Structural Analyses of Carbohydrate Receptors for Enterotoxins and Adhesins of Enterotoxigenic Escherichia Coli

Department of Medical Biochemistry and Cell Biology

Institute of Biomedicine

Sahlgrenska Academy, University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2023

Cover illustration: Down the pore by Dani Zalem.

dani.zalem@gu.se

ISBN 978-91-8069-533-6 (PRINT)
ISBN 978-91-8069-534-3 (PDF)

Printed in Borås, Sweden 2023
Printed by Stema Specialtryck AB

To my parents.

Structural Analyses of Carbohydrate Receptors for Enterotoxins and Adhesins of Enterotoxigenic Escherichia Coli

Dani Zalem

Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine
Sahlgrenska Academy, University of Gothenburg
Gothenburg, Sweden

ABSTRACT

Carbohydrate-binding proteins expressed by microbes are key determinants in initiating and sustaining infections that account for millions of deaths each year. This thesis focused on proteins integral to infections instigated by enterotoxigenic *Escherichia coli* (ETEC); estimated as the largest bacterial cause of diarrhea in the world with hundreds of millions of cases each year. ETEC infections are mediated by two primary carbohydrate-binding proteins; 1) Colonization factors (CF), which facilitate host cell attachment, and 2) Enterotoxins, which penetrate host cells to induce a potentially lethal diarrheal response. By employing biochemical techniques, such as chromatogram binding assays, mass spectrometry and NMR, we dissected the precise mechanisms fundamental for the interactions of ETEC carbohydrate-binding proteins.

In the presented papers, the novel colonization factor CS30, and the enterotoxins LT-IIb, and LT-IIc where investigated. Our findings identified the sulfatide glycosphingolipid as the principal receptor for CS30 and emphasized the significance of the carbohydrate-presenting lipid moiety in binding. The diarrhea-inducing toxins, LT-IIb and LT-IIc, demonstrated distinct binding specificity to sialic acid presenting glycosphingolipids, and the presence of such receptors were confirmed in the human intestine. Lastly, structural studies detailed the atomic framework of these binding interactions and quantified the binding affinities.

By revealing the specific carbohydrate interactions underpinning both adhesion and toxin action, our study uncovers the intricate processes governing pathogenic infection mechanisms, which may inform the design of next-generation anti-bacterial therapeutics, vaccines and diagnostical tools.

Keywords: Carbohydrate recognition, glycosphingolipid, diarrheagenic *E. coli*, ETEC, heat-labile enterotoxin, colonization factor

ISBN 978-91-8069-533-6 (PRINT)
ISBN 978-91-8069-534-3 (PDF)

SAMMANFATTNING PÅ SVENSKA

Varje år drabbas hundratals miljoner människor och tiotusentals mister livet på grund av enterotoxinbildande *Escherichia coli* (ETEC). Infektionen etableras och upprätthålls med två typer av proteiner hos bakterien; *kolonisationsfaktorer* som medger inbindning av bakterier till tunntarmens celler, och *toxiner* som tar sig in i tarmcellerna och påbörjar diarréreaktionen.

Syftet med denna avhandling var att identifiera strukturer på våra celler som kolonisationsfaktorer och toxiner binder för att eventuellt orsaka infektion, samt kartlägga de molekylära förutsättningar som medger inbindning – allt med slutmålet att bana väg för framtida behandlingsstrategier. ETEC är nämligen en komplex bakterie som kan producera en uppsjö av olika kolonisationsfaktorer och toxiner, vilket sammantaget har hindrat vaccinutvecklingen.

I tre delarbeten så riktar vi vårt fokus mot ETEC:s toxiner och undersöker de senast upptäckta toxinerna LT-IIb och LT-IIc. Här upptäckte vi att ETEC:s toxiner binder till en viss form av kolhydratstrukturer som heter glykosfingolipider. Med masspektrometri lyckades vi fastställa att sådana glykosfingolipider uttrycks i människans tunntarm.

Vidare så använde vi oss av röntgenkristallografi för att i detalj avgöra toxinernas tredimensionella struktur tillsammans med kolhydraten. Vi upptäckte till vår förvåning att LT-IIb uppvisar totalt tio bindningssäten, medan LT-IIc enbart har fem. Kärnmagnetisk resonans hjälpte oss därefter att ytterligare kartlägga bindningsinteraktionen för LT-IIb. Dessa experiment gav stöd för bindningssätenas lokalisation, fastställde vilka aminosyror som ansvarar för kolhydratbindningen, samt uppskattade kvantitativt bindningsstyrkan.

Den nyupptäckta kolonisationsfaktorn CS30 studerades i ett delarbete. Vi visar här att ETEC som uttrycker CS30 binder till sulfatid, ännu en glykosfingolipid som finns på både människans och grisens tunntarmsceller. Frågan väcktes således huruvida ETEC-stammar kan infektera både människa och gris – något som inte har påvisats tidigare.

Sammantaget så klarlades många för ETEC grundläggande bindningsfenomen som kan användas för vidare forskning, vilket kan leda till nya terapeutiska och diagnostiska strategier.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. **Zalem D**, Ribeiro JP, Varrot A, Lebens M, Imberty A, Teneberg S. *Biochemical and structural characterization of the novel sialic acid-binding site of Escherichia coli heat-labile enterotoxin LT-IIb.*

Biochemical Journal. 2016 Nov 1;473(21):3923-3936.

II. Von Mentzer A, **Zalem D**, Chrienova Z, Teneberg S. *Colonization factor CS30 from enterotoxigenic Escherichia coli binds to sulfatide in human and porcine small intestine.*

Virulence. 2020 Dec;11(1):381-390.

III. **Zalem D**, Juhás M, Terrinoni M, King-Lyons N, Lebens M, Varrot A, Connell TD, Teneberg S. *Characterization of the ganglioside recognition profile of Escherichia coli heat-labile enterotoxin LT-IIc.*

Glycobiology. 2022 Apr 21;32(5):391-403.

IV. **Zalem D**, Lebens M, Teneberg S, Karlsson G. *Structural and dynamic studies of Escherichia coli heat-labile enterotoxin LT-IIb B-subunits by NMR spectroscopy.*

Manuscript.

CONTENTS

ABBREVIATIONS

CBA	Chromatogram binding assay
CF	Colonization factor
CFTR	Cystic fibrosis transmembrane regulator
CFU	Colony-forming units
CT	Cholera toxin
CSP	Chemical shift perturbation
CTB	Cholera toxin B-subunit
DEAE	Diethylaminoethyl
EGCase	Endoglycoceraminidase
EI MS	Electron impact mass spectrometry
ESI	Electrospray ionization mass spectrometry
ESI-MS/MS	Electrospray ionization tandem mass spectrometry
ETEC	Enterotoxigenic *E. coli*
ER	Endoplasmic reticulum
GAG	Glycosaminoglycan
GBP	Glycan-binding protein
cGMP	Cyclic GMP
GPI	Glycosylphosphatidylinositol
GSL	Glycosphingolipid

LC-ESI-MS	Liquid chromatography electrospray ionization mass spectrometry
LT	Heat-labile enterotoxin
LT-I	Type I heat-labile enterotoxin
LT-II	Type II heat-labile enterotoxin
MS	Mass spectrometry
NMR	Nuclear magnetic resonance spectroscopy
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
ST	Heat-stable enterotoxin
TLC	Thin-layer chromatography

1 INTRODUCTION

1.1 GENERAL OVERVIEW

The complex relationship of microorganisms and hosts is a balance between symbiotic harmony and parasitic disruption. On one end of the spectrum, microorganisms coexist with their host, establishing an environment where both entities thrive. On the opposite end, the interaction causes disease. Central to these interactions is a molecular interface where both entities meet and engage. In this thesis, we delve into the complexities of this interface from the perspective of enterotoxigenic *Escherichia coli* (ETEC).

Adherence of microorganisms to their host is a necessary precursor for colonization and subsequent infection. Microorganisms in this context are commonly referred to as pathogens. In the bacterial realm, adherence is conferred by surface polysaccharide or polypeptide structures, such as fimbriae, curli or non-fimbriae adhesins [1]. By anchoring themselves to specific host cell receptors, bacteria can get in proximity of the host to effectively proliferate and deliver toxic products. Indeed, the host receptor specificity of these adherence molecules determine which tissue of which host the pathogen will colonize – a phenomenon known as tropism. These host receptors are, more often than not, carbohydrate structures (glycans) presented on the cells of infected tissues [2-4].

Pathogenic capabilities to cause infection are known as virulence factors, and what usually follows in the process leading to infection, is the disruption of target cell functions. This can be by either direct cell invasion or toxin secretion, the latter of which is the case for ETEC. After adherence, ETEC produces enterotoxins that binds glycans in the small intestine, which causes the severe diarrhea synonymous with the disease [5, 6].

This thesis aims to elucidate the intricate interactions between ETEC's virulence factors and host glycan receptors, specifically glycosphingolipids. By dissecting the molecular and cellular determinants governing bacterial adherence, we may come closer to understanding the mechanisms that dictate the course of bacterial infections.

1.2 GLYCOCONJUGATES

Present on all cells, glycoconjugates are complex macromolecules composed of carbohydrates covalently bonded to non-carbohydrate entities such as proteins and lipids (figure 1). These biomolecules play a multitude of roles in multicellular organisms, impacting processes ranging from cellular recognition and signaling to endothelial homeostasis and immune responses [3, 7-9].

The building blocks of the glycan moiety are monosaccharide units, usually consisting of five or six carbons as D-enantiomers (except for L-fucose) of either aldoses or ketoses [10]. These units are connected *via* glycosidic bonds. A glycosidic bond can occur on a number of different atomical positions on each monosaccharide and is configurated in either an α- or β-form. In addition, each monosaccharide may sustain more than one glycosidic linkage, resulting in branching glycans, causing the overall possible complexity of polysaccharides to far exceed other cellular macromolecules, such as nucleotides or amino acids [8]. However, only a handful of monosaccharide species dominate the glycosylation of vertebrates, as represented in figure 1 [11].

The glycosidic linkage is catalyzed by glycosyltransferases, and their synthesis is therefore not directly regulated by gene expression [11, 12]. Furthermore, the availability of sugar donors and acceptors, competition among precursors, and the spatial distribution of glycosyltransferases also affect the total glycan expression of a cell (= glycome).

Eukaryotic cells are surrounded by a protective glycoconjugate layer, that may or may not be anchored to the cell membrane, known as the glycocalyx, which microbes must navigate to establish infection [13-15]. In addition to binding glycoconjugates during colonization, some microbes may adorn themselves with host-like glycans to evade immune response. For instance, the densely glycosylated *N*-linked protein, gp120, on the surface of the human immunodeficiency virus is believed to help disguise the underlying viral polypeptide [16]. Hyaluronan, an extracellular non-sulfated glycosaminoglycan present in vertebrate species, may also be expressed by some bacterial strains [17]. This not only forms a protective capsule, but may also be used to evade host immune response by molecular mimicry [18], and aid in bacterial adhesion [19].

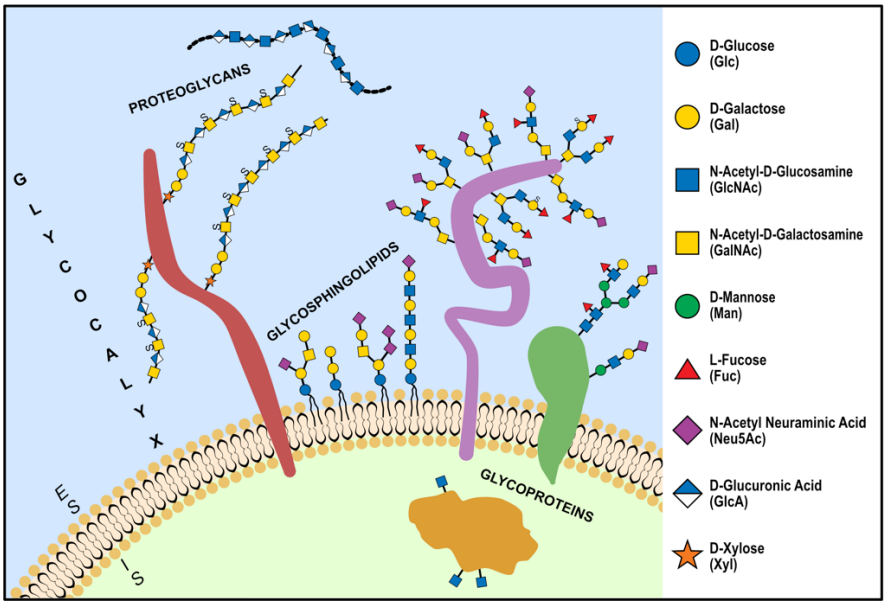

Figure 1. The glycoconjugates of the eukaryotic cell. *ES = extracellular space, IS = intracellular space.*

1.2.1 GLYCOPROTEINS

Glycoproteins play a myriad of roles vital to biological processes, including aiding in protein folding and functionality, and contributing to cell-cell and cell-matrix interactions [20]. During inflammatory response, endothelial cells express glycan-binding proteins (GBPs) known as selectins (E-, L- and P-selectins). These interact with sialyl-Lewis x (sLe^X) related glycans on white blood cell glycoproteins, such as P-selectin glycoprotein ligand-1, to aid in directing them to the inflammatory site [21, 22].

1.2.1.1 *N*-LINKED GLYCOPROTEINS

The most well-studied form of glycosylation among eukaryotes is the *N*-linked GlcNAcβ1-Asn-X-Ser/Thr sequence, in which "X" cannot be a proline. One glycoprotein may therefore carry several glycan chains, which in turn may vary in sequence and position of glycosylation [23].

Synthesis begins with the oligosaccharide formation on a lipid carrier (dolichol-phosphate) in the endoplasmic reticulum (ER). The oligosaccharide structure is then transferred to the sidechain amide of select asparagine residues, after which continued processing of the glycan occurs in the ER and Golgi apparatus [11, 20]. The *N*-linked glycoprotein follows the secretion pathway until it either ends up in the extracellular space or is incorporated into the plasma membrane.

1.2.1.2 *O*-LINKED GLYCOPROTEINS

Several forms of *O*-linked glycosylations have been described among eukaryotes. These include modifications by O-GlcNAc, an important and dynamical intracellular regulatory process akin to protein phosphorylation [24], as well as O-GalNAc, O-Fucose, O-Glucose, and O-Mannose [20]. Many of these are found on glycoproteins involved in cell signaling or as constituents of the extracellular matrix.

O-GalNAc modification, although being prevalent in many glycoproteins, has been widely studied as a constituent of mucin which carries the most numerous such modification of any glycoprotein class.

1.2.1.3 MUCINS

Mucins are large, highly *O*-glycosylated (O-GalNAc) glycoproteins that forms the mucus which is central in maintaining epithelial integrity [25] and modulating interactions with microorganisms [26]. Intestinal mucins, mainly MUC2 in the lower intestines of humans, are produced by goblet cells, where mucin proteins undergo extensive *O*-glycosylation to Ser/Thr residues in the Golgi apparatus. Following glycosylation, mucins are transported in secretory vesicles to the epithelial surface, where they are secreted and interact with water to form a protective, gel-like mucus barrier.

The small intestine forms a single, loose mucus layer which does not pose much resistance to bypassing bacteria (but makes nutrient uptake possible) [27]. Instead, the continuous mucus production, cell regeneration, intestinal peristaltic and secretion of antimicrobial peptides and IgA-antibodies helps rid the lumen from pathogens.

In contrast, the colon forms two mucus layers which creates much more of a physical barrier. The proximal, viscous mucus layer travels from the crypts of the epithelium and loosens up to form a second, outer mucus layer. The

commensal microbiota may harbor in the outer mucus layer, where they sustain on digested food or degraded mucin polysaccharides.

The protective barrier function formed by mucins is further enhanced by their ability to act as decoy molecules during infections. In-vitro cultures of gastric cells have shown that the membrane-bound MUC1 is shed from the cell surface upon interaction with *Helicobacter pylori*, thus physically removing the pathogen from the vicinity of the gastric cells [28].

1.2.1.4 GPI-ANCHORED PROTEINS

Glycosylphosphatidylinositol (GPI) -anchored proteins are a diverse group of membrane proteins that are attached to the cell surface via a glycolipid anchor. GPI-anchored proteins are involved in various biological processes, such as signal transduction, immune response modulation, and cellular adhesion, by virtue of their ability to associate with specific lipid rafts and facilitate localized signaling events [29].

GPI-anchored proteins are also involved in host-pathogen interactions. For instance, the trypanosome variant surface glycoprotein – a GPI-anchored protein – allows the parasite to evade the host immune response by creating a protective shield [30]. Furthermore, certain bacterial toxins target GPI-anchored proteins, such as aerolysin from *Aeromonas hydrophilia* [31].

1.2.2 PROTEOGLYCANS

Proteoglycans are comprised of a core protein and one or more covalently attached glycosaminoglycan (GAG) chains [32]. The GAG-chains are long, linear carbohydrate polymers that are negatively charged due to the presence of sulfate groups and uronic acids, forming a highly hydrated matrix, which resists compressive forces and provides structural support to the connective tissue.

One of the more well-studied proteoglycans is heparan sulfate, which plays an important role in blood coagulation and cell-signaling [33, 34], but also serves as a receptor for pathogens such as herpes simplex virus [35] and respiratory syncytial virus [36].

1.2.3 GLYCOSPHINGOLIPIDS

Glycolipids are present in all eukaryotic cells and in some prokaryotic [37]. These specialized molecules, found principally in the outer leaflet of cellular membranes, dictate numerous cellular interactions and are fundamental in the orchestration of various physiological processes; from providing the immune cells with a sense of self [38], to optimizing the efficiency of photosynthesis in plants [39].

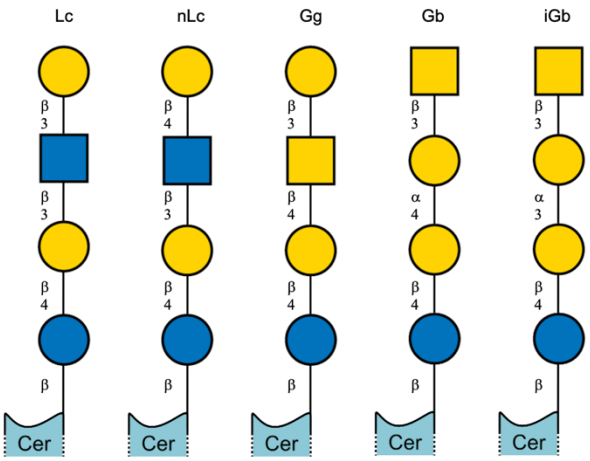

Figure 2. The five major mammalian glycosphingolipid core structures. *Lacto (Lc), neolacto (nLc), ganglio (Gg), globo (Gb) or isoglobo (Gb) core structures are linked to a ceramide (Cer) by a glycosidic bond.*

At their core, glycolipids are amphipathic molecules formed by the covalent attachment of a lipid tail to a glycan. Almost all glycolipids of vertebrates are glycosphingolipids [20], where the lipid moiety consist of a long-chain base with an attached lipid, known as ceramide (as opposed to a glycerolipid that is found in most phospholipids) [40-42]. The most common long-chain bases are sphingosine, phytosphingosine and sphinganine.

1.2.3.1 BIOSYNTHESIS AND CLASSIFICATION

The ceramide portion of a glycosphingolipid is synthesized in the endoplasmic reticulum (ER) [43], where a sphingosine is joined to a fatty acid chain by an amide bond. The fatty acid is typically between 16 to 24 carbon atoms long and may be hydroxylated.

The ceramide formation is followed by the addition of a galactose in the ER, or a glucose in the Golgi apparatus. These single-sugar compounds form the simplest of glycosphingolipids, called cerebrosides (since they were first isolated from brain tissue). Further saccharide elongation occurs by glycosyltransferases on the luminal side of the Golgi apparatus, followed by transportation to primarily the outer leaflet of the cell membrane.

The number of monosaccharide residues in a glycosphingolipid typically range from one to ten, but glycosphingolipids with more than 50 residues have been described [44]. Glycosphingolipids can be divided into several classes based on the sequence and stereochemistry of the monosaccharides closest to the ceramide (figure 2) [40]. Classification can also be made according to saccharide sequences associated with blood group determinants, such as the ABO, Lewis and P systems [45]. Furthermore, glycosphingolipids are often grouped as either acid or non-acid (or neutral). Acid glycosphingolipids are adorned with sialic acids or sulfate groups, commonly referred to as gangliosides and sulfatides, respectively.

A handful of sialic acids are known, but the most common one found in healthy human tissue is the *N*-actetylneuraminic acid (Neu5Ac). The *CMAH* gene encoding the enzyme CMP-Neu5Ac hydroxylase that converts Neu5Ac to *N*-Glycolylneuraminic acid (Neu5Gc) [46] was lost during evolution but remains in most other mammals [47].

1.2.3.2 LIPID RAFTS

Lipid rafts, conceptualized in the mid-1990s [48], are microdomains within the cellular membrane [49]. These are characterized by their distinct lipid composition, rich in cholesterol, sphingomyelins and glycosphingolipids. They may also contain certain transmembrane proteins and GPI-anchored proteins.

This order within the plasma membrane which defines lipid rafts, is seemingly transient and has made their observation challenging [50]. Nonetheless, lipid rafts have been proposed to serve as organizing centers for the assembly of signaling molecules, modulating various cellular processes including signal transduction, membrane trafficking, and cellular polarization. The lipid raft-associated ganglioside GM1 has played a pivotal role in lipid raft research, largely because of its detectability by cholera toxin B-subunits [49].

1.2.3.3 FUNCTION AND PATHOPHYSIOLOGY

Glycosphingolipids contribute to the structural integrity of membranes, impacting fluidity and facilitating lipid raft formation [49]. They are involved in transmitting extracellular signals by acting as receptors for signaling molecules [51], and play a crucial role in cell-to-cell recognition [52] and inflammatory processes [53]. Certain glycosphingolipids are associated with cell differentiation stages and may play a role in embryogenesis [54], as well as oncogenesis [55, 56]. Defects in glycosphingolipid metabolism are associated with several congenital diseases, that are heterogenous in symptomology and severity [57]. Glycosphingolipids may be targeted by the immune system, causing auto immune disorder [58], and they are also often targeted by pathogens [4], which is the focus of this thesis.

1.2.3.4 INTERACTION WITH MICROORGANISMS

Many pathogens recognize and bind specific glycosphingolipid structures with surface-presenting GBPs as a first step to colonization or infection. A well-studied such system is the P-fimbriated uropathogenic *Escherichia coli*, where the PapG adhesin at the tip of the P-fimbriae binds to the Galα1-4Gal sequence of globo series glycosphingolipids in the kidney and contributes to pyelonephritis [59-62]. This critical first step of infection, can in turn be used as a therapeutic strategy. In a murine model, reinfection of uropathogenic *E. coli* was effectively hindered by targeting its mannose-binding FimH fimbriae [63].

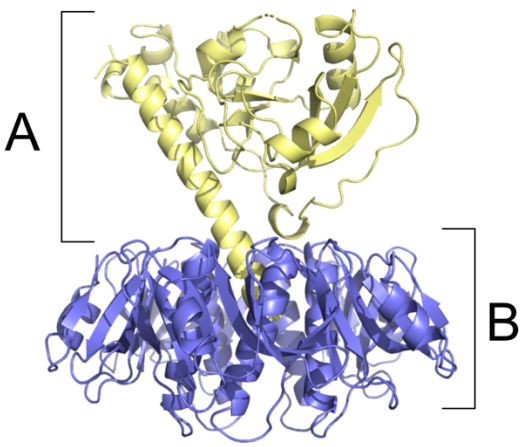

Figure 3. The A- and B-subunits of AB_5-toxins. *The A-subunit (yellow) carries the enzymatic property, and the B-subunit binds receptors. PDB code 1TII.*

After gaining entry to host tissues, pathogens may modify host glycosphingolipids to create a favorable environment for survival and proliferation [64]. Some bacteria possess enzymes that can alter the structure of glycosphingolipids, affecting cellular processes and immune responses, ultimately benefiting the pathogen.

Several bacterial toxins also display affinities for specific glycosphingolipid structures [65, 66], using them as entry points or disruptors of cellular functions. In fact, some of the strongest known, non-covalent molecular interactions are toxin-glycan interactions, such as the interaction of cholera toxin (CT) and the GM1 ganglioside [67, 68].

Bacterial AB_5 toxins, are defined by their bifunctional AB_5 architecture (figure 3), where "A" signifies an enzymatically active subunit and "B_5" designates a pentamer that mediates host cell binding [66, 69]. While various AB_5 toxins exhibit certain disparities in their specific cellular targets, pathogenic strategies, and induced pathologies, common structural and functional are shared among them. The A subunit universally disrupts cellular mechanisms, often through ADP-ribosylation of specific proteins, whereas the B subunit invariably facilitates host-cell binding and entry, interacting predominantly with surface glycolipids or glycoproteins. Notable bacterial species producing these toxins include *Vibrio cholerae*, *Bordetella pertussis*, *Shigella dysenteriae*, *Salmonella enterica, Citrobacter freundii* and certain *E. coli* strains (table 1) [66, 69, 70].

Table 1. Examples of AB_5-toxin producing pathogens.

Bacteria	AB_5-toxin	A-subunit activity	Toxin family
V. cholerae	Cholera toxin	ADP-ribosyltransferase	Cholera
E. coli	Heat-labile	ADP-ribosyltransferase	Cholera
C. freundii	CfxAB	Metalloprotease	Cholera
E. coli	Shiga-like toxin	RNA N-glycosidase	Shiga
S. dysenteriae	Shiga toxin	RNA N-glycosidase	Shiga
E. coli	Subtilase toxin	Serine protease	Subtilase
B. Pertussis	Pertussis toxin	ADP-ribosyltransferase	Pertussis
S. enterica	Typhoid toxin*	ADP-ribosyltransferase and DNase	Pertussis

*Several variants exist

1.3 ENTEROTOXIGENIC E. COLI

Belonging to the *Enterobacteriaceae* family, the coliform enterotoxigenic *E. coli* is a gram-negative bacteria defined by its ability to express enterotoxins that can cause immense diarrhea [71]. Notably, ETEC is the most predominant among the five principal diarrheal *E. coli* strains, which also encompass enterohemorrhagic, enteropathogenic, enteroinvasive, and enteroaggregative [72].

The hundreds of millions of annual cases makes ETEC a leading cause of bacterial diarrhea, and the most common cause of travelers' diarrhea [73-77]. Similar to other diarrheal diseases, an ETEC infection is propagated through fecal-oral transmission and can lead to extensive outbreaks if drinking water becomes contaminated, and is endemic to areas where proper sanitation may not be sufficiently upheld [78].

The prevalence among farm animals further exacerbates the challenges faced by developing countries [79, 80], where livestock commonly represent a primary source of income. Furthermore, in resource-limited settings, close-knit living with farm animals may contribute to fecal contamination of water supplies and transmission of diarrheal pathogens. However, there is no definitive evidence to categorize ETEC as a zoonotic disease as of yet, despite the similarities in some virulence factors expressed by isolates from animals and humans.

Symptomatic similarities with other infections, and a scarcity of both resources and efficient diagnostical tests [81], make diagnosis a challenge and hinders the ability to estimate the true extent of the disease. Although ETEC disease is chiefly self-limiting, tens of thousands of individuals die due to the severe dehydration every year [82]. Most cases and fatalities occur among children under the age of five, with peak incidence between six to 36 months [83, 84]. While recovery typically occurs within three to four days, the accompanying loss of nutrients and intestinal tissue damage can have enduring repercussions, including stunted growth and disruption of the commensal flora [85].

Antibiotic treatment has no place in the treatment of ETEC other than in the most severe cases. Instead, the primary intervention involves the swift administration of liquid and electrolyte replacement therapy, either orally or intravenously, which includes water, glucose, and salts [6]. While this

treatment is relatively straightforward, it is contingent upon access to clean drinking water, highlighting a potential challenge in certain environments.

1.3.1 INFECTION BY ETEC

Typically, more than 10^6 ETEC colony-forming units (CFUs) are required for successful transmission [86]. This substantial inoculum size aligns with that of *V. cholerae*. Studies involving human volunteers suggest that this high quantity is necessary to overcome the normal acidity of the stomach, and as few as 10^4 cells might suffice at elevated gastric pH levels [87].

The primary site of infection is the small intestine, where ETEC will try to surpass the protective mucus layer and gain access to the epithelial cells. For instance, ETEC may deploy a serine protease (EatA) to cleave MUC2 [88], thereby facilitating its path to the epithelial surface.

Once in place, two principal carbohydrate-binding virulence factors of ETEC initiate and mediate the infection. These host-specific surface adhesion molecules are known as colonization factors (CFs) and enterotoxins.

1.3.2 COLONIZATION FACTORS

To date, over 25 known human ETEC CFs have been described [5, 6, 89, 90]. With the exception of CS2, these CFs are encoded on plasmids, alongside other virulence genes such as those for enterotoxins [91, 92]. CFs are typically grouped based on antigenic and genetic differences, leading to the formation of distinct categories such as the CFA/I-like, CS5-like, and class 1b groups [5]. Additionally, a collection of miscellaneous CFs that do not fit into these main categories is also recognized.

Generally, CFs are composed of a repeating major subunit, which forms the bulk of the protruding adhesin, and a minor subunit commonly located at the tip of the complex [93]. Some ETEC strains express multiple CFs, potentially enhancing their adhesive capacity and thereby increasing their colonization potential. Moreover, CFs are associated with the co-expression of either heat-stable or heat-labile enterotoxins, or in some instances, both (approximately one-third each) [94-96].

While the receptor specificity of many ETEC CFs remain incompletely characterized, studies have predominantly shown affinity for glycoconjugates [97-100]. For instance, the ubiquitous non-fimbrial CS6 exhibits affinity for

sulfatides in human small intestinal epithelial cells [101]. Furthermore, CFA/I, the first described CF and among the most common, adheres to glycoproteins and neutral glycosphingolipids present in the human small intestine [98]. The interactions with glycosphingolipids were found to be mediated by the major subunit (CfaB) [99], and a notable binding-active glycan-motif was the Le^a determinant. In a following clinical study, it was demonstrated a significantly higher incidence for CFA/I-expressing ETEC infection among Le(a+b-) children than those for Le(a-b+) children [102].

1.3.2.1 CS30

Despite the identification of a substantial number of ETEC CFs from human isolates, previous studies have indicated that 30-50% of isolates did not exhibit an identifiable CF [94, 96]. In a study by von Mentzer et al. (2017) [90], whole-genome sequencing was employed on 94 such isolates, leading to the discovery of a set of genes (*csmA-G*) proposed to belong to a new CF, termed CS30. The expression of this novel protein was confirmed by SDS-PAGE and mass spectrometry (MS), and CS30+ ETEC strains demonstrated an affinity towards Caco-2 cells. However, the specific host receptors to which CS30 binds remained unidentified, and was further investigated in paper II.

1.3.3 ENTEROTOXINS

Once ETEC has gained access to the small intestinal epithelia, enterotoxins are released. The enterotoxins of ETEC are either heat-stable (STs) or heat-labile (LTs) and are the main reasons for the diarrheal symptoms [103, 104].

1.3.3.1 HEAT-STABLE ENTEROTOXINS

Two subtypes of ST can be formed according to structure and host-specificity; STa, which is found in isolates infecting humans, and STb, which is found in isolates from pig and cattle where it has been shown to bind sulfatide receptors [65, 103, 105]. STh and STp are, in turn, subtypes of STa, both of which may infect humans, but the latter was originally found to infect pigs [106].

The encoding gene for ST, *estA*, is found on plasmids and the expressed protein consists of a single, non-immunogenic oligopeptide of 18-19 amino acid residues [103, 107]. Its small size and stable structure make it resistant to heat and proteolytic degradation, which contributes to its potency as a toxin.

ST binds to the transmembrane receptor guanylate cyclase C on the intestinal epithelial cells, triggering increased levels of cyclic GMP (cGMP) [107, 108]. The elevated cGMP levels induce chloride ion secretion by activating Cystic

Fibrosis Transmembrane Regulator (CFTR), and inhibit sodium ion absorption, resulting in a net fluid secretion into the intestinal lumen and diarrhea. However, a recent study on human jejunal organoids showed elevated cGMP mainly in the basolateral space of the cells upon infection with ST, suggesting extracellular effects from cGMP [109].

1.3.3.2 HEAT-LABILE ENTEROTOXINS

Many important bacterial pathogens deploy structurally related AB_5-toxins as their main way of pathogenicity (table 1), collectively responsible for more than one million deaths per year [66, 69, 70]. LTs belong to this superfamily of toxins, which are defined by their molecular composition, with a unique A-subunit responsible for toxicity and a pentameric B-subunit enabling glycan binding and cell entry. Each of these components play a distinct role in the functionality and pathogenicity of the toxin.

The AB_5-toxin superfamily can further be divided into four families, depending on A-subunit function and sequence homology, of which LTs belong to the cholera toxin family [66]. The LTs of ETEC bear significant resemblance to CT in terms of structure, function, and, to a certain degree, receptor specificity. This includes powerful adjuvant properties [110].

LTs can be divided into two antigenically different groups, type I (LT-I) and type II (LT-II), with LT-I sharing the most resemblance to CT [111, 112]. Both human and porcine variants of LT-I exist (LT-Ih and LT-Ip, respectively). LT-II, in turn, can be further divided into LT-IIa, LT-IIb and LT-IIc, reflecting the order of discovery. LT-I is encoded on plasmid elements, while LT-IIs are located on the bacterial chromosome [113].

Although LT-I is the primary heat-labile enterotoxin commonly associated with human infections, type II LTs, especially LT-IIc, have been linked to clinically significant human infections [104, 114]. Nonetheless, the relative novelty of LT-II toxins means they have been less extensively studied.

1.3.3.2.1 MECHANISM OF ACTION

While the specific actions of different AB_5 toxins may vary, they generally follow a similar scheme. For CT and LTs, after ganglioside binding of the B-pentamer and endocytosis, the holotoxin is transferred to the ER via retrograde trafficking through the Golgi apparatus [115]. Once in the ER, the A-subunit is dissociated from the B-pentamer. The A-subunit is unfolded and exported to the cytoplasm, where it refolds and catalyzes ADP-ribosylation of $G_s\alpha$, which

in turn locks the heterotrimeric G-protein in an active state [116, 117]. This stimulates adenylate cyclase which causes the continued production of cAMP, and in turn activates CFTR [66]. The subsequent immense efflux of ions pulls water out of the intestinal cells, resulting in symptomatic diarrhea.

1.3.3.2.2 THE STRUCTURE OF LTS

A-subunit

The 28 kDa A-subunit consist of one polypeptide made up of two domains; The A1 domain which carries the ADP-ribosylating properties, and the A2, which is anchored to the B-pentamer [118]. The A1 chain is cleaved from A2 by the activity of protein disulfide isomerase in the ER.

The A-subunits of ETEC and *V. cholerae* share similar structural morphology and more than 50% amino acid homology across the board. This converts to similar enzymatic properties [66].

B-subunit

Crystal structures reveal a striking resemblance of the AB_5-toxins from *V. cholerae* and ETEC (figure 4) [119-122]. Focusing on the B-pentamers, CT and LT-I shares approximately 80% amino acid homology, while the LT-IIs share about 50% identity. However, there is remarkably almost no sequence similarity between type II LTs and LT-I/CT.

Each of the five monomers of a B-pentamer consist of approximately 100 amino acids (~11.5 kDa), folded into a globular shape of two α-helixes and two β-sheets. Studies on crystal structures of B-pentamers together with glycan receptors reveal that the primary binding sites are located in a loop region between two monomers, resulting in a total of five binding sites per B-pentamer [123, 124]. The five binding sites allow for two advantages; 1) although the affinity for one binding site may be in the micro- or even millimolar range (which for CTB has been enough to trigger endocytosis [125]), the overall avidity of the binding sites can add up to the nanomolar level [126, 127], and 2) the position of the binding sites forces a crypt-like formation in the plasma membrane of target cells when receptors are bound, triggering vesicle formation and endocytosis [128-130].

Both LT-I/CT and LT-IIs exhibit glycosphingolipid-binding properties. GM1 is the main receptor for LT-I and CT [131], although LT-I also binds to glycoconjugates with terminal *N*-acetyllactosamine [132, 133]. Additionally, CT has recently been shown to interact with human blood group antigens on

glycoproteins [134]. The type II LTs display preference to a wider array of ganglioside receptors [135], which are presented in papers I and III.

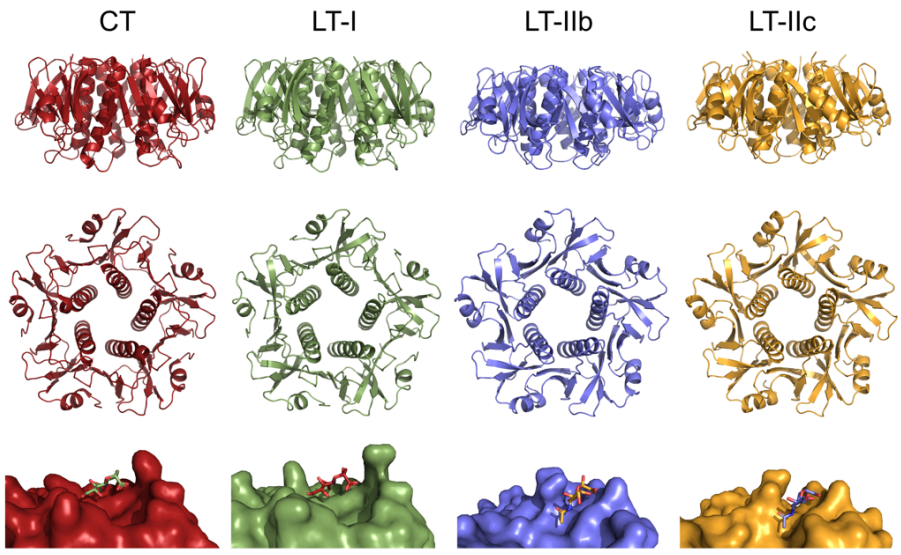

Figure 4. B-subunits of the cholera toxin family. *From top to bottom: Side-view, bottom-view, and focused view on the sialic acid binding sites. Although very similar, the configuration of the binding pocket, and mode of the glycan unit, differ between the CT/LT-I class and the type II LTs. PDB codes are 3CHB, 2XRQ, 5G3L and 7PRS.*

1.4 THERAPEUTIC IMPLICATIONS

The interaction between glycans and microbial pathogens is multifaceted, influencing various stages of infection. Many of these infections are mediated by common pathogenic themes, such as the glycan-binding of AB_5 toxins. Understanding such interactions holds significant therapeutic potential [136, 137] and may aid to prevent microbial adherence and entry, as well as modulate immune responses, and disrupting microbial survival mechanisms [138-140].

Previous studies on the enterotoxins of ETEC and *V. cholera* helped elucidate fundamental cellular and pathological processes, such as the formation of lipid rafts and vesicles, chart the glycan expression of eukaryotic cells, deepen our understanding of the molecular basis for receptor-ligand interactions, and pave way for potent vaccines and adjuvants [139, 141].

2 AIM

2.1 GENERAL AIM

The overall aim of this thesis was to deepen our understanding of the pathogenic mechanisms leading to ETEC – and ETEC-related – infections, as well as characterizing the fundamental processes governing protein-carbohydrate interactions. In addition, our goals were to further characterize the glycosphingolipid expression of the human small intestine, and search for potential receptors involved in ETEC pathogenicity.

2.2 SPECIFIC AIMS

Paper I: To characterize the carbohydrate-binding specificity of ETEC type II enterotoxin LT-IIb and structurally define the interactions involved in the carbohydrate binding.

Paper II: To characterize the carbohydrate-binding specificity of the novel ETEC colonization factor CS30.

Paper III: To characterize the carbohydrate-binding specificity of the novel ETEC type II enterotoxin LT-IIc and structurally define the interactions involved in the carbohydrate binding.

Paper IV: To study the dynamics governing the protein-carbohydrate interactions of LT-IIb B-pentamers to allow for comparison with other, related B-pentamers of the AB_5-toxin family.

3 MATERIAL AND METHODS

3.1 BACTERIAL CULTURING AND RECOMBINANT PROTEIN EXPRESSION

Expression and purification of recombinant proteins are essential steps in structural and functional studies. The *E. coli* expression system is favored for its rapid growth, ease of genetic manipulation, and cost-effectiveness [142-144]. The *E. coli* strain BL21 is particularly suitable for its properties that enhance protein expression and was therefore used in paper I, III and IV.

Synthetic genes that encode the glycan binding B-subunits of LT-IIb and -IIc were acquired from ATG-Biosynthetics and cloned into the locally developed and highly efficient expression plasmid pML-λcI857 [145, 146], initially formulated for expressing cholera toxin B-subunits (CTB) in *V. cholerae*. The plasmid is based on the commercially available pBR322 expression vector and modified with a lambda promotor and temperature-sensitive cI857 repressor derived from the bacteriophage *Escherichia virus Lambda.*

The plasmids were transformed into electrocompetent *E. coli* BL21-strains, which were selected for ampicillin resistance, conveyed by the plasmid. Following bacterial growth in Luria Bertani (LB) broth at 30°C, protein production is induced simply by increasing temperature to 42°C which inactivates the cI857 repressor. Initially, a suboptimal expression of the native proteins was obtained. Therefore, the signal peptide, responsible for directing the proteins for periplasmic secretion, was removed. Consequently, B-monomers were expressed intracellularly, accumulating as inclusion bodies due to the rapid expression rate and elevated temperature. This allowed for verification of protein expression through phase-contrast microscopy, wherein the inclusion bodies were easily identified.

However, the intracellular protein expression necessitates extraction, which was achieved by extensive lysis of the bacterial cells using both lysozyme and ultrasonic treatment. The high-density inclusion bodies were thereafter separated from cell debris by centrifugation. Repeated washing steps eliminated cell debris such as lipopolysaccharides, after which the inclusion bodies were dissolved in high concentrations of urea to dissociate the protein aggregates. This was followed by gradual dialysis against a physiological

buffer (50 mM sodium carbonate buffer at pH 9) to promote proper folding into pentameric units.

Final purification steps were performed by ion exchange and size-exclusion chromatography until samples generated single bands at both ~11.5 kDa and ~55 kDa on 14% Tris/glycine gel, corresponding to the molecular mass of monomeric and pentameric protein, respectively.

The isolated wild-type ETEC strain E873 expressing CS30 [90], was cultured on either CFA agar plates or in CFA broth, containing 0.15% crude bile at 37°C. The same conditions, with addition of kanamycin 0.05 mg/ml, were used for culture and labeling of the mutant ETEC strain E873ΔcsmA with disrupted csmA (major subunit) gene.

3.2 LABELING

3.2.1 ^{125}I

Protein labeling is often necessary for their detection and study. In papers I and III, LT-II B-pentamers were radiolabeled by the isotope iodine-125 (^{125}I); a popular isotope due to its relatively long half-life (approximately 60 days) and its ability to emit gamma radiation [147], which can be easily detected with high sensitivity. Radiolabeling is especially useful for binding assays since the radioactive decay allows for direct and highly sensitive detection of protein by film or a gamma counter. However, the isotope decay of ^{125}I may cause protein degradation and loss of function. Therefore, proteins should be used promptly after radiolabeling.

The labeling procedure requires a catalyzing agent, such as the iodination reagent IODO-GEN, that initiates the iodation reaction from Na^{125}I to the side chains of tyrosine [148]. Excess of Na^{125}I, is then separated from radiolabeled protein by a gel filtration column.

3.2.2 ^{35}S

In paper II, bacterial cells were grown in media containing ^{35}S-methionine. Under these conditions, the bacteria metabolically incorporate ^{35}S-methionine into their proteins during synthesis, enabling detection of radiolabeled cells in a manner analogous to proteins labeled with ^{125}I. Notably, the half-life of this

isotope is almost three months. A bacterial concentration of 1×10^8 CFU/ml was used on thin-layer chromatograms during CBAs.

3.2.3 ^{13}C, ^{15}N AND ^{2}H LABELING OF RECOMBINANT B-SUBUNITS

^{13}C and ^{15}N labeling of proteins is commonly employed for their study using nuclear magnetic resonance (NMR) spectroscopy, enabling two-dimensional (2D) and three-dimensional (3D) experiments. The magnetic moment of these isotopes allows them to be detected by NMR, unlike the more abundant ^{12}C and ^{14}N isotopes. With natural abundances of approximately 1.1% and 0.37% for ^{13}C and ^{15}N respectively, enriching proteins with these isotopes significantly enhances the resultant spectrum [149]. This enhancement improves the signal-to-noise ratio and resolution of cross-peaks, facilitating clearer identification and assignment of amino acid residues.

While the proton (^{1}H) spin state is NMR-detectable and sufficient for characterizing smaller molecules (using 1D-NMR), it can be problematic when studying larger biomolecules like proteins [150]. This is due to the abundance of NMR signals from both homonuclear (^{1}H-^{1}H) and heteronuclear (^{1}H-^{15}N or ^{1}H-^{13}C) scalar couplings, which can overcrowd and complicate a spectrum. To alleviate this, proteins are sometimes labeled with deuterium (^{2}H), which, although detectable by NMR, has different resonances that do not typically interfere with ^{1}H-spectra. Deuterium at amide sites can also exchange with environmental protons, preserving detectable amide signals. Such spectra exhibit reduced complexity, resulting in narrower line widths and simplified peak interpretation across both 2D and 3D. Notably, even in multidimensional experiments, signal acquisition primarily focuses on protons due to their high gyromagnetic ratio, enhancing magnetization transferred from ^{13}C and ^{15}N through heteronuclear couplings.

Labeling of proteins in paper IV was achieved by exchanging the growth medium from LB to deuterated minimal medium, containing ^{13}C labeled glucose as the carbon source and ^{15}N labeled ammonium chloride as the nitrogen source. The concentration of deuterium was slowly increased to acclimate the BL21-strain before protein expression was induced.

3.3 GLYCOSPHINGOLIPID ISOLATION

The amphipathic character of glycosphingolipids, which facilitates their existence within the hydrophobic domain of cell membranes while simultaneously interacting with a hydrophilic environment, also enables their extraction (summarized in figure 5). First, as outlined by Karlsson (1987) [151], an organic solution mixture is employed in a Soxhlet apparatus to extract all lipids from a homogenized and lyophilized tissue.

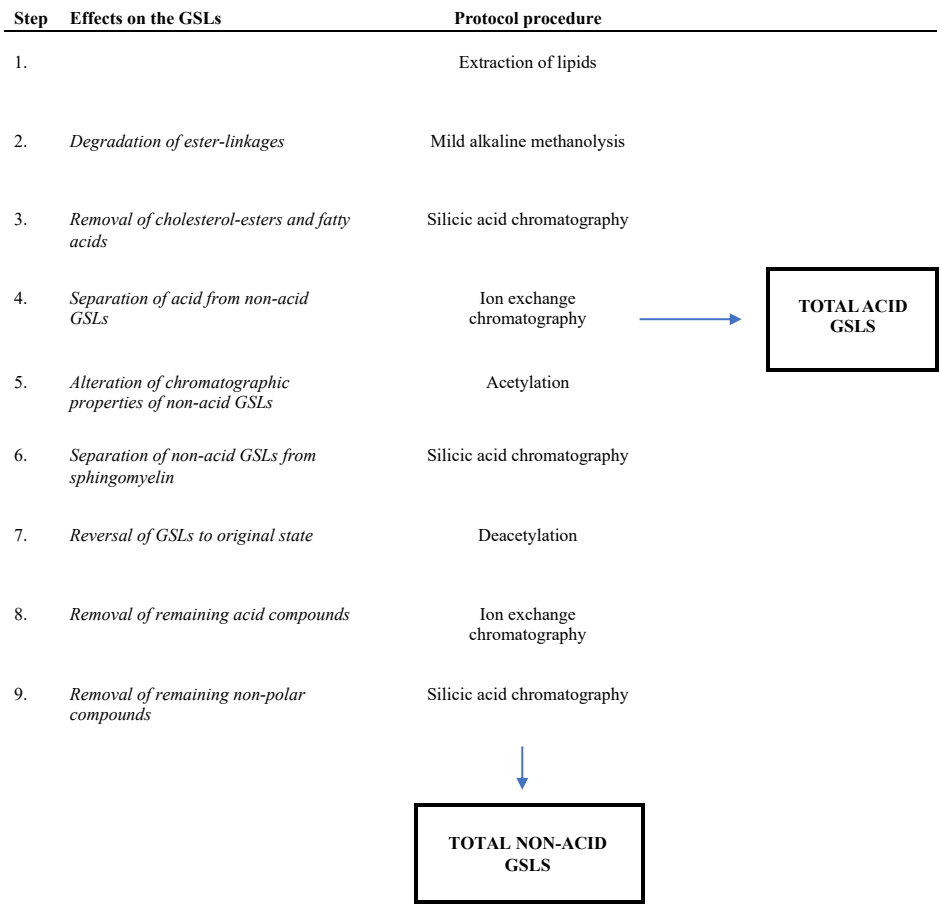

Figure 5. The main steps of glycosphingolipid isolation. *Starting with lipid extraction from a lyophilized tissue sample, the procedure results in two fractions of acid and non-acid glycosphingolipids. GSL = glycosphingolipid.*

Next is the task of separating glycosphingolipids from unwanted lipids, such as triglycerides, cholesterols, and phospholipids. Given that glycosphingolipids exhibit resistance to mild alkaline conditions – a trait not generally shared with non-sphingosine lipids – treatment with a mild alkaline enables hydrolysis of the glycerol-based phospholipids. After dialysis the non-polar compounds are separated from the glycosphingolipids by silicic acid column chromatography.

The isolated glycosphingolipid fraction is then divided into an acid and non-acid (neutral) fraction by ion-exchange chromatography, such as a diethylaminoethyl (DEAE) cellulose column. However, the non-acid fraction contains the alkali-stable phospholipid sphingomyelin and is therefore subjected to acetylation [152] followed by a second silicic acid column separation, deacetylation, and dialysis. The remaining non-acid fraction is purified by a final step of DEAE- and silicic acid-columns.

These mixtures of acid and non-acid glycosphingolipids generated from each tissue may be sufficient for the testing of protein-carbohydrate interactions. But for any in-depth analysis of glycan structure, further separation into single compound fraction is needed. Optionally, lectins may be utilized to segregate specific glycosphingolipids via affinity chromatography. Separation can also be achieved by chromatography on Iatrobeads^{R} silica gel columns [153] or high-pressure liquid chromatography, with simultaneous thin-layer chromatography (TLC) [154] to monitor the separation. Within the TLC process, mixtures of chloroform, methanol, and water serve as the mobile phase, facilitating the separation of glycosphingolipids across the silica gel plate based on the polarity and size of both carbohydrate and ceramide moieties. For staining, anisaldehyde [155] or resorcinol [156] is utilized to detect total glycosphingolipid content or gangliosides only, respectively.

All in all, this is a highly reproducible and straightforward method, albeit a bit time-consuming, which allows for detailed functional and structural studies of glycosphingolipids.

3.4 CHROMATOGRAM BINDING ASSAYS

Leveraging the simplicity of TLC to separate glycosphingolipids is a notable advantage for their use in chromatogram binding assays (CBA) [157]. Post-separation of glycosphingolipids on a silica gel plate, the entirety of the plate is soaked in polyisobutylmethacrylate, dissolved in an *n*-hexane/diethyl ether

solution, for a one-minute duration, serving to immobilize the silica gel. Unspecific binding is prevented through pre-incubation with bovine serum albumin, or an alternative blocking agent, whereafter the silica plate is overlaid and incubated with the studied adhesin. The adhesin can be either directly or indirectly detected (with the use of secondary antibodies). Direct labeling with radioactive isotopes was used in this thesis, followed by autoradiography for the detection of bound adhesins.

The strategy for elucidating the glycan binding specificity of the adhesins in papers I-III, involved the initial screening of known and unknown glycosphingolipids amassed over the years at our department from various animal species and tissues. Binding assays were performed with fractions containing either characterized single compounds glycosphingolipid, or acid and non-acid glycosphingolipid mixtures from various tissues.

Upon achieving binding to any of the glycosphingolipid mixtures, further separation into a refined, single-compound fraction can be performed by repeated Iatrobead^{R} silica gel columns. The structural determination of binding-active glycosphingolipids is primarily achieved through analytical techniques such as mass spectrometry (MS), NMR, or by glycan-binding proteins.

An inherent strength of employing glycosphingolipids for discerning carbohydrate binding activity lies in the fact that each lipid possesses only a singular saccharide chain – unlike glycoproteins which may be glycosylated with numerous heterogeneous chains. Moreover, some saccharide structures, such as the lactose sequence, are exclusively present in glycosphingolipids. While glycosphingolipids do not contain mannose in higher animals, this can be mitigated by using mannose containing glycosphingolipid mixtures from alternative sources, like plants and sea animals [158]. Thirdly, glycosphingolipids have commendable stability and are amenable to storage. Lastly, the ease, speed, and cost-effectiveness of the TLC-based method render it ideal for repeated experiments.

Nonetheless, like any experimental model, a few drawbacks warrant acknowledgment. Due to the static nature of the studied, immobilized glycosphingolipids, the impact of avidity for multivalent binding sites, and the importance of epitope presentation, might be underestimated. For instance, while binding of cholera toxin B-subunits to fucosylated glycoproteins has been observed to a non-canonical binding site [134], no such binding is evident

to similar epitopes on glycosphingolipids [159]. The proposed idea is that the shorter length of the glycosphingolipids does not access all possible binding sites on the B-pentamer, thereby initiating only a limited number of binding interactions that are too weak to detect.

3.4.1 INHIBITION STUDIES

When a specific glycan motif is suspected to mediate an observed binding interaction to glycosphingolipids, inhibition studies may be used to strengthen such suspicions. In this context, the adhesin of interest is pre-incubated with an oligosaccharide analogue, resembling the proposed binding epitope, prior to proceeding with the CBA as previously described. Absent glycosphingolipid interaction through this procedure serves to confirm the saccharide epitope requisite for binding. Nonetheless, this approach may warrant caution as there are documented instances where saccharides, although not constituents of the recognized glycan structure, have inhibited interaction with the protein. For instance, the lectins E-PHA and L-PHA isolated from kidney beans are inhibited by GalNAc, a sugar that is not part of the *N*-glycan structures that they recognize [160, 161].

3.4.2 MICROTITER WELLS

Once the initial screening has provided evidence for binding active glycosphingolipid receptors, a more semi-quantitative and comparative binding analysis can be performed by microtiter wells assays. Increasing amounts of glycosphingolipids are immobilized in microtiter wells, followed by the addition of radiolabeled adhesin. The plates are then washed to remove any unbound material and a gamma counter will provide information equivalent to the bound adhesin.

3.5 GLYCOSPHINGOLIPID CHARACTERIZATION

Several strategies exist for the structural identification of glycosphingolipids, each varying in complexity and precision. The most straightforward approach entails utilizing TLC to compare the retention factor of known reference glycosphingolipids with that of an unknown sample, or by running the isolated fraction alongside known reference samples on a silica gel plate and inspecting for co-migration. However, since different glycosphingolipids can co-migrate and manifest as a single band due to subtle variations in *e.g.* ceramide

composition, this approach leaves room for uncertainty. Luckily, alternate methods are available to supplement or validate findings derived from this primary strategy, ensuring a comprehensive and accurate analysis of glycosphingolipid structures.

3.5.1 GLYCAN BINDING PROTEINS

A key tool in characterizing glycan structures is glycan-binding proteins (GBPs). Their ability to distinguish glycans can range from a single monosaccharide, to complex, branching oligosaccharides, including the discrimination of α- and β-anomeric linkages. In addition to recognizing terminal glycan structures, some GBPs instead recognize core saccharide sequences, allowing for a more comprehensive structural glycan characterization when different GBPs are combined.

Commonly, GBPs are classified as either lectins, antibodies, microbial adhesins or viral agglutinins. Historically, lectins have been extensively isolated from plants ever since their first description at the end of the 19th century [162]. They are however found in all living organisms, but it was not until the 1960's that the first animal lectin was isolated [163]. The hemagglutinating properties of some lectins played a pivotal role in investigating the structural foundation of the ABO blood group antigens. Although the *in vivo* function of numerous lectins is yet to be established, many animal lectins play important roles in *e.g.* the immune response and glycoprotein synthesis.

3.5.2 MASS SPECTROMETRY

The immense structural diversity that arises from multiple saccharides linked together far exceeds that of amino acids or nucleotides and requires precise analytical tools for their study. The single most revolutionizing technique for carbohydrate characterization is MS which has been used to structurally determine glycosphingolipids over the past 50 years [164, 165]. At its core, MS is an analytical technique where molecules are ionized to form charged species. These ions are then sorted based on their mass-to-charge ratio (*m/z*) in an analyzer and detected in a manner proportional to their abundance.

Different MS modalities exist, divergent mainly in the ionization step, which gives the option to use the technique best suited for the analyte in mind. Electrospray ionization (ESI) MS, often used for polar and larger molecules, typically produces intact ions and fewer fragment ions than other forms of MS,

such as electron impact (EI) MS, and is therefore considered to be a “soft” ionization technique. This is particularly suitable for the study of glycosphingolipids since it allows for the study of intact molecules. Furthermore, tandem MS analysis (ESI-MS/MS or MS^2) may generate and detect more fragmented ions if needed, with the purpose of increasing resolution and sensitivity. Furthermore, connecting ESI-MS to an initial liquid chromatography step (LC-ESI-MS) allows for an on-line separation of glycosphingolipids, which facilitates the study of more heterogeneous samples.

Characterization of the glycan part of glycosphingolipids can be achieved by LC-ESI-MS after removal of the ceramide by endoglycoceraminidase digestion (EGCase) [166]. The oligosaccharides are separated on a graphitized carbon column, whereby resolution of isomeric oligosaccharides is obtained, and MS^2 gives C-type ion series from which the carbohydrate sequence can be deduced. Information about linkage positions is given by cross-ring ^{0,2}A-type fragment ions which are present when HexNAc or Hex is substituted at C-4.

This method has however been hampered by the somewhat restricted hydrolytic capacity of EGCase II, which does not fully digest globo series glycosphingolipids and gangliosides. However, recently EGCase I, with a more universal hydrolytic capacity [167] has become commercially available.

3.5.3 PROTON NMR

While NMR spectroscopy is occasionally hindered by the substantial material quantities required to obtain a spectrum of adequate quality (typically between 100 and 200 micrograms) its increased resolution enables a detailed determination of monosaccharide composition, anomeric configuration and linkage positions [168]. Proton NMR spectroscopy, utilizing its inherent capability to analyze chemical shifts exclusively on ^{1}H nuclei, does not necessitate any labeling, rendering it particularly well-suited for the study of purified glycosphingolipid isolates from tissues. Additionally, the non-destructive nature of NMR analysis allows for recovery of the analyte if further investigation is desired.

3.5.4 ADDITIONAL METHODS

Information about monosaccharide types and glycosidic linkage positions can be obtained by degradation studies [169]. However, this method is not often

used nowadays. Stepwise glycosidase digestion can also be used for structural characterization.

3.6 STRUCTURAL AND DYNAMIC CHARACTERIZATION OF PROTEIN-CARBOHYDRATE INTERACTIONS

Once binding-active glycosphingolipids are discovered and structurally characterized, a more detailed analysis of the underlying molecular foundation governing the protein-glycan interaction can be made. To leverage structural techniques, such as X-ray crystallography and NMR spectroscopy, it is imperative to first attain sufficiently pure samples of the recombinant protein. In paper II, recombinant protein expression was not successful, thus preventing such structural analysis.

3.6.1 CO-CRYSTALLIZATION STUDIES

X-ray crystallography stands as one of the most powerful techniques for determining the three-dimensional structures of biological macromolecules, providing pivotal insights into the molecular frameworks of proteins. This technique was first used to solve the structure of myoglobin derived from sperm whale [170], garnering the Nobel Prize in Chemistry in 1962. Ever since, X-ray crystallography has firmly established itself as an indispensable tool for the structural characterization of molecules due to its exceptional resolution capabilities [171].

At its essence, X-ray crystallography involves directing an X-ray beam toward a crystallized sample which generates a diffraction pattern, emerging from the interaction between the X-rays and the electron clouds of the atoms within the crystal. It is therefore crucial to produce high-quality crystals, which in turn is influenced by factors such as protein size and chemical properties. The detected diffraction pattern contains data pertaining to the electron density within the crystal, which is subsequently translated into a molecular model through a series of computational transformations.

In protein-ligand interaction studies, the protein of interest is crystallized in conjunction with the ligand, and the resulting diffraction pattern can be compared with the ligand-free protein [172]. The co-crystallization studies of

papers I and III were carried out by collaborators from the laboratory of Annabelle Varrot and Anne Imberty at the University of Grenoble Alpes.

The advantage of this mature technique is its near atomic resolution and lack of size limitations, as long as quality crystals can be acquired [173]. Conversely, the main drawbacks of X-ray crystallography revolve around crystal formation and the inability to determine dynamic and transient states. The complexity of achieving crystal formation tends to escalate with increasing protein size or insolubility and may necessitate a broad screening procedure to identify the optimum conditions for crystallization, which typically involve experimenting with a variety of buffers and pH levels.

3.6.2 NMR FOR PROTEIN-LIGAND INTERACTIONS

NMR spectroscopy is a powerful tool to not only determine the structure of a protein and its binding sites, but also reveal information about dynamic changes, binding affinity, and molecular contributions to binding interactions [174]. When placed in a magnetic field, certain nuclei (such as the ^{1}H nucleus) resonate (or precesses) at specific frequencies, which can be detected. The frequency of this resonance can be influenced by the atom's local environment, including its interactions with neighboring atoms and molecules. This causes a variation in signal (chemical shift) which can be leveraged to identify molecular structures.

For the study of protein-ligand interactions, either the protein ("protein-observed") or the ligand ("ligand-observed") is detected. The protein-observed technique detects the ligand's effects on the protein and was used in paper IV.

3.6.2.1 PROTEIN-OBSERVED NMR

Through the detection of specific atomic nuclei, commonly ^{1}H, ^{13}C, and ^{15}N due to their specific magnetic properties and natural abundance, NMR offers the advantage of monitoring molecular motion over a range of timescales. This allows for the revelation of not just static structures but also the inherent flexibility and dynamics of proteins, which are crucial for their function. As a result, protein NMR has become an invaluable tool in the fields of biochemistry, molecular biology, and drug discovery. Several different two- and three-dimensional experiments can be employed for protein detection, varying mainly in the isotopic labeling requirement of the protein, the pulse program applied for magnetization transfer, and the method of signal detection.

Special consideration must be given to the size of the protein, as the reduced tumbling rates of larger molecules lead to increased relaxation rates, resulting in diminished NMR signal strength [174]. A significant advancement in the NMR study of large proteins was the introduction of transverse relaxation optimized spectroscopy (TROSY) [175] – a pulse sequence that has enabled the analysis of proteins up to 1 MDa in size [176]. However, this often necessitates extensive labeling of the protein with ^{2}H, ^{13}C, and ^{15}N isotopes [149]. The combination of a 2D HN-TROSY spectrum with spectra from triple resonance experiments, such as HNCA, HN(CO)CA, HNCO, HNCACB, and HN(CO)CACB, frequently allows for comprehensive protein assignment. Nuclear Overhauser enhancement spectroscopy (NOESY) [177] provides information of through-space interactions, typically up to 5 Å, and is an essential tool for determining 3D-structures by NMR. Yet, the efficacy of NOESY diminishes for larger proteins or complexes due to the associated decrease in NMR signal stemming from accelerated relaxation rates. Additionally, for these larger structures, the interpretation of NOESY spectra can become challenging because of signal overlap.

Following the structural characterization of a protein and assignment of peaks in a TROSY spectrum, one can explore ligand interactions using chemical shift perturbation (CSP) analysis [178-181]. This technique is essential for probing the dynamics of protein-ligand interactions, pinpointing the location of binding sites, and determining binding affinity. In a typical CSP experiment, a consistent amount of protein in a physiological buffer undergoes titration with increasing quantities of ligand – preferably exceeding the binding sites by at least 20-fold to ensure saturation. At each titration step, NMR spectra are recorded. These are commonly a combination of HN- and HC-heteronuclear single quantum coherence (HSQC) experiments. Although HSQC is the 2D experiment foundation of TROSY, it tends to offer lower resolution for larger proteins. However, its relatively quicker acquisition time makes it a suitable choice when the protein's backbone assignment is already established.

4 RESULTS AND DISCUSSION

Infectious diarrhea, primarily transmitted via the fecal-oral route, ranks second in global microbial-related mortality, following pneumonia. Proximity to domestic livestock without rigorous infrastructure and sanitation measures, enhances the risk of contamination with feces, thereby elevating the potential for transmission of pathogens between humans and animals. Numerous colonization factors (CFs) in ETEC utilize tip-localized adhesins to bind to glycan receptors, facilitating colonization of host tissues, with various glycosphingolipid receptors being characterized in human and porcine ETEC infections. Although CFs are considered to be host-specific, the relatively new Class 1B CFs in ETEC from humans with diarrhea exhibit structural and homological similarities to adhesin F6 (987P) in porcine ETEC strains [182]. This class includes the recently discovered CS30, which was expressed by ETEC strains isolated from symptomatic children worldwide [90]. The amino acid homology of the major subunit of CS30 (CsmA) with the major subunit of F6 (FasA) is more than 50%. To understand the prerequisites of infection, paper II studies the glycan binding preferences of these novel CS30 expressing ETEC strain.

While type II heat-labile enterotoxins of ETEC are commonly associated with animal infections [79], they have also been identified in symptomatic humans [114]. However, due in part to their relative novelty, LT-IIb and LT-IIc have not been studied to the same extent as CT and LT-I. Prior to the studies conducted in papers I and III, binding assays were conducted exclusively with a restricted selection of commercially available ganglio core chain gangliosides, showing interactions with gangliosides GD1a and GT1b for LT-IIb [135, 183], and with GM1, GM2, GM3 and GD1a for LT-IIc [184, 185]. Furthermore, structural data on the two pentamers were very limited, especially in conjunction with a ligand. Therefore, these two studies sought out to exhaustively elucidate the carbohydrate binding specificity of the B-pentamers from LT-IIb and -IIc, as well as structurally characterize these interactions. The last paper (IV) dives into the detailed ligand-interaction of LT-IIb, in an attempt to understand the dynamics of the binding and further characterize the binding sites.

4.1 PAPER II – CS30

In this study, a comprehensive examination was performed to explore potential binding of the novel ETEC colonization factor CS30 to glycosphingolipids. The wild type ETEC strain E873, which expresses CS30, and its mutant E873ΔcsmA, with a disrupted major subunit, were subjected to diverse carbohydrate structures from mixtures of acid and non-acid glycosphingolipid fractions. Notably, a specific binding of CS30-expressing ETEC to a fast-migrating compound within acid glycosphingolipid fractions was consistently observed, a phenomenon not observed with the mutant strain. Slow-migrating compounds in acid fractions or non-acid glycosphingolipids were not recognized by the CS30-positive bacteria.

Moreover, the CS30-positive strain demonstrated binding to a fast-migrating compound in acid glycosphingolipid fractions of both human and porcine small intestines, typically where cholesterol sulfate and sulfatide (SO3-3Galβ1Cer) co-migrate. Subsequent binding assays to reference fractions containing sulfatide revealed exclusive binding with the CS30-expressing strain, with no detectable binding to cholesterol sulfate.

Further isolation and analysis via LC-ESI-MS of binding-active compounds from human small intestinal samples displayed molecular ions corresponding to glycosphingolipids with sulfated hexose (SO_3-Hex) and a set of variant ceramides. Anti-SO_3-3Galβ monoclonal antibodies, employed in binding assays, confirmed the presence of sulfatide in these samples.

To investigate the significance of the ceramide portion for CS30 binding, a number of sulfatide species – varying in ceramide composition – were isolated from human stomach and human meconium. Characterization of CS30-binding sulfatides by LC-ESI-MS, revealed a pronounced binding preference for sulfatide featuring d18:1-h24:0 ceramide.

Glycosphingolipids related to sulfatide, such as galactosylceramide (Galβ1Cer), were not recognized CS30-positive *E. coli*, underscoring the sulfate moiety's necessity in the binding process. Similarly, no binding was achieved to sulf-gangliotetraosylceramide (SO_3-3Galβ3GalNAcβ4Galβ4Glcβ1Cer), suggesting an integral role of ceramide in the interaction, either as part of the binding epitope or in optimizing the presentation of the relatively succinct SO_3-3Galβ sequence.

CS30-expressing ETEC were incubated with saccharides and anionic polysaccharides before conducting binding to serially diluted sulfatide on chromatograms. A notable inhibition of the interaction between CS30-expressing ETEC and sulfatide occurred with dextran sulfate at 10 mg/ml, while a concentration of 5 mg/ml exerted no inhibitory effect. Dextran, sodium octadecyl-sulfate, and galactose-4-sulfate failed to inhibit binding under the tested conditions.

In conclusion, these findings reveal a sulfatide binding specificity of CS30 positive ETEC, of which binding is influenced by the ceramide moiety. Sulfatide recognition, noted also in pathogens such as *Mycoplasma pneumoniae*, *Bordetella pertussis*, and *H. pylori*, underscores its biological relevance.

Given that the porcine CF F6 (987P), a Class 1b-related variant, shares substantial amino acid sequence homology with CS30 (CsmA), and similarly binds to sulfatide, albeit only occasionally, indicate that CS30 positive strains might possess the capability to infect multiple hosts.

4.2 PAPER I – LT-IIB

To deepen the understanding of toxins produced by ETEC, an exhaustive investigation into the carbohydrate binding specificity of the LT-IIb B-subunit was performed. The recombinant LT-IIb B-subunit was effectively expressed in the lambda expression model following method optimization. To maximize the variation in glycan epitope presentation to the B-pentamers, the initial strategy was to conduct binding assays to a comprehensive collection of acid and non-acid glycosphingolipid mixtures, isolated from various animal species and tissues.

Distinct binding interactions to slow-migrating compounds in acid glycosphingolipid mixtures originating from rabbit thymus, human small intestine, and neutrophils were obtained. Major components of rabbit thymus and human neutrophils, which migrate at this level on a silica gel plate, include sialyl-neolactotetraosylceramide and sialyl-neolactohexaosylceramide [186-188]. Both, as reference compounds, were also recognized by the LT-IIb B-subunits, underscoring the pivotal role of gangliosides with a neolacto core chain in the interaction.

Next, CBAs were performed with a diverse and extensive set of previously isolated and structurally determined glycosphingolipids to decipher carbohydrate sequences vital for the binding of LT-IIb B-subunits. The GD1a ganglioside was reaffirmed as a binding glycosphingolipid, although no binding to GT1b was observed. Additional interactions, primarily with neolacto core gangliosides, were also revealed. Taken together this suggests that the binding epitope is Neu5Acα3Galβ3/4GlcNAc/GalNAc but that access is hindered by the di-sialyl unit of GT1b. Furthermore, terminal di-sialyl motifs and Neu5Ac in an α6-linkage, as opposed to an α3-linkage, went unrecognized, as did non-acid glycosphingolipids.

The sensitivity of radioisotope-detected CBAs, suitable for estimating relative affinity of protein-glycan interactions, highlighted a detection limit of less than 20 ng per lane for Neu5Gcα3-neolactohexaosylceramide, while the detection limits for Neu5Acα3-neolactotetraosylceramide and the GD1a ganglioside were approximately 800 ng per lane. This 40-fold preference for Neu5Gcα3-neolactohexaosylceramide might partially be attributed to the elongated oligosaccharide structure of Neu5Gcα3-neolactohexaosylceramide, or a favored binding to Neu5Gc gangliosides over Neu5Ac.

Although in limited quantities, GD1a is expressed in the human small intestine. Nonetheless, binding-active slow-migrating compounds from human small intestinal isolates migrated on a level corresponding to that of neolacto core gangliosides on silica gel plates. These fractions were isolated into singular compounds and analyzed by LC-ESI-MS, revealing the presence of Neu5Acα3-neolactohexaosylceramide. Thus, the human small intestine expresses receptors for LT-IIb as both GD1a and Neu5Acα3-neolactohexaosylceramide.

Moreover, despite limited data regarding the acid glycosphingolipids presented by lymphocytes, their expression of neolacto core gangliosides has been documented and may designate them as target cells of LT-IIb [189, 190]. This could, in turn, mediate or amplify the strong adjuvant properties demonstrated by the protein. Past studies investigating immune responses induced by the cholera toxin family, some of the most potent known mucosal adjuvants, have attributed the elicited immune response profile to variations in ganglioside recognition by the B-subunits on lymphocytes.

Lastly, to fully elucidate the binding interaction on a molecular level, the recombinant B-subunit was co-crystallized with the oligosaccharide Neu5Ac-nLT (Neu5Acα3Galβ4GlcNAcβ3Galβ4Glc). The subsequently obtained diffraction data, resolved at 1.72 Å, exhibited intact B-pentamers with electron density corresponding to the sugar ligand at the putative binding site. The binding site is situated between two monomers, analogous to the binding sites of CT and LT-I. However, the ligand-binding mode is distinct from those observed for GM1 in LT-I and CT.

Additionally, a second binding site was identified on the same side as the primary/putative binding site. In both binding sites, the Neu5Ac unit facilitated most binding interactions, and due to the shallower groove of the secondary binding site, no saccharide besides the Neu5Ac moiety was accommodated therein.

Attempts to fit the two Neu5Ac residues of a single GD1a oligosaccharide into the two binding sites of LT-IIb by molecular modeling were unsuccessful. Hence, it is more plausible that LT-IIb binds to separate sialylated glycosphingolipids on cell surfaces. This is the first description of a member from the cholera toxin family exhibiting two binding sites within a monomer, resulting in a total of ten binding sites per B-subunit to the same ligand.

4.3 PAPER III – LT-IIC

Initially, the carbohydrate-binding specificity of LT-IIc was explored through solid-phase glycosphingolipid-binding studies, employing mixtures of both acid and non-acid glycosphingolipid mixtures in the screening process. Contrasting with other B-subunits from ETEC and *V. cholerae*, LT-IIc B-subunits demonstrated a notably promiscuous binding, revealing bands in numerous mid- to slow-migrating acid compounds on silica gel plates, primarily migrating at and below the level of the GM3 ganglioside. No binding was observed to sulfatides, the major glycosphingolipid in the acid fraction of human small intestine. Mixtures which were considerably composed of neolacto core gangliosides, notably those isolated from rabbit thymus and human neutrophils, exhibited particularly intensive binding. LT-IIc B-subunits displayed no binding to GD3 during comparative analysis using TLC-plates overlaid with anti-GD3 antibodies.

Subsequent investigations with a set of known, single-compound reference glycosphingolipids, aiming for a more detailed understanding of LT-IIc's glycan binding specificity. A discernible pattern emerged where LT-IIc bound to terminal Neu5Acα3- or Neu5Gcα3-Gal gangliosides, excluding ganglioside GM4, Neu5Acα3-lactotetraosylceramide, and the Neu5Ac-globopenta/SSEA-4 ganglioside.

The inability to bind to GM4 ganglioside may be attributed to its relatively short carbohydrate chain, potentially rendering it inaccessible to the B-subunits. Non-binding to Neu5Acα3-lactotetraosylceramide and the Neu5Ac-globopenta/SSEA-4 ganglioside could stem from conformational constraints.

In additional binding experiments, LT-IIc demonstrated no affinity for α6-linked Neu5Ac or gangliosides with a disialo sequence. Moreover, further experiments with derivatives of Neu5Gc-GM3, wherein the carboxyl group of the sialic acid was modified to methylamide, ethylamide, or propylamide, elucidated the pivotal role of the carboxyl group in the binding process as all modifications ceased binding.

To discern the comparative binding affinity of the binding-active gangliosides, serial dilutions of these compounds on silica gel plates and in microtiter wells demonstrated LT-IIc B-subunits binding with high affinity, comparative to that of the GM1 ganglioside binding of CTB, to gangliosides with a terminal Neu5Acα3-Gal or Neu5Gcα3-Gal epitope, such as GD1a ganglioside and the

Neu5Ac-neolactohexaosylceramide. Inhibition studies with Neu5Acα3-lactose (Neu5Acα3Galβ4Glc) markedly diminished binding to gangliosides in microtiter wells, thereby demonstrating a clear dependency on the sialic acid moiety for binding capacity.

As the minimal sequence necessary for LT-IIc B-subunit binding to glycosphingolipids is also prevalent on numerous glycoproteins, the binding of LT-IIc B-subunits with this glycan epitope on fetuin (though not asialo fetuin) was anticipated. Consequently, the binding of LT-IIc to either gangliosides or sialylated glycoproteins in the small intestine may play a pivotal role in the binding of the toxin and the overarching infection mechanism.

In addition to studies on glycosphingolipids from human small intestinal cells – which illustrated binding to slow-migrating compounds and to the gangliosides GM3, GD1a, and sialyl-neolactohexaosylceramide – binding assays to glycosphingolipids from immune cells were also conducted. These binding assays displayed similar binding, recognizing compounds co-migrating with GM3 ganglioside and below, such as sialyl-neolactotetraosylceramide and sialyl-neolactohexaosylceramide.

Mirroring its related B-subunits, LT-IIc exhibits robust adjuvant properties [191, 192]. It's noteworthy that, while immune cells lack gangliosides that are based on ganglio core chains [193], they are documented to express GM3 and gangliosides featuring neolacto core chains [189, 190]. Consequently, the discernible binding of LT-IIc to Neu5Acα3-neolacto gangliosides, complemented by Neu5Ac-GM3 binding, may underscore a pivotal role in steering the immunomodulatory activities of the protein.

Structurally, despite a mere 53% amino acid homology, the crystal structure of LT-IIc is nearly identical to LT-IIb, preserving all secondary structure elements and almost all amino acids engaged in the sialic acid interaction of the primary binding sites. Indeed, electron density revealed a binding site analogue to the primary binding site of LT-IIb, where the sugar ligand (Neu5Ac-nLT) was modeled. Only two differences in the core of this conserved binding site were noted, potentially explaining the divergence in glycan binding specificity from LT-IIb B-subunits; a lysine residue instead of asparagine at position 32 of LT-IIc, and a tyrosine instead of proline at position 86.

Unlike LT-IIb, LT-IIc did not have a secondary binding site. This could be explained by major side chain alterations, such as substitution of the LT-IIb B-subunit amino acid residues Ser50, Arg51, Ala52, Lys53 and Asp54, thereby altering binding and interaction potentials.

While LT-IIc exhibited no binding to non-acid glycosphingolipids, it is noteworthy that CTB binding to non-acid fucosylated glycans has been exclusively observed when these moieties are expressed on glycoproteins. It is plausible that the few fucosylated gangliosides recognized by LT-IIc in this study is attributable to their terminal Neu5Acα3Gal sequence.

4.4 PAPER IV – DYNAMICS OF LT-IIB

The highly resolved crystal complex attained in paper I revealed the presence of two ligand binding regions on the surface of B-subunit monomers and described the atomic interactions of the protein-glycan interaction. However, based on the static nature of crystallography, the interaction visualized is a temporary moment stuck in time, which does not provide any dynamic information and may over- or underestimate binding interactions. NMR spectroscopy has been employed extensively to not only structurally determine a protein and its binding sites, but also reveal information about transient states, binding affinity, and molecular contributions to the binding [149]. However, protein-NMR does present with some challenges. The isotope labeling of proteins, as well as the backbone assignments, can be tedious and time consuming. Relatively large amounts of proteins are needed and the study of larger polypeptides is challenging due to slower tumbling, causing a fast decay of NMR signal, leading to decreased resolution and sensitivity [174]. However recent advancements have allowed for the analysis of much larger proteins than previously [175]. For instance, NMR-studies of the Shiga-like toxin from *E. coli* helped pave way for the development of effective inhibitors based on receptor analogues [138, 194].

This study attempted to unravel the intricate dynamics of the sialic acid binding mechanism associated with the B-pentamer from LT-IIb by the techniques of NMR-spectroscopy. Initially, the recombinant, ^{13}C, ^{15}N and ^{2}H labeled protein allowed for an almost complete backbone assignment, which establishes a significant stepping stone for deeper structural and dynamic explorations. Besides proline residues, asparagine 11 and arginine 12 were the only two amino acids not resolved in the 2D-TROSY spectrum. These residues are situated within an exposed loop region neighboring the primary binding site, and their missing signal possibly roots from a slow to intermediate exchange of the amide protons with the surrounding aqueous environment.

Next, chemical shift perturbation analysis allowed for the discrimination of amino acids that were most reactive upon Neu5Ac-nLT addition. Pointing once again to two separate sialic acid binding sites. The K_d-value to the primary site was estimated to be approximately 22.1 mM, while the secondary binding site had a K_d-value of approximately 31.3 mM. A challenge here was the insufficient amount of oligosaccharides added in the titration experiment, which did not completely saturate the binding sites.

An interesting observation was the major chemical shift change tied to residue methionine 37. Intriguingly, this residue did not seem to interact directly with the ligand in prior co-crystallization studies. A potential explanation could be its indirect influence, possibly stemming from conformational adjustments when the ligand binds to either of the binding sites. However, no other substantial sign of cooperative interaction was apparent.

The intricate involvement of specific amino acid residues, such as threonine 14 and the tryptophan residues, showcases the depth and complexity of the ligand binding dynamics. For instance, T14's prompt disappearance during ligand titration may be caused by strong interactions with the ligand, which was previously suggested in paper I. Although apt for weaker binding affinities, chemical shift perturbation studies do not lend itself well to stronger interactions. The slow exchange that occurs when the affinity is very high, limits the possibility of a traceable chemical shift evolution, which in turn hinders the determination of a K_d-value.

In conclusion, the spatial segregation of amino acids influenced by the ligand in this study, in tandem with the disparity in K_d-values, supports the notion of two sialic acid-binding sites on the surface of LT-II B-subunit monomers.

While the affinity of individual glycan interactions is relatively low, here reported as 22.1 mM for the primary and 31.3 mM for the secondary site, the accumulated avidity emerging from ten binding sites per B-pentamer may still infer physiologically relevant binding interactions.

This study offers a novel insight into the type II heat-labile enterotoxin family and lays the foundation for further investigations by providing an almost complete backbone assignment of the LT-IIb pentamer and dynamically profiling the sialic acid binding interactions.

5 CONCLUSION

Glycans represent a complex and crucial facet of biology that has historically been underexplored, largely due to the challenges associated with their research methodologies. Nonetheless, glycosphingolipids have served as premier tools in the study of protein-carbohydrate interactions, such as microbial adhesins. Expressed on the surfaces of eukaryotic cells, glycosphingolipids act as anchoring points for countless of microbes, thereby playing a pivotal role in infection mechanisms. Thus, research into protein-glycan interactions specific to ETEC infections not only sheds light on the intricacies of ETEC pathogenesis, but may also offer a broader model for understanding various other infections.

ETEC infections are rendered complex due to the numerous colonization factors and toxin variants expressed, with new ones being continually identified. In papers I and III, the carbohydrate binding profiles of the most recently discovered ETEC toxins, LT-IIb and LT-IIc, were extensively examined. Novel human intestinal glycosphingolipid receptors were characterized, and the ligand binding sites on the B-subunits were structurally determined. This study underlines the importance of a diverse and comprehensive screening procedure for receptor assessments. The study also highlights the extensive variability in glycan expression in nature and reinforces the call for their characterization. Additionally, it was demonstrated that subtle differences in either amino acid or glycan composition can profoundly influence their interactions.

AB_5 toxins are expressed among many different pathogens and share structural similarities [66, 69]. The receptor-binding B-subunits are especially intriguing; they can have striking similarities even at the level of binding site morphology, despite often negligible sequence homology. These similarities have allowed for the construct of hybrid B-pentamers with monomers belonging to different pathogens [195]. Research into the function and glycan-binding specificity of B-subunits has been linked to a number of areas, ranging from the development of inhibitory drugs [63] and immunomodulatory adjuvants [141, 196, 197] to their potential in drug delivery [198], therapeutics [137], cancer targeting, and diagnostic tools [199]. Beyond these applications, such studies have deepened our comprehension of microbial dynamics [140], cell membrane composition [49], vesicle genesis [129], and evolutionary facets of our glycome [3], while also equipping us with methodologies to further probe into these areas.

Thus, paper IV set out to provide more detailed insights on the receptor binding interactions of LT-IIb, identifying specific backbone assignments and pinpointing the amino acids involved in the binding. Additionally, further evidence for a secondary sialic acid-binding site was detected, as well as an estimate of the K_d-values involved in the bindings. Some signs of conformational change outside of the binding sites were observed, but more studies are needed to determine if any cooperative binding occurs.

The tropism of ETEC, mediated by its colonization factors, has historically been described as species specific. However, the main receptor for the novel CF CS30, as shown in paper II, is sulfatide – an acid glycosphingolipid that constitutes a significant portion of both human and porcine intestinal glycosphingolipid composition. This finding specifies the need to investigate further the potential of ETEC as a zoonotic pathogen and raises the question of whether treatment strategies aimed at humans should also encompass animals. Furthermore, the intricate nature of glycoconjugate interactions was emphasized as the sulfatide-binding ability of CS30-expressing *E. coli* was affected by the composition of the underlying ceramide moiety. Moreover, the sulfatide binding was negated by the competitive binding of sulfated sugars. This not only reaffirms the sulfatide binding proficiency of CS30-expressing cells but also showcases the potential of such a strategy for future therapeutic approaches that could reduce antibiotic usage.

6 FUTURE PERSPECTIVES

These studies adopted a fundamental research approach for the study of ETEC adhesins, and several future perspectives can be derived from these studies. In terms of glycosphingolipid expression, further characterization of both newly examined and previously studied tissues and cultured cells is warranted. Such investigations could offer insights into tissue and species-specific expressions, as well as variations in glycan expression over time and between individuals.

The luminal expression of the proposed glycosphingolipid receptors in these studies is a prerequisite for their accessibility to ETEC. This is worth investigating given the proposed polar expression pattern of glycosphingolipids on cell surfaces [200], and their intracellular localization in some cells [201, 202]. It is also worth noting that binding does not necessarily equate to a pathophysiological response. The role of the known receptors for type II LTs in this context remains largely unexplored.

One approach for validating the presence of a secondary binding site and investigating its contribution to binding is by NMR spectroscopy, but additional *in vitro* experiments can be conducted by site-directed mutagenesis, followed by binding and structural studies.

Paper IV, while essentially prepared for submission, remains an active project since similar NMR experiments for the B-subunits of LT-IIc and CT have been performed for comparative purposes. However, a few additional experiments would enhance these comparisons, notably using ligand-observed NMR techniques to scrutinize the specific atoms of the sugar ligands involved in the interaction. Incorporating other methods to verify dissociation constants, such as surface plasmon resonance and isothermal titration calorimetry, would also be beneficial.

The novel CF CS30, which was found commonly among ETEC isolates in symptomatic children, may also be prevalent among animals. Epidemiological efforts would therefore be of importance and such findings would have implications for future therapeutical strategies and vaccine development. However, the binding interactions in paper II were contingent on CS30 being expressed on the surface of a wild-type ETEC strain. Further research would greatly benefit from the use of a recombinantly expressed CS30, enabling more in-depth investigations, including structural studies.

ACKNOWLEDGEMENT

This journey has been a wonderful experience thanks to all the people whom I am grateful to have encountered along the way. Many of you contributed directly to my research, while others kept me company in the lunchrooms.

First of all, since my first semester of medical school, your unwavering guidance and mentorship have been paramount in shaping both my personal and professional development. Susann, you have been an amazing supervisor.

To my main co-supervisor, Mike, it has been an absolute joy to pick your brain and learn from you. And to my co-supervisors Ulf and Nicklas, thank you for your support.

My heartfelt thanks to Göran and the people at the Swedish NMR Center – Zoltán, Ulrika, and Anders – for your eagerness to teach me the ways of NMR.

I am immensely grateful for the collaborative efforts of many, particularly in the laboratory of Annabelle, Anne, Terry, and Astrid (who constantly steered me in the right direction).

Manuela, Stefan, and Susanne, you have always been more than willing to answer every question of mine, no matter its level or the frequency of which I ask it.

Marianne, your company, wisdom, and contagious outlook on life have been enriching.

To my lovely colleagues: Frida for your workout sessions, Nikita's ever-present positivity, Janarthan's creative ideas, Andrew's beer treats, Angelica's companionship, and Gudrun for standing tall as a role model. To Jin and Tina, and to everyone in the groups of Aishe, Linda and Joan – thank you for your help, companionship and coffee. Carola, Iwona, Marta, Paolo, Damasus, Finn, Ana, Simon, Petra, Ayaan, Sara, and many more – you have filled my days with laughter.

Thank you, Sara, for being an inspiration, and to everyone in your group for "adopting" me: Sinan, Stephie, Maca and Mattias for treating me as family, and most of all Licínia, for being such a good influence and keeping me grounded.

Our previous neighbors, “the people upstairs”, thank you for all the good talks, fika and advice. Especially Ana Sofia, Anandi and Taher for letting me be part of your research. And to Erik for teaching me everything I know about laboratory safety. I am also grateful to Ingela and Eva for assisting me with my teaching assignments.

To my past and present lab mates: Miralda, Klara, John, Angela, Benjamin, Catharina, Farzana and Karel, for all the fun memories and for creating an inspiring work environment.

To all my friends who still wants to hang out even though I have not given them much consideration lately. And lastly, to my dear wife, parents, and siblings – your support has been my strongest pillar.

REFERENCES

1. Kline, K.A., et al., *Bacterial adhesins in host-microbe interactions.* Cell Host Microbe, 2009. **5**(6): p. 580-92.
2. Pieters, R.J., *Carbohydrate mediated bacterial adhesion.* Adv Exp Med Biol, 2011. **715**: p. 227-40.
3. Varki, A., *Nothing in glycobiology makes sense, except in the light of evolution.* Cell, 2006. **126**(5): p. 841-5.
4. Karlsson, K.A., *Animal glycosphingolipids as membrane attachment sites for bacteria.* Annu Rev Biochem, 1989. **58**: p. 309-50.
5. Gaastra, W. and A.M. Svennerholm, *Colonization factors of human enterotoxigenic Escherichia coli (ETEC).* Trends Microbiol, 1996. **4**(11): p. 444-52.
6. Qadri, F., et al., *Enterotoxigenic Escherichia coli in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention.* Clin Microbiol Rev, 2005. **18**(3): p. 465-83.
7. Cummings, R.D., *"Stuck on sugars - how carbohydrates regulate cell adhesion, recognition, and signaling".* Glycoconj J, 2019. **36**(4): p. 241-257.
8. Varki, A., *Biological roles of glycans.* Glycobiology, 2017. **27**(1): p. 3-49.
9. Suzuki, K., T. Miura, and H. Okada, *The endothelial glycocalyx-All the same? No, it is not.* Acute Med Surg, 2023. **10**(1): p. e896.
10. Rabuka, D., et al., *Noncovalent cell surface engineering: incorporation of bioactive synthetic glycopolymers into cellular membranes.* J Am Chem Soc, 2008. **130**(18): p. 5947-53.
11. Moremen, K.W., M. Tiemeyer, and A.V. Nairn, *Vertebrate protein glycosylation: diversity, synthesis and function.* Nat Rev Mol Cell Biol, 2012. **13**(7): p. 448-62.
12. Ohtsubo, K. and J.D. Marth, *Glycosylation in cellular mechanisms of health and disease.* Cell, 2006. **126**(5): p. 855-67.
13. Linden, S.K., et al., *Mucins in the mucosal barrier to infection.* Mucosal Immunol, 2008. **1**(3): p. 183-97.
14. Johansson, M.E., et al., *Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis.* Gut, 2014. **63**(2): p. 281-91.
15. Tarbell, J.M. and L.M. Cancel, *The glycocalyx and its significance in human medicine.* J Intern Med, 2016. **280**(1): p. 97-113.
16. Wei, X., et al., *Antibody neutralization and escape by HIV-1.* Nature, 2003. **422**(6929): p. 307-12.
17. DeAngelis, P.L., J. Papaconstantinou, and P.H. Weigel, *Molecular cloning, identification, and sequence of the hyaluronan synthase gene*

from group A Streptococcus pyogenes. J Biol Chem, 1993. **268**(26): p. 19181-4.
18. Prehm, S., et al., *Antibodies against proteins of streptococcal hyaluronate synthase bind to human fibroblasts and are present in patients with rheumatic fever.* J Anat, 1995. **187 (Pt 2)**(Pt 2): p. 271-7.
19. Limoli, D.H., C.J. Jones, and D.J. Wozniak, *Bacterial Extracellular Polysaccharides in Biofilm Formation and Function.* Microbiol Spectr, 2015. **3**(3).
20. Varki A, C.R., Esko JD, Stanley P, Hart GW, Aebi M, Mohnen D, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH, editors. , *Essentials of Glycobiology 4th ed.* 2022: Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
21. Zimmerman, G.A., *Two by two: the pairings of P-selectin and P-selectin glycoprotein ligand 1.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10023-4.
22. Carlos, T.M. and J.M. Harlan, *Leukocyte-endothelial adhesion molecules.* Blood, 1994. **84**(7): p. 2068-101.
23. Apweiler, R., H. Hermjakob, and N. Sharon, *On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database.* Biochim Biophys Acta, 1999. **1473**(1): p. 4-8.
24. Hart, G.W., et al., *Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease.* Annu Rev Biochem, 2011. **80**: p. 825-58.
25. Corfield, A.P., *Mucins: a biologically relevant glycan barrier in mucosal protection.* Biochim Biophys Acta, 2015. **1850**(1): p. 236-52.
26. Hansson, G.C., *Mucins and the Microbiome.* Annu Rev Biochem, 2020. **89**: p. 769-793.
27. Luis, A.S. and G.C. Hansson, *Intestinal mucus and their glycans: A habitat for thriving microbiota.* Cell Host Microbe, 2023. **31**(7): p. 1087-1100.
28. Lindén, S.K., et al., *MUC1 limits Helicobacter pylori infection both by steric hindrance and by acting as a releasable decoy.* PLoS Pathog, 2009. **5**(10): p. e1000617.
29. Paulick, M.G. and C.R. Bertozzi, *The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins.* Biochemistry, 2008. **47**(27): p. 6991-7000.
30. Ferguson, M.A., *The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research.* J Cell Sci, 1999. **112 (Pt 17)**: p. 2799-809.
31. Abrami, L., et al., *A pore-forming toxin interacts with a GPI-anchored protein and causes vacuolation of the endoplasmic reticulum.* J Cell Biol, 1998. **140**(3): p. 525-40.

32. Kuettner, K.E. and J.H. Kimura, *Proteoglycans: an overview.* J Cell Biochem, 1985. **27**(4): p. 327-36.
33. Bourin, M.C. and U. Lindahl, *Glycosaminoglycans and the regulation of blood coagulation.* Biochem J, 1993. **289 (Pt 2)**(Pt 2): p. 313-30.
34. Bernfield, M., et al., *Functions of cell surface heparan sulfate proteoglycans.* Annu Rev Biochem, 1999. **68**: p. 729-77.
35. Shukla, D. and P.G. Spear, *Herpesviruses and heparan sulfate: an intimate relationship in aid of viral entry.* Journal of Clinical Investigation, 2001. **108**(4): p. 503-510.
36. Hallak, L.K., et al., *Glycosaminoglycan sulfation requirements for respiratory syncytial virus infection.* J Virol, 2000. **74**(22): p. 10508-13.
37. Curatolo, W., *Glycolipid function.* Biochim Biophys Acta, 1987. **906**(2): p. 137-60.
38. Duan, S. and J.C. Paulson, *Siglecs as Immune Cell Checkpoints in Disease.* Annu Rev Immunol, 2020. **38**: p. 365-395.
39. Holzl, G. and P. Dormann, *Structure and function of glycoglycerolipids in plants and bacteria.* Prog Lipid Res, 2007. **46**(5): p. 225-43.
40. Guo, Z., *The Structural Diversity of Natural Glycosphingolipids (GSLs).* J Carbohydr Chem, 2022. **41**(2-3): p. 63-154.
41. Hakomori, S.I., *Structure and function of glycosphingolipids and sphingolipids: recollections and future trends.* Biochim Biophys Acta, 2008. **1780**(3): p. 325-46.
42. Levy, M. and A.H. Futerman, *Mammalian ceramide synthases.* IUBMB Life, 2010. **62**(5): p. 347-56.
43. Fishman, P.H. and R.O. Brady, *Biosynthesis and function of gangliosides.* Science, 1976. **194**(4268): p. 906-15.
44. Miller-Podraza, H., et al., *Screening for the presence of polyglycosylceramides in various tissues: partial characterization of blood group-active complex glycosphingolipids of rabbit and dog small intestines.* Glycoconj J, 1997. **14**(2): p. 231-9.
45. Lindström, K., et al., *Blood group glycosphingolipid expression in kidney of an individual with the rare blood group A1 Le(a-b+) p phenotype: absence of blood group structures based on the globoseries.* Glycoconj J, 1996. **13**(2): p. 307-13.
46. Irie, A., et al., *The molecular basis for the absence of N-glycolylneuraminic acid in humans.* J Biol Chem, 1998. **273**(25): p. 15866-71.
47. Varki, A., *Multiple changes in sialic acid biology during human evolution.* Glycoconj J, 2009. **26**(3): p. 231-45.
48. Simons, K. and E. Ikonen, *Functional rafts in cell membranes.* Nature, 1997. **387**(6633): p. 569-72.

49. Lingwood, D. and K. Simons, *Lipid rafts as a membrane-organizing principle.* Science, 2010. **327**(5961): p. 46-50.
50. Levental, I., K.R. Levental, and F.A. Heberle, *Lipid Rafts: Controversies Resolved, Mysteries Remain.* Trends Cell Biol, 2020. **30**(5): p. 341-353.
51. Hakomori, S. and Y. Igarashi, *Functional role of glycosphingolipids in cell recognition and signaling.* J Biochem, 1995. **118**(6): p. 1091-103.
52. Lopez, P.H. and R.L. Schnaar, *Gangliosides in cell recognition and membrane protein regulation.* Curr Opin Struct Biol, 2009. **19**(5): p. 549-57.
53. Zhang, T., et al., *The Role of Glycosphingolipids in Immune Cell Functions.* Front Immunol, 2019. **10**: p. 90.
54. Breimer, M.E., et al., *Glycosphingolipids of human embryonic stem cells.* Glycoconj J, 2017. **34**(6): p. 713-723.
55. Cumin, C., et al., *Glycosphingolipids are mediators of cancer plasticity through independent signaling pathways.* Cell Rep, 2022. **40**(7): p. 111181.
56. Russo, D., et al., *Glycosphingolipid metabolism in cell fate specification.* J Cell Sci, 2018. **131**(24).
57. Breiden, B. and K. Sandhoff, *Lysosomal Glycosphingolipid Storage Diseases.* Annu Rev Biochem, 2019. **88**: p. 461-485.
58. Ho, T.W., et al., *Guillain-Barré syndrome in northern China. Relationship to Campylobacter jejuni infection and anti-glycolipid antibodies.* Brain, 1995. **118 (Pt 3)**: p. 597-605.
59. Bock, K., et al., *Specificity of binding of a strain of uropathogenic Escherichia coli to Gal alpha 1----4Gal-containing glycosphingolipids.* J Biol Chem, 1985. **260**(14): p. 8545-51.
60. Spaulding, C.N. and S.J. Hultgren, *Adhesive Pili in UTI Pathogenesis and Drug Development.* Pathogens, 2016. **5**(1).
61. Cusumano, Z.T., R.D. Klein, and S.J. Hultgren, *Innovative Solutions to Sticky Situations: Antiadhesive Strategies for Treating Bacterial Infections.* Microbiol Spectr, 2016. **4**(2).
62. Klein, R.D. and S.J. Hultgren, *Urinary tract infections: microbial pathogenesis, host-pathogen interactions and new treatment strategies.* Nat Rev Microbiol, 2020. **18**(4): p. 211-226.
63. Spaulding, C.N., et al., *Selective depletion of uropathogenic E. coli from the gut by a FimH antagonist.* Nature, 2017. **546**(7659): p. 528-532.
64. Jennings, M.P., C.J. Day, and J.M. Atack, *How bacteria utilize sialic acid during interactions with the host: snip, snatch, dispatch, match and attach.* Microbiology (Reading), 2022. **168**(3).

65. Roussel, É.H., J.; Dubreuil, J. D. , *Sulfatide from the Pig Jejunum Brush Border Epithelial Cell Surface Is Involved in Binding of Escherichia coli Enterotoxin b.* Infect Immun, 1998. **66**(12): p. 5650-5668.
66. Beddoe, T., et al., *Structure, biological functions and applications of the AB5 toxins.* Trends Biochem Sci, 2010. **35**(7): p. 411-8.
67. MacKenzie, C.R., et al., *Quantitative analysis of bacterial toxin affinity and specificity for glycolipid receptors by surface plasmon resonance.* J Biol Chem, 1997. **272**(9): p. 5533-8.
68. Kuziemko, G.M., M. Stroh, and R.C. Stevens, *Cholera toxin binding affinity and specificity for gangliosides determined by surface plasmon resonance.* Biochemistry, 1996. **35**(20): p. 6375-84.
69. Merritt, E.A.H., W. G., *AB5 toxins.* Curr Opin Struct Biol, 1995. **5**: p. 165-171.
70. Brown, P.I., et al., *The diverse landscape of AB5-type toxins.* Engineering Microbiology, 2023. **3**(4).
71. Sack, R.B., *Enterotoxigenic Escherichia coli Isolated from Patients with Severe Cholera-like Disease.* J Infect Dis, 1971. **123**(4): p. 85.
72. Gomes, T.A., et al., *Diarrheagenic Escherichia coli.* Braz J Microbiol, 2016. **47 Suppl 1**: p. 3-30.
73. Lamberti, L.M., et al., *Estimating diarrheal illness and deaths attributable to Shigellae and enterotoxigenic Escherichia coli among older children, adolescents, and adults in South Asia and Africa.* PLoS Negl Trop Dis, 2014. **8**(2): p. e2705.
74. Tickell, K.D., et al., *The effect of acute malnutrition on enteric pathogens, moderate-to-severe diarrhoea, and associated mortality in the Global Enteric Multicenter Study cohort: a post-hoc analysis.* The Lancet Global Health, 2020. **8**(2): p. e215-e224.
75. Ashkenazi, S. and E. Schwartz, *Traveler's diarrhea in children: New insights and existing gaps.* Travel Med Infect Dis, 2019: p. 101503.
76. Steffen, R., D.R. Hill, and H.L. DuPont, *Traveler's diarrhea: a clinical review.* JAMA, 2015. **313**(1): p. 71-80.
77. Svennerholm, A.M. and A. Lundgren, *Developments in oral enterotoxigenic Escherichia coli vaccines.* Curr Opin Immunol, 2023. **84**: p. 102372.
78. Lozano, R., et al., *Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010.* The Lancet, 2012. **380**(9859): p. 2095-2128.
79. Dubreuil, J.D., R.E. Isaacson, and D.M. Schifferli, *Animal Enterotoxigenic Escherichia coli.* EcoSal Plus, 2016. **7**(1).
80. Chekole, W.S., et al., *Occurrence of Escherichia coli Pathotypes in Diarrheic Calves in a Low-Income Setting.* Pathogens, 2022. **12**(1).

81. Liu, J., et al., *Development and assessment of molecular diagnostic tests for 15 enteropathogens causing childhood diarrhoea: a multicentre study.* The Lancet Infectious Diseases, 2014. **14**(8): p. 716-724.
82. Khalil, I.A., et al., *Morbidity and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: the Global Burden of Disease Study 1990-2016.* Lancet Infect Dis, 2018. **18**(11): p. 1229-1240.
83. Kotloff, K.L., et al., *Global burden of diarrheal diseases among children in developing countries: Incidence, etiology, and insights from new molecular diagnostic techniques.* Vaccine, 2017. **35**(49 Pt A): p. 6783-6789.
84. Kotloff, K.L., et al., *Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study.* The Lancet, 2013. **382**(9888): p. 209-222.
85. Rogawski, E.T., et al., *Early Antibiotic Exposure in Low-resource Settings Is Associated With Increased Weight in the First Two Years of Life.* J Pediatr Gastroenterol Nutr, 2017. **65**(3): p. 350-356.
86. Porter, C.K., et al., *A systematic review of experimental infections with enterotoxigenic Escherichia coli (ETEC).* Vaccine, 2011. **29**(35): p. 5869-85.
87. Nalin, D.R., et al., *Cholera, non-vibrio cholera, and stomach acid.* Lancet, 1978. **2**(8095): p. 856-9.
88. Kumar, P., et al., *EatA, an immunogenic protective antigen of enterotoxigenic Escherichia coli, degrades intestinal mucin.* Infect Immun, 2014. **82**(2): p. 500-8.
89. Madhavan, T.P. and H. Sakellaris, *Colonization factors of enterotoxigenic Escherichia coli.* Adv Appl Microbiol, 2015. **90**: p. 155-97.
90. von Mentzer, A., et al., *Identification and characterization of the novel colonization factor CS30 based on whole genome sequencing in enterotoxigenic Escherichia coli (ETEC).* Sci Rep, 2017. **7**(1): p. 12514.
91. Caron, J.C., L. M..; Scott J. R., *A plasmid-encoded regulatory gene, rns, required for expression of the CS1 and CS2 adhesins of enterotoxigenic Escherichia coli.* Proc Natl Acad Sci U S A, 1989. **86**: p. 963-967.
92. Svennerholm, A.M. and J. Tobias, *Vaccines against enterotoxigenic Escherichia coli.* Expert Rev Vaccines, 2008. **7**(6): p. 795-804.
93. Nataro, J.P. and J.B. Kaper, *Diarrheagenic Escherichia coli.* Clin Microbiol Rev, 1998. **11**(1): p. 142-201.

94. Isidean, S.D., et al., *A systematic review of ETEC epidemiology focusing on colonization factor and toxin expression.* Vaccine, 2011. **29**(37): p. 6167-78.
95. Shaheen, H.I., et al., *Phenotypic and genotypic analysis of enterotoxigenic Escherichia coli in samples obtained from Egyptian children presenting to referral hospitals.* J Clin Microbiol, 2009. **47**(1): p. 189-97.
96. Rivera, F.P., et al., *Genotypic and phenotypic characterization of enterotoxigenic Escherichia coli strains isolated from Peruvian children.* J Clin Microbiol, 2010. **48**(9): p. 3198-203.
97. Pieroni, P.W., E.; Paranchych, W.; Armstrong, G., *Identification of a Human Erythrocyte Receptorfor Colonization Factor Antigen I Pili Expressed by H10407 Enterotoxigenic Escherichia coli.* Infect Immun, 1988. **56**(5): p. 1334-1340.
98. Wenneras, C., J. Holmgren, and A.M. Svennerholm, *The binding of colonization factor antigens of enterotoxigenic Escherichia coli to intestinal cell membrane proteins.* FEMS Microbiol Lett, 1990. **54**(1-3): p. 107-12.
99. Jansson, L., et al., *The major subunit, CfaB, of colonization factor antigen i from enterotoxigenic Escherichia coli is a glycosphingolipid binding protein.* Infect Immun, 2006. **74**(6): p. 3488-97.
100. Turner, S.M., et al., *Weapons of mass destruction: virulence factors of the global killer enterotoxigenic Escherichia coli.* FEMS Microbiol Lett, 2006. **263**(1): p. 10-20.
101. Jansson, L., et al., *Sulfatide recognition by colonization factor antigen CS6 from enterotoxigenic Escherichia coli.* PLoS One, 2009. **4**(2): p. e4487.
102. Ahmed, T., et al., *Children with the Le(a+b-) blood group have increased susceptibility to diarrhea caused by enterotoxigenic Escherichia coli expressing colonization factor I group fimbriae.* Infect Immun, 2009. **77**(5): p. 2059-64.
103. Wang, H., et al., *Heat-Stable Enterotoxins of Enterotoxigenic Escherichia coli and Their Impact on Host Immunity.* Toxins (Basel), 2019. **11**(1).
104. Jobling, M.G. and R.K. Holmes, *Heat-Labile Enterotoxins.* EcoSal Plus, 2006. **2**(1).
105. Labrie, V.B.H.H., J.; Dubreuil, D., *Binding to sulfatide and enterotoxicity of various Escherichia coli STb mutants.* Microbiology, 2001. **147**: p. 2141-3148.
106. Bölin, I., et al., *Enterotoxigenic Escherichia coli with STh and STp genotypes is associated with diarrhea both in children in areas of endemicity and in travelers.* J Clin Microbiol, 2006. **44**(11): p. 3872-7.

107. Fleckenstein, J.M., et al., *Molecular mechanisms of enterotoxigenic Escherichia coli infection.* Microbes Infect, 2010. **12**(2): p. 89-98.
108. Field, M., et al., *Heat-stable enterotoxin of Escherichia coli: in vitro effects on guanylate cyclase activity, cyclic GMP concentration, and ion transport in small intestine.* Proc Natl Acad Sci U S A, 1978. **75**(6): p. 2800-4.
109. Foulke-Abel, J., et al., *Phosphodiesterase 5 (PDE5) restricts intracellular cGMP accumulation during enterotoxigenic Escherichia coli infection.* Gut Microbes, 2020. **12**(1): p. 1752125.
110. Lycke, N. and C. Lebrero-Fernandez, *ADP-ribosylating enterotoxins as vaccine adjuvants.* Curr Opin Pharmacol, 2018. **41**: p. 42-51.
111. Holmgren, J.S.A.M., *Immunological cross-reactivity betweenEscherichia coli heat-labile enterotoxin and cholera toxin A and B subunits.* Current Microbiology, 1979. **2**: p. 55-58.
112. Clements, J.D. and R.A. Finkelstein, *Immunological cross-reactivity between a heat-labile enterotoxin(s) of Escherichia coli and subunits of Vibrio cholerae enterotoxin.* Infect Immun, 1978. **21**(3): p. 1036-9.
113. Jobling, M.G., *The chromosomal nature of LT-II enterotoxins solved: a lambdoid prophage encodes both LT-II and one of two novel pertussis-toxin-like toxin family members in type II enterotoxigenic Escherichia coli.* Pathog Dis, 2016. **74**(3).
114. Jobling, M.G. and R.K. Holmes, *Type II heat-labile enterotoxins from 50 diverse Escherichia coli isolates belong almost exclusively to the LT-IIc family and may be prophage encoded.* PLoS One, 2012. **7**(1): p. e29898.
115. Mukhopadhyay, S. and A.D. Linstedt, *Retrograde trafficking of AB(5) toxins: mechanisms to therapeutics.* J Mol Med (Berl), 2013. **91**(10): p. 1131-41.
116. Tsai, S.C., et al., *Enhancement of choleragen ADP-ribosyltransferase activities by guanyl nucleotides and a 19-kDa membrane protein.* Proc Natl Acad Sci U S A, 1987. **84**(15): p. 5139-42.
117. Sánchez, J. and J. Holmgren, *Cholera toxin - A foe & a friend.* Indian Journal of Medical Research, 2011. **133**(2): p. 153-163.
118. Fan, E., et al., *AB(5) toxins: structures and inhibitor design.* Curr Opin Struct Biol, 2000. **10**(6): p. 680-6.
119. Sixma, T.K., et al., *Crystal structure of a cholera toxin-related heat-labile enterotoxin from E. coli.* Nature, 1991. **351**(6325): p. 371-377.
120. Holmes, R.K., E.M. Twiddy, and C.L. Pickett, *Purification and characterization of type II heat-labile enterotoxin of Escherichia coli.* Infect Immun, 1986. **53**(3): p. 464-73.
121. van den Akker, F., et al., *Crystal structure of a new heat-labile enterotoxin, LT-IIb.* Structure, 1996. **4**(6): p. 665-78.

122. Zalem, D., et al., *Characterization of the ganglioside recognition profile of Escherichia coli heat-labile enterotoxin LT-IIc.* Glycobiology, 2022. **32**(5): p. 391-403.
123. Merritt, E.A., et al., *The 1.25 A resolution refinement of the cholera toxin B-pentamer: evidence of peptide backbone strain at the receptor-binding site.* J Mol Biol, 1998. **282**(5): p. 1043-59.
124. Merritt, E.A., et al., *Structural foundation for the design of receptor antagonists targeting Escherichia coli heat-labile enterotoxin.* Structure, 1997. **5**(11): p. 1485-99.
125. Jobling, M.G., et al., *A single native ganglioside GM1-binding site is sufficient for cholera toxin to bind to cells and complete the intoxication pathway.* mBio, 2012. **3**(6).
126. Collins, B.E. and J.C. Paulson, *Cell surface biology mediated by low affinity multivalent protein-glycan interactions.* Curr Opin Chem Biol, 2004. **8**(6): p. 617-25.
127. Williams, T.L. and A.T. Jenkins, *Measurement of the binding of cholera toxin to GM1 gangliosides on solid supported lipid bilayer vesicles and inhibition by europium (III) chloride.* J Am Chem Soc, 2008. **130**(20): p. 6438-43.
128. Chong, D.C., et al., *Clathrin-dependent trafficking of subtilase cytotoxin, a novel AB5 toxin that targets the endoplasmic reticulum chaperone BiP.* Cell Microbiol, 2008. **10**(3): p. 795-806.
129. Johannes, L., *Shiga Toxin-A Model for Glycolipid-Dependent and Lectin-Driven Endocytosis.* Toxins (Basel), 2017. **9**(11).
130. Kabbani, A.M., et al., *Structured clustering of the glycosphingolipid GM1 is required for membrane curvature induced by cholera toxin.* Proc Natl Acad Sci U S A, 2020. **117**(26): p. 14978-14986.
131. Holmgren, J., *Comparison of the tissue receptors for Vibrio cholerae and Escherichia coli enterotoxins by means of gangliosides and natural cholera toxoid.* Infect Immun, 1973. **8**(6): p. 851-9.
132. Orlandi, P.A., D.R. Critchley, and P.H. Fishman, *The heat-labile enterotoxin of Escherichia coli binds to polylactosaminoglycan-containing receptors in CaCo-2 human intestinal epithelial cells.* Biochemistry, 1994. **33**(43): p. 12886-95.
133. Teneberg, S., et al., *Comparison of the glycolipid-binding specificities of cholera toxin and porcine Escherichia coli heat-labile enterotoxin: identification of a receptor-active non-ganglioside glycolipid for the heat-labile toxin in infant rabbit small intestine.* Glycoconj J, 1994. **11**(6): p. 533-40.
134. Wands, A.M., et al., *Fucosylation and protein glycosylation create functional receptors for cholera toxin.* Elife, 2015. **4**: p. e09545.
135. Fukuta, S., et al., *Comparison of the carbohydrate-binding specificities of cholera toxin and Escherichia coli heat-labile*

enterotoxins LTh-I, LT-IIa, and LT-IIb. Infect Immun, 1988. **56**(7): p. 1748-53.
136. Aerts, J., et al., *Glycosphingolipids and Infection. Potential New Therapeutic Avenues.* Front Cell Dev Biol, 2019. **7**: p. 324.
137. Lingwood, C., *Therapeutic Uses of Bacterial Subunit Toxins.* Toxins (Basel), 2021. **13**(6).
138. Kitov, P.I., et al., *Shiga-like toxins are neutralized by tailored multivalent carbohydrate ligands.* Nature, 2000. **403**(6770): p. 669-72.
139. Lebens, M., et al., *Construction and preclinical evaluation of mmCT, a novel mutant cholera toxin adjuvant that can be efficiently produced in genetically manipulated Vibrio cholerae.* Vaccine, 2016. **34**(18): p. 2121-8.
140. Patry, R.T., et al., *Bacterial AB5 toxins inhibit the growth of gut bacteria by targeting ganglioside-like glycoconjugates.* Nat Commun, 2019. **10**(1): p. 1390.
141. Norton, E.B., et al., *Characterization of a mutant Escherichia coli heat-labile toxin, LT(R192G/L211A), as a safe and effective oral adjuvant.* Clin Vaccine Immunol, 2011. **18**(4): p. 546-51.
142. Hannig, G. and S.C. Makrides, *Strategies for optimizing heterologous protein expression in Escherichia coli.* Trends Biotechnol, 1998. **16**(2): p. 54-60.
143. Baneyx, F., *Recombinant protein expression in Escherichia coli.* Curr Opin Biotechnol, 1999. **10**(5): p. 411-21.
144. Mergulhao, F.J., D.K. Summers, and G.A. Monteiro, *Recombinant protein secretion in Escherichia coli.* Biotechnol Adv, 2005. **23**(3): p. 177-202.
145. Sanchez, J. and J. Holmgren, *Recombinant system for overexpression of cholera toxin B subunit in Vibrio cholerae as a basis for vaccine development.* Proc Natl Acad Sci U S A, 1989. **86**(2): p. 481-5.
146. Lebens, M., et al., *Large-scale production of Vibrio cholerae toxin B subunit for use in oral vaccines.* Biotechnology (N Y), 1993. **11**(13): p. 1574-8.
147. Fitch, F.W., J. Winebright, and P.V. Harper, *Iodine-125 as a protein label in immunology.* Science, 1962. **135**(3508): p. 1068-9.
148. Salacinski, P.R., et al., *Iodination of proteins, glycoproteins, and peptides using a solid-phase oxidizing agent, 1,3,4,6-tetrachloro-3 alpha,6 alpha-diphenyl glycoluril (Iodogen).* Anal Biochem, 1981. **117**(1): p. 136-46.
149. Gardner, K.H. and L.E. Kay, *The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins.* Annu Rev Biophys Biomol Struct, 1998. **27**: p. 357-406.
150. Kainosho, M., et al., *Optimal isotope labelling for NMR protein structure determinations.* Nature, 2006. **440**(7080): p. 52-7.

151. Karlsson, K.A., *Preparation of total nonacid glycolipids for overlay analysis of receptors for bacteria and viruses and for other studies.* Methods Enzymol, 1987. **138**: p. 212-20.
152. Handa, S., *BLOOD GROUP ACTIVE GLYCOLIPID FROM HUMAN ERYTHROCYTES.* Jpn J Exp Med, 1963. **33**: p. 347-60.
153. Ando, S., M. Isobe, and Y. Nagai, *High performance preparative column chromatography of lipids using a new porous silica, Iatrobeads. I. Separation of molecular species of sphingoglycolipids.* Biochim Biophys Acta, 1976. **424**(1): p. 98-105.
154. Schnaar, R.L. and L.K. Needham, *Thin-layer chromatography of glycosphingolipids.* Methods Enzymol, 1994. **230**: p. 371-89.
155. Waldi, D., *[Simple quantitative rapid determination of pregnanediol on thin layer plates for early pregnancy demonstration and for supervision of the menstrual cycle in women].* Klin Wochenschr, 1962. **40**: p. 827-30.
156. Svennerholm, L., *CHROMATOGRAPHIC SEPARATION OF HUMAN BRAIN GANGLIOSIDES.* J Neurochem, 1963. **10**: p. 613-23.
157. Magnani, J.L., D.F. Smith, and V. Ginsburg, *Detection of gangliosides that bind cholera toxin: direct binding of 125I-labeled toxin to thin-layer chromatograms.* Anal Biochem, 1980. **109**(2): p. 399-402.
158. Sugita, M., et al., *The occurrence of glycosphingolipids containing mannose in the sea-water bivalve, Meretrix lusoria (Hamaguri).* J Biochem, 1985. **98**(1): p. 27-34.
159. Cervin, J., et al., *GM1 ganglioside-independent intoxication by Cholera toxin.* PLoS Pathog, 2018. **14**(2): p. e1006862.
160. Bojar, D., et al., *A Useful Guide to Lectin Binding: Machine-Learning Directed Annotation of 57 Unique Lectin Specificities.* ACS Chem Biol, 2022. **17**(11): p. 2993-3012.
161. Green, E.D. and J.U. Baenziger, *Oligosaccharide specificities of Phaseolus vulgaris leukoagglutinating and erythroagglutinating phytohemagglutinins. Interactions with N-glycanase-released oligosaccharides.* J Biol Chem, 1987. **262**(25): p. 12018-29.
162. Polito, L., et al., *Ricin: An Ancient Story for a Timeless Plant Toxin.* Toxins (Basel), 2019. **11**(6).
163. Sharon, N. and H. Lis, *History of lectins: from hemagglutinins to biological recognition molecules.* Glycobiology, 2004. **14**(11): p. 53R-62R.
164. Samuelsson, B.E., W. Pimlott, and K.A. Karlsson, *Mass spectrometry of mixtures of intact glycosphingolipids.* Methods Enzymol, 1990. **193**: p. 623-46.
165. Karlsson, K.A., *On the chemistry and occurrence of sphingolipid long-chain bases.* Chem Phys Lipids, 1970. **5**(1): p. 6-43.

166. Karlsson, H., A. Halim, and S. Teneberg, *Differentiation of glycosphingolipid-derived glycan structural isomers by liquid chromatography/mass spectrometry.* Glycobiology, 2010. **20**(9): p. 1103-16.
167. Albrecht, S., et al., *Comprehensive Profiling of Glycosphingolipid Glycans Using a Novel Broad Specificity Endoglycoceramidase in a High-Throughput Workflow.* Anal Chem, 2016. **88**(9): p. 4795-802.
168. Acquotti, D., L. Mauri, and S. Sonnino, *Nuclear Magnetic Resonance of Gangliosides.* Methods Mol Biol, 2018. **1804**: p. 241-284.
169. Geyer, R., et al., *Methylation analysis of complex carbohydrates in small amounts: capillary gas chromatography-mass fragmentography of methylalditol acetates obtained from N-glycosidically linked glycoprotein oligosaccharides.* Anal Biochem, 1983. **133**(1): p. 197-207.
170. Kendrew, J.C., et al., *A three-dimensional model of the myoglobin molecule obtained by x-ray analysis.* Nature, 1958. **181**(4610): p. 662-6.
171. Smyth, M.S. and J.H. Martin, *x ray crystallography.* Mol Pathol, 2000. **53**(1): p. 8-14.
172. Maveyraud, L. and L. Mourey, *Protein X-ray Crystallography and Drug Discovery.* Molecules, 2020. **25**(5).
173. Acharya, K.R. and M.D. Lloyd, *The advantages and limitations of protein crystal structures.* Trends Pharmacol Sci, 2005. **26**(1): p. 10-4.
174. Jiang, Y. and C.G. Kalodimos, *NMR Studies of Large Proteins.* J Mol Biol, 2017. **429**(17): p. 2667-2676.
175. Pervushin, K., et al., *Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution.* Proc Natl Acad Sci U S A, 1997. **94**(23): p. 12366-71.
176. Sprangers, R. and L.E. Kay, *Quantitative dynamics and binding studies of the 20S proteasome by NMR.* Nature, 2007. **445**(7128): p. 618-622.
177. Aguirre, C., O. Cala, and I. Krimm, *Overview of Probing Protein-Ligand Interactions Using NMR.* Curr Protoc Protein Sci, 2015. **81**: p. 17.18.1-17.18.24.
178. Williamson, M.P., *Using chemical shift perturbation to characterise ligand binding.* Prog Nucl Magn Reson Spectrosc, 2013. **73**: p. 1-16.
179. Nerli, S., A.C. McShan, and N.G. Sgourakis, *Chemical shift-based methods in NMR structure determination.* Prog Nucl Magn Reson Spectrosc, 2018. **106-107**: p. 1-25.

180. Hobbs, B., J. Drant, and M.P. Williamson, *The measurement of binding affinities by NMR chemical shift perturbation.* J Biomol NMR, 2022. **76**(4): p. 153-163.
181. Schumann, F.H., et al., *Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions.* J Biomol NMR, 2007. **39**(4): p. 275-89.
182. Nada, R.A., et al., *Discovery and phylogenetic analysis of novel members of class b enterotoxigenic Escherichia coli adhesive fimbriae.* J Clin Microbiol, 2011. **49**(4): p. 1403-10.
183. Berenson, C.S., et al., *Mammalian cell ganglioside-binding specificities of E. coli enterotoxins LT-IIb and variant LT-IIb(T13I).* Glycobiology, 2010. **20**(1): p. 41-54.
184. Berenson, C.S., et al., *Ganglioside-binding specificities of E. coli enterotoxin LT-IIc: Importance of long-chain fatty acyl ceramide.* Glycobiology, 2013. **23**(1): p. 23-31.
185. Nawar, H.F., et al., *LT-IIc, a new member of the type II heat-labile enterotoxin family encoded by an Escherichia coli strain obtained from a nonmammalian host.* Infect Immun, 2010. **78**(11): p. 4705-13.
186. Iwamori, M. and Y. Nagai, *Ganglioside composition of rabbit thymus.* Biochim Biophys Acta, 1981. **665**(2): p. 205-13.
187. Stroud, M.R., et al., *Monosialogangliosides of human myelogenous leukemia HL60 cells and normal human leukocytes. 1. Separation of E-selectin binding from nonbinding gangliosides, and absence of sialosyl-Le(x) having tetraosyl to octaosyl core.* Biochemistry, 1996. **35**(3): p. 758-69.
188. Stroud, M.R., et al., *Monosialogangliosides of human myelogenous leukemia HL60 cells and normal human leukocytes. 2. Characterization of E-selectin binding fractions, and structural requirements for physiological binding to E-selectin.* Biochemistry, 1996. **35**(3): p. 770-8.
189. Macher, B.A., et al., *Isolation and structural characterization of human lymphocyte and neutrophil gangliosides.* J Biol Chem, 1981. **256**(4): p. 1968-74.
190. Kiguchi, K., C.B. Henning-Chubb, and E. Huberman, *Glycosphingolipid patterns of peripheral blood lymphocytes, monocytes, and granulocytes are cell specific.* J Biochem, 1990. **107**(1): p. 8-14.
191. Nawar, H.F., et al., *LT-IIc, a new member of the type II heat-labile enterotoxin family, exhibits potent immunomodulatory properties that are different from those induced by LT-IIa or LT-IIb.* Vaccine, 2011. **29**(4): p. 721-7.
192. Hu, J.C., et al., *Intradermal administration of the Type II heat-labile enterotoxins LT-IIb and LT-IIc of enterotoxigenic Escherichia coli*

enhances humoral and CD8+ T cell immunity to a co-administered antigen. PLoS One, 2014. **9**(12): p. e113978.
193. Yohe, H.C., et al., *The major gangliosides of human peripheral blood monocytes/macrophages: absence of ganglio series structures.* Glycobiology, 2001. **11**(10): p. 831-41.
194. Shimizu, H., et al., *Solution structure of the complex between the B-subunit homopentamer of verotoxin VT-1 from Escherichia coli and the trisaccharide moiety of globotriaosylceramide.* Biochemistry, 1998. **37**(31): p. 11078-82.
195. Angstrom, J., et al., *Novel carbohydrate binding site recognizing blood group A and B determinants in a hybrid of cholera toxin and Escherichia coli heat-labile enterotoxin B-subunits.* J Biol Chem, 2000. **275**(5): p. 3231-8.
196. Nawar, H.F., et al., *Binding to gangliosides containing N-acetylneuraminic acid is sufficient to mediate the immunomodulatory properties of the nontoxic mucosal adjuvant LT-IIb(T13I).* Clin Vaccine Immunol, 2010. **17**(6): p. 969-78.
197. Connell, T.D., *Cholera toxin, LT-I, LT-IIa and LT-IIb: the critical role of ganglioside binding in immunomodulation by type I and type II heat-labile enterotoxins.* Expert Rev Vaccines, 2007. **6**(5): p. 821-34.
198. Lichtenstein, B.R. and B. Hocker, *Engineering an AB(5) Protein Carrier.* Sci Rep, 2018. **8**(1): p. 12643.
199. Shewell, L.K., et al., *Serum Neu5Gc biomarkers are elevated in primary cutaneous melanoma.* Biochem Biophys Res Commun, 2022. **642**: p. 162-166.
200. Hansson, G.C., *The subcellular localization of the glycosphingolipids in the epithelial cells of rat small intestine.* Biochim Biophys Acta, 1983. **733**(2): p. 295-9.
201. Barone, A., et al., *Sialyl-lactotetra, a novel cell surface marker of undifferentiated human pluripotent stem cells.* J Biol Chem, 2014. **289**(27): p. 18846-59.
202. Symington, F.W., et al., *Intracellular localization of lactosylceramide, the major human neutrophil glycosphingolipid.* J Biol Chem, 1987. **262**(23): p. 11356-63.