, 0 min), and addition of biotin leads to its release and transport to the Golgi (Fig. 6.14C, 10 min), then to the plasma membrane, where it accumulates at cell-cell contact sites (Fig. 6.14C, 60-150 min) through homophilic interactions with the ECD on adjacent cells[298]. In the absence of contact with neighbouring cells overexpressing the ECD, the ECD is internalised as documented previously[301,311] (Fig. 6.14D). Finally, the Fn3-4 domain of PTPRM also successfully traffics from the ER (Fig. 6.14E, 0 min) to the Golgi after biotin addition (Fig. 6.14E, 10 min), eventually diffusing throughout the plasma membrane (Fig. 6.14E, 150 min). This diffuse plasma membrane localisation is expected since the Fn3-4 domains do not participate in homodimeric cell junction contacts[298].

To understand whether PTPRM is mistrafficked to an intracellular compartment instead of the plasma membrane in GALC KO cells, PTPRM localisation in GALC KO oligodendrocytes was assessed in a pilot experiment. I overexpressed the PTPRM ECD and Fn3-4 constructs described above (Fig. 6.14A) in GALC KO and EV control MO3.13 oligodendrocytes in the absence of the Strep-KDEL hook peptide. Thus, PTPRM constructs were not retained in the ER but trafficked directly towards their destination compartments. Although this experimental set up does not facilitate the study of trafficking kinetics, it may allow the comparison of PTPRM destination compartments in GALC KO cells versus EV control cells. As observed in HeLa cells, the ECD accumulated at cell-cell junctions when this construct was expressed by neighbouring cells or internalised into intracellular compartments in the absence of homophilic contacts in EV as well as GALC KO oligodendrocytes (Fig. 6.15, left panel). The Fn3-4 domain was expressed diffusely at the cell surface as well as in intracellular compartments in both EV and GALC KO oligodendrocytes (Fig. 6.15, right panel). Overall, no difference was observed in the localisation of overexpressed PTPRM constructs in WT and GALC KO cells.

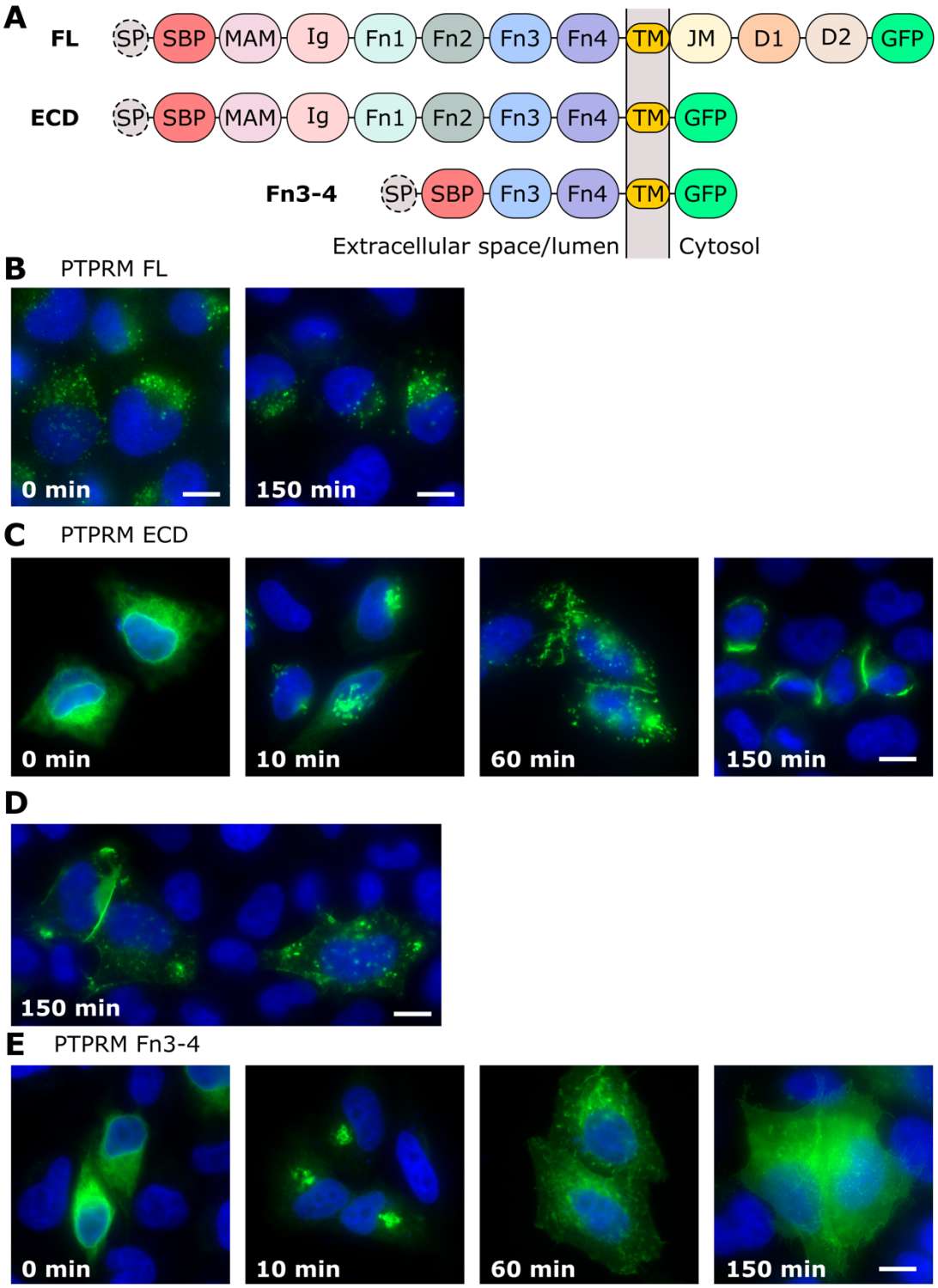

Figure 6.14. PTPRM trafficking in HeLa cells

A. Full-length (FL) PTPRM, the extracellular domain (ECD) including the transmembrane (TM) region or the Fn3-4 domain including the TM region were fused to an ER targeting signal peptide (SP), an N-terminal Streptavidin-binding peptide (SBP) and a C-terminal GFP tag. These constructs were transiently expressed in HeLa cells stably expressing a Streptavidin-KDEL hook peptide, retaining the SBP-fused constructs in the ER. At t=0, biotin was added to the cells, causing the synchronous release of the SBP-fused constructs from the ER. (B-E) Trafficking of these constructs was monitored at the indicated time points by fluorescence microscopy. Nuclei were stained with DAPI. Scale bar: 10 µm.

Establishing GALC KO and EV MO3.13 cell lines stably expressing the hook peptide Strep-KDEL will allow synchronous release and trafficking of PTPRM constructs and may permit detection of time-dependent differences in PTPRM trafficking. A caveat of this experiment is that PTPRM is overexpressed at variable levels much higher than endogenous levels, which prevents quantitative analysis and may mask subtle changes in protein localisation. Alternatively, immunofluorescence microscopy may be used to observe the localisation of endogenous PTPRM, however this will be dependent on antibodies sensitive enough to detect minute amounts of PTPRM. This work will continue to be pursued by our group.

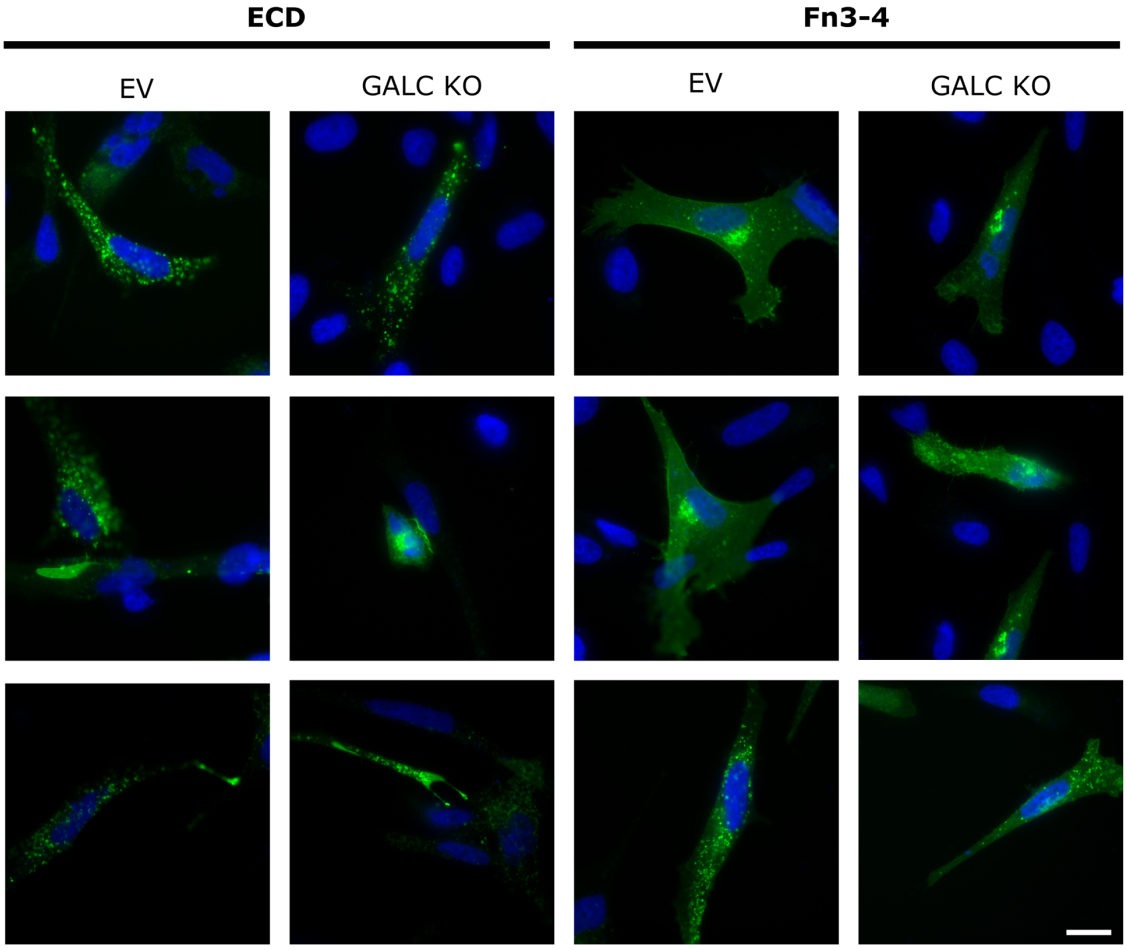

Figure 6.15. PTPRM localisation in GALC KO oligodendrocytes
GFP-tagged extracellular domain (ECD) and Fn3-4 PTPRM transmembrane constructs described in Fig. 6.14A were transiently expressed in EV control and GALC KO MO3.13 oligodendrocytes in the absence of the Streptavidin-KDEL peptide, allowing GFP-tagged PTPRM constructs to traffic towards their destination compartments, observed by fluorescence microscopy. Nuclei were stained with DAPI. Scale bar: 10 µm.

6.4 Discussion

Following up on previous work in our lab showing a specific interaction between the membrane-proximal Fn3-4 domains of PTPRM and sulfatide, I attempted to identify the molecular determinants of this interaction by co-crystallisation and crystal soaking with the GalCer and sulfatide analogues galactose and Gal-4S. This approach did not yield the desired structure of ligand-bound PTPRM, and subsequent experiments performed by Alex Nicholson showed that PTPRM Fn3-4 does not interact with galactose. Further work in our lab will be performed with Gal-3S, a closer analogue of the sulfatide head group produced by our collaborators. Future interaction and structural studies will be essential to understand how PTPRM interacts with sulfatide.

Fibronectin III domains are a common modular domain in the ectodomain of adhesion molecules[312]. Intriguingly, the fibronectin III domains of the adhesion proteins Ephrin A2 (EphA2) and Sidekick-1 and -2 (SDK1 and SDK2) have been shown to mediate binding to lipids[313,314]. EphA2 is a receptor tyrosine kinase binding to Ephrin A ligands on the apposed cell membrane. The membrane-proximal fibronectin III domain of EphA2 binds anionic lipids and stabilises this domain in a "sideways on" conformation at an angle to the remainder of the ectodomain protruding away from the membrane[313,315] (Fig. 6.16A). The residues mediating this interaction have been identified by molecular dynamics simulations and validated by liposome pull-down assays[313]. SDK1 and SDK2 participate in homophilic interactions. Both membrane-proximal and membrane-distal fibronectin III domains of SDK1 and SDK2 interact with lipids, allowing the entire ectodomain to "lie down" on the membrane to reduce intermembrane spacing of adhesion interfaces[314] (Fig. 6.16B). This indicates that lipid binding can help orient and stabilise adhesion molecules for optimal function. The published structure of the ectodomain of PTPRM did not have continuous electron density for the Fn4 domain[298], suggesting that this domain may have some flexibility relative to the rest of the ECD. However, our structures of the Fn3-4 domains of PTPRM shows that the interface between the Fn3 and Fn4 domain is stabilised by a network of hydrogen bond and hydrophobic interactions, suggesting that the entire ectodomain of PTPRM adopts a rigid rod-like structure (Fig. 6.16C). It is currently unclear whether lipid binding may alter the orientation of the PTPRM ectodomain relative to the membrane.

Considering the striking specificity of PTPRM binding to sulfatide, but not other negatively charged lipids such as lysophosphatidic acid and GM1 (Fig. 6.1A), it is likely that lipid binding of PTPRM has roles beyond orienting this protein relative to the membrane. Based on data obtained in our group showing loss of PTPRM at the plasma membrane of GALC-deficient oligodendrocytes, it appears likely that lipid binding controls the trafficking or localisation of PTPRM. Pilot validation experiments showed no depletion of overexpressed, GFP-tagged PTPRM from the plasma membrane of GALC KO

oligodendrocytes. However, overexpression of recombinant PTPRM is expected to be variable across cells and this may mask any subtle changes in PTPRM localisation. Future experiments using RUSH microscopy in GALC KO oligodendrocytes or immunofluorescence microscopy of endogenous PTPRM will be key to identify changes in PTPRM trafficking caused by loss of GALC. These experiments must be supplemented with staining for organelle markers to determine the localisation of mistrafficked PTPRM. Alternatively, loss of endogenous PTPRM at the cell surface of GALC-deficient oligodendrocytes may be validated by flow cytometry, a more sensitive technique than immunofluorescence microscopy, and the localisation of mistrafficked PTPRM may be probed by subcellular protein fractionation and Western blotting.

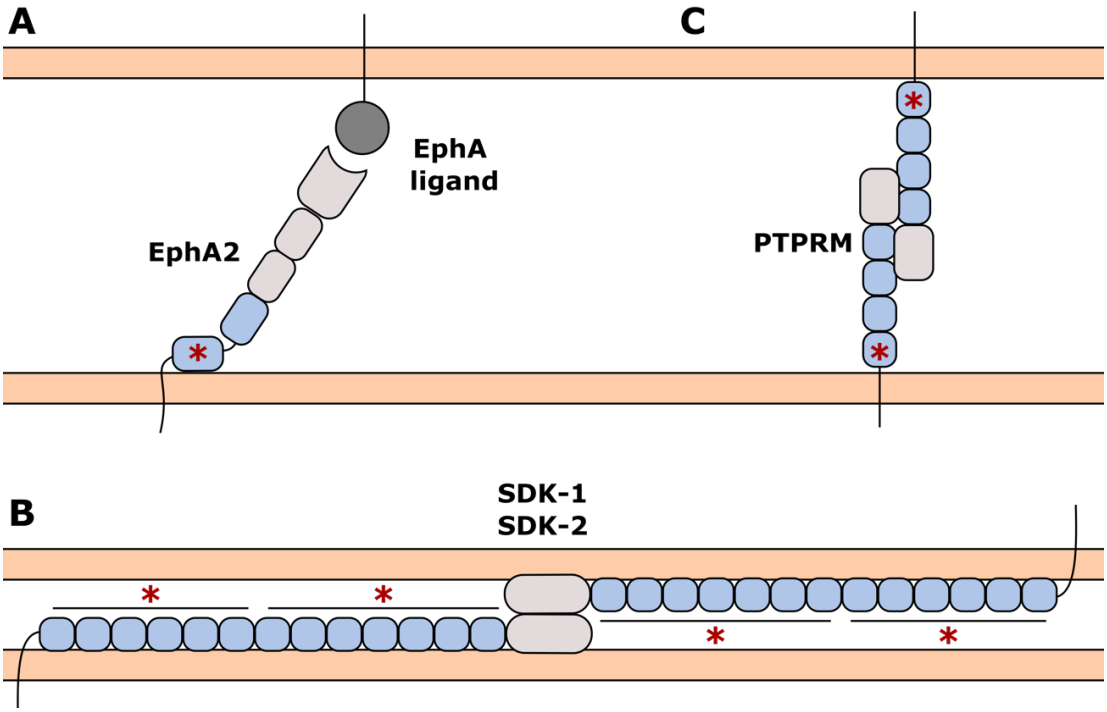

Figure 6.16. Lipid-binding adhesion molecules
Schematic representations of the lipid-binding adhesion molecules EphA2, SDK1, SDK2 and PTPRM. Lipid bilayers are represented in orange. Fibronectin III domains are coloured in blue. Lipid-binding domains or regions are highlighted by a red (*) symbol. A. EphA2 binds ephrin A ligands on the apposed cell membrane. The membrane-proximal fibronectin III domain is stabilised in a "sideways on" conformation by binding to anionic lipids. B. SDK1 and SDK2 form *trans* homodimers. The fibronectin III domains 1-6 and 7-13 bind lipids, allowing SDK1/2 to lie on the membrane to reduce the distance between lipid bilayers. C. PTPRM forms *trans* homodimers in a rigid rod conformation.

The membrane-proximal Fn4 domain of PTPRM has an unusual structure including a large, flexible loop protruding away from the main fibronectin III fold. This loop appears to be present in all receptor protein tyrosine phosphatases of the R2B family, but not in other fibronectin III domains based on

distant homology and structure-based alignments. The structure described in this chapter is the first available structure of the Fn4 domain of an R2B receptor. Interestingly, furin/PC5 cleavage of PTPRM and other R2B members occurs on this loop, thus named the F-loop. The functional consequences of proprotein convertase cleavage are unclear. Approximately half of PTPRM molecules endogenously expressed at the cell surface are present the cleaved form[301], a proportion increased at high cell densities[300]. The Fn3 and Fn4 domains do not participate in cell-cell adhesion homodimer formation[298] and F-loop cleavage is therefore not expected to modify the affinity of this interaction. An intriguing hypothesis based on our data is that proprotein convertase cleavage regulates lipid binding affinity or specificity. This may allow dynamic targeting of PTPRM to specific membrane domains, perhaps acting as a regulatory mechanism of PTPRM function by driving PTPRM near or away from substrates. Future work in our lab will investigate the effect of proprotein convertase cleavage on the lipid-binding properties of PTPRM.

Cleavage at the furin site is a necessary pre-condition for the initiation of a proteolytic cascade involving cleavage of the extracellular domain of PTPRM by ADAM family metalloproteinases, resulting in shedding of a soluble extracellular domain fragment, and intramembrane cleavage by γ-secretases, potentially leading to internalisation of the intracellular phosphatase fragment[316]. An alternative hypothesis explaining the specific loss of PTPRM at the surface of GALC-deficient oligodendrocytes is that that ADAM activity might be enhanced in GALC KO cells, resulting in increased shedding of PTPRM. However, this would also result in increased shedding of PTPRK[316], which is not down-regulated at the plasma membrane of GALC KO oligodendrocytes. Nonetheless, this hypothesis should be considered if further results do not support for our initial hypothesis that loss of GALC results in mistrafficking of PTPRM.

7 Conclusions and open questions

How sphingolipids are extracted from membranes and passed from one protein to another are fascinating questions that I have tried to address during my PhD. A second key focus has been the impact of sphingolipid imbalances on immune cell function and cell death. Finally, I investigated how interactions with sphingolipids may regulate protein trafficking and function. The model systems I used to study these questions were sphingolipid hydrolases, their helper proteins saposins and the lysosomal storage diseases caused by loss of function of these proteins. These diseases are powerful models because they are monogenic and tractable, and may provide insights into more complex, late-onset neurodegenerative diseases such as Parkinson's disease and Alzheimer's disease also associated with mutations in sphingolipid hydrolases[104,317]. However, considerable uncertainty remains regarding the pathological mechanisms underlying sphingolipidoses. Several areas I have been investigating involve questions that have conflicting evidence in the literature and hypotheses that have raised some controversy, including the validity of the psychosine hypothesis and the definition of lipid rafts.

In Chapter 3, I aimed to understand how SapB loads lipids onto CD1d. Using structural and biochemical methods, I attempted to characterise the interaction between CD1d and SapB facilitating the transfer of lipids between these two proteins. Surprisingly, I did not detect an interaction between CD1d and SapB using several sensitive biochemical methods, despite evidence from cell-based and *in vitro* experiments showing that SapB directly loads lipids onto CD1d[165,166,169]. Although the reasons for these inconsistencies are still unclear, my findings strongly suggest that another component is necessary for lipid transfer between CD1d and SapB. The identification of this component will be an exciting challenge. The interaction between CD1d, SapB and any partner protein(s) occurs in the lumen of endolysosomal compartments, limiting the direct use of available techniques such as proximity labelling methods. However, the development of such methods for use in the acidic and degradative environment of endolysosomal compartments may be possible. This may allow the identification of the protein complex facilitating lipid loading onto CD1d, which could then be investigated through structural studies. Furthermore, a lysosomal proximity labelling method would provide a tool to investigate other aspects of lysosomal biology, for example to validate whether sphingolipid-degrading enzymes form multi-enzyme complexes facilitating the sequential transfer of substrate from one enzyme to the next in the lysosome[129,318].

Sphingolipidoses are often accompanied by inflammation[61]. The mechanisms of immune dysregulation by accumulated sphingolipids are still largely undefined. Many non-exclusive factors

may contribute to inflammation, including inflammatory cell death, tissue damage and stress, defective immune signalling, inflammasome activation, as well as aberrant lipid presentation. The consequences of sphingolipidoses on lipid presentation by CD1 molecules have been scarcely investigated. In a recent study by Pereira *et al*., antigen presenting cells from patients with lysosomal storage diseases were shown to retain their capacity to present CD1d and CD1b antigens exogenously added to these cells[319]. Investigating the changes in the self-lipid repertoire presented by CD1 molecules in the context of sphingolipidoses may yield new insights into the immune pathologies of these diseases.

Several immune receptors are known to be dependent on the correct lipid environment for their function and trafficking. Due to the immune abnormalities observed in Krabbe disease, including the formation of globoid phagocytic cells and elevated levels of pro-inflammatory cytokines, I hypothesised that accumulation of galactosphingolipids may cause changes in the membrane proteome of GALC-deficient monocytes/phagocytes. This was investigated in Chapter 4. Generating the cell lines for this study proved more challenging than expected due to wide-ranging off-target effects in the control cell lines, likely caused by lentiviral transduction or off-target Cas9 activity in these cells. This illustrates that genetic modification of cell lines can cause substantial and unpredictable changes, which may only be detected through extensive control experiments. As a consequence of these complications, I could not draw firm conclusions about the effect of GALC loss on the plasma membrane proteome of monocytes. However, I observed that loss of GALC alone was not sufficient to induce a globoid phenotype in monocyte/macrophage cells. This was confirmed in a study published by Weinstock *et al*. showing that globoid cell formation is triggered by accumulation of undigested myelin debris in GALC-deficient macrophages[276]. This demonstrates the complex interplay of different cell types in the pathology of Krabbe disease.

Prior to the start of this project, there was some uncertainty about whether monocytes and macrophages synthesise galactosphingolipids due to the technical difficulties in quantifying these lipids. I observed that the expression of GALC and the enzyme UGT8 responsible for GalCer synthesis is dynamically regulated by differentiation and stimulation of monocyte/macrophage cells, suggesting that galactosphingolipids are present in these cells and that their levels are regulated to adapt for different cell functions required in these states. Weinstock *et al*. showed that no detectable amounts of GalCer are present in mouse bone-marrow derived macrophages, but small amounts of psychosine are detectable[276]. This suggests that some GalCer may be synthesised but processed to psychosine in

these cells. Whether psychosine plays an active role in normal monocyte/macrophage function remains an open question that could be investigated by knockout of UGT8.

Abnormal accumulation of psychosine in oligodendrocytes is thought to drive Krabbe disease pathogenesis, but the enzyme metabolising GalCer to psychosine has long remained elusive[108,109]. Recent evidence showing that the enzyme AC can process substrates other than ceramide[142-144] prompted me to investigate whether AC deacylates GalCer to psychosine, thus contributing to Krabbe disease pathogenesis. This hypothesis formed the basis of the preliminary investigations described in Chapter 5 and was proved to be correct in a study published by Li *et al*.[284]. Genetic ablation of AC in a mouse model of Krabbe disease eliminates psychosine accumulation as well as the clinical symptoms of Krabbe disease. This provides strong support for the idea that psychosine accumulation drives the pathogenesis of Krabbe disease and opens up an exciting therapeutic avenue for this disease.

AC inhibition holds promise for Krabbe disease treatment as well as cancer therapy[139]. However, current AC inhibitors are associated with high toxicity. An improved understanding of AC function may facilitate the design of improved inhibitors. The fundamental mechanism of membrane-embedded substrate access by AC is still unknown and was investigated in Chapter 5. I performed structural studies with SapA lipoprotein nanodiscs providing a membrane environment for interaction with AC. However, instead of SapA nanodiscs, my protocol yielded a different type of tetrameric SapA assembly never observed previously. Unfortunately, this assembly did not provide a lipid bilayer surface suitable for interactions with AC. Other lipid nanodiscs systems have been recently developed, including polymer-based systems suitable for use in the acidic conditions necessary for the interaction of AC with membranes. These tools may help determine the mechanism of substrate access by AC. Furthermore, it is still unclear how SapD promotes this mechanism. A structure of lipid nanodisc-bound AC and SapD would be very informative to understand this process. Finally, the catalytic mechanism of GalCer hydrolysis by AC is currently unknown. I attempted to obtain a crystal structure illustrating this mechanism. Although this aim was not reached, such a structure would be extremely valuable for structure-based design of inhibitors specifically targeting psychosine formation.

Psychosine accumulation in GALC-deficient oligodendrocytes is thought to directly cause cell death and subsequent demyelination[108-110]. I aimed to determine whether loss of AC reduces the levels of psychosine-mediated cell death in GALC KO oligodendrocytes. In contradiction with previously published observations[214,288], I observed that the GALC KO oligodendrocyte cell line generated by our lab did not present increased levels of cell death compared to WT. This challenges our current

understanding of the psychosine hypothesis and opens questions about how loss of GALC in oligodendrocytes causes demyelination. This work highlights that it is likely the interplay between oligodendrocytes, neurons and other cell types in the brain that drives demyelination and cell death.

The formation and maintenance of myelin is dependent on interactions between neurons and oligodendrocytes, as well as interactions between myelin layers[104]. Loss of adhesion molecules at the cell surface of oligodendrocytes may be a mechanism contributing to the demyelination observed in Krabbe disease. Work performed by our group showed that the adhesion protein PTPRM is lost at the surface of GALC-deficient oligodendrocytes but not intracellularly, suggesting that sphingolipid imbalances may cause dysregulation of PTPRM trafficking. Our group showed that the membrane-proximal fibronectin III domain of PTPRM directly and specifically interacts with sulfatide. This suggests that the sphingolipid-binding properties of PTPRM result in dysfunctional trafficking of this protein in GALC-deficient cells. In Chapter 6, I attempted to obtain a structure of the membrane-proximal domains of PTPRM associated with a sulfatide analogue to identify the molecular determinants of this interaction. The sulfatide analogues used in these experiments did not yield a structure of lipid head group-bound PTPRM; future work in our lab will be performed using a better sulfatide analogue synthesised by our collaborators. Nonetheless, I was able to improve our structure model of PTPRM. I observed that the membrane-proximal domain has a unique arch-like loop cleaved in its middle by proprotein convertases. The function of this loop and the effect of cleavage are unknown. An intriguing hypothesis is that this loop may contribute to sulfatide binding and its cleavage may alter the lipid-binding properties of PTPRM, thereby regulating its localisation. Future work in our group will address the interaction of PTPRM with sulfatide and the consequences of sulfatide binding for PTPRM function. These investigations may reveal a role for sphingolipid-protein interactions in Krabbe disease pathogenesis.

Taken together, the work presented in this thesis challenges the current understanding of the fundamental molecular mechanisms of sphingolipid metabolism and presentation to the immune system. This work sets the foundation for new insights into these mechanisms, leading towards improved therapeutic interventions.

8 References

1 Hannun, Y. A. & Obeid, L. M. Sphingolipids and their metabolism in physiology and disease. *Nat Rev Mol Cell Biol* **19**, 175-191, doi:10.1038/nrm.2017.107 (2018).

2 Ogretmen, B. Sphingolipid metabolism in cancer signalling and therapy. *Nat Rev Cancer* **18**, 33-50, doi:10.1038/nrc.2017.96 (2018).

3 van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol* **9**, 112-124, doi:10.1038/nrm2330 (2008).

4 Huang, J. & Feigenson, G. W. A microscopic interaction model of maximum solubility of cholesterol in lipid bilayers. *Biophys J* **76**, 2142-2157, doi:10.1016/S0006-3495(99)77369-8 (1999).

5 Simons, K. & Ikonen, E. Functional rafts in cell membranes. *Nature* **387**, 569-572, doi:10.1038/42408 (1997).

6 Munro, S. Lipid rafts: elusive or illusive? *Cell* **115**, 377-388, doi:10.1016/s0092-8674(03)00882-1 (2003).

7 Levental, I., Levental, K. R. & Heberle, F. A. Lipid Rafts: Controversies Resolved, Mysteries Remain. *Trends Cell Biol* **30**, 341-353, doi:10.1016/j.tcb.2020.01.009 (2020).

8 Levental, I., Lingwood, D., Grzybek, M., Coskun, U. & Simons, K. Palmitoylation regulates raft affinity for the majority of integral raft proteins. *Proc Natl Acad Sci U S A* **107**, 22050-22054, doi:10.1073/pnas.1016184107 (2010).

9 Lorent, J. H. *et al.* Structural determinants and functional consequences of protein affinity for membrane rafts. *Nat Commun* **8**, 1219, doi:10.1038/s41467-017-01328-3 (2017).

10 Fiedler, K., Kobayashi, T., Kurzchalia, T. V. & Simons, K. Glycosphingolipid-enriched, detergent-insoluble complexes in protein sorting in epithelial cells. *Biochemistry* **32**, 6365-6373, doi:10.1021/bi00076a009 (1993).

11 Bieberich, E. Sphingolipids and lipid rafts: Novel concepts and methods of analysis. *Chem Phys Lipids* **216**, 114-131, doi:10.1016/j.chemphyslip.2018.08.003 (2018).

12 Diaz-Rohrer, B. B., Levental, K. R., Simons, K. & Levental, I. Membrane raft association is a determinant of plasma membrane localization. *Proc Natl Acad Sci U S A* **111**, 8500-8505, doi:10.1073/pnas.1404582111 (2014).

13 Stone, M. B., Shelby, S. A., Nunez, M. F., Wisser, K. & Veatch, S. L. Protein sorting by lipid phase-like domains supports emergent signaling function in B lymphocyte plasma membranes. *Elife* **6**, doi:10.7554/eLife.19891 (2017).

14 Triantafilou, M., Miyake, K., Golenbock, D. T. & Triantafilou, K. Mediators of innate immune recognition of bacteria concentrate in lipid rafts and facilitate lipopolysaccharide-induced cell activation. *J Cell Sci* **115**, 2603-2611, doi:10.1242/jcs.115.12.2603 (2002).

15 Plociennikowska, A., Hromada-Judycka, A., Borzecka, K. & Kwiatkowska, K. Co-operation of TLR4 and raft proteins in LPS-induced pro-inflammatory signaling. *Cell Mol Life Sci* **72**, 557-581, doi:10.1007/s00018-014-1762-5 (2015).

16 Hailemariam, T. K. *et al.* Sphingomyelin synthase 2 deficiency attenuates NFkappaB activation. *Arterioscler Thromb Vasc Biol* **28**, 1519-1526, doi:10.1161/ATVBAHA.108.168682 (2008).

17 Li, Z. *et al.* Impact of sphingomyelin synthase 1 deficiency on sphingolipid metabolism and atherosclerosis in mice. *Arterioscler Thromb Vasc Biol* **32**, 1577-1584, doi:10.1161/ATVBAHA.112.251538 (2012).
18 Mobarak, E. *et al.* Glucosylceramide modifies the LPS-induced inflammatory response in macrophages and the orientation of the LPS/TLR4 complex in silico. *Sci Rep* **8**, 13600, doi:10.1038/s41598-018-31926-0 (2018).
19 Köberlin, Marielle S. *et al.* A Conserved Circular Network of Coregulated Lipids Modulates Innate Immune Responses. *Cell* **162**, 170-183, doi:10.1016/j.cell.2015.05.051 (2015).
20 Zhu, X. *et al.* Macrophage ABCA1 reduces MyD88-dependent Toll-like receptor trafficking to lipid rafts by reduction of lipid raft cholesterol. *J Lipid Res* **51**, 3196-3206, doi:10.1194/jlr.M006486 (2010).
21 Stancevic, B. & Kolesnick, R. Ceramide-rich platforms in transmembrane signaling. *FEBS Lett* **584**, 1728-1740, doi:10.1016/j.febslet.2010.02.026 (2010).
22 Shamseddine, A. A., Airola, M. V. & Hannun, Y. A. Roles and regulation of neutral sphingomyelinase-2 in cellular and pathological processes. *Adv Biol Regul* **57**, 24-41, doi:10.1016/j.jbior.2014.10.002 (2015).
23 Megha & London, E. Ceramide selectively displaces cholesterol from ordered lipid domains (rafts): implications for lipid raft structure and function. *J Biol Chem* **279**, 9997-10004, doi:10.1074/jbc.M309992200 (2004).
24 Ira & Johnston, L. J. Sphingomyelinase generation of ceramide promotes clustering of nanoscale domains in supported bilayer membranes. *Biochim Biophys Acta* **1778**, 185-197, doi:10.1016/j.bbamem.2007.09.021 (2008).
25 Cremesti, A. *et al.* Ceramide enables fas to cap and kill. *J Biol Chem* **276**, 23954-23961, doi:10.1074/jbc.M101866200 (2001).
26 Grassme, H., Schwarz, H. & Gulbins, E. Molecular mechanisms of ceramide-mediated CD95 clustering. *Biochem Biophys Res Commun* **284**, 1016-1030, doi:10.1006/bbrc.2001.5045 (2001).
27 Grassme, H., Jendrossek, V., Bock, J., Riehle, A. & Gulbins, E. Ceramide-rich membrane rafts mediate CD40 clustering. *J Immunol* **168**, 298-307, doi:10.4049/jimmunol.168.1.298 (2002).
28 Taha, T. A., Mullen, T. D. & Obeid, L. M. A house divided: ceramide, sphingosine, and sphingosine-1-phosphate in programmed cell death. *Biochim Biophys Acta* **1758**, 2027-2036, doi:10.1016/j.bbamem.2006.10.018 (2006).
29 Nagahashi, M. *et al.* Sphingosine-1-phosphate transporters as targets for cancer therapy. *Biomed Res Int* **2014**, 651727, doi:10.1155/2014/651727 (2014).
30 Hanson, M. A. *et al.* Crystal structure of a lipid G protein-coupled receptor. *Science* **335**, 851-855, doi:10.1126/science.1215904 (2012).
31 Cartier, A. & Hla, T. Sphingosine 1-phosphate: Lipid signaling in pathology and therapy. *Science* **366**, doi:10.1126/science.aar5551 (2019).
32 Garcia-Pacios, M. *et al.* Sphingosine-1-phosphate as an amphipathic metabolite: its properties in aqueous and membrane environments. *Biophys J* **97**, 1398-1407, doi:10.1016/j.bpj.2009.07.001 (2009).
33 Murata, N. *et al.* Interaction of sphingosine 1-phosphate with plasma components, including lipoproteins, regulates the lipid receptor-mediated actions. *Biochem J* **352 Pt 3**, 809-815, doi:10.1042/bj3520809 (2000).

34 Contreras, F. X. *et al.* Molecular recognition of a single sphingolipid species by a protein's transmembrane domain. *Nature* **481**, 525-529, doi:10.1038/nature10742 (2012).
35 Bjorkholm, P. *et al.* Identification of novel sphingolipid-binding motifs in mammalian membrane proteins. *Biochim Biophys Acta* **1838**, 2066-2070, doi:10.1016/j.bbamem.2014.04.026 (2014).
36 Blom, T. *et al.* LAPTM4B facilitates late endosomal ceramide export to control cell death pathways. *Nat Chem Biol* **11**, 799-806, doi:10.1038/nchembio.1889 (2015).
37 Zhou, K. *et al.* A Ceramide-Regulated Element in the Late Endosomal Protein LAPTM4B Controls Amino Acid Transporter Interaction. *ACS Cent Sci* **4**, 548-558, doi:10.1021/acscentsci.7b00582 (2018).
38 Ruysschaert, J. M. & Lonez, C. Role of lipid microdomains in TLR-mediated signalling. *Biochim Biophys Acta* **1848**, 1860-1867, doi:10.1016/j.bbamem.2015.03.014 (2015).
39 Jafurulla, M., Bandari, S., Pucadyil, T. J. & Chattopadhyay, A. Sphingolipids modulate the function of human serotonin1A receptors: Insights from sphingolipid-deficient cells. *Biochim Biophys Acta Biomembr* **1859**, 598-604, doi:10.1016/j.bbamem.2016.10.016 (2017).
40 Prasanna, X., Jafurulla, M., Sengupta, D. & Chattopadhyay, A. The ganglioside GM1 interacts with the serotonin1A receptor via the sphingolipid binding domain. *Biochim Biophys Acta* **1858**, 2818-2826, doi:10.1016/j.bbamem.2016.08.009 (2016).
41 Coskun, U., Grzybek, M., Drechsel, D. & Simons, K. Regulation of human EGF receptor by lipids. *Proc Natl Acad Sci U S A* **108**, 9044-9048, doi:10.1073/pnas.1105666108 (2011).
42 Yoon, S. J., Nakayama, K., Hikita, T., Handa, K. & Hakomori, S. I. Epidermal growth factor receptor tyrosine kinase is modulated by GM3 interaction with N-linked GlcNAc termini of the receptor. *Proc Natl Acad Sci U S A* **103**, 18987-18991, doi:10.1073/pnas.0609281103 (2006).
43 Silva, M., Videira, P. A. & Sackstein, R. E-Selectin Ligands in the Human Mononuclear Phagocyte System: Implications for Infection, Inflammation, and Immunotherapy. *Front Immunol* **8**, 1878, doi:10.3389/fimmu.2017.01878 (2017).
44 Pronker, M. F. *et al.* Structural basis of myelin-associated glycoprotein adhesion and signalling. *Nat Commun* **7**, 13584, doi:10.1038/ncomms13584 (2016).
45 Popovic, Z. V. *et al.* Sulfated glycosphingolipid as mediator of phagocytosis: SM4s enhances apoptotic cell clearance and modulates macrophage activity. *J Immunol* **179**, 6770-6782, doi:10.4049/jimmunol.179.10.6770 (2007).
46 Schneider-Schaulies, J. & Schneider-Schaulies, S. Sphingolipids in viral infection. *Biol Chem* **396**, 585-595, doi:10.1515/hsz-2014-0273 (2015).
47 Belotserkovsky, I. *et al.* Glycan-Glycan Interaction Determines Shigella Tropism toward Human T Lymphocytes. *mBio* **9**, doi:10.1128/mBio.02309-17 (2018).
48 Sprong, H. *et al.* UDP-galactose:ceramide galactosyltransferase is a class I integral membrane protein of the endoplasmic reticulum. *J Biol Chem* **273**, 25880-25888, doi:10.1074/jbc.273.40.25880 (1998).
49 Kumagai, K. & Hanada, K. Structure, functions and regulation of CERT, a lipid-transfer protein for the delivery of ceramide at the ER-Golgi membrane contact sites. *FEBS Lett* **593**, 2366-2377, doi:10.1002/1873-3468.13511 (2019).

50 Kudo, N. *et al.* Structural basis for specific lipid recognition by CERT responsible for nonvesicular trafficking of ceramide. *Proc Natl Acad Sci U S A* **105**, 488-493, doi:10.1073/pnas.0709191105 (2008).
51 Yamaji, T. & Hanada, K. Sphingolipid metabolism and interorganellar transport: localization of sphingolipid enzymes and lipid transfer proteins. *Traffic* **16**, 101-122, doi:10.1111/tra.12239 (2015).
52 Ochoa-Lizarralde, B. *et al.* Structural analyses of 4-phosphate adaptor protein 2 yield mechanistic insights into sphingolipid recognition by the glycolipid transfer protein family. *J Biol Chem* **293**, 16709-16723, doi:10.1074/jbc.RA117.000733 (2018).
53 Kolter, T. & Sandhoff, K. Lysosomal degradation of membrane lipids. *FEBS Lett* **584**, 1700-1712, doi:10.1016/j.febslet.2009.10.021 (2010).
54 Sillence, D. J. & Platt, F. M. Glycosphingolipids in endocytic membrane transport. *Semin Cell Dev Biol* **15**, 409-416, doi:10.1016/j.semcdb.2004.03.004 (2004).
55 Rudnik, S. & Damme, M. The lysosomal membrane-export of metabolites and beyond. *FEBS J*, doi:10.1111/febs.15602 (2020).
56 Kitatani, K., Idkowiak-Baldys, J. & Hannun, Y. A. The sphingolipid salvage pathway in ceramide metabolism and signaling. *Cell Signal* **20**, 1010-1018, doi:10.1016/j.cellsig.2007.12.006 (2008).
57 Lloyd-Evans, E. *et al.* Niemann-Pick disease type C1 is a sphingosine storage disease that causes deregulation of lysosomal calcium. *Nat Med* **14**, 1247-1255, doi:10.1038/nm.1876 (2008).
58 Fuller, M. *et al.* Glucosylceramide accumulation is not confined to the lysosome in fibroblasts from patients with Gaucher disease. *Mol Genet Metab* **93**, 437-443, doi:10.1016/j.ymgme.2007.11.011 (2008).
59 Hein, L. K., Meikle, P. J., Hopwood, J. J. & Fuller, M. Secondary sphingolipid accumulation in a macrophage model of Gaucher disease. *Mol Genet Metab* **92**, 336-345, doi:10.1016/j.ymgme.2007.08.001 (2007).
60 Platt, F. M., d'Azzo, A., Davidson, B. L., Neufeld, E. F. & Tifft, C. J. Lysosomal storage diseases. *Nat Rev Dis Primers* **4**, 27, doi:10.1038/s41572-018-0025-4 (2018).
61 Vitner, E. B., Platt, F. M. & Futerman, A. H. Common and uncommon pathogenic cascades in lysosomal storage diseases. *J Biol Chem* **285**, 20423-20427, doi:10.1074/jbc.R110.134452 (2010).
62 van Eijk, M., Ferraz, M. J., Boot, R. G. & Aerts, J. Lyso-glycosphingolipids: presence and consequences. *Essays Biochem* **64**, 565-578, doi:10.1042/EBC20190090 (2020).
63 Sprong, H., van der Sluijs, P. & van Meer, G. How proteins move lipids and lipids move proteins. *Nat Rev Mol Cell Biol* **2**, 504-513, doi:10.1038/35080071 (2001).
64 Matsuda, J. *et al.* A mutation in the saposin A domain of the sphingolipid activator protein (prosaposin) gene results in a late-onset, chronic form of globoid cell leukodystrophy in the mouse. *Hum Mol Genet* **10**, 1191-1199, doi:10.1093/hmg/10.11.1191 (2001).
65 Spiegel, R. *et al.* A mutation in the saposin A coding region of the prosaposin gene in an infant presenting as Krabbe disease: first report of saposin A deficiency in humans. *Mol Genet Metab* **84**, 160-166, doi:10.1016/j.ymgme.2004.10.004 (2005).
66 Kuchar, L. *et al.* Prosaposin deficiency and saposin B deficiency (activator-deficient metachromatic leukodystrophy): report on two patients detected by analysis of urinary sphingolipids and carrying novel PSAP gene mutations. *Am J Med Genet A* **149A**, 613-621, doi:10.1002/ajmg.a.32712 (2009).

67 Vaccaro, A. M. *et al.* Saposin C mutations in Gaucher disease patients resulting in lysosomal lipid accumulation, saposin C deficiency, but normal prosaposin processing and sorting. *Hum Mol Genet* **19**, 2987-2997, doi:10.1093/hmg/ddq204 (2010).
68 Motta, M. *et al.* Gaucher disease due to saposin C deficiency is an inherited lysosomal disease caused by rapidly degraded mutant proteins. *Hum Mol Genet* **23**, 5814-5826, doi:10.1093/hmg/ddu299 (2014).
69 Kang, L. *et al.* A rare form of Gaucher disease resulting from saposin C deficiency. *Blood Cells Mol Dis* **68**, 60-65, doi:10.1016/j.bcmd.2017.04.001 (2018).
70 Hazkani-Covo, E., Altman, N., Horowitz, M. & Graur, D. The evolutionary history of prosaposin: two successive tandem-duplication events gave rise to the four saposin domains in vertebrates. *J Mol Evol* **54**, 30-34, doi:10.1007/s00239-001-0014-0 (2002).
71 Lefrancois, S., Zeng, J., Hassan, A. J., Canuel, M. & Morales, C. R. The lysosomal trafficking of sphingolipid activator proteins (SAPs) is mediated by sortilin. *EMBO J* **22**, 6430-6437, doi:10.1093/emboj/cdg629 (2003).
72 Canuel, M., Korkidakis, A., Konnyu, K. & Morales, C. R. Sortilin mediates the lysosomal targeting of cathepsins D and H. *Biochem Biophys Res Commun* **373**, 292-297, doi:10.1016/j.bbrc.2008.06.021 (2008).
73 Hiraiwa, M. *et al.* Lysosomal proteolysis of prosaposin, the precursor of saposins (sphingolipid activator proteins): its mechanism and inhibition by ganglioside. *Arch Biochem Biophys* **341**, 17-24, doi:10.1006/abbi.1997.9958 (1997).
74 Xiong, Z. J., Huang, J., Poda, G., Pomes, R. & Prive, G. G. Structure of Human Acid Sphingomyelinase Reveals the Role of the Saposin Domain in Activating Substrate Hydrolysis. *J Mol Biol* **428**, 3026-3042, doi:10.1016/j.jmb.2016.06.012 (2016).
75 Wright, C. S., Zhao, Q. & Rastinejad, F. Structural analysis of lipid complexes of GM2-activator protein. *J Mol Biol* **331**, 951-964, doi:10.1016/s0022-2836(03)00794-0 (2003).
76 Wendeler, M. *et al.* The enzyme-binding region of human GM2-activator protein. *FEBS J* **273**, 982-991, doi:10.1111/j.1742-4658.2006.05126.x (2006).
77 Li, X., Saha, P., Li, J., Blobel, G. & Pfeffer, S. R. Clues to the mechanism of cholesterol transfer from the structure of NPC1 middle lumenal domain bound to NPC2. *Proc Natl Acad Sci U S A* **113**, 10079-10084, doi:10.1073/pnas.1611956113 (2016).
78 Kishimoto, Y., Hiraiwa, M. & O'Brien, J. S. Saposins: structure, function, distribution, and molecular genetics. *J Lipid Res* **33**, 1255-1267, doi:10.1016/S0022-2275(20)40540-1 (1992).
79 Ahn, V. E., Leyko, P., Alattia, J. R., Chen, L. & Prive, G. G. Crystal structures of saposins A and C. *Protein Sci* **15**, 1849-1857, doi:10.1110/ps.062256606 (2006).
80 Hill, C. H. *et al.* The mechanism of glycosphingolipid degradation revealed by a GALC-SapA complex structure. *Nat Commun* **9**, 151, doi:10.1038/s41467-017-02361-y (2018).
81 Popovic, K., Holyoake, J., Pomes, R. & Prive, G. G. Structure of saposin A lipoprotein discs. *Proc Natl Acad Sci U S A* **109**, 2908-2912, doi:10.1073/pnas.1115743109 (2012).
82 Ahn, V. E., Faull, K. F., Whitelegge, J. P., Fluharty, A. L. & Prive, G. G. Crystal structure of saposin B reveals a dimeric shell for lipid binding. *Proc Natl Acad Sci U S A* **100**, 38-43, doi:10.1073/pnas.0136947100 (2003).

83 Rossmann, M. *et al.* Crystal structures of human saposins C and D: implications for lipid recognition and membrane interactions. *Structure* **16**, 809-817, doi:10.1016/j.str.2008.02.016 (2008).
84 Hawkins, C. A., de Alba, E. & Tjandra, N. Solution structure of human saposin C in a detergent environment. *J Mol Biol* **346**, 1381-1392, doi:10.1016/j.jmb.2004.12.045 (2005).
85 Popovic, K. & Prive, G. G. Structures of the human ceramide activator protein saposin D. *Acta Crystallogr D Biol Crystallogr* **64**, 589-594, doi:10.1107/S0907444908003120 (2008).
86 Gebai, A., Gorelik, A. & Nagar, B. Crystal structure of saposin D in an open conformation. *J Struct Biol* **204**, 145-150, doi:10.1016/j.jsb.2018.07.011 (2018).
87 Remmel, N., Locatelli-Hoops, S., Breiden, B., Schwarzmann, G. & Sandhoff, K. Saposin B mobilizes lipids from cholesterol-poor and bis(monoacylglycero)phosphate-rich membranes at acidic pH - Unglycosylated patient variant saposin B lacks lipid-extraction capacity. *FEBS J* **274**, 3405-3420, doi:10.1111/j.1742-4658.2007.05873.x (2007).
88 Alattia, J. R., Shaw, J. E., Yip, C. M. & Prive, G. G. Molecular imaging of membrane interfaces reveals mode of beta-glucosidase activation by saposin C. *Proc Natl Acad Sci U S A* **104**, 17394-17399, doi:10.1073/pnas.0704998104 (2007).
89 Alattia, J. R., Shaw, J. E., Yip, C. M. & Prive, G. G. Direct visualization of saposin remodelling of lipid bilayers. *J Mol Biol* **362**, 943-953, doi:10.1016/j.jmb.2006.08.009 (2006).
90 Linke, T. *et al.* Interfacial regulation of acid ceramidase activity. Stimulation of ceramide degradation by lysosomal lipids and sphingolipid activator proteins. *J Biol Chem* **276**, 5760-5768, doi:10.1074/jbc.M006846200 (2001).
91 Tatti, M. *et al.* Structural and membrane-binding properties of saposin D. *Eur J Biochem* **263**, 486-494, doi:10.1046/j.1432-1327.1999.00521.x (1999).
92 Ciaffoni, F. *et al.* Saposin D solubilizes anionic phospholipid-containing membranes. *J Biol Chem* **276**, 31583-31589, doi:10.1074/jbc.M102736200 (2001).
93 Chien, C. H. *et al.* An Adaptable Phospholipid Membrane Mimetic System for Solution NMR Studies of Membrane Proteins. *J Am Chem Soc* **139**, 14829-14832, doi:10.1021/jacs.7b06730 (2017).
94 Frauenfeld, J. *et al.* A saposin-lipoprotein nanoparticle system for membrane proteins. *Nat Methods* **13**, 345-351, doi:10.1038/nmeth.3801 (2016).
95 Flayhan, A. *et al.* Saposin Lipid Nanoparticles: A Highly Versatile and Modular Tool for Membrane Protein Research. *Structure* **26**, 345-355 e345, doi:10.1016/j.str.2018.01.007 (2018).
96 Jain, M. & De Jesus, O. Krabbe Disease. In: *StatPearls* (StatPearls Publishing, 2021).
97 Allewelt, H. *et al.* Long-Term Functional Outcomes after Hematopoietic Stem Cell Transplant for Early Infantile Krabbe Disease. *Biol Blood Marrow Transplant* **24**, 2233-2238, doi:10.1016/j.bbmt.2018.06.020 (2018).
98 Suzuki, K. & Suzuki, K. The twitcher mouse: a model for Krabbe disease and for experimental therapies. *Brain Pathol* **5**, 249-258, doi:10.1111/j.1750-3639.1995.tb00601.x (1995).
99 Hirth, R. S. & Nielsen, S. W. A familial canine globoid cell leukodystrophy ("Krabbe Type"). *J Small Anim Pract* **8**, 569-575, doi:10.1111/j.1748-5827.1967.tb04499.x (1967).

100 Johnson, K. H. Globoid leukodystrophy in the cat. *J Am Vet Med Assoc* **157**, 2057-2064 (1970).
101 Luzi, P., Rafi, M. A., Victoria, T., Baskin, G. B. & Wenger, D. A. Characterization of the rhesus monkey galactocerebrosidase (GALC) cDNA and gene and identification of the mutation causing globoid cell leukodystrophy (Krabbe disease) in this primate. *Genomics* **42**, 319-324, doi:10.1006/geno.1997.4744 (1997).
102 Pritchard, D. H., Napthine, D. V. & Sinclair, A. J. Globoid cell leucodystrophy in polled Dorset sheep. *Vet Pathol* **17**, 399-405, doi:10.1177/030098588001700402 (1980).
103 Poitelon, Y., Kopec, A. M. & Belin, S. Myelin Fat Facts: An Overview of Lipids and Fatty Acid Metabolism. *Cells* **9**, doi:10.3390/cells9040812 (2020).
104 Olsen, A. S. B. & Faergeman, N. J. Sphingolipids: membrane microdomains in brain development, function and neurological diseases. *Open Biol* **7**, doi:10.1098/rsob.170069 (2017).
105 Charles, P. *et al.* Neurofascin is a glial receptor for the paranodin/Caspr-contactin axonal complex at the axoglial junction. *Curr Biol* **12**, 217-220, doi:10.1016/s0960-9822(01)00680-7 (2002).
106 Boggs, J. M. *et al.* Participation of galactosylceramide and sulfatide in glycosynapses between oligodendrocyte or myelin membranes. *FEBS Lett* **584**, 1771-1778, doi:10.1016/j.febslet.2009.11.074 (2010).
107 Stewart, R. J. & Boggs, J. M. A carbohydrate-carbohydrate interaction between galactosylceramide-containing liposomes and cerebroside sulfate-containing liposomes: dependence on the glycolipid ceramide composition. *Biochemistry* **32**, 10666-10674, doi:10.1021/bi00091a017 (1993).
108 Miyatake, T. & Suzuki, K. Globoid cell leukodystrophy: additional deficiency of psychosine galactosidase. *Biochem Biophys Res Commun* **48**, 539-543, doi:10.1016/0006-291x(72)90381-6 (1972).
109 Suzuki, K. Twenty five years of the "psychosine hypothesis": a personal perspective of its history and present status. *Neurochem Res* **23**, 251-259, doi:10.1023/a:1022436928925 (1998).
110 Feltri, M. L. *et al.* Mechanisms of demyelination and neurodegeneration in globoid cell leukodystrophy. *Glia*, doi:10.1002/glia.24008 (2021).
111 Hawkins-Salsbury, J. A. *et al.* Psychosine, the cytotoxic sphingolipid that accumulates in globoid cell leukodystrophy, alters membrane architecture. *J Lipid Res* **54**, 3303-3311, doi:10.1194/jlr.M039610 (2013).
112 Won, J. S., Singh, A. K. & Singh, I. Biochemical, cell biological, pathological, and therapeutic aspects of Krabbe's disease. *J Neurosci Res* **94**, 990-1006, doi:10.1002/jnr.23873 (2016).
113 Nicaise, A. M., Bongarzone, E. R. & Crocker, S. J. A microglial hypothesis of globoid cell leukodystrophy pathology. *J Neurosci Res* **94**, 1049-1061, doi:10.1002/jnr.23773 (2016).
114 Ida, H., Rennert, O. M., Watabe, K., Eto, Y. & Maekawa, K. Pathological and biochemical studies of fetal Krabbe disease. *Brain Dev* **16**, 480-484, doi:10.1016/0387-7604(94)90013-2 (1994).
115 Martin, J. J. *et al.* Fetal Krabbe leukodystrophy. A morphologic study of two cases. *Acta Neuropathol* **53**, 87-91, doi:10.1007/BF00689987 (1981).

116 Potter, G. B. *et al.* Missense mutation in mouse GALC mimics human gene defect and offers new insights into Krabbe disease. *Hum Mol Genet* **22**, 3397-3414, doi:10.1093/hmg/ddt190 (2013).
117 LeVine, S. M. & Brown, D. C. IL-6 and TNFα expression in brains of twitcher, quaking and normal mice. *J Neuroimmunol* **73**, 47-56, doi:10.1016/S0165-5728(96)00166-X (1997).
118 Borda, J. T. *et al.* Clinical and immunopathologic alterations in rhesus macaques affected with globoid cell leukodystrophy. *Am J Pathol* **172**, 98-111, doi:10.2353/ajpath.2008.070404 (2008).
119 Jurewicz, A., Matysiak, M., Tybor, K. & Selmaj, K. TNF-induced death of adult human oligodendrocytes is mediated by c-jun NH2-terminal kinase-3. *Brain* **126**, 1358-1370, doi:10.1093/brain/awg146 (2003).
120 Nakazawa, T. *et al.* Tumor necrosis factor-alpha mediates oligodendrocyte death and delayed retinal ganglion cell loss in a mouse model of glaucoma. *J Neurosci* **26**, 12633-12641, doi:10.1523/JNEUROSCI.2801-06.2006 (2006).
121 Pasqui, A. L., Di Renzo, M., Auteri, A., Federico, G. & Puccetti, L. Increased TNF-alpha production by peripheral blood mononuclear cells in patients with Krabbe's disease: effect of psychosine. *Eur J Clin Invest* **37**, 742-745, doi:10.1111/j.1365-2362.2007.01850.x (2007).
122 Gane, J. M., Stockley, R. A. & Sapey, E. TNF-alpha Autocrine Feedback Loops in Human Monocytes: The Pro- and Anti-Inflammatory Roles of the TNF-alpha Receptors Support the Concept of Selective TNFR1 Blockade In Vivo. *J Immunol Res* **2016**, 1079851, doi:10.1155/2016/1079851 (2016).
123 Zhang, T., de Waard, A. A., Wuhrer, M. & Spaapen, R. M. The Role of Glycosphingolipids in Immune Cell Functions. *Front Immunol* **10**, 90, doi:10.3389/fimmu.2019.00090 (2019).
124 Eliyahu, E., Park, J. H., Shtraizent, N., He, X. & Schuchman, E. H. Acid ceramidase is a novel factor required for early embryo survival. *FASEB J* **21**, 1403-1409, doi:10.1096/fj.06-7016com (2007).
125 Li, C. M. *et al.* Insertional mutagenesis of the mouse acid ceramidase gene leads to early embryonic lethality in homozygotes and progressive lipid storage disease in heterozygotes. *Genomics* **79**, 218-224, doi:10.1006/geno.2002.6686 (2002).
126 Ferlinz, K. *et al.* Human acid ceramidase: processing, glycosylation, and lysosomal targeting. *J Biol Chem* **276**, 35352-35360, doi:10.1074/jbc.M103066200 (2001).
127 Shtraizent, N. *et al.* Autoproteolytic cleavage and activation of human acid ceramidase. *J Biol Chem* **283**, 11253-11259, doi:10.1074/jbc.M709166200 (2008).
128 Gebai, A., Gorelik, A., Li, Z., Illes, K. & Nagar, B. Structural basis for the activation of acid ceramidase. *Nat Commun* **9**, 1621, doi:10.1038/s41467-018-03844-2 (2018).
129 He, X. *et al.* Purification and characterization of recombinant, human acid ceramidase. Catalytic reactions and interactions with acid sphingomyelinase. *J Biol Chem* **278**, 32978-32986, doi:10.1074/jbc.M301936200 (2003).
130 Okino, N. *et al.* The reverse activity of human acid ceramidase. *J Biol Chem* **278**, 29948-29953, doi:10.1074/jbc.M303310200 (2003).
131 Azuma, N., O'Brien, J. S., Moser, H. W. & Kishimoto, Y. Stimulation of acid ceramidase activity by saposin D. *Arch Biochem Biophys* **311**, 354-357, doi:10.1006/abbi.1994.1248 (1994).

132 Klein, A. *et al.* Sphingolipid activator protein D (sap-D) stimulates the lysosomal degradation of ceramide in vivo. *Biochem Biophys Res Commun* **200**, 1440-1448, doi:10.1006/bbrc.1994.1612 (1994).
133 Matsuda, J. *et al.* Mutation in saposin D domain of sphingolipid activator protein gene causes urinary system defects and cerebellar Purkinje cell degeneration with accumulation of hydroxy fatty acid-containing ceramide in mouse. *Hum Mol Genet* **13**, 2709-2723, doi:10.1093/hmg/ddh281 (2004).
134 Yu, F. P. S., Amintas, S., Levade, T. & Medin, J. A. Acid ceramidase deficiency: Farber disease and SMA-PME. *Orphanet J Rare Dis* **13**, 121, doi:10.1186/s13023-018-0845-z (2018).
135 Ehlert, K. *et al.* Allogeneic hematopoietic cell transplantation in Farber disease. *J Inherit Metab Dis* **42**, 286-294, doi:10.1002/jimd.12043 (2019).
136 Liu, X. *et al.* Acid ceramidase upregulation in prostate cancer: role in tumor development and implications for therapy. *Expert Opin Ther Targets* **13**, 1449-1458, doi:10.1517/14728220903357512 (2009).
137 Cheng, J. C. *et al.* Radiation-induced acid ceramidase confers prostate cancer resistance and tumor relapse. *J Clin Invest* **123**, 4344-4358, doi:10.1172/JCI64791 (2013).
138 Beckham, T. H. *et al.* Acid ceramidase induces sphingosine kinase 1/S1P receptor 2-mediated activation of oncogenic Akt signaling. *Oncogenesis* **2**, e49, doi:10.1038/oncsis.2013.14 (2013).
139 Cho, S. M. & Kwon, H. J. Acid ceramidase, an emerging target for anti-cancer and anti-angiogenesis. *Arch Pharm Res* **42**, 232-243, doi:10.1007/s12272-019-01114-3 (2019).
140 Kuzuhara, S. *et al.* Subacute leucoencephalopathy induced by carmofur, a 5-fluorouracil derivative. *J Neurol* **234**, 365-370, doi:10.1007/BF00314079 (1987).
141 Dementiev, A. *et al.* Molecular Mechanism of Inhibition of Acid Ceramidase by Carmofur. *J Med Chem* **62**, 987-992, doi:10.1021/acs.jmedchem.8b01723 (2019).
142 Ferraz, M. J. *et al.* Lysosomal glycosphingolipid catabolism by acid ceramidase: formation of glycosphingoid bases during deficiency of glycosidases. *FEBS Lett* **590**, 716-725, doi:10.1002/1873-3468.12104 (2016).
143 Kim, M. J., Jeon, S., Burbulla, L. F. & Krainc, D. Acid ceramidase inhibition ameliorates alpha-synuclein accumulation upon loss of GBA1 function. *Hum Mol Genet* **27**, 1972-1988, doi:10.1093/hmg/ddy105 (2018).
144 Yamaguchi, Y., Sasagasako, N., Goto, I. & Kobayashi, T. The synthetic pathway for glucosylsphingosine in cultured fibroblasts. *J Biochem* **116**, 704-710, doi:10.1093/oxfordjournals.jbchem.a124584 (1994).
145 Siddiqui, S., Visvabharathy, L. & Wang, C. R. Role of Group 1 CD1-Restricted T Cells in Infectious Disease. *Front Immunol* **6**, 337, doi:10.3389/fimmu.2015.00337 (2015).
146 Chaudhry, M. S. & Karadimitris, A. Role and regulation of CD1d in normal and pathological B cells. *J Immunol* **193**, 4761-4768, doi:10.4049/jimmunol.1401805 (2014).
147 Layre, E., de Jong, A. & Moody, D. B. Human T cells use CD1 and MR1 to recognize lipids and small molecules. *Curr Opin Chem Biol* **23**, 31-38, doi:10.1016/j.cbpa.2014.09.007 (2014).
148 Moody, D. B. & Cotton, R. N. Four pathways of CD1 antigen presentation to T cells. *Curr Opin Immunol* **46**, 127-133, doi:10.1016/j.coi.2017.07.013 (2017).

149 de la Salle, H. *et al.* Assistance of microbial glycolipid antigen processing by CD1e. *Science* **310**, 1321-1324, doi:10.1126/science.1115301 (2005).
150 Wieland Brown, L. C. *et al.* Production of alpha-galactosylceramide by a prominent member of the human gut microbiota. *PLoS Biol* **11**, e1001610, doi:10.1371/journal.pbio.1001610 (2013).
151 Brennan, P. J., Brigl, M. & Brenner, M. B. Invariant natural killer T cells: an innate activation scheme linked to diverse effector functions. *Nat Rev Immunol* **13**, 101-117, doi:10.1038/nri3369 (2013).
152 Oleinika, K. *et al.* CD1d-dependent immune suppression mediated by regulatory B cells through modulations of iNKT cells. *Nat Commun* **9**, 684, doi:10.1038/s41467-018-02911-y (2018).
153 Kawakami, K. *et al.* Critical role of Valpha14+ natural killer T cells in the innate phase of host protection against Streptococcus pneumoniae infection. *Eur J Immunol* **33**, 3322-3330, doi:10.1002/eji.200324254 (2003).
154 Opasawatchai, A. & Matangkasombut, P. iNKT Cells and Their Potential Lipid Ligands during Viral Infection. *Front Immunol* **6**, 378, doi:10.3389/fimmu.2015.00378 (2015).
155 Jahng, A. *et al.* Prevention of autoimmunity by targeting a distinct, noninvariant CD1d-reactive T cell population reactive to sulfatide. *J Exp Med* **199**, 947-957, doi:10.1084/jem.20031389 (2004).
156 Cuevas-Zuviria, B., Minguez-Toral, M., Diaz-Perales, A., Garrido-Arandia, M. & Pacios, L. F. Structural Dynamics of the Lipid Antigen-Binding Site of CD1d Protein. *Biomolecules* **10**, doi:10.3390/biom10040532 (2020).
157 Cox, D. *et al.* Determination of cellular lipids bound to human CD1d molecules. *PLoS One* **4**, e5325, doi:10.1371/journal.pone.0005325 (2009).
158 Yuan, W., Kang, S. J., Evans, J. E. & Cresswell, P. Natural lipid ligands associated with human CD1d targeted to different subcellular compartments. *J Immunol* **182**, 4784-4791, doi:10.4049/jimmunol.0803981 (2009).
159 Haig, N. A. *et al.* Identification of Self-lipids Presented by CD1c and CD1d Proteins. *J Biol Chem* **286**, 37692-37701, doi:10.1074/jbc.M111.267948 (2011).
160 Teyton, L. Role of lipid transfer proteins in loading CD1 antigen-presenting molecules. *J Lipid Res* **59**, 1367-1373, doi:10.1194/jlr.R083212 (2018).
161 Prigozy, T. I. *et al.* Glycolipid antigen processing for presentation by CD1d molecules. *Science* **291**, 664-667, doi:10.1126/science.291.5504.664 (2001).
162 Darmoise, A. *et al.* Lysosomal alpha-galactosidase controls the generation of self lipid antigens for natural killer T cells. *Immunity* **33**, 216-228, doi:10.1016/j.immuni.2010.08.003 (2010).
163 Zhou, D. *et al.* Editing of CD1d-bound lipid antigens by endosomal lipid transfer proteins. *Science* **303**, 523-527, doi:10.1126/science.1092009 (2004).
164 Kang, S. J. & Cresswell, P. Saposins facilitate CD1d-restricted presentation of an exogenous lipid antigen to T cells. *Nat Immunol* **5**, 175-181, doi:10.1038/ni1034 (2004).
165 Yuan, W. *et al.* Saposin B is the dominant saposin that facilitates lipid binding to human CD1d molecules. *Proc Natl Acad Sci U S A* **104**, 5551-5556, doi:10.1073/pnas.0700617104 (2007).
166 Salio, M. *et al.* Saposins modulate human invariant Natural Killer T cells self-reactivity and facilitate lipid exchange with CD1d molecules during antigen presentation. *Proc Natl Acad Sci U S A* **110**, E4753-4761, doi:10.1073/pnas.1310050110 (2013).

167 Schrantz, N. *et al.* The Niemann-Pick type C2 protein loads isoglobotrihexosylceramide onto CD1d molecules and contributes to the thymic selection of NKT cells. *J Exp Med* **204**, 841-852, doi:10.1084/jem.20061562 (2007).
168 Winau, F. *et al.* Saposin C is required for lipid presentation by human CD1b. *Nat Immunol* **5**, 169-174, doi:10.1038/ni1035 (2004).
169 Leon, L. *et al.* Saposins utilize two strategies for lipid transfer and CD1 antigen presentation. *Proc Natl Acad Sci U S A* **109**, 4357-4364, doi:10.1073/pnas.1200764109 (2012).
170 Wong, L. H., Gatta, A. T. & Levine, T. P. Lipid transfer proteins: the lipid commute via shuttles, bridges and tubes. *Nat Rev Mol Cell Biol* **20**, 85-101, doi:10.1038/s41580-018-0071-5 (2019).
171 Ryu, J. K. *et al.* Reconstruction of LPS Transfer Cascade Reveals Structural Determinants within LBP, CD14, and TLR4-MD2 for Efficient LPS Recognition and Transfer. *Immunity* **46**, 38-50, doi:10.1016/j.immuni.2016.11.007 (2017).
172 Hussain, M. M., Rava, P., Walsh, M., Rana, M. & Iqbal, J. Multiple functions of microsomal triglyceride transfer protein. *Nutr Metab (Lond)* **9**, 14, doi:10.1186/1743-7075-9-14 (2012).
173 Li, Z., Michael, I. P., Zhou, D., Nagy, A. & Rini, J. M. Simple piggyBac transposon-based mammalian cell expression system for inducible protein production. *Proc Natl Acad Sci U S A* **110**, 5004-5009, doi:10.1073/pnas.1218620110 (2013).
174 Heckman, K. L. & Pease, L. R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc* **2**, 924-932, doi:10.1038/nprot.2007.132 (2007).
175 Lund, M. E., To, J., O'Brien, B. A. & Donnelly, S. The choice of phorbol 12-myristate 13-acetate differentiation protocol influences the response of THP-1 macrophages to a pro-inflammatory stimulus. *J Immunol Methods* **430**, 64-70, doi:10.1016/j.jim.2016.01.012 (2016).
176 Pellicci, D. G. *et al.* Differential recognition of CD1d-alpha-galactosyl ceramide by the V beta 8.2 and V beta 7 semi-invariant NKT T cell receptors. *Immunity* **31**, 47-59, doi:10.1016/j.immuni.2009.04.018 (2009).
177 Lopez-Sagaseta, J., Kung, J. E., Savage, P. B., Gumperz, J. & Adams, E. J. The molecular basis for recognition of CD1d/alpha-galactosylceramide by a human non-Valpha24 T cell receptor. *PLoS Biol* **10**, e1001412, doi:10.1371/journal.pbio.1001412 (2012).
178 Fairhead, M. & Howarth, M. Site-specific biotinylation of purified proteins using BirA. *Methods Mol Biol* **1266**, 171-184, doi:10.1007/978-1-4939-2272-7_12 (2015).
179 Walter, T. S. *et al.* Semi-automated microseeding of nanolitre crystallization experiments. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **64**, 14-18, doi:10.1107/S1744309107057260 (2008).
180 Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**, 186-190, doi:10.1107/S0021889809045701 (2010).
181 Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol* **74**, 85-97, doi:10.1107/S2059798317017235 (2018).
182 Tickle, I. J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., & Vonrhein, C., Bricogne, G. STARANISO. Cambridge, United Kingdom: Global Phasing Ltd. (2018).
183 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214, doi:10.1107/S0907444913000061 (2013).

184 Lopez-Sagaseta, J., Sibener, L. V., Kung, J. E., Gumperz, J. & Adams, E. J. Lysophospholipid presentation by CD1d and recognition by a human Natural Killer T-cell receptor. *EMBO J* **31**, 2047-2059, doi:10.1038/emboj.2012.54 (2012).
185 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
186 Bunkoczi, G. & Read, R. J. Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr D Biol Crystallogr* **67**, 303-312, doi:10.1107/S0907444910051218 (2011).
187 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).
188 Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367, doi:10.1107/S0907444912001308 (2012).
189 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).
190 Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519-530, doi:10.1107/S2059798318002425 (2018).
191 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21, doi:10.1107/S0907444909042073 (2010).
192 Emsley, P. & Crispin, M. Structural analysis of glycoproteins: building N-linked glycans with Coot. *Acta Crystallogr D Struct Biol* **74**, 256-263, doi:10.1107/S2059798318005119 (2018).
193 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82, doi:10.1002/pro.3943 (2021).
194 Laguerre, M., Saux, M., Dubost, J. P. & Carpy, A. MLPP: A Program for the Calculation of Molecular Lipophilicity Potential in Proteins. *Pharm. Pharmacol. Commun* **3**, 217-222, doi:10.1111/j.2042-7158.1997.tb00257.x (1997).
195 Ghose, A. K., Viswanadhan, V. N. & Wendoloski, J. J. Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *J Phys Chem A* **102**, 3762-3772, doi:10.1021/jp980230o (1998).
196 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
197 Xu, D., Jaroszewski, L., Li, Z. & Godzik, A. FFAS-3D: improving fold recognition by including optimized structural features and template re-ranking. *Bioinformatics* **30**, 660-667, doi:10.1093/bioinformatics/btt578 (2014).
198 Holm, L. DALI and the persistence of protein shape. *Protein Sci* **29**, 128-140, doi:10.1002/pro.3749 (2020).
199 Kumar, M. *et al.* ELM-the eukaryotic linear motif resource in 2020. *Nucleic Acids Res* **48**, D296-D306, doi:10.1093/nar/gkz1030 (2020).
200 Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881-10890, doi:10.1093/nar/16.22.10881 (1988).

201 Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191, doi:10.1093/bioinformatics/btp033 (2009).
202 Spratley, S. J. *et al.* Molecular Mechanisms of Disease Pathogenesis Differ in Krabbe Disease Variants. *Traffic* **17**, 908-922, doi:10.1111/tra.12404 (2016).
203 Momoi, T., Ben-Yoseph, Y. & Nadler, H. L. Substrate-specificities of acid and alkaline ceramidases in fibroblasts from patients with Farber disease and controls. *Biochem J* **205**, 419-425, doi:10.1042/bj2050419 (1982).
204 Bernardo, K. *et al.* Purification, characterization, and biosynthesis of human acid ceramidase. *J Biol Chem* **270**, 11098-11102, doi:10.1074/jbc.270.19.11098 (1995).
205 Tani, M., Okino, N., Mitsutake, S. & Ito, M. Specific and sensitive assay for alkaline and neutral ceramidases involving C12-NBD-ceramide. *J Biochem* **125**, 746-749, doi:10.1093/oxfordjournals.jbchem.a022345 (1999).
206 Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* **42**, e168, doi:10.1093/nar/gku936 (2014).
207 Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* **12**, 115-121, doi:10.1038/nmeth.3252 (2015).
208 Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3, doi:10.2202/1544-6115.1027 (2004).
209 Schwammle, V., Leon, I. R. & Jensen, O. N. Assessment and improvement of statistical tools for comparative proteomics analysis of sparse data sets with few experimental replicates. *J Proteome Res* **12**, 3874-3883, doi:10.1021/pr400045u (2013).
210 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).
211 Benes, V., Blake, J. & Doyle, K. Ribo-Zero Gold Kit: improved RNA-seq results after removal of cytoplasmic and mitochondrial ribosomal RNA. *Nat Methods* **8**, iii-iv, doi:10.1038/nmeth.f.352 (2011).
212 Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084-2092, doi:10.1093/bioinformatics/bty895 (2019).
213 Strimmer, K. A unified approach to false discovery rate estimation. *BMC Bioinformatics* **9**, 303, doi:10.1186/1471-2105-9-303 (2008).
214 Haq, E., Giri, S., Singh, I. & Singh, A. K. Molecular mechanism of psychosine-induced cell death in human oligodendrocyte cell line. *J Neurochem* **86**, 1428-1440, doi:10.1046/j.1471-4159.2003.01941.x (2003).
215 Shamin, M., Benedyk, T. H., Graham, S. C. & Deane, J. E. The lipid transfer protein Saposin B does not directly bind CD1d for lipid antigen loading. *Wellcome Open Res* **4**, 117, doi:10.12688/wellcomeopenres.15368.2 (2019).
216 Koch, M. *et al.* The crystal structure of human CD1d with and without alpha-galactosylceramide. *Nat Immunol* **6**, 819-826, doi:10.1038/ni1225 (2005).
217 Gumperz, J. E. *et al.* Murine CD1d-restricted T cell recognition of cellular lipids. *Immunity* **12**, 211-221 (2000).

218 Kim, H. S. *et al.* Biochemical characterization of CD1d expression in the absence of beta2-microglobulin. *J Biol Chem* **274**, 9289-9295, doi:10.1074/jbc.274.14.9289 (1999).
219 Hill, C. H., Read, R. J. & Deane, J. E. Structure of human saposin A at lysosomal pH. *Acta Crystallogr F Struct Biol Commun* **71**, 895-900, doi:10.1107/S2053230X15008584 (2015).
220 Elbein, A. D., Tropea, J. E., Mitchell, M. & Kaushal, G. P. Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. *J Biol Chem* **265**, 15599-15605 (1990).
221 Chang, V. T. *et al.* Glycoprotein structural genomics: solving the glycosylation problem. *Structure* **15**, 267-273, doi:10.1016/j.str.2007.01.011 (2007).
222 Brennan, P. J. *et al.* Structural determination of lipid antigens captured at the CD1d-T-cell receptor interface. *Proc Natl Acad Sci U S A* **114**, 8348-8353, doi:10.1073/pnas.1705882114 (2017).
223 Muenzner, J., Traub, L. M., Kelly, B. T. & Graham, S. C. Cellular and viral peptides bind multiple sites on the N-terminal domain of clathrin. *Traffic* **18**, 44-57, doi:10.1111/tra.12457 (2017).
224 Pollard, T. D. A guide to simple and informative binding assays. *Mol Biol Cell* **21**, 4061-4067, doi:10.1091/mbc.E10-08-0683 (2010).
225 Paduraru, C. *et al.* An N-linked glycan modulates the interaction between the CD1d heavy chain and beta 2-microglobulin. *J Biol Chem* **281**, 40369-40378, doi:10.1074/jbc.M608518200 (2006).
226 Robin, T., Bairoch, A., Muller, M., Lisacek, F. & Lane, L. Large-Scale Reanalysis of Publicly Available HeLa Cell Proteomics Data in the Context of the Human Proteome Project. *J Proteome Res* **17**, 4160-4170, doi:10.1021/acs.jproteome.8b00392 (2018).
227 Hussain, M. M., Bakillah, A. & Jamil, H. Apolipoprotein B binding to microsomal triglyceride transfer protein decreases with increases in length and lipidation: implications in lipoprotein biosynthesis. *Biochemistry* **36**, 13060-13067, doi:10.1021/bi971395a (1997).
228 Li, S. C. *et al.* Activator protein required for the enzymatic hydrolysis of cerebroside sulfate. Deficiency in urine of patients affected with cerebroside sulfatase activator deficiency and identity with activators for the enzymatic hydrolysis of GM1 ganglioside and globotriaosylceramide. *J Biol Chem* **260**, 1867-1871 (1985).
229 Sun, Y. *et al.* Neurological deficits and glycosphingolipid accumulation in saposin B deficient mice. *Hum Mol Genet* **17**, 2345-2356, doi:10.1093/hmg/ddn135 (2008).
230 Harzer, K. *et al.* Saposins (sap) A and C activate the degradation of galactosylceramide in living cells. *FEBS Lett* **417**, 270-274, doi:10.1016/s0014-5793(97)01302-1 (1997).
231 Harzer, K., Hiraiwa, M. & Paton, B. C. Saposins (sap) A and C activate the degradation of galactosylsphingosine. *FEBS Lett* **508**, 107-110, doi:10.1016/s0014-5793(01)03044-7 (2001).
232 Dong, G., Wearsch, P. A., Peaper, D. R., Cresswell, P. & Reinisch, K. M. Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer. *Immunity* **30**, 21-32, doi:10.1016/j.immuni.2008.10.018 (2009).
233 Blees, A. *et al.* Structure of the human MHC-I peptide-loading complex. *Nature* **551**, 525-528, doi:10.1038/nature24627 (2017).

234 Bosshart, H. & Heinzelmann, M. THP-1 cells as a model for human monocytes. *Ann Transl Med* **4**, 438, doi:10.21037/atm.2016.08.53 (2016).
235 Wenger, D. A., Rafi, M. A., Luzi, P., Datto, J. & Costantino-Ceccarini, E. Krabbe disease: genetic aspects and progress toward therapy. *Mol Genet Metab* **70**, 1-9, doi:10.1006/mgme.2000.2990 (2000).
236 Liu, H., Shi, B., Huang, C. C., Eksarko, P. & Pope, R. M. Transcriptional diversity during monocyte to macrophage differentiation. *Immunol Lett* **117**, 70-80, doi:10.1016/j.imlet.2007.12.012 (2008).
237 Smythies, L. E. *et al.* Inflammation anergy in human intestinal macrophages is due to Smad-induced IkappaBalpha expression and NF-kappaB inactivation. *J Biol Chem* **285**, 19593-19604, doi:10.1074/jbc.M109.069955 (2010).
238 Spratley, S. J. & Deane, J. E. New therapeutic approaches for Krabbe disease: The potential of pharmacological chaperones. *J Neurosci Res* **94**, 1203-1219, doi:10.1002/jnr.23762 (2016).
239 Liu, F. *et al.* TLR-Induced SMPD3 Defects Enhance Inflammatory Response of B Cell and Macrophage in the Pathogenesis of SLE. *Scand J Immunol* **86**, 377-388, doi:10.1111/sji.12611 (2017).
240 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).
241 Deane, J. E. *et al.* Insights into Krabbe disease from structures of galactocerebrosidase. *Proc Natl Acad Sci U S A* **108**, 15169-15173, doi:10.1073/pnas.1105639108 (2011).
242 Im, D. S., Heise, C. E., Nguyen, T., O'Dowd, B. F. & Lynch, K. R. Identification of a molecular target of psychosine and its role in globoid cell formation. *J Cell Biol* **153**, 429-434, doi:10.1083/jcb.153.2.429 (2001).
243 Ijichi, K. *et al.* MMP-3 mediates psychosine-induced globoid cell formation: implications for leukodystrophy pathology. *Glia* **61**, 765-777, doi:10.1002/glia.22471 (2013).
244 Aldo, P. B., Craveiro, V., Guller, S. & Mor, G. Effect of culture conditions on the phenotype of THP-1 monocyte cell line. *Am J Reprod Immunol* **70**, 80-86, doi:10.1111/aji.12129 (2013).
245 Nakamura, S. *et al.* Expression and responsiveness of human interleukin-18 receptor (IL-18R) on hematopoietic cell lines. *Leukemia* **14**, 1052-1059, doi:10.1038/sj.leu.2401789 (2000).
246 Benedito, R. *et al.* The notch ligands Dll4 and Jagged1 have opposing effects on angiogenesis. *Cell* **137**, 1124-1135, doi:10.1016/j.cell.2009.03.025 (2009).
247 Sato, H., Fujiwara, K., Sagara, J. & Bannai, S. Induction of cystine transport activity in mouse peritoneal macrophages by bacterial lipopolysaccharide. *Biochem J* **310 (Pt 2)**, 547-551, doi:10.1042/bj3100547 (1995).
248 Lu, H. K. *et al.* Leukocyte Ig-like receptor B4 (LILRB4) is a potent inhibitor of FcgammaRI-mediated monocyte activation via dephosphorylation of multiple kinases. *J Biol Chem* **284**, 34839-34848, doi:10.1074/jbc.M109.035683 (2009).
249 Rothlin, C. V., Ghosh, S., Zuniga, E. I., Oldstone, M. B. & Lemke, G. TAM receptors are pleiotropic inhibitors of the innate immune response. *Cell* **131**, 1124-1136, doi:10.1016/j.cell.2007.10.034 (2007).

250 Horiuchi, T., Mitoma, H., Harashima, S., Tsukamoto, H. & Shimoda, T. Transmembrane TNF-alpha: structure, function and interaction with anti-TNF agents. *Rheumatology (Oxford)* **49**, 1215-1228, doi:10.1093/rheumatology/keq031 (2010).
251 Khan, M., Haq, E., Giri, S., Singh, I. & Singh, A. K. Peroxisomal participation in psychosine-mediated toxicity: implications for Krabbe's disease. *J Neurosci Res* **80**, 845-854, doi:10.1002/jnr.20529 (2005).
252 Giri, S., Khan, M., Rattan, R., Singh, I. & Singh, A. K. Krabbe disease: psychosine-mediated activation of phospholipase A2 in oligodendrocyte cell death. *J Lipid Res* **47**, 1478-1492, doi:10.1194/jlr.M600084-JLR200 (2006).
253 Hawkins-Salsbury, J. A., Qin, E. Y., Reddy, A. S., Vogler, C. A. & Sands, M. S. Oxidative stress as a therapeutic target in globoid cell leukodystrophy. *Exp Neurol* **237**, 444-452, doi:10.1016/j.expneurol.2012.07.013 (2012).
254 Voccoli, V., Tonazzini, I., Signore, G., Caleo, M. & Cecchini, M. Role of extracellular calcium and mitochondrial oxygen species in psychosine-induced oligodendrocyte cell death. *Cell Death Dis* **5**, e1529, doi:10.1038/cddis.2014.483 (2014).
255 Conrad, M. & Sato, H. The oxidative stress-inducible cystine/glutamate antiporter, system x (c) (-) : cystine supplier and beyond. *Amino Acids* **42**, 231-246, doi:10.1007/s00726-011-0867-5 (2012).
256 Galaris, D., Barbouti, A. & Pantopoulos, K. Iron homeostasis and oxidative stress: An intimate relationship. *Biochim Biophys Acta Mol Cell Res* **1866**, 118535, doi:10.1016/j.bbamcr.2019.118535 (2019).
257 Cao, Z. *et al.* Methionine sulfoxide reductase B3 requires resolving cysteine residues for full activity and can act as a stereospecific methionine oxidase. *Biochem J* **475**, 827-838, doi:10.1042/BCJ20170929 (2018).
258 Ling, X. B. *et al.* Mammalian Metallothionein-2A and Oxidative Stress. *Int J Mol Sci* **17**, doi:10.3390/ijms17091483 (2016).
259 Zeng, C. M. *et al.* Aldo-Keto Reductase AKR1C1-AKR1C4: Functions, Regulation, and Intervention for Anti-cancer Therapy. *Front Pharmacol* **8**, 119, doi:10.3389/fphar.2017.00119 (2017).
260 Habib, E., Linher-Melville, K., Lin, H. X. & Singh, G. Expression of xCT and activity of system xc(-) are regulated by NRF2 in human breast cancer cells in response to oxidative stress. *Redox Biol* **5**, 33-42, doi:10.1016/j.redox.2015.03.003 (2015).
261 Harada, N. *et al.* Nrf2 regulates ferroportin 1-mediated iron efflux and counteracts lipopolysaccharide-induced ferroportin 1 mRNA suppression in macrophages. *Arch Biochem Biophys* **508**, 101-109, doi:10.1016/j.abb.2011.02.001 (2011).
262 Penning, T. M. Aldo-Keto Reductase Regulation by the Nrf2 System: Implications for Stress Response, Chemotherapy Drug Resistance, and Carcinogenesis. *Chem Res Toxicol* **30**, 162-176, doi:10.1021/acs.chemrestox.6b00319 (2017).
263 Silva-Islas, C. A. & Maldonado, P. D. Canonical and non-canonical mechanisms of Nrf2 activation. *Pharmacol Res* **134**, 92-99, doi:10.1016/j.phrs.2018.06.013 (2018).
264 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).
265 Gene Ontology, C. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **49**, D325-D334, doi:10.1093/nar/gkaa1113 (2021).
266 Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, RESEARCH0034, doi:10.1186/gb-2002-3-7-research0034 (2002).

267 Zheng, T. *et al.* Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence. *Sci Rep* **7**, 40638, doi:10.1038/srep40638 (2017).
268 Zhang, L. *et al.* Systematic in vitro profiling of off-target affinity, cleavage and efficiency for CRISPR enzymes. *Nucleic Acids Res* **48**, 5037-5053, doi:10.1093/nar/gkaa231 (2020).
269 Wang, J. & Wilkinson, M. F. Site-directed mutagenesis of large (13-kb) plasmids in a single-PCR procedure. *Biotechniques* **29**, 976-978, doi:10.2144/00295bm09 (2000).
270 Timms, R. T., Tchasovnikarova, I. A. & Lehner, P. J. Position-effect variegation revisited: HUSHing up heterochromatin in human cells. *Bioessays* **38**, 333-343, doi:10.1002/bies.201500184 (2016).
271 Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res* **49**, D545-D551, doi:10.1093/nar/gkaa970 (2021).
272 Skinner, B. M. & Johnson, E. E. Nuclear morphologies: their diversity and functional relevance. *Chromosoma* **126**, 195-212, doi:10.1007/s00412-016-0614-5 (2017).
273 Pereira, M. *et al.* Common signalling pathways in macrophage and osteoclast multinucleation. *J Cell Sci* **131**, doi:10.1242/jcs.216267 (2018).
274 Spano, A., Barni, S., Bertone, V. & Sciola, L. Changes on lysosomal compartment during PMA-induced differentiation of THP-1 monocytic cells: Influence of type I and type IV collagens. *Adv Biosci Biotechnol* **4**, doi:10.4236/abb.2013.48A3002 (2013).
275 Tominaga, T. *et al.* Establishment of an activated macrophage cell line, A-THP-1, and its properties. *Tohoku J Exp Med* **186**, 99-119, doi:10.1620/tjem.186.99 (1998).
276 Weinstock, N. I. *et al.* Macrophages Expressing GALC Improve Peripheral Krabbe Disease by a Mechanism Independent of Cross-Correction. *Neuron* **107**, 65-81 e69, doi:10.1016/j.neuron.2020.03.031 (2020).
277 Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822-826, doi:10.1038/nbt.2623 (2013).
278 Papapetrou, E. P. & Schambach, A. Gene Insertion Into Genomic Safe Harbors for Human Gene Therapy. *Mol Ther* **24**, 678-684, doi:10.1038/mt.2016.38 (2016).
279 Chen, S. J. Minimizing off-target effects in CRISPR-Cas9 genome editing. *Cell Biol Toxicol* **35**, 399-401, doi:10.1007/s10565-019-09486-4 (2019).
280 Kondo, Y., Adams, J. M., Vanier, M. T. & Duncan, I. D. Macrophages counteract demyelination in a mouse model of globoid cell leukodystrophy. *J Neurosci* **31**, 3610-3624, doi:10.1523/JNEUROSCI.6344-10.2011 (2011).
281 Leney, A. C. *et al.* Picodiscs for facile protein-glycolipid interaction analysis. *Anal Chem* **87**, 4402-4408, doi:10.1021/acs.analchem.5b00170 (2015).
282 Li, J. *et al.* Characterizing the Size and Composition of Saposin A Lipoprotein Picodiscs. *Anal Chem* **88**, 9524-9531, doi:10.1021/acs.analchem.6b02097 (2016).
283 Urzhumtseva, L., Afonine, P. V., Adams, P. D. & Urzhumtsev, A. Crystallographic model quality at a glance. *Acta Crystallogr D Biol Crystallogr* **65**, 297-300, doi:10.1107/S0907444908044296 (2009).
284 Li, Y. *et al.* Genetic ablation of acid ceramidase in Krabbe disease confirms the psychosine hypothesis and identifies a new therapeutic target. *Proc Natl Acad Sci U S A* **116**, 20097-20103, doi:10.1073/pnas.1912108116 (2019).
285 Hill, C. H., Graham, S. C., Read, R. J. & Deane, J. E. Structural snapshots illustrate the catalytic cycle of beta-galactocerebrosidase, the defective enzyme in Krabbe disease. *Proc Natl Acad Sci U S A* **110**, 20479-20484, doi:10.1073/pnas.1311990110 (2013).

286 Bedia, C., Camacho, L., Abad, J. L., Fabrias, G. & Levade, T. A simple fluorogenic method for determination of acid ceramidase activity and diagnosis of Farber disease. *J Lipid Res* **51**, 3542-3547, doi:10.1194/jlr.D010033 (2010).
287 van Echten-Deckert, G. Sphingolipid extraction and analysis by thin-layer chromatography. *Methods Enzymol* **312**, 64-79, doi:10.1016/s0076-6879(00)12900-3 (2000).
288 Won, J. S., Kim, J., Paintlia, M. K., Singh, I. & Singh, A. K. Role of endogenous psychosine accumulation in oligodendrocyte differentiation and survival: implication for Krabbe disease. *Brain Res* **1508**, 44-52, doi:10.1016/j.brainres.2013.02.024 (2013).
289 Craighead, M. W., Tiwari, P., Keynes, R. G. & Waters, C. M. Human oligodendroglial cell line, MO3.13, can be protected from apoptosis using the general caspase inhibitor zVAD-FMK. *J Neurosci Res* **57**, 236-243, doi:10.1002/(SICI)1097-4547(19990715)57:2<236::AID-JNR9>3.0.CO;2-D (1999).
290 Puthenveetil, R., Nguyen, K. & Vinogradova, O. Nanodiscs and Solution NMR: preparation, application and challenges. *Nanotechnol Rev* **6**, 111-126, doi:10.1515/ntrev-2016-0076 (2017).
291 Chen, A., Majdinasab, E. J., Fiori, M. C., Liang, H. & Altenberg, G. A. Polymer-Encased Nanodiscs and Polymer Nanodiscs: New Platforms for Membrane Protein Research and Applications. *Front Bioeng Biotechnol* **8**, 598450, doi:10.3389/fbioe.2020.598450 (2020).
292 Hall, S. C. L. *et al.* An acid-compatible co-polymer for the solubilization of membranes and proteins into lipid bilayer-containing nanoparticles. *Nanoscale* **10**, 10609-10619, doi:10.1039/c8nr01322e (2018).
293 Nowak, A., Beuschlein, F., Sivasubramaniam, V., Kasper, D. & Warnock, D. G. Lyso-Gb3 associates with adverse long-term outcome in patients with Fabry disease. *J Med Genet*, doi:10.1136/jmedgenet-2020-107338 (2021).
294 Choi, L. *et al.* The Fabry disease-associated lipid Lyso-Gb3 enhances voltage-gated calcium currents in sensory neurons and causes pain. *Neurosci Lett* **594**, 163-168, doi:10.1016/j.neulet.2015.01.084 (2015).
295 Sidransky, E. *et al.* Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N Engl J Med* **361**, 1651-1661, doi:10.1056/NEJMoa0901281 (2009).
296 Taguchi, Y. V. *et al.* Glucosylsphingosine Promotes alpha-Synuclein Pathology in Mutant GBA-Associated Parkinson's Disease. *J Neurosci* **37**, 9617-9631, doi:10.1523/jneurosci.1525-17.2017 (2017).
297 Beerepoot, S. *et al.* Peripheral neuropathy in metachromatic leukodystrophy: current status and future perspective. *Orphanet J Rare Dis* **14**, 240, doi:10.1186/s13023-019-1220-4 (2019).
298 Aricescu, A. R. *et al.* Structure of a tyrosine phosphatase adhesive interaction reveals a spacer-clamp mechanism. *Science* **317**, 1217-1220, doi:10.1126/science.1144646 (2007).
299 Craig, S. E. & Brady-Kalnay, S. M. Regulation of development and cancer by the R2B subfamily of RPTPs and the implications of proteolysis. *Semin Cell Dev Biol* **37**, 108-118, doi:10.1016/j.semcdb.2014.09.004 (2015).

300 Campan, M. *et al.* Increased proteolytic processing of protein tyrosine phosphatase mu in confluent vascular endothelial cells: the role of PC5, a member of the subtilisin family. *Biochemistry* **35**, 3797-3802, doi:10.1021/bi952552d (1996).
301 Gebbink, M. F. *et al.* Cell surface expression of receptor protein tyrosine phosphatase RPTP mu is regulated by cell-cell contact. *J Cell Biol* **131**, 251-260, doi:10.1083/jcb.131.1.251 (1995).
302 Remacle, A. G. *et al.* Substrate cleavage analysis of furin and related proprotein convertases. A comparative study. *J Biol Chem* **283**, 20897-20906, doi:10.1074/jbc.M803762200 (2008).
303 Garman, E. F. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr D Biol Crystallogr* **66**, 339-351, doi:10.1107/S0907444910008656 (2010).
304 Lutteke, T. Analysis and validation of carbohydrate three-dimensional structures. *Acta Crystallogr D Biol Crystallogr* **65**, 156-168, doi:10.1107/S0907444909001905 (2009).
305 Schneider, M., Al-Shareffi, E. & Haltiwanger, R. S. Biological functions of fucose in mammals. *Glycobiology* **27**, 601-618, doi:10.1093/glycob/cwx034 (2017).
306 Koide, A., Bailey, C. W., Huang, X. & Koide, S. The fibronectin type III domain as a scaffold for novel binding proteins. *J Mol Biol* **284**, 1141-1151, doi:10.1006/jmbi.1998.2238 (1998).
307 Bucher, R. M., Svergun, D. I., Muhle-Goll, C. & Mayans, O. The structure of the FnIII Tandem A77-A78 points to a periodically conserved architecture in the myosin-binding region of titin. *J Mol Biol* **401**, 843-853, doi:10.1016/j.jmb.2010.06.011 (2010).
308 Seals, D. F. & Courtneidge, S. A. The ADAMs family of metalloproteases: multidomain proteins with multiple functions. *Genes Dev* **17**, 7-30, doi:10.1101/gad.1039703 (2003).
309 Boncompain, G. *et al.* Synchronization of secretory protein traffic in populations of cells. *Nat Methods* **9**, 493-498, doi:10.1038/nmeth.1928 (2012).
310 Chen, Y., Gershlick, D. C., Park, S. Y. & Bonifacino, J. S. Segregation in the Golgi complex precedes export of endolysosomal proteins in distinct transport carriers. *J Cell Biol* **216**, 4141-4151, doi:10.1083/jcb.201707172 (2017).
311 Del Vecchio, R. L. & Tonks, N. K. The conserved immunoglobulin domain controls the subcellular localization of the homophilic adhesion receptor protein-tyrosine phosphatase mu. *J Biol Chem* **280**, 1603-1612, doi:10.1074/jbc.M410181200 (2005).
312 Rudenko, G. Dynamic Control of Synaptic Adhesion and Organizing Molecules in Synaptic Plasticity. *Neural Plast* **2017**, 6526151, doi:10.1155/2017/6526151 (2017).
313 Chavent, M., Seiradake, E., Jones, E. Y. & Sansom, M. S. Structures of the EphA2 Receptor at the Membrane: Role of Lipid Interactions. *Structure* **24**, 337-347, doi:10.1016/j.str.2015.11.008 (2016).
314 Tang, H. *et al.* Architecture of cell-cell adhesion mediated by sidekicks. *Proc Natl Acad Sci U S A* **115**, 9246-9251, doi:10.1073/pnas.1801810115 (2018).
315 Seiradake, E., Harlos, K., Sutton, G., Aricescu, A. R. & Jones, E. Y. An extracellular steric seeding mechanism for Eph-ephrin signaling platform assembly. *Nat Struct Mol Biol* **17**, 398-402, doi:10.1038/nsmb.1782 (2010).
316 Anders, L. *et al.* Furin-, ADAM 10-, and gamma-secretase-mediated cleavage of a receptor tyrosine phosphatase and regulation of beta-catenin's transcriptional

activity. *Mol Cell Biol* **26**, 3917-3934, doi:10.1128/MCB.26.10.3917-3934.2006 (2006).
317 Robak, L. A. *et al.* Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. *Brain* **140**, 3191-3203, doi:10.1093/brain/awx285 (2017).
318 Gorelik, A., Illes, K., Hasan, S. M. N., Nagar, B. & Mazhab-Jafari, M. T. Structure of the murine lysosomal multienzyme complex core. *Sci Adv* **7**, doi:10.1126/sciadv.abf4155 (2021).
319 Pereira, C. S. *et al.* Lipid Antigen Presentation by CD1b and CD1d in Lysosomal Storage Disease Patients. *Front Immunol* **10**, 1264, doi:10.3389/fimmu.2019.01264 (2019).

Appendix I

Publication arising from this work.

Wellcome Open Research

Wellcome Open Research 2019, 4:117 Last updated: 12 FEB 2020

Check for updates

RESEARCH ARTICLE

REVISED The lipid transfer protein Saposin B does not directly bind CD1d for lipid antigen loading [version 2; peer review: 3 approved]

Maria Shamin [1,2], Tomasz H. Benedyk [2], Stephen C. Graham [2], Janet E. Deane [1,3]

[1]Cambridge Institute for Medical Research, University of Cambridge, Cambridge, CB2 0XY, UK
[2]Department of Pathology, University of Cambridge, Cambridge, CB2 1QP, UK
[3]Department of Clinical Neurosciences, University of Cambridge, Cambridge, CB2 0QQ, UK

v2 **First published:** 02 Aug 2019, **4**:117 (https://doi.org/10.12688/wellcomeopenres.15368.1)
Latest published: 18 Oct 2019, **4**:117 (https://doi.org/10.12688/wellcomeopenres.15368.2)

Abstract

Background: Lipid antigens are presented on the surface of cells by the CD1 family of glycoproteins, which have structural and functional similarity to MHC class I molecules. The hydrophobic lipid antigens are embedded in membranes and inaccessible to the lumenal lipid-binding domain of CD1 molecules. Therefore, CD1 molecules require lipid transfer proteins for lipid loading and editing. CD1d is loaded with lipids in late endocytic compartments, and lipid transfer proteins of the saposin family have been shown to play a crucial role in this process. However, the mechanism by which saposins facilitate lipid binding to CD1 molecules is not known and is thought to involve transient interactions between protein components to ensure CD1-lipid complexes can be efficiently trafficked to the plasma membrane for antigen presentation. Of the four saposin proteins, the importance of Saposin B (SapB) for loading of CD1d is the most well-characterised. However, a direct interaction between CD1d and SapB has yet to be described.
Methods: In order to determine how SapB might load lipids onto CD1d, we used purified, recombinant CD1d and SapB and carried out a series of highly sensitive binding assays to monitor direct interactions. We performed equilibrium binding analysis, chemical cross-linking and co-crystallisation experiments, under a range of different conditions.
Results: We could not demonstrate a direct interaction between SapB and CD1d using any of these binding assays.
Conclusions: This work strongly indicates that the role of SapB in lipid loading does not involve direct binding to CD1d. We discuss the implication of this for our understanding of lipid loading of CD1d and propose several factors that may influence this process.

Keywords

CD1, prosaposin, lipid transfer protein, lipid antigens, lipid loading, lipid exchange, lipid binding proteins, sphingolipid storage disorder

Open Peer Review

Reviewer Status ✓ ✓ ✓

	Invited Reviewers		
	1	**2**	**3**
version 2 (revision) 18 Oct 2019	✓ report	✓ report	✓ report
version 1 02 Aug 2019	? report	? report	? report

1 **Weiming Yuan**, Keck School of Medicine, University of Southern California, Los Angeles, USA

2 **Tim P. Levine**, University College London, London, UK

3 **Mariolina Salio**, University of Oxford, Oxford, UK

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Janet E. Deane (jed55@cam.ac.uk)

Author roles: Shamin M: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Benedyk TH**: Methodology, Writing – Review & Editing; **Graham SC**: Funding Acquisition, Investigation, Methodology, Writing – Review & Editing; **Deane JE**: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: M.S. is supported by a Wellcome Trust PhD studentship [203984]. S.C.G. is supported by a Sir Henry Dale Fellowship co-funded by the Royal Society and Wellcome Trust [098406]. J.E.D. is supported by a Royal Society University Research Fellowship [UF100371].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Shamin M, Benedyk TH, Graham SC and Deane JE. **The lipid transfer protein Saposin B does not directly bind CD1d for lipid antigen loading [version 2; peer review: 3 approved]** Wellcome Open Research 2019, **4**:117 (https://doi.org/10.12688/wellcomeopenres.15368.2)

First published: 02 Aug 2019, **4**:117 (https://doi.org/10.12688/wellcomeopenres.15368.1)

REVISED **Amendments from Version 1**

We thank the reviewers for their insightful comments. We have revised the manuscript following their helpful suggestions as follows:

We acknowledge that the lack of a detectable interaction does not necessarily mean that there is no interaction. We have edited the abstract and discussed why an interaction too weak to be detected by the assays used in this study is unlikely to be physiologically relevant. This is supported by the fact that strong direct protein-protein interactions are crucial for several different lipid transfer pathways as detailed in the Discussion.

We have performed an additional equilibrium binding experiment where SapB was pre-loaded with the lipid α-GalCer while CD1d remained loaded with the endogenous human lipids with which it co-purifies. In agreement with our earlier results, no interaction was detected in this new experiment. This additional data is now included as part of Figure 2.

We have clarified the nature and relevance of the lipids expected to be associated with CD1d and SapB in the experiments reported in this manuscript and clarified that glycosylation of SapB is not required for lipid loading of CD1d.

Any further responses from the reviewers can be found at the end of the article

Introduction

The presentation of antigenic molecules is fundamental to host defence and immune regulation. The MHC class I and II molecules present peptide antigens to T cells and the molecular mechanisms of peptide loading and editing are relatively well understood[1–3]. However, the mechanism by which lipid antigens are loaded onto the CD1 family of antigen-presenting molecules remains poorly understood. CD1 molecules have structural similarity to MHC class I, including association with β-2-microglobulin (β_2m), but unlike MHC molecules, CD1 molecules are non-polymorphic and the antigen-binding grooves of CD1 molecules are extremely hydrophobic, allowing the binding of most classes of lipids including phospholipids, glycerolipids, lysolipids and glycolipids possessing a huge variety of glycosylated head groups[4].

Of the five CD1 family members, antigen presentation by CD1d is best characterised. CD1d undergoes complex cellular trafficking in a manner similar to that of MHC class II. Following synthesis in the endoplasmic reticulum, CD1d traffics via the Golgi apparatus to the plasma membrane, and is then internalised and trafficked through early and late endocytic compartments before being returned to the plasma membrane[5], where CD1d-lipid complexes are recognized by the T cell receptor (TCR) of natural killer T (NKT) cells. Invariant NKT (iNKT) cells, which bear an invariant TCR and are characteristically potently activated by recognition of CD1d loaded with the synthetic glycolipid α-galactosylceramide (α-GalCer), are innate-like T cells involved in the orchestration of immune responses[6].

Lipid antigens are embedded in cellular membranes[7] and thus are not directly accessible to the lumenal, lipid-binding groove of CD1 molecules. The loading of lipid antigen onto CD1 molecules therefore requires the action of lipid transfer proteins (LTPs) for efficient antigen presentation. The loading of lipids onto CD1d can occur in the ER, at the cell surface and in endocytic compartments. Relatively little is known about how lipid specificity is determined, and studies using secreted or recycling, surface-cleavable human CD1d found differences in the lipid repertoire bound to CD1d, reflecting the lipids present in the compartments through which it had trafficked[8–10]. Different LTPs are present in the different cellular locations where lipid antigen loading occurs and are likely to determine the lipid repertoire bound to CD1 molecules[11]. Interestingly, the late endocytic/lysosomal compartment is the site of both glycolipid catabolism and lipid loading of CD1d and the resident LTPs function in both these processes. Importantly, the antigenicity of glycolipids can be regulated by processing of their glycan headgoups prior to presentation to T cells, highlighting the complex interplay between these two pathways[12,13].

The LTPs present in the lysosomal compartments include ganglioside monosialic 2 activator (GM2A), Niemann-Pick type C 2 (NPC2) and four saposin (Sap) proteins. SapA, SapB, SapC and SapD are the products of the proteolytic cleavage of a prosaposin (PSAP) precursor and are each small, non-enzymatic proteins possessing three disulphide bonds that stabilise an extremely heat-resistant helical structure. The saposins can exist in a "closed" monomeric, globular conformation or, at low pH, adopt a more "open" conformation, forming higher-order oligomers enclosing a hydrophobic cavity into which lipid acyl chains can be buried[14–19]. Each saposin functions in conjunction with specific hydrolases to facilitate the degradation of different glycosphingolipids and the loss of saposin function phenotypically resembles the loss of the associated hydrolases[20–25]. There are two proposed mechanisms for how saposins assist in lipid presentation to hydrolases: the "solubiliser" model, whereby saposins encapsulate lipids within a hydrophobic oligomeric complex for presentation to soluble enzymes, and the "liftase" model, where saposins bind directly to membranes destabilising them, allowing membrane-associated hydrolases access to lipid substrates. Structural and cell-based evidence exists to support both these models[19,26–29]. These two models are also relevant for the proposed mechanisms by which saposins load lipids onto CD1 molecules[30]. SapB is thought to behave as a "solubiliser", forming protein-lipid complexes that enhance lipid loading of CD1 molecules, while SapC associates with membranes and may facilitate lipid loading at the membrane surface.

The evidence for saposin-mediated lipid loading of CD1 molecules comes from animal knockout studies, cell-based assays and *in vitro* assays. Knockout of PSAP in mice results in the dramatic loss of presentation of endogenous iNKT cell ligands by CD1d[31]. The presentation of exogenous lipid antigens, such as α-GalCer, is impaired in PSAP knockout or knockdown cells[31–34]. *In vitro* assays have identified that all the saposins, as well as the other lysosomal LTPs, GM2A and NPC2, are capable of transferring lipids onto CD1 molecules[31,35]. However, some LTPs appear to be more efficient than others at loading specific lipid antigens onto different CD1 molecules[30,33,36]. Specificity within the saposin family in this process has been explored via *in vitro* assays identifying SapB as the dominant saposin for the loading of lipid antigens onto CD1d. Specifically, incubation of CD1d with SapB and α-GalCer has been shown to enhance the stimulation of iNKT cells[33,34] and SapB

can mediate lipid binding to CD1d in T cell independent assays[33] suggesting a direct role for this specific saposin in lipid loading of CD1d. The ability of SapB to facilitate and enhance lipid exchange on CD1d has also been monitored *in vitro* via iNKT TCR binding to CD1d following incubation with lipids alone or lipids plus SapB[34]. These *in vitro* assays implicate SapB as a "lipid editor" that facilitates the loading and unloading of lipid antigens onto CD1d.

Due to the hydrophobic property of lipids, their exposure to aqueous solutions is thermodynamically unfavourable. Therefore, lipid transfer between hydrophobic lipid-binding cavities of proteins requires direct interaction of the protein components to provide continuous shielding of the lipid from the aqueous phase[37]. This hypothesis is supported by the fact that other known lipid transfer proteins directly interact with the protein receiving the lipid[38–40]. However, several publications report a failure to detect a direct interaction between saposins and human CD1d[11,33,34]. Unfortunately, these reports do not show any data or detail the approaches used to monitor binding and the implication is that these interactions exist but are weak and very transitory. Therefore we have used a number of robust and sensitive biochemical and structural techniques to monitor direct interactions between SapB and CD1d. Using multiple approaches in a range of conditions we do not observe any evidence of a direct interaction between CD1d and SapB.

Methods

Cloning and cell line production

Codon-optimised cDNA was synthesized (GeneArt) encoding human β_2m (residues 21-119, Uniprot P61769) and the extracellular domain of human CD1d (residues 20-301, Uniprot P15813) with a C-terminal hexahistiding tag (CD1d-H_6). A modified version of the piggyBac target protein plasmid PB-T-PAF[41] was constructed retaining the N-terminal secretion signal but with the Protein A fusion removed. CD1d-H_6 and β_2m were individually cloned into this modified vector, using *Spe*I and *Asc*I restriction-endonuclease sites, to produce PB-T-CD1d-H_6 and untagged PB-T-β_2m. For production of an inducible, stable co-expression cell line, HEK293F cells were quadruple-transfected with PB-T-CD1d-H_6, PB-T-β_2m, PB-RN and PBase using a DNA mass ratio of 5:5:1:1 and transfected cells were selected with geneticin using the established protocol for the piggyBac expression system[41].

Codon-optimised cDNA was synthesized (GeneArt) encoding the segment of human PSAP corresponding to SapB (residues 195-274, Uniprot P07602) and subcloned into the bacterial expression vector pET-15b using *Nco*I and *Xho*I restriction endonuclease sites to produce untagged protein. SapB was also subcloned into the mammalian expression vector pHLsec[42] using the restriction enzymes *Age*I and *Kpn*I to encode a C-terminally H_6-tagged protein.

Protein expression and purification

Protein expression of CD1d-β_2m complex in the HEK293F co-expression cell line was induced with 2 µg/mL doxycycline and expression media was collected over a 2 month period. CD1d-H_6 was purified from conditioned media by nickel affinity chromatography in phosphate-buffered saline (PBS) pH 7.4 and eluted using PBS containing 300 mM imidazole. CD1d-β_2m was further purified by size exclusion chromatography on a Superdex 200 column (GE Healthcare) equilibrated in cross-linking buffer (20 mM HEPES pH 7.0, 150 mM NaCl) for interaction assays or distinct buffers (described below) for crystallisation experiments. CD1d-β_2m was concentrated to 2 mg/mL and stored at 4 °C for up to one month.

To produce deglycosylated CD1d-β_2m, cells were treated with 5 µM kifunensine at the same time as protein expression was induced. CD1d-β_2m was purified by nickel affinity purification and eluted with 100 mM citrate pH 4. CD1d-β_2m protein was buffer exchanged into 100 mM citrate pH 5.5, using repeated rounds of dilution and concentration using a centrifugal concentrator, concentrated to 2 mg/mL and incubated with 3125 units of Endoglycosidase H (Endo H; New England Biolabs) per milligram of protein for 3 hours at room temperature. Deglycosylated CD1d-β_2m was purified by size exclusion chromatography (as described above).

Untagged SapB was expressed in *Escherichia coli* Origami (DE3) cells and purified as described previously[43]. Briefly, cleared lysate was heat-treated, precipitated proteins were cleared by centrifugation and supernatant containing SapB was dialysed overnight in the presence of 20 µg/mL DNAse against anion exchange buffer (50 mM Tris pH 7.4, 25 mM NaCl). SapB was further purified by anion exchange chromatography (HiTrap QSepharose column) followed by size-exclusion chromatography (HiLoad 16/600 Superdex 75 column) in 50 mM Tris pH 7.4, 150 mM NaCl. Purified SapB was concentrated to 16 mg/mL and stored at 4 °C.

His_6-tagged SapB was expressed in HEK293F cells by polyethylenimine-based transient transfection. After 4 days, SapB-H_6 was purified from conditioned media by nickel affinity and size exclusion chromatography with a Superdex 75 column (GE Healthcare) in cross-linking buffer. Prior to crystallisation trials (but not cross-linking assays), the tag was removed by incubation with 1.2 U/mL Carboxypeptidase A-agarose at 25°C overnight followed by incubation with Ni-NTA resin to remove tagged protein. The supernatant containing untagged SapB was separated from Carboxypeptidase A-agarose and Ni-NTA agarose by passing through a gravity flow column.

α-GalCer loading

Endogenously purified lipids were exchanged for the lipid α-GalCer (Avanti Lipids) based on a protocol used to obtain crystal structures of α-GalCer-loaded CD1d[44,45]. Lipid was resuspended in DMSO at 1 mg/mL and dissolved by heating the solution at 80°C for 10 min, with brief sonication and vortexing every 3 minutes. CD1d-β_2m and SapB were incubated with a 3-fold molar excess of α-GalCer in PBS overnight at room temperature. α-GalCer-loaded proteins were then separated from free α-GalCer by size exclusion chromatography with a Superdex 200 10/300 column equilibrated in 20 mM HEPES pH 7.0, 150 mM NaCl. α-GalCer-loaded proteins were used within 24 hours after size exclusion chromatography to avoid significant hydrolysis of the lipid.

Equilibrium binding assay

Ni-NTA agarose beads were loaded with a saturating amount of CD1d-β_2m, and complete capture of CD1d-β_2m was verified by measuring the absorbance of the supernatant at 280 nm. Saturated beads were washed three times in equilibrium assay buffer (50 mM HEPES pH 7.0, 150 mM NaCl, 0.05% digitonin, 2.5 mM $CaCl_2$, 2.5 mM $MgCl_2$). Serial dilutions (1:1) of CD1d-β_2m-saturated beads were made in equilibrium assay buffer supplemented with unloaded Ni-NTA beads to equalize the bead volume across all samples. SapB was added at a final concentration of 680 nM in equilibrium assay buffer. The reaction was incubated with shaking at room temperature for 1 hour. The supernatant containing unbound SapB was separated from the beads by centrifugation (800*g*, 2 min) followed by separation with a Micro-Spin column (Pierce) (800*g*, 2 min) to remove all beads. The supernatant was analysed by SDS-PAGE followed by staining with SYPRO Ruby. For analysis of bead samples, these were boiled in SDS-PAGE loading dye, separated by SDS-PAGE and detected by Coomassie staining.

Cross-linking assay

Amine-reactive cross-linking agents disuccinimidyl sulfoxide (DSSO) and PEGylated bis(sulfosuccinimidyl)suberate ($BS(PEG)_5$) were dissolved in DMSO and added to 10 µM CD1d-β_2m alone, 20 µM SapB alone or 10 µM CD1d + 20 µM SapB with final cross-linker concentrations of 0, 200 µM or 1 mM in cross-linking buffer. The reaction was incubated for 30 min at room temperature and terminated by incubation with 20 mM Tris pH 8 for 15 min. The reaction products were analysed by SDS-PAGE followed by Coomassie staining.

Crystallography

Crystallization experiments were performed in 96-well nanolitre-scale sitting drops (200 nL protein plus 200 nL of precipitant) equilibrated at 20°C against 80 µL reservoirs of precipitant. The crystals reported here were grown in drops containing the following components:

Crystal form (i) (Figure 4B) contained glycosylated CD1d-H_6-β_2m (13.5 mg/mL) + 2-fold molar excess of unglycosylated SapB (4.7 mg/mL) in 150 mM NaCl, 50 mM Tris pH 7.4 and grew against a reservoir containing 20% w/v polyethylene glycol (PEG) 6000, 0.1 M HEPES pH 7 and 0.2 M calcium chloride.

Crystal form (ii) (Figure 4B) contained glycosylated CD1d-H_6-β_2m (12.4 mg/mL) + 2-fold molar excess of glycosylated SapB (after tag removal) (4.3 mg/mL) in 20 mM HEPES pH 7.0, 150 mM NaCl and grew against a reservoir containing 20 mM 1,6-hexanediol, 20 mM 1-butanol, 20 mM 1,2-propanediol, 20 mM 2-propanol, 20 mM 1,4-butanediol, 20 mM 1,3-propanediol, 0.1 M bicine/Trizma base pH 8.5, 12.5% w/v PEG 1000, 12.5% w/v PEG 3350 and 12.5% v/v 2-methyl-2,4-pentanediol.

Crystal form (iii) (Figure 4B) contained α-GalCer-loaded glycosylated CD1d-H_6-β_2m (11.9 mg/mL) + 2-fold molar excess of α-GalCer-loaded unglycosylated SapB (4.2 mg/mL) in 50 mM Tris pH 7.4, 150 mM NaCl and grew against a reservoir containing 0.1 M sodium acetate pH 5.0, 20% w/v PEG 6000 and 0.2 M ammonium chloride.

Crystal form (iv) (Figure 4B) contained deglycosylated CD1d-H_6-β_2m (13.3 mg/mL) + 2-fold molar excess of unglycosylated SapB (4.7 mg/mL) in 20 mM citrate pH 6, 150 mM NaCl, and grew against a reservoir containing 16% v/v PEG 500 monomethylether, 8% w/v PEG 20000, 0.1 M bicine/Trizma base pH 8.5, 30 mM sodium nitrate, 30 mM sodium phosphate dibasic and 30 mM ammonium sulfate.

Crystal form (v) (Figure 4D) contained deglycosylated CD1d-H_6-β_2m (6.5 mg/mL) + 2-fold molar excess of unglycosylated SapB (2.3 mg/mL) in 100 mM citrate pH 6.0, 150 mM NaCl and grew against a reservoir containing 20% w/v PEG 3350, 0.1 M bis-Tris propane pH 7.5 and 0.2 M di-sodium malonate.

X-ray data collection and molecular replacement

Crystals were cryoprotected in reservoir solution supplemented with 20% v/v glycerol and flash-cooled by plunging into liquid nitrogen. Diffraction data were recorded at Diamond Light Source (DLS) on beamlines I03, I04 and I04-1. Diffraction data were indexed and integrated using the automated data processing pipelines at DLS[46] implementing XIA2 DIALS[47], XIA2 3dii or autoPROC and STARANISO[48] depending on the extent of anisotropic diffraction, then scaled and merged using AIMLESS[49]. Version numbers for crystallography data processing software are detailed in Table 1. The model for molecular replacement was generated using the published structure of CD1d-β_2m bound to lysophosphatidylcholine[50] (PDB ID 3U0P, chains A and B), using phenix.sculptor (Phenix v1.14-3260)[51,52] to remove the lipid ligand from the model and revert the glycosylation site mutations present in this structure. Molecular replacement was performed using phenix.phaser[53] and was

Table 1. Crystallography data processing software.

	Crystal (iii)	Crystal (iv)	Crystal (v)
Processing software	autoPROC (v1.0.5) and STARANISO (v1.10.15)	XIA2 (v0.5.771) 3dii (version January 26, 2018)	XIA2 (v0.5.764) DIALS (v1.12.2)
Scaling software	AIMLESS (v0.7.2)	AIMLESS (v0.7.3)	AIMLESS (v0.7.3)

followed by rigid body refinement using phenix.refine[54]. Molecular replacement into the highly anisotropic data (Crystal form (iii)) was carried out using both the isotropic and anisotropically processed data, resulting in the same solution. Maps were inspected using COOT (v0.8.9)[55]. Structure figures were rendered using PyMOL (v2.1.0, Schrödinger LLC).

SDS-PAGE analysis of crystals

For each crystallisation condition tested, several crystals were harvested from the crystallisation drop and washed by transferring to a drop of reservoir solution to remove uncrystallised protein components. Crystals were then transferred to a second drop of reservoir solution and dissolved in SDS-PAGE loading buffer. Alternatively, large crystals from which data had been collected at the synchrotron were recovered into SDS-PAGE loading buffer from the crystal-mounting loop. Individual crystal components were analysed by SDS-PAGE followed by Coomassie staining.

Results

Purification of CD1d-β_2m complex from a human cell line

Several strategies have been developed previously for the purification of CD1d including refolding from *E. coli* inclusion bodies[56], baculovirus-insect cell expression[50] and mammalian expression systems[57]. An important consideration for CD1d protein expression and purification is the inclusion of the essential partner protein β_2m. In a manner similar to the heavy chain of MHC class I, CD1d non-covalently associates with β_2m and this association is required for maturation of its four glycan chains[58]. In order to obtain high quality, near-native CD1d protein for *in vitro* studies, we chose to co-express human CD1d with β_2m in the mammalian suspension cell line HEK293F as this would confer correct folding, native post-translational modification and full glycosylation of CD1d. Initial trials using polyethylenimine-based transient co-transfection of CD1d and β_2m resulted in very low yields. Thus we used a piggyBac-expression system[41] to establish stable and inducible co-expression of full-length, untagged β_2m with the ectodomain of CD1d possessing a C-terminal hexahistidine tag (H_6) (Figure 1A). Both proteins were expressed with an N-terminal secretion signal sequence targeting synthesis to the endoplasmic reticulum, resulting in protein secretion into the media. This expression strategy results in CD1d molecules that are bound to endogenous human lipids acquired during folding in the endoplasmic reticulum and trafficking through the Golgi apparatus. CD1d-H_6 was purified from conditioned media by nickel affinity chromatography followed by size exclusion chromatography (Figure 1B and C). Untagged β_2m co-purified with CD1d, indicating that both proteins are correctly folded and form a stable complex. Although some CD1d alone was purified, we were able to separate this from CD1d-β_2m following size exclusion (Figure 1C). Using this expression system, we obtained a high yield (1 mg/100 mL media) of very pure CD1d-β_2m protein. The yield was over 10 times higher than that obtained with transient co-transfection in the same cell line.

Purification of SapB

The four saposin proteins (SapA-D) are produced from the proteolytic cleavage of the precursor protein PSAP which normally occurs upon delivery to the lysosome. To produce recombinant SapB alone, the segment of human PSAP corresponding to SapB (Figure 1D) was expressed in *E. coli*. Untagged SapB was purified following a well-established protocol[43] that exploits the heat stability of saposins followed by anion exchange and size exclusion chromatography (Figure 1E and F). Previous structural studies with *E. coli*-expressed SapB identified that it exists as a tight dimer that co-purifies with *E. coli* phosphatidylethanolamine[14].

CD1d binding to SapB is not detectable using *in vitro* interaction assays

Classical pull-down experiments involve the capture and immobilisation onto resin of a tagged bait protein followed by incubation with prey protein. Subsequent wash steps remove unbound protein leaving behind protein complexes for detection. However, as the concentration of free prey protein is reduced at each wash, the equilibrium is pushed towards complex dissociation, meaning this method is not well suited for the capture of weak or transient interactions. Using this approach, no interaction was detectable between immobilised CD1d-β_2m and SapB. Therefore, an alternative approach was adopted to overcome the limitations of prey-depletion by using an *in vitro* equilibrium binding assay (Figure 2A). In this assay, the bait is bound to beads at a range of concentrations and the prey is added at a low, limiting concentration. Following incubation, and no wash steps, the supernatant is separated from the beads and unbound prey is detected using the sensitive protein dye SYPRO Ruby. Monitoring the depletion of prey from the supernatant by the large excess of bait protein, rather than monitoring complex formation directly, represents a more sensitive measure of binding and thus is optimised for weak or transient interactions[59].

Equilibrium binding assays were carried out using CD1d-β_2m as bait at a concentration range of 0-40 μM and with SapB as the prey at 680 nM. Previous *in vitro* studies have identified that SapB can load α-GalCer onto CD1d at a pH range between 5 and 8[33]. Although this exchange appears to be most efficient at pH 6, this pH causes partial dissociation of the CD1d His-tag from the Ni-NTA resin. Therefore, the equilibrium binding assay was performed at pH 7. Previous work, and our preliminary experiments, showed that β_2m dissociates from CD1d in the presence of detergent[60]. However, the gentle detergent digitonin was found to cause minimal disruption of the CD1d-β_2m complex[60] and was used in this assay. Despite the use of this optimised approach, no binding of SapB to CD1d could be detected (Figure 2B). In these conditions, at high CD1d-β_2m concentrations, a small amount of β_2m is detectable in the supernatant due to dissociation from CD1d (Figure 2B). By comparing how little appears to be lost based on the Coomassie-stained bead samples this highlights the sensitivity of this technique and of the SYPRO Ruby stain for protein detection.

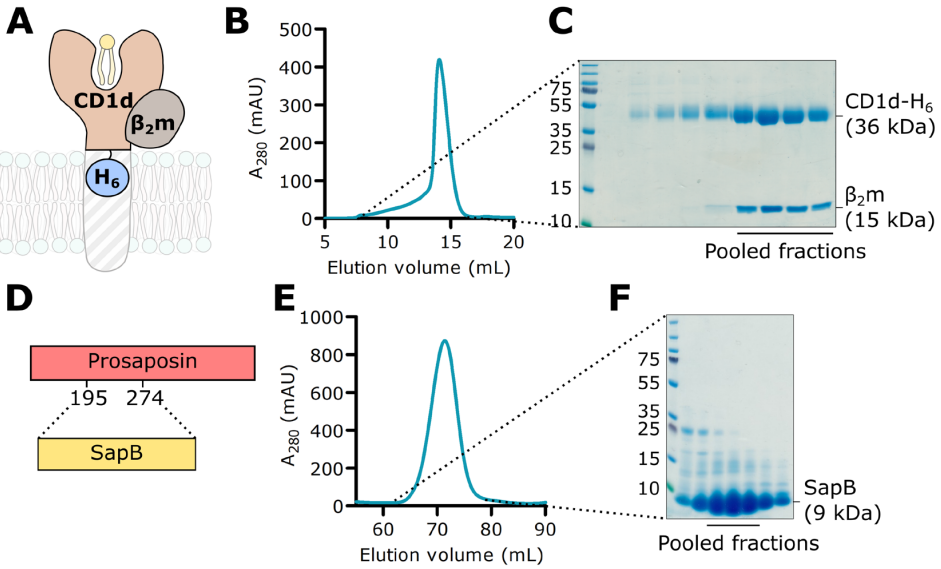

Figure 1. Purification of CD1d-β_2m and SapB. A. A His-tagged construct encoding the extracellular domain of CD1d (CD1d-H_6) was co-expressed with untagged full-length β_2m in mammalian cells. **B**. Size exclusion chromatography (SEC) showing co-purification of CD1d-H_6 with β_2m. **C**. SDS-PAGE of SEC fractions. The indicated fractions containing β_2m-associated CD1d were used in subsequent experiments. **D**. Untagged SapB (residues 195-274 of the precursor protein PSAP) was expressed in *E. coli*. **E**. SEC of SapB after multi-step purification. **F**. SDS-PAGE of SEC fractions. The indicated fractions were used in subsequent experiments.

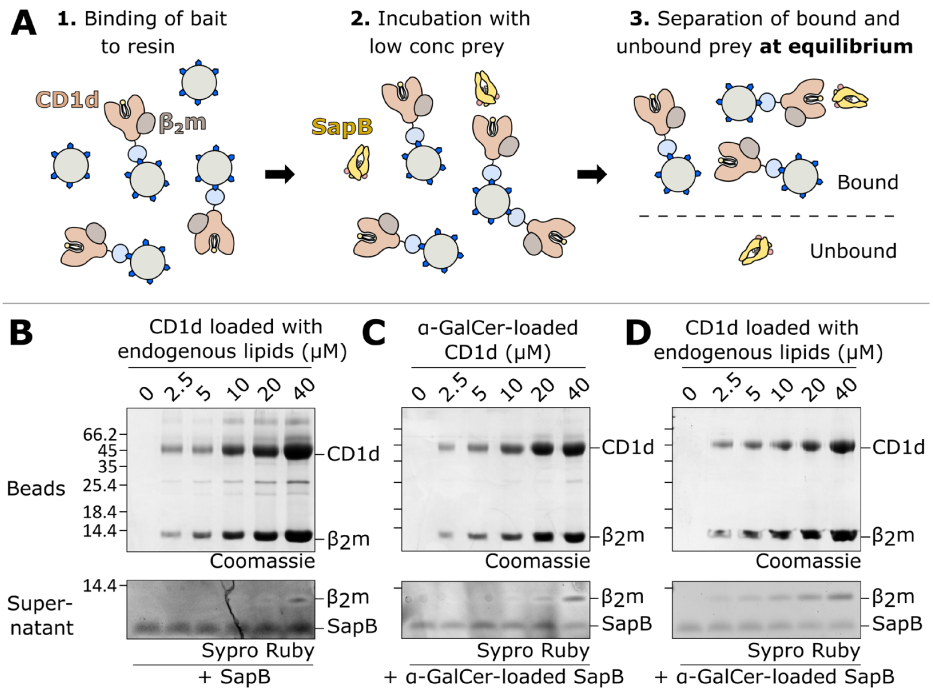

Figure 2. Equilibrium binding assay of CD1d-β_2m and SapB. A. Schematic diagram illustrating the equilibrium binding assay protocol. **B**. Increasing concentrations of CD1d-β_2m were bound to beads and detected by Coomassie-stained SDS-PAGE. After incubation with SapB, unbound SapB was detected in supernatants by SYPRO Ruby-stained SDS-PAGE. The lipids bound to CD1d-β_2m and SapB in this experiment are the endogenous lipids they co-purify with following expression in human cells and *E. coli*, respectively. **C**. As for panel **B** but with CD1d-β_2m and SapB both pre-loaded with the lipid α-GalCer. **D**. As for panel **B** but with only SapB pre-loaded with the lipid α-GalCer.

Previous studies have shown that the ectodomain of CD1d co-purifies with a broad range of cellular lipids following synthesis and trafficking through the secretory pathway, but is likely to be primarily loaded with sphingomyelin[8–10]. SapB co-purifies with *E. coli* phosphatidylethanolamine[14], a lipid species that has also been shown to bind CD1d[61]. Although it seems likely that this lipid composition for SapB and CD1d should be compatible with protein-protein interactions required for lipid exchange, lipid loading onto CD1d has only been formally demonstrated *in vitro* for a few lipids, and in particular α-GalCer[33,34]. Therefore, both proteins were pre-loaded with α-GalCer prior to performing the equilibrium binding assay. However, this variation also did not result in a detectable interaction between SapB and CD1d (Figure 2C). To exclude the possibility that a strong interaction between CD1d and SapB only occurs when SapB exchanges a lower affinity antigen for α-GalCer, we repeated the equilibrium binding assay where SapB was pre-loaded with α-GalCer while CD1d retained the co-purified endogenous lipids (Figure 2D). Again, there was no substantial depletion of SapB from the supernatant in this assay. These results strongly suggest that CD1d and SapB do not directly interact *in vitro*, or that their interaction is extremely weak with a $K_d > 40$ μM.

In vitro cross-linking does not capture an interaction between CD1d and SapB

Cross-linking of proteins using short chemical tethers that react with surface-exposed lysine side-chains can be used to capture transient or weak interactions. Both CD1d and SapB have several surface-exposed lysines, and of particular importance, lysine residues are present near the lipid-binding groove of CD1d where we would predict SapB to bind (Figure 3A and B). Therefore, it is possible that CD1d and SapB interact in a way that is amenable to intermolecular cross-linking of surface lysines by amine-reactive cross-linkers. We used cross-linkers of different lengths to maximise the chances of capturing an interaction: DSSO, with a 10.3 Å spacer arm, and BS(PEG)$_5$, with a 21.7 Å spacer arm. CD1d-β_2m and SapB were incubated with these cross-linkers either alone or following mixing (Figure 3C). For the CD1d-β_2m samples alone, cross-linking between CD1d and β_2m is clearly evident, as expected. There is also some evidence of additional higher order oligomers, probably due to the large number of surface lysines on both CD1d and β_2m. However, no additional cross-linking products were observed when CD1d-β_2m and SapB were incubated together with the cross-linkers. This indicates that CD1d-β_2m and SapB do not interact together in a conformation that is amenable to cross-linking.

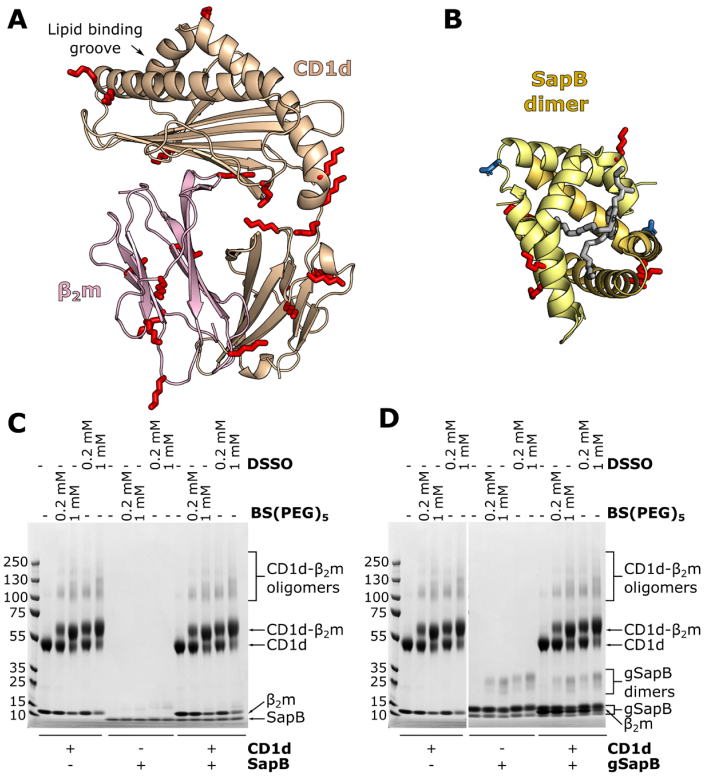

Figure 3. *In vitro* cross-linking of CD1d-β_2m and SapB. A. Ribbon diagram of the structure of CD1d (beige) and β_2m (pink)(PDB ID 3U0P) with the lipid antigen groove labelled (arrow). Surface-accessible lysine residues available for cross-linking are highlighted (red sticks). **B**. Ribbon diagram of the structure of a SapB dimer (yellow and orange, PDB ID 1N69) with bound lipid (in gray). Lysine residues are highlighted as in panel **A**. The N-linked glycosylation site of SapB is highlighted (blue sticks). **C**. Coomassie-stained SDS-PAGE of CD1d-β_2m and unglycosylated SapB samples following incubation with cross-linkers of different lengths: DSSO (10.3 Å) or BS(PEG)$_5$ (21.7 Å). **D**. As for panel **C** but using CD1d-β_2m and glycosylated SapB (gSapB).

SapB possesses one N-linked glycosylation site on residue Asn21 (Figure 3B), which remains unglycosylated when expressed in *E. coli*. Although unglycosylated SapB loads lipids onto CD1d[33,34], it is possible that this glycan on SapB may increase its affinity for CD1d. Therefore, we repeated the cross-linking experiment following expression of His-tagged SapB in mammalian HEK293F cells, enabling the purification of glycosylated protein. Following purification, two SapB-H_6 species were present, with molecular weights of approximately 10 and 13 kDa, likely corresponding to unglycosylated and glycosylated SapB-H_6 (Figure 3D). It was not possible to separate these two SapB species, likely due to SapB dimerization, and so cross-linking experiments were carried out using this mixed sample (Figure 3D). However, we again did not observe formation of any additional cross-linked products when CD1d-β_2m was incubated with glycosylated SapB-H_6. We note that although the same amount of unglycosylated SapB (Figure 3C) and glycosylated SapB (Figure 3D) were used in cross-linking assays, the Coomassie-stained bands corresponding to glycosylated SapB appear much stronger. This increased staining may be due to the additional histidine residues in the affinity tag increasing the amount of Coomassie reagent binding to this small protein.

CD1d-β_2m and SapB do not co-crystallise

An alternative method to capture weak, transient protein-protein interactions is to attempt co-crystallisation. In this experiment both protein components are mixed together at a range of concentrations and different molar ratios, then subjected to crystallisation trials. The benefit of using the sitting-drop vapour diffusion approach is that during the experiment the drop containing protein and precipitant gradually equilibrates with the reservoir, generally resulting in reduced drop volume and the progressive increase in concentration of the drop components. In this way, the protein components are steadily concentrated together with the aim of capturing an interaction that is otherwise extremely hard to measure using biochemical and biophysical assays[62].

Extensive crystallisation trials containing both CD1d-β_2m and SapB were performed, varying the following parameters: (1) protein concentration (6-13 mg/mL CD1d-β_2m, 2-6 mg/mL SapB), (2) molar ratio of SapB to CD1d, taking into account the predicted dimeric state of SapB (2:1-3:1 SapB to CD1d); (3) full glycosylation or deglycosylation of protein components; and (4) lipid exchange for α-GalCer. Figure 4A shows an example of the high purity of the protein components that went into crystallisation trials. These experiments yielded several different crystals in a range of conditions, a subset of which are shown in Figure 4B. To determine the composition of these crystals, two approaches were used depending on the quality of the crystals grown. The first approach, which can be used for any protein crystal, involves harvesting and washing the crystals in reservoir solution to remove residual uncrystallised components, then separating the proteins by SDS-PAGE followed by detection using Coomassie staining (Figure 4C, each gel lane corresponds to the above crystal in panel 4B). This revealed that both SapB and CD1d-β_2m were capable of forming crystals but in none of the crystals grown did both components crystallise together. The second approach used was to collect diffraction data for any crystals that diffracted to 5 Å resolution or better, then use molecular replacement to confirm the contents of the asymmetric unit (Table 2). Three crystal forms satisfied this criteria, including two that had also

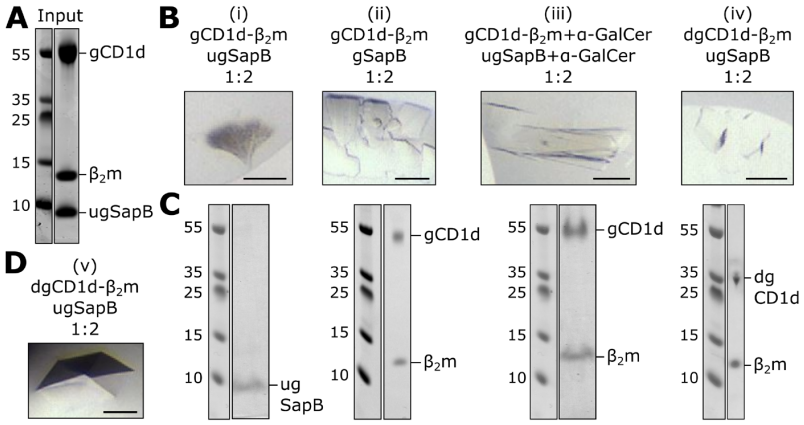

Figure 4. Crystallisation trials of CD1d-β_2m and SapB. A. Coomassie-stained SDS-PAGE example of the purified protein sample used in crystallisation trials. **B**. Examples of four crystals (i)–(iv) grown in a range of conditions using different protein compositions. Scale bar represents 100 µm. **C**. Coomassie-stained SDS-PAGE of harvested and washed crystals shown in panel **B**. **D**. Image of the crystal that diffracted to highest resolution as described in Table 2. Scale bar represents 100 µm. Abbreviations: gCD1d-glycosylated CD1d, dgCD1d-deglycosylated CD1d, gSapB-glycosylated SapB, and ugSapB-unglycosylated SapB.

Table 2. Data collection statistics and molecular replacement solutions. For Crystal 4B (iii) statistics have been included for data from a single crystal before (isotropic) and after anisotropic truncation and scaling using STARANISO. Values in parentheses are for the highest-resolution shell.

Data collection	Crystal Fig 4B (iii)		Crystal Fig. 4B (iv)	Crystal Fig. 4D (v)
Beamline	I04		I04-1	I03
Wavelength (Å)	0.9999		0.9159	0.9793
Processing package	autoPROC + STARANISO		XIA2 3dii	XIA2 DIALS
Space group	*C* 2		*P* 2_1	*C* 2
Cell dimensions				
a,*b*,*c* (Å)	273.3, 88.0, 127.9		123.8, 122.9, 159.9	186.2, 57.1, 129.2
α,β,γ (°)	90, 114, 90		90, 95, 90	90, 100, 90
	Isotropic	*Anisotropic*		
Resolution (Å)	124.71 – 4.97 (5.05 – 4.97)	124.71 – 4.55 (5.00 – 4.55)	123.27 – 3.49 (3.55 – 3.49)	54.53 – 2.52 (2.56 – 2.52)
R_{merge}	0.438 (2.187)	0.198 (0.507)	0.161 (1.572)	0.072 (2.438)
R_{pim}	0.279 (1.366)	0.127 (0.323)	0.067 (0.663)	0.031 (1.028)
$CC_{1/2}$	0.967 (0.322)	0.987 (0.704)	0.992 (0.714)	0.999 (0.813)
$I/\sigma I$	1.9 (0.5)	3.4 (1.8)	6.9 (1.0)	10.9 (0.8)
Completeness (%)	99.8 (99.8)	e* 80.0 (56.7) s* 39.2 (8.1)	99.8 (99.9)	99.8 (98.0)
Redundancy	3.4 (3.5)	3.4 (3.4)	6.9 (6.6)	6.7 (6.6)
Molecular replacement				
# CD1-β_2m/ASU	4		6	2

*ellipsoidal (e) and spherical (s) completeness are quoted for anisotropic data processing using STARANISO.

been tested using SDS-PAGE (Figure 4B (iii) and (iv) and 4D (v)). These three samples all contained CD1d-β_2m only as confirmed by molecular replacement trials using multiple copies of all potential components as well as manual inspection of maps. In all datasets additional density was observed for unmodelled glycans on CD1d (deglycosylation of CD1d using Endo H leaves a residual acetylated glycosamine), confirming the validity of the molecular replacement solution. For several CD1d molecules, additional unmodelled density near the lipid-binding groove was evident that likely corresponds to partially-ordered lipid components. Further inspection of electron density maps revealed no evidence for the presence of SapB. For most molecules of CD1d in all three different crystal forms the predicted binding site for SapB (over the lipid-binding groove) was occupied by another copy of CD1d. The use of both SDS-PAGE analysis of all grown crystals and molecular replacement solution of higher quality crystals confirmed that, despite extensive efforts with a broad range of CD1d-β_2m/SapB combinations, these do not form a stable, crystallisable complex.

Discussion

Using highly purified components we have been unable to identify a direct interaction between CD1d-β_2m and SapB *in vitro*. Several biochemical and structural approaches were implemented that have the capacity to capture weak and transient interactions, including binding equilibrium assays, protein cross-linking, and co-crystallisation trials. Various factors that may influence the strength or specificity of these interactions were tested, including different pH and buffer systems, exchange of the lipids loaded onto both protein components, and changing the glycosylation state of each protein. None of these variants resulted in the detection of a direct interaction. Our previous work studying saposin interactions with lysosomal hydrolases showed that although these interactions are sensitive to pH and salt concentrations, they were always detectable, and while detergent/lipid was required, neither the nature of the lipid/detergent nor the glycosylation state mattered[19].

Direct protein-protein interactions are crucial for several different lipid transfer pathways. The lipopolysaccharide (LPS) transfer cascade resulting in Toll-like receptor 4 (TLR4) activation involves several lipid transfer steps. LPS binding protein (LBP) extracts LPS and transfers it onto the receptor CD14. CD14 then transfers LPS to the TLR4 co-receptor MD-2. A direct, high affinity (K_d = 5.52 nM) interaction between LBP and CD14 is necessary for lipid transfer, and lipid transfer promotes rapid dissociation of CD14:LPS from LBP. Similarly, a direct interaction between CD14 and the TLR4:MD-2 complex is required for lipid transfer[38]. The transfer of lipids between proteins is also seen in the formation of Apolipoprotein B

(apoB)-containing lipoproteins. Microsomal Triglyceride Transfer protein assists this process by binding to apoB with high affinity (K_d 10-30 nM)[63] and transferring lipids onto apoB through an unknown mechanism. Both functions are required for lipoprotein assembly[39]. As another example, Niemann–Pick C1 (NPC1) and Niemann–Pick C2 (NPC2) proteins collaborate to export cholesterol from the lysosome. Soluble NPC2 binds cholesterol in the lysosomal lumen, interacts with the middle lumenal domain of the multipass transmembrane protein NPC1 and transfers cholesterol to NPC1 via a transfer tunnel linking the cholesterol-binding pockets of both proteins. The interaction between NPC2 and NPC1 (K_d 0.4-2 μM) is required for cholesterol export[40]. In all cases, a high-to-medium affinity direct interaction between proteins engaged in lipid transfer is detectable. In this study, we were unable to detect a direct interaction between CD1d and SapB despite the assays being capable of detecting very weak interactions. Specifically, the equilibrium assay used in this study can detect protein interactions where the dissociation constant is as weak as 40 μM. In the absence of avidity effects, interactions weaker than 40 μM are unlikely to be physiologically relevant. Furthermore, the cross-linking assay can detect short-lived, transient interactions and the co-crystallography technique uses very high concentrations of protein components allowing for the detection of very weak interactions. It therefore seems unlikely that an interaction that is undetectable by these methods could permit direct lipid exchange between SapB and CD1d.

The study of lipid loading of CD1 molecules is technically challenging as biochemical methods often require the inclusion of detergents, which may interfere with lipid solubilisation and binding. Also, the detection of changes to CD1d-lipid complexes is limited by available reagents including specific antibodies and iNKT TCRs. However, the inability to identify a direct interaction between CD1d and SapB raises a number of important questions regarding the role of saposins in lipid antigen presentation.

It is well established that knockout of the saposin precursor protein PSAP results in impaired presentation of lipid antigens by CD1d. Due to the dual role of saposins in lipid loading and lipid processing, the absence of PSAP will not only alter CD1d lipid loading, but will also significantly alter the composition of the lysosomal lipidome. Indeed, the loss of a single saposin results in the cellular accumulation of multiple, different GSLs. For example, loss of SapB alone, in mice or in humans, results in the accumulation of sulfatides, globotriaosylceramide, and lactosylceramide[64,65]. Therefore, in saposin knockout and rescue models, the effects on antigen processing and the changes to local lipid concentrations may have indirect and unpredictable effects on CD1d loading in the lysosome. Similarly, cell-based assays that involve recognition of CD1d-lipid complexes by NKT cells are complicated by the requirement of some glycolipid headgroups to be processed by lysosomal hydrolases[12], a process that is itself dependent upon saposins.

In vitro assays testing the role of different saposins in facilitating lysosomal hydrolase activity identify some of the difficulties in determining specificity of saposin-lipid binding. For example, *in vitro* assays have identified that SapC and SapA are both able to enhance the activity of β-galactosylceramidase (GALC) to promote degradation of the galactosphingolipids[66,67]. However, this does not reflect the specificity of saposins in whole organisms as loss of SapA causes Krabbe disease, similar to that of loss of GALC, while mutations in SapC do not[20,21,23–25]. Why these *in vitro* assays do not recapitulate the specificity of saposins in a whole organism remains unclear but may suggest a non-specific detergent-like effect of saposins when present at relatively high concentrations. It is unclear whether a similar effect may be partially confounding the interpretation of SapB influence on CD1d lipid loading. While *in vitro* studies demonstrating the ability of SapB to enhance CD1d stimulation of iNKT cells or binding to iNKT TCRs are very compelling regarding the importance of SapB in this process, these do represent an indirect way of monitoring how SapB may be involved in lipid exchange[33,34]. In light of our data showing no detectable interaction between CD1d and SapB, the question remains: how can SapB facilitate lipid editing without directly binding CD1d? Modelling of how SapB might facilitate lipid exchange identified a role for SapB in enhancing the kinetics of lipid exchange in a manner similar to that of Tapasin in peptide exchange on MHC class I molecules[34]. This comparison is driven by the crucial roles that Tapasin and SapB play in the binding and exchange of peptides or lipids, respectively, to the antigen-binding grooves of MHC class I or CD1d molecules. In order for Tapasin to carry out this role, it binds directly to MHC class I but as part of a larger complex including additional proteins[2,68]. Although the *in vitro* assays demonstrating the importance of SapB for CD1d lipid presentation did not require other proteins, it is possible that there may be additional factors involved in enhancing the affinity of SapB for CD1d in cells.

Additional factors may influence the efficiency of lipid loading onto CD1d in cells. For example, the relative abundance of different LTPs in the lysosome may influence the lysosomal lipid composition, altering exchange kinetics. Therefore, it may be important that SapA-D exist in equimolar concentrations in the endolysosome, as ensured by their production from a common precursor. If this is the case, studies involving individual saposins or partial PSAP rescues will be difficult to interpret, potentially explaining some experimental inconsistencies. The lysosomal lipidome is not only influenced by the activity of saposins and sphingolipid processing hydrolases but also by the careful balance of other membrane components such as cholesterol and bis(monoacylglycero)phosphate that are crucial for endosomal vesicle formation. Lipid loading of CD1 in endolysosomal compartments is highly complex, requiring the delicate interplay of lipid binding, lipid processing and lipid exchange. For now, the mechanism of transfer of lipids onto CD1 molecules, as well as the molecular determinants of LTP binding specificity, remains unclear.

Data availability

Underlying data

Figshare: Original data used in Figure 1, https://doi.org/10.6084/m9.figshare.8926709[69].

Figshare: Original equilibrium binding assay gels used in Figure 2, https://doi.org/10.6084/m9.figshare.8864384[70].

Figshare: Original cross-linking gels used in Figure 3, https://doi.org/10.6084/m9.figshare.8932856[71].

Figshare: Original gels and images used in Figure 4, https://doi.org/10.6084/m9.figshare.8932898[72].

Apollo: Research data supporting "The lipid transfer protein Saposin B does not directly bind CD1d for lipid antigen loading". https://doi.org/10.17863/CAM.41683[73].

This project contains the following underlying data:

- Original X-ray diffraction datasets, processed diffraction data, molecular replacement solutions for crystals (iii), (iv) and (v)

Acknowledgements

We acknowledge Diamond Light Source for time on beamlines I03, I04 and I04-1 under proposal MX15916. Remote access was supported in part by the EU FP7 infrastructure grant BIOSTRUCT-X (Contract No. 283570).

References

1. Pos W, Sethi DK, Call MJ, *et al.*: **Crystal structure of the HLA-DM-HLA-DR1 complex defines mechanisms for rapid peptide selection.** *Cell.* 2012; **151**(7): 1557–68.
PubMed Abstract | Publisher Full Text | Free Full Text
2. Blees A, Januliene D, Hofmann T, *et al.*: **Structure of the human MHC-I peptide-loading complex.** *Nature.* 2017; **551**(7681): 525–528.
PubMed Abstract | Publisher Full Text
3. Ilca FT, Neerincx A, Hermann C, *et al.*: **TAPBPR mediates peptide dissociation from MHC class I using a leucine lever.** *eLife.* 2018; **7**: pii: e40126.
PubMed Abstract | Publisher Full Text | Free Full Text
4. Layre E, de Jong A, Moody DB: **Human T cells use CD1 and MR1 to recognize lipids and small molecules.** *Curr Opin Chem Biol.* 2014; **23**: 31–8.
PubMed Abstract | Publisher Full Text
5. Moody DB, Cotton RN: **Four pathways of CD1 antigen presentation to T cells.** *Curr Opin Immunol.* 2017; **46**: 127–133.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Brennan PJ, Brigl M, Brenner MB: **Invariant natural killer T cells: an innate activation scheme linked to diverse effector functions.** *Nat Rev Immunol.* 2013; **13**(2): 101–17.
PubMed Abstract | Publisher Full Text
7. Sprong H, van der Sluijs P, van Meer G: **How proteins move lipids and lipids move proteins.** *Nat Rev Mol Cell Biol.* 2001; **2**(7): 504–13.
PubMed Abstract | Publisher Full Text
8. Cox D, Fox L, Tian R, *et al.*: **Determination of cellular lipids bound to human CD1d molecules.** *PLoS One.* 2009; **4**(5): e5325.
PubMed Abstract | Publisher Full Text | Free Full Text
9. Yuan W, Kang SJ, Evans JE, *et al.*: **Natural lipid ligands associated with human CD1d targeted to different subcellular compartments.** *J Immunol.* 2009; **182**(8): 4784–91.
PubMed Abstract | Publisher Full Text | Free Full Text
10. Haig NA, Guan Z, Li D, *et al.*: **Identification of self-lipids presented by CD1c and CD1d proteins.** *J Biol Chem.* 2011; **286**(43): 37692–37701.
PubMed Abstract | Publisher Full Text | Free Full Text
11. Teyton L: **Role of lipid transfer proteins in loading CD1 antigen-presenting molecules.** *J Lipid Res.* 2018; **59**(8): 1367–1373.
PubMed Abstract | Publisher Full Text | Free Full Text
12. Prigozy TI, Naidenko O, Qasba P, *et al.*: **Glycolipid antigen processing for presentation by CD1d molecules.** *Science.* 2001; **291**(5504): 664–7.
PubMed Abstract | Publisher Full Text
13. Darmoise A, Teneberg S, Bouzonville L, *et al.*: **Lysosomal alpha-galactosidase controls the generation of self lipid antigens for natural killer T cells.** *Immunity.* 2010; **33**(2): 216–28.
PubMed Abstract | Publisher Full Text | Free Full Text
14. Ahn VE, Faull KF, Whitelegge JP, *et al.*: **Crystal structure of saposin B reveals a dimeric shell for lipid binding.** *Proc Natl Acad Sci U S A.* 2003; **100**(1): 38–43.
PubMed Abstract | Publisher Full Text | Free Full Text
15. de Alba E, Weiler S, Tjandra N: **Solution structure of human saposin C: pH-dependent interaction with phospholipid vesicles.** *Biochemistry.* 2003; **42**(50): 14729–40.
PubMed Abstract | Publisher Full Text
16. Ahn VE, Leyko P, Alattia JR, *et al.*: **Crystal structures of saposins A and C.** *Protein Sci.* 2006; **15**(8): 1849–57.
PubMed Abstract | Publisher Full Text | Free Full Text
17. Popovic K, Privé GG: **Structures of the human ceramide activator protein saposin D.** *Acta Crystallogr D Biol Crystallogr.* 2008; **64**(Pt 5): 589–94.
PubMed Abstract | Publisher Full Text
18. Popovic K, Holyoake J, Pomès R, *et al.*: **Structure of saposin A lipoprotein discs.** *Proc Natl Acad Sci U S A.* 2012; **109**(8): 2908–12.
PubMed Abstract | Publisher Full Text | Free Full Text
19. Hill CH, Cook GM, Spratley SJ, *et al.*: **The mechanism of glycosphingolipid degradation revealed by a GALC-SapA complex structure.** *Nat Commun.* 2018; **9**(1): 151.
PubMed Abstract | Publisher Full Text | Free Full Text
20. Matsuda J, Vanier MT, Saito Y, *et al.*: **A mutation in the saposin A domain of the sphingolipid activator protein (prosaposin) gene results in a late-onset, chronic form of globoid cell leukodystrophy in the mouse.** *Hum Mol Genet.* 2001; **10**(11): 1191–9.
PubMed Abstract | Publisher Full Text
21. Spiegel R, Bach G, Sury V, *et al.*: **A mutation in the saposin A coding region of the prosaposin gene in an infant presenting as Krabbe disease: first report of saposin A deficiency in humans.** *Mol Genet Metab.* 2005; **84**(2): 160–6.
PubMed Abstract | Publisher Full Text
22. Kuchar L, Ledvinová J, Hrebicek M, *et al.*: **Prosaposin deficiency and saposin B deficiency (activator-deficient metachromatic leukodystrophy): report on two patients detected by analysis of urinary sphingolipids and carrying novel PSAP gene mutations.** *Am J Med Genet A.* 2009; **149A**(4): 613–21.
PubMed Abstract | Publisher Full Text | Free Full Text
23. Vaccaro AM, Motta M, Tatti M, *et al.*: **Saposin C mutations in Gaucher disease patients resulting in lysosomal lipid accumulation, saposin C deficiency, but normal prosaposin processing and sorting.** *Hum Mol Genet.* 2010; **19**(15): 2987–97.
PubMed Abstract | Publisher Full Text
24. Motta M, Camerini S, Tatti M, *et al.*: **Gaucher disease due to saposin C deficiency is an inherited lysosomal disease caused by rapidly degraded mutant proteins.** *Hum Mol Genet.* 2014; **23**(21): 5814–26.
PubMed Abstract | Publisher Full Text
25. Kang L, Zhan X, Ye J, *et al.*: **A rare form of Gaucher disease resulting from saposin C deficiency.** *Blood Cells Mol Dis.* 2018; **68**: 60–65.
PubMed Abstract | Publisher Full Text
26. Locatelli-Hoops S, Remmel N, Klingenstein R, *et al.*: **Saposin A mobilizes lipids from low cholesterol and high bis(monoacylglycerol)phosphate-containing membranes: patient variant Saposin A lacks lipid extraction capacity.** *J Biol Chem.* 2006; **281**(43): 32451–60.
PubMed Abstract | Publisher Full Text
27. Remmel N, Locatelli-Hoops S, Breiden B, *et al.*: **Saposin B mobilizes lipids from cholesterol-poor and bis(monoacylglycero)phosphate-rich membranes at**

acidic pH. Unglycosylated patient variant saposin B lacks lipid-extraction capacity. *FEBS J.* 2007; **274**(13): 3405–3420.
PubMed Abstract | Publisher Full Text

28. Alattia JR, Shaw JE, Yip CM, *et al.*: **Direct visualization of saposin remodelling of lipid bilayers.** *J Mol Biol.* 2006; **362**(5): 943–53.
PubMed Abstract | Publisher Full Text

29. Alattia JR, Shaw JE, Yip CM, *et al.*: **Molecular imaging of membrane interfaces reveals mode of beta-glucosidase activation by saposin C.** *Proc Natl Acad Sci U S A.* 2007; **104**(44): 17394–9.
PubMed Abstract | Publisher Full Text | Free Full Text

30. León L, Tatituri RV, Grenha R, *et al.*: **Saposins utilize two strategies for lipid transfer and CD1 antigen presentation.** *Proc Natl Acad Sci U S A.* 2012; **109**(12): 4357–64.
PubMed Abstract | Publisher Full Text | Free Full Text

31. Zhou D, Cantu C 3rd, Sagiv Y, *et al.*: **Editing of CD1d-bound lipid antigens by endosomal lipid transfer proteins.** *Science.* 2004; **303**(5657): 523–7.
PubMed Abstract | Publisher Full Text | Free Full Text

32. Kang SJ, Cresswell P: **Saposins facilitate CD1d-restricted presentation of an exogenous lipid antigen to T cells.** *Nat Immunol.* 2004; **5**(2): 175–81.
PubMed Abstract | Publisher Full Text

33. Yuan W, Qi X, Tsang P, *et al.*: **Saposin B is the dominant saposin that facilitates lipid binding to human CD1d molecules.** *Proc Natl Acad Sci U S A.* 2007; **104**(13): 5551–6.
PubMed Abstract | Publisher Full Text | Free Full Text

34. Salio M, Ghadbane H, Dushek O, *et al.*: **Saposins modulate human invariant Natural Killer T cells self-reactivity and facilitate lipid exchange with CD1d molecules during antigen presentation.** *Proc Natl Acad Sci U S A.* 2013; **110**(49): E4753–61.
PubMed Abstract | Publisher Full Text | Free Full Text

35. Schrantz N, Sagiv Y, Liu Y, *et al.*: **The Niemann-Pick type C2 protein loads isoglobotrihexosylceramide onto CD1d molecules and contributes to the thymic selection of NKT cells.** *J Exp Med.* 2007; **204**(4): 841–52.
PubMed Abstract | Publisher Full Text | Free Full Text

36. Winau F, Schwierzeck V, Hurwitz R, *et al.*: **Saposin C is required for lipid presentation by human CD1b.** *Nat Immunol.* 2004; **5**(2): 169–74.
PubMed Abstract | Publisher Full Text

37. Wong LH, Gatta AT, Levine TP: **Lipid transfer proteins: the lipid commute via shuttles, bridges and tubes.** *Nat Rev Mol Cell Biol.* 2019; **20**(2): 85–101.
PubMed Abstract | Publisher Full Text

38. Ryu JK, Kim SJ, Rah SH, *et al.*: **Reconstruction of LPS Transfer Cascade Reveals Structural Determinants within LBP, CD14, and TLR4-MD2 for Efficient LPS Recognition and Transfer.** *Immunity.* 2017; **46**(1): 38–50.
PubMed Abstract | Publisher Full Text

39. Hussain MM, Rava P, Walsh M, *et al.*: **Multiple functions of microsomal triglyceride transfer protein.** *Nutr Metab (Lond).* 2012; **9**: 14.
PubMed Abstract | Publisher Full Text | Free Full Text

40. Li X, Saha P, Li J, *et al.*: **Clues to the mechanism of cholesterol transfer from the structure of NPC1 middle lumenal domain bound to NPC2.** *Proc Natl Acad Sci U S A.* 2016; **113**(36): 10079–84.
PubMed Abstract | Publisher Full Text | Free Full Text

41. Li Z, Michael IP, Zhou D, *et al.*: **Simple *piggyBac* transposon-based mammalian cell expression system for inducible protein production.** *Proc Natl Acad Sci U S A.* 2013; **110**(13): 5004–9.
PubMed Abstract | Publisher Full Text | Free Full Text

42. Aricescu AR, Lu W, Jones EY: **A time- and cost-efficient system for high-level protein production in mammalian cells.** *Acta Crystallogr D Biol Crystallogr.* 2006; **62**(Pt 10): 1243–50.
PubMed Abstract | Publisher Full Text

43. Hill CH, Read RJ, Deane JE: **Structure of human saposin A at lysosomal pH.** *Acta Crystallogr F Struct Biol Commun.* 2015; **71**(Pt 7): 895–900.
PubMed Abstract | Publisher Full Text | Free Full Text

44. Pellicci DG, Patel O, Kjer-Nielsen L, *et al.*: **Differential recognition of CD1d-alpha-galactosyl ceramide by the V beta 8.2 and V beta 7 semi-invariant NKT T cell receptors.** *Immunity.* 2009; **31**(1): 47–59.
PubMed Abstract | Publisher Full Text | Free Full Text

45. López-Sagaseta J, Kung JE, Savage PB, *et al.*: **The molecular basis for recognition of CD1d/α-galactosylceramide by a human non-Vα24 T cell receptor.** *PLoS Biol.* 2012; **10**(10): e1001412.
PubMed Abstract | Publisher Full Text | Free Full Text

46. Winter G: ***xia2*: an expert system for macromolecular crystallography data reduction.** *J Appl Crystallogr.* 2010; **43**: 186–190.
Publisher Full Text

47. Winter G, Waterman DG, Parkhurst JM, *et al.*: ***DIALS*: implementation and evaluation of a new integration package.** *Acta Crystallogr D Struct Biol.* 2018; **74**(Pt 2): 85–97.
PubMed Abstract | Publisher Full Text | Free Full Text

48. Tickle IJ, Flensburg C, Keller P, *et al.*: **STARANISO.** Cambridge, United Kingdom: Global Phasing Ltd. 2018.
Reference Source

49. Evans PR, Murshudov GN: **How good are my data and what is the resolution?** *Acta Crystallogr D Biol Crystallogr.* 2013; **69**(Pt 7): 1204–1214.
PubMed Abstract | Publisher Full Text | Free Full Text

50. López-Sagaseta J, Sibener LV, Kung JE, *et al.*: **Lysophospholipid presentation by CD1d and recognition by a human Natural Killer T-cell receptor.** *EMBO J.* 2012; **31**(8): 2047–59.
PubMed Abstract | Publisher Full Text | Free Full Text

51. Adams PD, Afonine PV, Bunkóczi G, *et al.*: ***PHENIX*: a comprehensive Python-based system for macromolecular structure solution.** *Acta Crystallogr D Biol Crystallogr.* 2010; **66**(Pt 2): 213–21.
PubMed Abstract | Publisher Full Text | Free Full Text

52. Bunkóczi G, Read RJ: **Improvement of molecular-replacement models with *Sculptor*.** *Acta Crystallogr D Biol Crystallogr.* 2011; **67**(Pt 4): 303–12.
PubMed Abstract | Publisher Full Text | Free Full Text

53. McCoy AJ, Grosse-Kunstleve RW, Adams PD, *et al.*: ***Phaser* crystallographic software.** *J Appl Crystallogr.* 2007; **40**(Pt 4): 658–674.
PubMed Abstract | Publisher Full Text | Free Full Text

54. Afonine PV, Grosse-Kunstleve RW, Echols N, *et al.*: **Towards automated crystallographic structure refinement with *phenix.refine*.** *Acta Crystallogr D Biol Crystallogr.* 2012; **68**(Pt 4): 352–67.
PubMed Abstract | Publisher Full Text | Free Full Text

55. Emsley P, Lohkamp B, Scott WG, *et al.*: **Features and development of Coot.** *Acta Crystallogr D Biol Crystallogr.* 2010; **66**(Pt 4): 486–501.
PubMed Abstract | Publisher Full Text | Free Full Text

56. Koch M, Stronge VS, Shepherd D, *et al.*: **The crystal structure of human CD1d with and without alpha-galactosylceramide.** *Nat Immunol.* 2005; **6**(8): 819–826.
PubMed Abstract | Publisher Full Text

57. Gumperz JE, Roy C, Makowska A, *et al.*: **Murine CD1d-restricted T cell recognition of cellular lipids.** *Immunity.* 2000; **12**(2): 211–21.
PubMed Abstract | Publisher Full Text

58. Kim HS, Garcia J, Exley M, *et al.*: **Biochemical characterization of CD1d expression in the absence of beta2-microglobulin.** *J Biol Chem.* 1999; **274**(14): 9289–95.
PubMed Abstract | Publisher Full Text

59. Pollard TD: **A guide to simple and informative binding assays.** *Mol Biol Cell.* 2010; **21**(23): 4061–7.
PubMed Abstract | Publisher Full Text | Free Full Text

60. Paduraru C, Spiridon L, Yuan W, *et al.*: **An *N*-linked glycan modulates the interaction between the CD1d heavy chain and and beta 2-microglobulin.** *J Biol Chem.* 2006; **281**(52): 40369–78.
PubMed Abstract | Publisher Full Text

61. Brennan PJ, Cheng TY, Pellicci DG, *et al.*: **Structural determination of lipid antigens captured at the CD1d-T-cell receptor interface.** *Proc Natl Acad Sci U S A.* 2017; **114**(31): 8348–8353.
PubMed Abstract | Publisher Full Text | Free Full Text

62. Muenzner J, Traub LM, Kelly BT, *et al.*: **Cellular and viral peptides bind multiple sites on the N-terminal domain of clathrin.** *Traffic.* 2017; **18**(1): 44–57.
PubMed Abstract | Publisher Full Text | Free Full Text

63. Hussain MM, Bakillah A, Jamil H: **Apolipoprotein B binding to microsomal triglyceride transfer protein decreases with increases in length and lipidation: implications in lipoprotein biosynthesis.** *Biochemistry.* 1997; **36**(42): 13060–7.
PubMed Abstract | Publisher Full Text

64. Li SC, Kihara H, Serizawa S, *et al.*: **Activator protein required for the enzymatic hydrolysis of cerebroside sulfate. Deficiency in urine of patients affected with cerebroside sulfatase activator deficiency and identity with activators for the enzymatic hydrolysis of GM1 ganglioside and globotriaosylceramide.** *J Biol Chem.* 1985; **260**(3): 1867–71.
PubMed Abstract

65. Sun Y, Witte DP, Ran H, *et al.*: **Neurological deficits and glycosphingolipid accumulation in saposin B deficient mice.** *Hum Mol Genet.* 2008; **17**(15): 2345–56.
PubMed Abstract | Publisher Full Text | Free Full Text

66. Harzer K, Paton BC, Christomanou H, *et al.*: **Saposins (*sap*) A and C activate the degradation of galactosylceramide in living cells.** *FEBS Lett.* 1997; **417**(3): 270–4.
PubMed Abstract | Publisher Full Text

67. Harzer K, Hiraiwa M, Paton BC: **Saposins (sap) A and C activate the degradation of galactosylsphingosine.** *FEBS Lett.* 2001; **508**(1): 107–10.
PubMed Abstract | Publisher Full Text

68. Dong G, Wearsch PA, Peaper DR, *et al.*: **Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer.** *Immunity.* 2009; **30**(1): 21–32.
PubMed Abstract | Publisher Full Text | Free Full Text

69. Shamin M, Benedyk TH, Graham SC, *et al.*: **Original data used in Figure 1.** *figshare.* Figure. 2019.
http://www.doi.org/10.6084/m9.figshare.8926709.v1

70. Shamin M, Benedyk TH, Graham SC, *et al.*: **Original equilibrium binding assay gels used in Figure 2.** *figshare.* Figure. 2019.
http://www.doi.org/10.6084/m9.figshare.8864384.v1

71. Shamin M, Benedyk TH, Graham SC, *et al.*: **Original cross-linking gels used in Figure 3.** *figshare.* Figure. 2019.
http://www.doi.org/10.6084/m9.figshare.8932856.v1

72. Shamin M, Benedyk TH, Graham SC, *et al.*: **Original gels and images used in Figure 4.** *figshare.* Figure. 2019.
http://www.doi.org/10.6084/m9.figshare.8932898.v1

73. Deane J, Graham S, Shamin M, *et al.*: **Research data supporting "The lipid transfer protein Saposin B does not directly bind CD1d for lipid antigen loading".** [Dataset]. 2019.
http://www.doi.org/10.17863/CAM.41683

Appendix II

Proteins significantly changed more than 2-fold at the plasma membrane of GALC KO1 and KO2 versus EV1 control THP-1 cells, quantified by plasma membrane profiling-mass spectrometry (Benjamini-Hochberg adjusted q-value < 0.05).

Gene ID	Log_2 (fold change KO1/EV1)	q-value	Log_2 (fold change KO2/EV1)	q-value
IL18R1	3.35	7.16E-04	1.98	5.56E-03
SLC7A11	3.22	8.89E-04	2.69	2.06E-03
FAS	2.85	8.89E-04	1.98	5.10E-03
TRHDE	2.61	2.34E-04	2.85	6.82E-05
GGT5	2.55	2.34E-04	1.40	3.82E-03
ANPEP	2.49	2.34E-04	1.66	1.65E-03
EPHB1	2.21	8.89E-04	1.48	6.19E-03
CIB1	2.20	2.16E-03	2.63	9.25E-04
SDK1	2.19	1.57E-03	1.92	2.73E-03
SDC2	2.10	8.18E-04	1.47	4.34E-03
SLC40A1	2.09	2.76E-03	1.90	5.22E-03
JAG1	1.83	2.34E-04	1.76	4.11E-04
TMC8	1.81	8.89E-04	2.10	4.22E-04
ACE	1.72	2.71E-03	2.22	9.25E-04
PROCR	1.63	8.89E-04	1.54	9.61E-04
SLC7A7	1.62	6.17E-03	1.74	4.18E-03
TMC6	1.62	9.05E-04	1.83	7.46E-04
MCEMP1	1.39	1.95E-03	1.39	2.04E-03
CD244	1.38	1.95E-03	1.64	9.61E-04
PCDH20	1.34	6.17E-03	1.42	4.34E-03
IL2RB	1.22	5.67E-03	1.11	9.37E-03
LPAR4	1.10	8.84E-03	1.11	8.12E-03
ICAM2	-1.08	9.10E-03	-1.17	5.56E-03
DLL1	-1.30	7.05E-03	-1.63	2.04E-03
FLT3	-1.31	6.60E-03	-1.27	6.14E-03
ITGAL	-1.61	1.46E-03	-1.57	1.66E-03
TMEM158	-1.62	3.70E-03	-1.97	1.52E-03
ALB	-2.01	3.09E-03	-1.99	3.89E-03
CNTN4	-2.16	6.17E-03	-2.01	8.12E-03
AXL	-2.52	1.57E-03	-1.81	8.91E-03
LILRB4	-2.65	2.34E-04	-1.47	5.00E-03
MDFIC	-2.74	1.65E-03	-3.09	9.25E-04

Appendix III

Proteins significantly changed more than 2-fold in lysates of GALC KO1 and KO2 versus EV1 control THP-1 cells, quantified by whole cell proteomics (Benjamini-Hochberg adjusted q-value < 0.05)

Gene ID	Log_2 (fold change KO1/EV1)	q-value	Log_2 (fold change KO2/EV1)	q-value
SLC7A11	2.93	5.75E-04	2.37	1.14E-03
HSD11B1	2.86	5.37E-03	4.57	1.46E-03
MSRB3	2.48	4.42E-04	3.02	2.10E-04
MARCKS	2.01	1.04E-03	3.75	2.10E-04
OAS3	2.01	5.42E-04	3.30	2.04E-04
MNDA	2.00	1.15E-03	2.52	6.51E-04
MYO6	1.88	3.98E-04	1.20	1.09E-03
H1F0	1.63	1.50E-03	1.40	2.49E-03
OCLN	1.55	1.17E-02	1.91	6.57E-03
CKAP4	1.46	4.58E-04	1.57	3.15E-04
AKR1C2	1.41	2.78E-03	1.62	2.13E-03
ARHGAP31	1.30	2.77E-03	1.20	6.42E-03
MT2A	1.27	4.37E-03	1.39	4.39E-03
SASH3	1.24	2.11E-03	1.19	2.18E-03
PLS1	1.24	3.98E-04	1.43	2.48E-04
RAB20	1.23	3.51E-03	1.04	5.27E-03
RNASE2	1.22	3.51E-03	1.53	1.46E-03
AP3S2	1.16	2.13E-03	1.11	2.55E-03
TNIK	1.16	1.12E-03	1.10	1.42E-03
ISG20	1.11	4.11E-03	1.84	1.03E-03
MCEMP1	1.09	6.33E-03	1.07	8.37E-03
HSPA4L	1.05	9.49E-04	1.10	8.16E-04
SLC9A7	-1.37	1.92E-02	-1.35	1.42E-02
FHL1	-1.53	2.64E-03	-1.01	1.38E-02
GALC	-2.02	5.37E-03	-2.78	1.46E-03

Appendix IV

Differentially expressed genes in GALC KO1 and KO2 versus EV1 control THP-1 cells quantified by RNA sequencing (> 2-fold change, False discovery rate adjusted q-value < 0.05)

Gene ID	Log_2 (fold change KO1/EV1)	q-value	Log_2 (fold change KO2/EV1)	q-value
MSRB3	2.98	1.01E-13	4.14	6.33E-14
OAS3	2.67	1.01E-13	4.38	6.33E-14
IFI6	2.07	1.01E-13	4.20	6.33E-14
H2BU1	1.98	1.01E-13	1.63	6.33E-14
SESN3	1.98	3.07E-11	2.51	1.88E-11
OAS2	1.86	1.01E-13	3.90	6.33E-14
ENC1	1.78	5.25E-08	1.74	1.46E-04
SLC7A11	1.74	1.01E-13	2.01	6.33E-14
CACNA1E	1.74	4.37E-12	1.97	3.20E-10
MAP4K1	1.73	2.18E-05	1.60	1.40E-02
CKAP4	1.72	1.01E-13	1.91	6.33E-14
CX3CR1	1.69	3.76E-11	1.40	1.80E-04
NLRC3	1.68	1.15E-07	1.62	4.38E-04
TRHDE	1.59	4.65E-03	2.67	4.36E-03
OAS1	1.56	3.85E-06	3.01	6.33E-14
C3orf80	1.56	1.01E-13	1.31	1.94E-07
SDK1	1.54	2.62E-05	1.38	1.78E-02
CTSG	1.54	2.00E-06	1.88	6.74E-06
H2AW	1.52	1.01E-13	1.12	7.54E-07
NEGR1	1.49	2.94E-03	1.61	4.30E-02
SYNE1	1.45	2.10E-04	1.32	4.07E-02
H3-2	1.44	1.01E-13	1.50	2.31E-11
MYO6	1.44	1.88E-05	1.15	3.59E-02
DGKH	1.42	7.89E-12	1.43	2.46E-08
TSPAN5	1.41	6.30E-04	1.80	3.41E-03
TRIM22	1.40	4.93E-05	2.58	3.45E-11
ABCA1	1.40	1.60E-10	1.40	1.60E-06
IRS2	1.38	4.31E-05	1.39	3.76E-03
PDGFD	1.34	3.14E-04	1.19	4.47E-02
TSC22D1	1.33	1.81E-05	1.34	1.78E-03
CMPK2	1.31	1.77E-05	3.48	6.33E-14
MNDA	1.31	5.18E-04	2.69	1.42E-10
ISG15	1.31	3.95E-06	3.18	6.33E-14
ANPEP	1.30	1.13E-11	1.22	2.19E-06
NPTXR	1.27	5.82E-04	1.47	3.38E-03

Gene ID	Log_2 (fold change KO1/EV1)	q-value	Log_2 (fold change KO2/EV1)	q-value
IFIT5	1.23	3.51E-04	2.43	1.74E-11
HOXA13	1.21	1.65E-04	1.32	2.02E-03
JAG1	1.20	2.32E-03	1.28	3.55E-02
NTNG2	1.20	5.00E-03	1.64	9.03E-03
GBP5	1.19	4.34E-03	1.70	3.54E-03
HELZ2	1.19	5.82E-08	2.33	6.33E-14
IFIT3	1.18	2.60E-05	2.90	6.33E-14
IRF9	1.16	8.53E-06	1.61	2.30E-08
SASH3	1.11	4.58E-06	1.18	1.12E-04
IFIT1	1.11	2.30E-03	3.20	6.33E-14
LACC1	1.10	3.06E-03	1.11	4.27E-02
MT2A	1.09	5.10E-07	1.48	6.09E-10
DUSP16	1.09	6.47E-03	1.24	3.24E-02
HSD11B1	1.09	1.36E-02	2.56	9.83E-05
IFITM1	1.09	4.59E-02	3.01	4.75E-03
DDX60L	1.06	1.77E-11	1.70	6.33E-14
PARP14	1.06	2.98E-09	2.21	6.33E-14
PARP9	1.05	8.45E-05	2.25	6.33E-14
KIAA0408	1.05	7.93E-03	1.43	4.64E-03
CYP1B1	1.03	2.91E-08	1.10	4.90E-07
JCAD	1.02	6.32E-03	1.29	6.25E-03
TCF23	1.01	7.63E-05	1.76	2.45E-11
NFATC2	1.00	3.97E-03	1.35	1.41E-03
DDX60	1.00	1.31E-04	1.97	3.44E-13
TENT5A	1.00	5.37E-04	1.79	2.62E-09
ARRDC3	-1.02	7.85E-07	-1.65	1.11E-10
DLL1	-1.21	1.38E-03	-1.22	2.03E-02
SLC9A7	-1.25	1.88E-04	-1.88	3.64E-05
KCNQ3	-1.34	8.86E-06	-2.09	2.15E-07
RASA1	-1.42	1.01E-13	-1.49	1.20E-13
AFAP1	-1.51	3.74E-07	-2.04	1.40E-07
MDFIC	-2.03	7.14E-07	-2.77	1.50E-05

Appendix V

Proteins significantly changed more than 2-fold in lysates of Cas9 versus WT THP-1 cells, quantified by whole cell proteomics (Benjamini-Hochberg adjusted q-value < 0.05)

Gene ID	Log_2 (fold change Cas9/WT)	q-value
MAGEA4	3.00	5.50E-11
RPS19BP1	2.89	1.06E-09
SCD	2.53	6.36E-09
C19orf53	2.49	4.93E-08
CCND3	2.46	6.35E-06
CNTNAP4	2.46	2.70E-07
BYSL	2.41	2.02E-09
TFEB	2.35	5.50E-11
MAP1B	2.34	1.89E-07
RRP36	2.29	1.91E-09
HIF1AN	2.19	1.93E-09
PACSIN1	2.17	2.90E-07
RPL22L1	2.17	2.03E-07
ASNS	2.14	9.14E-06
IGF2BP3	2.10	3.17E-09
GMPR	2.10	9.48E-08
PHGDH	2.08	4.70E-06
C16orf72	2.07	7.45E-08
ARRB2	2.06	2.02E-09
BCAT1	1.97	1.06E-09
AMOT	1.97	1.10E-09
OARD1	1.94	4.20E-10
DPYSL3	1.84	4.78E-08
FAM49A	1.82	1.23E-08
FOXP4	1.80	1.96E-08
MEF2C	1.76	2.18E-06
KCNK5	1.75	3.52E-06
BARX1	1.74	1.21E-06
SAYSD1	1.67	1.21E-06
YBX3	1.66	3.68E-09
CHPT1	1.65	3.50E-06
CFAP97	1.64	1.66E-05
ENKD1	1.62	3.47E-06
LPO	1.59	1.57E-05

Gene ID	Log_2 (fold change Cas9/WT)	q-value
FAM234B	1.58	1.06E-09
NOP53	1.54	2.98E-07
SLC7A5	1.53	2.77E-07
PHKA1	1.52	1.25E-07
SENP6	1.50	5.36E-09
SLITRK5	1.49	1.54E-07
SCD5	1.48	2.58E-07
SLC38A1	1.47	2.60E-07
NPM3	1.47	5.00E-06
GLUL	1.46	5.24E-07
PLA2G4A	1.46	6.09E-08
HK2	1.44	2.02E-09
NOB1	1.43	2.44E-08
C1QBP	1.42	1.11E-07
GSTM4	1.42	1.49E-07
RAB33A	1.41	1.93E-06
CDCA7L	1.40	2.91E-07
STX10	1.40	5.21E-08
REEP5	1.39	7.63E-08
GOT1	1.38	4.96E-07
MATK	1.38	3.47E-06
NDUFB8	1.38	1.16E-05
PHF23	1.36	2.68E-08
PSRC1	1.35	4.43E-06
FAIM	1.34	3.27E-07
MPHOSPH10	1.33	1.54E-08
RRP1	1.32	6.72E-09
FAM32A	1.32	1.64E-06
MAPT	1.30	4.88E-09
MSI2	1.30	1.37E-08
AKAP10	1.30	4.08E-08
SLF2	1.29	1.26E-09
ALKBH2	1.29	2.48E-06
COA8	1.29	1.88E-07

Gene ID	Log_2 (fold change Cas9/WT)	q-value
LTV1	1.27	3.43E-09
PCLAF	1.27	1.40E-04
ZNF280A	1.27	1.45E-07
RCN3	1.27	1.48E-08
LCOR	1.26	4.17E-07
GCSH	1.25	1.41E-07
CD109	1.25	7.93E-08
KCTD15	1.25	6.21E-07
FAM204A	1.24	5.48E-05
CCDC7	1.24	8.41E-06
CCDC137	1.23	9.92E-09
DESI2	1.23	2.27E-03
NUP42	1.21	5.19E-07
CEP72	1.21	3.88E-08
BIN2	1.20	1.57E-07
PSAT1	1.20	3.40E-05
NOP2	1.19	1.75E-09
UTP11	1.19	8.16E-09
PDLIM1	1.18	8.06E-08
PHB	1.18	1.01E-08
TENM3	1.17	2.12E-07
SLC16A1	1.16	3.68E-06
PRKAR1B	1.16	3.14E-07
PHB2	1.14	2.91E-07
CCDC86	1.14	1.61E-08
OTX1	1.14	9.29E-05
PTPRN2	1.14	1.68E-06
MYO10	1.13	1.20E-07
ALAS1	1.13	3.87E-07
HENMT1	1.13	5.29E-06
CCNB1	1.12	6.30E-06
OSBPL10	1.12	7.22E-05
NR2F1	1.12	3.11E-06
STAR	1.12	2.26E-05
AEBP2	1.11	2.50E-06
CWF19L1	1.11	2.12E-08
SDHAF1	1.11	6.74E-04
TPI1	1.11	8.93E-09
CAMKK1	1.11	1.22E-07

Gene ID	Log_2 (fold change Cas9/WT)	q-value
RAD18	1.11	2.68E-08
ZFAND5	1.10	7.37E-04
MLXIPL	1.10	5.47E-05
PCMTD2	1.10	3.01E-07
CD320	1.10	8.43E-06
WWP1	1.10	3.08E-03
ZNF697	1.10	2.14E-07
HMGN5	1.10	9.00E-08
NQO1	1.09	6.10E-07
REXO2	1.09	9.92E-06
CD55	1.09	2.26E-08
ADAT2	1.09	9.57E-06
SDHAF2	1.09	2.27E-05
ASS1	1.08	4.61E-04
RRM2	1.08	1.41E-04
AGPAT4	1.08	1.50E-05
SYNPO	1.07	8.14E-07
MTHFD1L	1.07	2.54E-08
UBE2S	1.07	2.69E-05
BBX	1.07	1.01E-06
NAP1L5	1.07	1.21E-06
CLCC1	1.06	1.10E-09
HPDL	1.06	3.66E-08
CEBPG	1.05	2.50E-04
PTMS	1.05	9.93E-04
MARCKS	1.05	2.58E-07
C14orf93	1.04	3.48E-03
NFATC2IP	1.04	1.83E-06
LBR	1.04	2.99E-08
ICE1	1.04	1.94E-07
CDK4	1.04	1.18E-06
MTRR	1.04	9.23E-08
ARHGAP11A	1.03	1.68E-06
STRAP	1.03	1.68E-08
GPT2	1.03	7.55E-04
GALK2	1.03	1.34E-08
COX15	1.03	6.98E-07
GSTM1	1.02	5.83E-07
C1orf131	1.02	2.42E-07

Gene ID	Log_2 (fold change Cas9/WT)	q-value
PARVB	1.02	1.44E-05
TSR3	1.02	1.44E-05
TRAF3IP3	1.02	4.58E-06
C7orf50	1.02	2.32E-08
SCO1	1.02	7.18E-08
LARP4	1.02	1.53E-05
CTTN	1.02	1.76E-05
FOXM1	1.01	9.93E-06
PLEKHF1	1.01	5.62E-07
C12orf45	1.01	3.35E-05
XPO5	1.00	1.64E-08
SFT2D3	-1.00	2.10E-07
MFSD1	-1.00	4.95E-07
RAB2A	-1.01	2.12E-08
FAM174B	-1.01	1.07E-05
CTBS	-1.01	1.80E-06
PACS1	-1.01	2.46E-07
MYO6	-1.01	1.21E-08
DENND4B	-1.01	4.08E-08
HGSNAT	-1.01	3.11E-07
MAP3K20	-1.02	1.02E-07
PTK2B	-1.02	1.38E-06
PHLPP1	-1.02	2.68E-08
LAMC1	-1.02	1.25E-05
RAB3IL1	-1.02	1.78E-06
ITGB2	-1.02	3.82E-08
ZBTB34	-1.02	1.48E-06
POMP	-1.02	4.32E-03
NAGA	-1.03	6.26E-09
SYNJ2BP	-1.03	8.59E-07
APLP2	-1.03	7.81E-08
IGSF10	-1.03	7.16E-04
SRSF11	-1.03	3.91E-06
GNAQ	-1.04	6.13E-09
GAMT	-1.04	3.24E-05
CHCHD7	-1.04	1.71E-04
TALDO1	-1.05	8.87E-08
DOCK1	-1.05	2.54E-06
CEBPD	-1.05	2.21E-04

Gene ID	Log_2 (fold change Cas9/WT)	q-value
PBXIP1	-1.05	2.10E-07
CTSV	-1.05	3.74E-08
VAT1L	-1.06	3.19E-05
GSTM3	-1.06	6.21E-08
MAN2B1	-1.07	4.93E-08
WDFY2	-1.08	4.51E-08
NIPA2	-1.08	6.45E-05
PTPRJ	-1.08	3.92E-08
EPB41L3	-1.09	3.25E-07
PEX11B	-1.09	3.90E-06
DDAH2	-1.10	3.43E-09
TTC28	-1.10	9.01E-04
DACH1	-1.10	9.50E-08
P2RY2	-1.10	4.28E-04
ALYREF	-1.11	8.38E-07
SLC22A18	-1.11	6.03E-07
FKBP9	-1.11	7.01E-09
PLD4	-1.11	1.48E-07
OSTC	-1.11	3.96E-05
PCYOX1L	-1.11	3.60E-08
ARC	-1.11	2.86E-02
BASP1	-1.11	3.26E-07
PARVA	-1.12	2.46E-07
AGK	-1.12	1.54E-09
MYO1G	-1.12	3.48E-04
TNFAIP8	-1.12	2.33E-06
TRMT2A	-1.13	1.41E-08
ORMDL1	-1.13	9.38E-05
NUDT18	-1.13	1.10E-06
UBA7	-1.14	1.08E-05
KIAA1755	-1.14	1.33E-04
FHL1	-1.15	1.52E-05
FGD2	-1.15	1.80E-05
GNPTG	-1.15	1.69E-06
MCOLN2	-1.15	1.02E-06
SIPA1	-1.16	2.41E-08
CSTB	-1.16	1.97E-07
CRYBG1	-1.16	7.60E-04
SH2D3C	-1.17	1.41E-05

Gene ID	Log_2 (fold change Cas9/WT)	q-value
RCSD1	-1.17	2.27E-06
CD84	-1.17	6.09E-06
SYK	-1.17	1.70E-09
TDRD7	-1.17	3.56E-07
ANXA2	-1.18	2.51E-05
SSH3	-1.18	3.54E-08
SAMHD1	-1.18	8.15E-06
MANBA	-1.19	1.02E-07
S100A6	-1.19	1.12E-03
ALDH6A1	-1.19	2.70E-07
GPR158	-1.19	1.15E-05
IGFBP7	-1.19	1.06E-07
AIF1	-1.19	8.89E-05
SLC17A5	-1.20	3.59E-06
SLC25A1	-1.20	9.25E-09
ARHGEF10L	-1.20	9.58E-07
LGALS3	-1.20	2.45E-08
PGD	-1.20	3.85E-07
NCOA4	-1.21	5.24E-04
MICAL1	-1.21	1.79E-07
SERPINB1	-1.21	1.36E-09
SIRPB1	-1.22	1.14E-06
AHNAK	-1.22	1.16E-05
PPT1	-1.23	1.77E-08
ARHGAP26	-1.24	2.82E-06
ECM1	-1.24	5.22E-06
SLC27A3	-1.24	3.29E-07
VPS35L	-1.24	6.03E-09
CYFIP2	-1.25	5.70E-06
ARSD	-1.25	4.63E-08
CAPZA2	-1.26	3.77E-08
DDT	-1.26	2.68E-07
NUDT7	-1.26	3.66E-08
CCNE2	-1.27	1.55E-03
PLBD1	-1.27	1.03E-04
SDSL	-1.27	9.08E-05
HDAC9	-1.27	7.51E-07
IRS2	-1.28	3.16E-06
NCSTN	-1.29	3.68E-09

Gene ID	Log_2 (fold change Cas9/WT)	q-value
OTULINL	-1.29	2.96E-07
A2M	-1.29	1.98E-05
TRIM32	-1.30	2.34E-08
PUDP	-1.30	3.16E-08
SIRT7	-1.31	2.23E-08
EDEM1	-1.31	3.97E-07
DPP7	-1.31	4.36E-08
ANXA1	-1.31	3.82E-08
ROBO1	-1.32	1.09E-06
CD82	-1.32	3.87E-07
CCNYL1	-1.32	8.38E-07
RNH1	-1.33	1.01E-07
EIF4A2	-1.33	1.12E-07
NME3	-1.33	5.36E-09
FMNL1	-1.33	1.24E-05
COMT	-1.35	6.13E-09
TPM2	-1.36	4.68E-07
HSP90AA4P	-1.36	4.12E-05
HSD17B8	-1.36	4.18E-05
ARHGAP9	-1.36	2.26E-08
MRC2	-1.36	3.68E-09
PSTPIP2	-1.37	1.96E-08
RUNX2	-1.37	4.95E-04
BLVRA	-1.37	2.68E-08
IRF7	-1.38	2.59E-03
TMEM59	-1.38	4.50E-06
CPM	-1.38	5.70E-08
C8orf82	-1.39	2.33E-02
NAXE	-1.39	2.02E-09
GCA	-1.41	5.79E-05
SGK3	-1.41	7.99E-09
LAPTM5	-1.41	2.08E-04
GATM	-1.41	9.35E-08
RUBCNL	-1.41	6.55E-07
FBXO6	-1.42	6.21E-08
TGFBI	-1.42	4.18E-09
ERMP1	-1.43	8.60E-09
ARSB	-1.44	6.22E-09
MELTF	-1.45	1.66E-07

Gene ID	Log_2 (fold change Cas9/WT)	q-value
SDCBP	-1.45	4.49E-09
CAP1	-1.46	3.46E-09
CAMK2D	-1.46	4.20E-10
HEXB	-1.47	1.04E-08
ENPP4	-1.47	5.99E-09
CD1D	-1.47	8.39E-05
SRRM1	-1.49	4.88E-09
GAA	-1.49	1.65E-08
ARHGAP1	-1.50	2.12E-09
CHKB	-1.50	3.46E-09
KCTD12	-1.51	3.17E-07
LRRC8C	-1.51	3.01E-09
NOMO1	-1.51	1.46E-06
ITGAX	-1.52	2.14E-05
ZNF408	-1.52	5.36E-09
STK32C	-1.55	2.33E-04
SERPINB10	-1.57	5.37E-07
EMB	-1.57	7.65E-09
SHARPIN	-1.58	4.77E-08
TNFAIP8L2	-1.60	1.32E-07
TLR2	-1.60	4.88E-09
ERLIN2	-1.61	5.82E-10
F13A1	-1.62	8.93E-09
NAPRT	-1.63	4.28E-10
TBXAS1	-1.63	1.21E-06
HCK	-1.64	4.66E-06
RET	-1.64	1.05E-08
NCF2	-1.65	1.60E-07
GPX1	-1.65	3.21E-07
FTL	-1.67	3.81E-07
GRN	-1.67	6.13E-09
AZU1	-1.67	8.67E-07
HSD11B1	-1.68	2.93E-05
HBE1	-1.68	1.17E-02
PRTN3	-1.70	1.34E-05
OCLN	-1.70	3.91E-09
LILRB4	-1.70	7.09E-07
CTNNB1	-1.71	1.69E-07
EMILIN2	-1.71	1.41E-08
GLUD2	-1.71	6.94E-08
DDAH1	-1.72	5.36E-09
PRSS12	-1.72	1.48E-08
GLO1	-1.72	3.83E-09
THEMIS2	-1.73	2.24E-08
TMOD2	-1.74	3.32E-03
CBR4	-1.75	2.89E-08
CD70	-1.75	7.41E-07
MNDA	-1.76	4.05E-08
NDRG2	-1.76	2.16E-04
CHID1	-1.77	4.20E-10
S100A4	-1.77	6.44E-06
FRY	-1.78	3.44E-09
RNASE2	-1.80	1.16E-06
TYMP	-1.81	1.56E-07
MVP	-1.81	1.52E-06
RBKS	-1.82	1.20E-08
PECAM1	-1.82	1.08E-08
IGF2BP2	-1.84	5.50E-11
TSPO	-1.85	1.03E-05
TPP1	-1.85	2.45E-08
PALM	-1.85	1.26E-08
WDFY1	-1.85	3.91E-09
TRIM21	-1.88	2.63E-08
CNN3	-1.88	2.03E-07
ALDH3B1	-1.91	5.36E-09
IFI16	-1.94	8.92E-06
FILIP1L	-1.95	6.03E-09
HYAL3	-1.96	3.25E-06
ACP3	-1.98	2.77E-07
CA2	-2.00	2.27E-08
RIPK3	-2.00	1.78E-07
EGFL7	-2.01	3.45E-09
NCAM2	-2.02	5.36E-09
COL10A1	-2.04	2.32E-06
GRAMD4	-2.06	2.02E-09
SDC2	-2.08	1.06E-09
LGALS1	-2.10	4.93E-08
BST1	-2.10	1.93E-09

Gene ID	Log_2 (fold change Cas9/WT)	q-value
CTNND1	-2.13	1.42E-08
LILRB2	-2.15	2.93E-06
RASSF4	-2.19	1.68E-08
NAIP	-2.22	6.44E-07
AKR1C2	-2.23	1.50E-05
DSG2	-2.27	4.93E-08
SLC44A1	-2.34	4.20E-10
DOK3	-2.36	1.54E-08
SERPINB8	-2.39	1.34E-08
FBP1	-2.40	9.05E-10
CD68	-2.40	2.79E-07
CD38	-2.47	1.03E-09
PRXL2A	-2.49	2.98E-07
NMNAT3	-2.54	5.84E-08
CLEC4M	-2.58	4.11E-06
TSPYL5	-2.61	7.17E-09
ALDH2	-2.62	8.16E-09
COL9A2	-2.65	1.64E-08
CKB	-2.67	1.71E-08
TES	-2.70	7.63E-08

Gene ID	Log_2 (fold change Cas9/WT)	q-value
FCER1G	-2.70	9.50E-08
CEACAM6	-2.78	7.62E-09
RENBP	-2.83	6.26E-09
DMRTA2	-2.85	2.70E-08
NCAM1	-2.93	6.53E-08
EPHX2	-2.99	1.06E-09
ITGAL	-3.02	1.88E-08
KYNU	-3.06	2.98E-07
HK3	-3.10	2.42E-07
VIM	-3.33	2.58E-07
S100A9	-3.42	1.39E-06
CAV1	-3.45	1.22E-07
ANXA6	-3.60	2.41E-08
S100A8	-3.61	2.11E-07
CGREF1	-3.79	1.83E-07
CRABP1	-3.86	1.13E-06
RPS4Y1	-4.28	3.88E-07

Appendix VI

Proteins significantly changed more than 1.4-fold in lysates of GALC KO1 and KO2 versus mixed population EV control (EVm) THP-1 cells, quantified by whole cell proteomics (Benjamini-Hochberg adjusted q-value < 0.05)

Gene ID	Log_2 (fold change KO1/EVm)	q-value	Log_2 (fold change KO2/EVm)	q-value
MSRB3	1.17	6.59E-03	1.11	3.63E-03
ANPEP	1.02	1.46E-05	0.33	1.21E-02
H1-0	0.97	1.07E-02	0.67	3.29E-02
RASSF4	0.80	6.59E-03	0.70	5.85E-03
MNDA	0.75	5.00E-03	0.62	6.17E-03
NDST2	0.70	5.48E-03	0.38	4.29E-02
CKAP4	0.61	2.19E-03	0.60	1.29E-03
OCLN	0.55	4.38E-03	0.49	3.91E-03
SASH3	0.51	7.65E-03	0.30	4.94E-02
HDAC9	-0.51	3.01E-02	-0.75	2.50E-03
NUDT12	-0.51	1.23E-03	-0.23	3.23E-02
PRKACB	-0.54	2.35E-04	-0.33	2.50E-03
ISOC1	-0.54	2.78E-04	-0.21	3.13E-02
PAIP2	-0.57	3.59E-02	-0.51	2.85E-02
TSPYL5	-0.60	2.77E-02	-0.58	1.91E-02
DDAH1	-0.66	1.37E-03	-0.69	7.75E-04
SMN1	-0.74	3.40E-02	-0.74	1.58E-02
SYNPO2	-0.79	1.16E-02	-0.60	2.07E-02
ENPP4	-0.94	4.43E-05	-0.54	1.29E-03
GALC	-1.03	9.88E-03	-0.75	2.01E-02