CHARLES UNIVERSITY
FACULTY OF PHARMACY IN HRADEC KRÁLOVÉ
Department of Pharmaceutical Chemistry and Pharmaceutical Analysis

UNIVERSITY OF GOTHENBURG
SAHLGRENSKA ACADEMY
Institute of Biomedicine
Department of Medicinal Biochemistry and Cell Biology
Department of Microbiology and Immunology

Carbohydrate recognition of the novel colonization factor CS30 of enterotoxigenic *E. coli*

DIPLOMA THESIS

Supervisors: PharmDr. Marta Kučerová, Ph.D.
Prof. Susann Teneberg, MD PhD
Astrid von Mentzer, PhD

Žofia Chrienová

ERASMUS Project
August 28th – December 31st
2017

DECLARATION

I declare that this thesis is my original copyrighted work. All literature and other sources of information I used are listed in references and properly cited. This thesis has not been used to gain different or equal degree.

Prehlasujem, že táto práca je mojím autorským dielom. Všetka literatúra a ďalšie zdroje, z ktorých som čerpala sú uvedené v zozname použitej literatúry a v práci riadne odcitované. Táto práca nebola použitá k získaniu iného či rovnakého titulu.

Hradec Králové 2018

Žofia Chrienová

ACKNOWLEDGEMENTS

I would like to thank the members of the research group of Prof. Susann Teneberg, Department of Medicinal Biochemistry and Cell Biology at the University of Gothenburg, for the great opportunity and lifetime experience I have been given.

Special thanks go to Prof. Susann Teneberg, MD PhD, for her guidance through the very new world of glycobiology and encouragement every time it was needed.

I would like to thank all the researchers from the Department of Microbiology and Immunology at the University of Gothenburg, especially members of ETEC and cholera lab, where I found a new and exciting world of gene cloning and where I felt very welcomed.

Special thanks go to Astrid von Mentzer, PhD, for her patience, guidance and knowledge, and for making every day in the lab special and exciting.

I would like to thank Stefan Nordqvist, PhD, Gudrun Wiklund, Susanne Källgård, and Michael Lebens, PhD, for helping me with every difficulty that occurred with my TOP10 strain.

My sincerest thanks go to PharmDr. Marta Kučerová, Ph.D., for her guidance and support.

Another thanks go to Mgr. Eugenia Nepovimová, Ph.D., for previous years in the lab together during which I found my love for science.

Last but not least, I would like to thank my parents, family and friends, for their care, love and support during my whole studies.

ABSTRACT

Charles University
Faculty of Pharmacy in Hradec Králové
Department of Pharmaceutical Chemistry and Pharmaceutical Analysis

University of Gothenburg
Sahlgrenska Academy
Institute of Biomedicine, Department of Medicinal Biochemistry and Cell Biology,
Department of Microbiology and Immunology

Student: Žofia Chrienová
Supervisors: PharmDr. Marta Kučerová, Ph.D.
Prof. Susann Teneberg, MD PhD
Astrid von Mentzer, PhD

Title of diploma thesis: Carbohydrate recognition of the novel colonization factor CS30 of enterotoxigenic *E. coli*

Enterotoxigenic *Escherichia coli* (ETEC) is a major cause of diarrhoea in developing countries and travellers to these areas, as well as in many farm animals. The main ETEC virulence factors are two enterotoxins and outer membrane proteins called colonization factors that mediate the adherence of the bacteria to the host epithelial cell surface in the small intestine. Carbohydrate recognition of the recently identified colonization factor CS30 was characterised using binding studies with the mixture of glycosphingolipids isolated from the human and porcine intestine. The CS30 binding compound was identified as a mixture of sulfatides with different ceramide species by mass spectrometry and a binding assay with monoclonal antibodies directed towards SO_3-3Galβ. Further testing confirmed that CS30 binds to the glycosphingolipid with a terminal SO_3-3Galβ and prefers sulfatides with the longer fatty acid chain. To prevent false positive results in the binding studies, a recombinant strain TOP10-CS30 containing CS30 operon that harbours seven genes *csmA-G* was constructed. RT-PCR confirmed transcription of all seven genes encoding the CS30 fimbriae, but transmission electron microscopic imaging showed no fimbrial structures on the cell surface, suggesting that the building of the fimbriae requires native regulatory genes.

Keywords: ETEC, diarrhoea, CS30, carbohydrate recognition, sulfatide, TOP10-CS30

ABSTRAKT

Univerzita Karlova

Farmaceutická fakulta v Hradci Králové

Katedra farmaceutickej chémie a farmaceutickej analýzy

Univerzita Göteborg

Sahlgrenska Academy

Ústav biomedicíny, Katedra lekárskej biochémie a bunečnej biológie,

Katedra mikrobiológie a imunológie

Študentka: Žofia Chrienová

Školitelia: PharmDr. Marta Kučerová, Ph.D.
Prof. Susann Teneberg, MD PhD
Astrid von Mentzer, PhD

Názov diplomovej práce: Rozpoznávanie nového kolonizačného faktora enterotoxigénnej *E. coli* CS30 sacharidmi

Enterotoxigénna *Escherichia coli* (ETEC) je najčastejší patogén spôsobujúci hnačku u ľudí žijúcich a cestujúcich do rozvojových krajín. Rovnako spôsobuje hnačku u mnohých domácich zvierat. Dvomi hlavnými virulentnými faktormi ETEC sú enterotoxíny a proteíny exprimované na jej vonkajšej membráne, nazývané kolonizačné faktory, pomocou ktorých baktéria adheruje na povrch epiteliálnych buniek hostiteľa v tenkom čreve. Väzba nedávno objaveného kolonizačného faktoru CS30 na sacharidovú zložku jeho receptoru bola charakterizovaná pomocou väzbových testov so zmesou glykosfingolipidov izolovaných z tenkého čreva človeka a prasaťa. Látka viažuca sa na CS30, bola pomocou hmotnostnej spektrometrie identifikovaná ako zmes sulfatidov s rôznou ceramidovou časťou. Výsledky väzbových testov potvrdili preferencie CS30 ku glykosfingolipidom s terminálnou SO_3-3Galβ skupinou a ďalšie testy poukázali na preferencie k sulfatidom s dlhším reťazcom mastnej kyseliny. Aby sa zabránilo falošne pozitívnym výsledkom, bola vyvinutá snaha o vytvorenie rekombinantného kmeňa TOP10-CS30, ktorý by exprimoval iba CS30 (kódovaný génmi *csmA-G*). RT-PCR síce potvrdila, že vytvorený kmeň je nositeľom *csmA-G* genotypovo, TEM však výskyt fimbrií na povrchu baktérie vyvrátila.

Kľúčové slová: ETEC, hnačka, CS30, rozpoznávanie sacharidmi, sulfatid, TOP10-CS30

TABLE OF CONTENTS

1 LIST OF ABBREVIATIONS 8

2 BACKGROUND 9

2.1 Introduction to Glycobiology 9

2.1.1 Glycosphingolipids 10

2.1.1.1 Structure and classification 10

2.1.1.2 Biosynthesis and degradation 12

2.1.1.3 Biological and pathological function 12

2.1.1.4 Glycan-binding proteins 13

2.2 *Escherichia coli* 13

2.2.1 Pathogenic *E. coli* 14

2.2.2 ETEC 16

2.2.2.1 Epidemiology 16

2.2.2.2 Symptoms, onset and treatment 16

2.2.2.3 Diagnosis 17

2.2.2.4 Pathogenesis 17

2.2.2.5 Serotypes 18

2.2.2.6 Heat-labile and heat-stable enterotoxins 18

2.2.2.7 Colonization factors 19

2.2.2.8 CS30 20

2.2.2.9 Other virulence structures 22

2.3 Theoretical background about gene cloning methods 23

2.3.1 Amplification of DNA fragments - Polymerase chain reaction (PCR) 24

2.3.2 Vectors 25

2.3.2.1 Plasmid vectors 25

2.3.3 Digestion of DNA-restriction enzymes 26

2.3.4 Ligation of inserted DNA and vector 27

2.3.5 Transport of recombinant DNA into the cell: transformation and electroporation 28
3 AIMS 29
4 MATERIALS AND METHODS 30
4.1 CS30 binding to glycosphingolipids 30
4.1.1 Thin-layer chromatography 30
4.1.2 Culture and labelling of bacteria 30
4.1.3 Chromatogram binding assay 31
4.1.4 Anti-sulfatide MAbs 31
4.1.5 Separation of the acid glycosphingolipids fractions from human intestine 32
4.1.6 Mass spectroscopy 32
4.2 Recombinant strain TOP10-CS30 33
4.2.1 Construction of recombinant strain TOP10-CS30 33
4.2.2 Fragment length and restriction enzyme size analyses 37
4.2.3 Expression of CS30 37
4.2.4 Negative staining of TOP10-CS30 and transmission electron microscopy (TEM) 39
5 RESULTS 40
6 DISCUSSION 50
7 CONCLUSION 53
8 REFERENCES 54

1 LIST OF ABBREVIATIONS

Cer	ceramide
CF	colonization factor
CFA/I	colonization factor antigen I
CFTR	cystic fibrosis transmembrane regulator
CS30	coli surface antigen 30
ER	endoplasmic reticulum
ESI	electrospray ionization
ETEC	enterotoxigenic *Escherichia coli*
Gal	galactose
GalNAc	*N*-acetylgalactosamine
gH_2O	genetic water
Glc	glucose
GlcNAc	*N*-acetylglucosamine
GM1	monosialotetrahexosylganglioside
GSL	glycosphingolipid
Hex	hexose
HGT	horizontal gene transfer
HPLC	high performance liquid chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Luria-Bertani
LC	liquid chromatography
LT	heat-labile enterotoxin
m/z	mass-to-charge
MGEs	mobile genetic elements
MS	mass spectrometry
PBS	phosphate buffered saline
BSA	bovine serum albumin
PCR	polymerase chain reaction
ST/STh/STp	heat-stable enterotoxin (STh: human; STp: porcine)
TLC	thin-layer chromatography

2 BACKGROUND

2.1 Introduction to Glycobiology

The central paradigm of molecular biology saying that biological information flows from DNA to RNA to protein may evoke the conclusion that these molecules play the most important role in the makeup of cells, tissues and organisms. In fact, the complexity of all living organisms requires lipids and carbohydrates as well. Alongside with their other functions (structural, signalling, energy preserving, etc.), lipids and carbohydrates are involved in the posttranslational modifications of proteins and thus explain how only small number of genes can generate the vast biological complexities in organisms.

Monosaccharides or oligosaccharides covalently attached on the surface of cells, which can be referred to as "glycans", mediate a variety of cell-cell, cell-matrix, and cell-molecule interactions critical to the functions of multicellular organisms, and they also mediate interaction between different organisms. All cells are covered with dense and diverse types of glycans, called glycocalyx.

Glycobiology is the study of the structure, biosynthesis, function, and biochemistry of saccharides, and their conjugates with proteins and lipids, and the proteins that recognise them. Mammalian glycans usually consist of the following monosaccharide units: mannose (Man), galactose (Gal), glucose (Glc), fucose (Fuc), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc), *N*-acetylneuraminic acid (Neu5Ac), and *N*-glycolylneuraminic acid (Neu5Gc). The glycans can be linear or branched, while all units are linked together by α- or β-glycosidic linkages. They can be modified by different substituents, for example, acetylation and sulfation, giving the molecule the possibility to form more structural combinations. The primary units can be synthesised in cells or regained from the environment. They are later activated into nucleotide sugars or lipid-linked sugars and transported to the place where they are used as donors for glycan synthesis (outer membrane or lumen of the endoplasmic reticulum (ER)-Golgi pathway). Glycans are recognised by specific intrinsic glycan-binding proteins as well as by extrinsic glycan-binding proteins of e.g. pathogens.

Their function falls into one of five categories:

1. Structural components of, for example, cell walls.
2. Modifying protein solubility and stability.
3. Directing intracellular or extracellular trafficking of glycoconjugates.
4. Mediating and modulating cell-cell or cell-matrix interactions.
5. Mediating and modulating intracellular and extracellular signalling.

The major classes that glycoconjugates can be divided into are glycoproteins, proteoglycans, and glycolipids. Glycoproteins are a group of proteins carrying one or more covalently linked glycan(s) via N or O linkages. N-glycans are covalently linked to the side chain of asparagines of a polypeptide chain and often consist of a pentasaccharide core region (oligomannose type, complex type, or hybrid type). O-glycans are linked to the polypeptide chain via *N*-acetylgalactosamine to the side chain of serine or threonine.

Proteoglycans have one or more glycosaminoglycan chains attached to a main protein via a core region ending with a xylose binding to a serine.

Glycolipids can be divided into several groups: glycosylphosphatidylinositols that serve as membrane anchors, lipopolysaccharides of Gram-negative bacteria, glycoglycerolipids that are predominant lipids in chloroplasts of plants and eukaryotic algae and have their glycans linked to the C-3 hydroxyl of diacylglycerol, and glycosphingolipids, the major glycolipids of animals. Glycosphingolipids (GSLs) consist of glycan that links to the terminal hydroxyl group of ceramide via glucose or galactose[1].

2.1.1 Glycosphingolipids

2.1.1.1 Structure and classification

Glycosphingolipids can be found in the vast range of organisms from bacteria to human. Their structure consists of the ceramide component: a long chain sphingoid base in amide linkage to a fatty acid, linked with different types of glycans. The sphingoid bases can be three types: sphingosine, sphinganine and phytosphingosine (Fig. 1). Although, the ceramide can vary in sphingoid base, length, saturation of fatty acid or hydroxylation, the major classification of the glycosphingolipids is based on the glycans. The first carbohydrate attached to ceramide can be β-linked glucose,

galactose or its analogue with sulfate at the C-3 hydroxyl. Another sub-classification divides glycosphingolipids into neutral (no ionic groups or charged sugars), and acid, which can be sialylated (also known as “gangliosides” regardless of their core structure mentioned later) or sulfated. Further extension of glycan chain produces series of core structure subfamilies establishing the nomenclature of glycosphingolipids (Table 1).

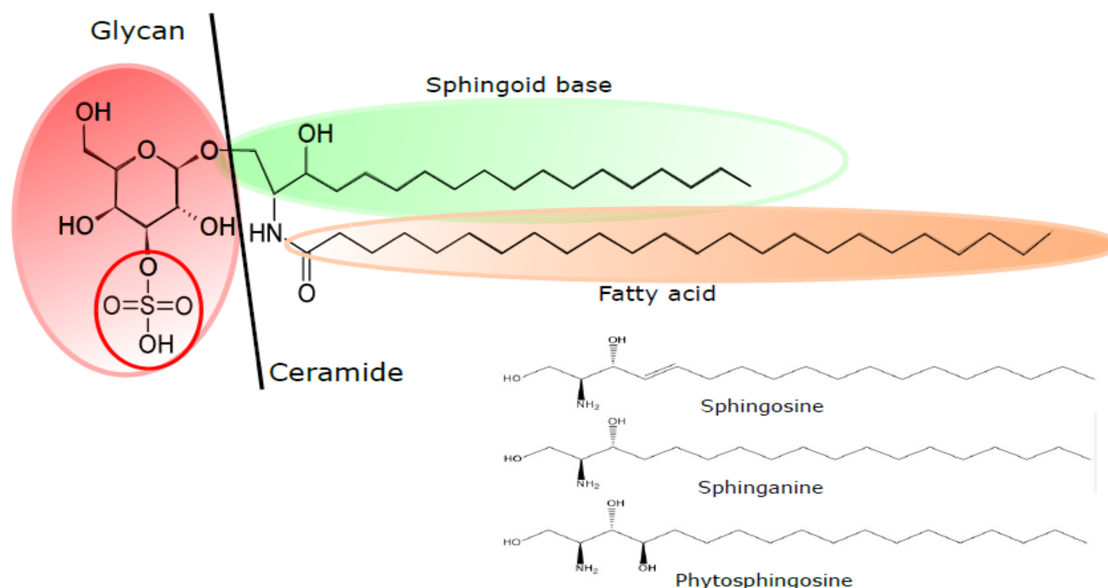

Figure 1: Structure of sulfatide. Three components of glycosphingolipid molecule shown on the structure of sulfatide: fatty acid part (orange), sphingoid base (green), and glycan part (red). Different types of sphingoid bases are shown at the bottom of the picture.

Table 1: Glycan core structures. GSL subfamilies show tissue-type specificity, for example in mammals, lacto-series occur in secretory organs, neolacto-series are expressed on hematopoietic cells, ganglio-GSLs are typically found in brain, and isoglobo/globo-series might be expressed on erythrocytes (table adapted from Schnaar *et al.*[2]).

Subfamily	Structure	Abbreviation
Lacto	GlcNAcβ1-3Galβ1-4GlcβCer Galβ1-3GlcNAcβ1-3Galβ1-4GlcβCer	Lc_3Cer Lc_4Cer
Neolacto	Galβ1-4GlcNAcβ1-3Galβ1-4GlcβCer Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcβCer	nLc_4Cer nLc_6Cer
Ganglio	GalNAcβ1-4Galβ1-4GlcβCer Galβ1-3GalNAcβ1-4Galβ1-4GlcβCer	Gg_3Cer Gg_4Cer
Globo	Galα1-4Galβ1-4GlcβCer GalNAcβ1-3Galα1-4Galβ1-4GlcβCer	Gb_3Cer Gb_4Cer
Isoglobo	Galα1-3Galβ1-4GlcβCer GalNAcβ1-3Galα1-3Galβ1-4GlcβCer	iGb_3Cer iGb_4Cer

2.1.1.2 Biosynthesis and degradation

The biosynthesis of glycosphingolipids occurs in two different ways, depending on the first sugar added to ceramide. GlcCer synthesis occurs on the cytoplasmic face of the ER and then flips into the Golgi lumen where the GlcCer is elongated by different glycosyltransferases that transfer individual sugars from nucleotide sugar donors. GalCer synthesis starts on the luminal face of the ER and then without previous flipping traffics through the Golgi apparatus where it is elongated by glycosyltransferases or where it may be sulfated. The ceramide is synthesised by condensation of serine and palmityl-CoA to form 3-ketosphinganine, which is reduced to sphinganine. The reaction is catalysed by the enzyme serine palmitoyltransferase. The composition of glycosphingolipids occurring on a cell depends on the glycosyltransferases expressed by the cell.

Degradation of the glycosphingolipids on the outer side of the plasma membrane begins with their internalization, along with other membrane structures, in vesicles that then fuse with endosomes. These structures fuse with primary lysosomes containing several hydrolases. They break glycosphingolipids into smaller structures with the assistance of the GM2-activator protein and four saposins, which are then available for reuse[2].

2.1.1.3 Biological and pathological function

The glycosphingolipids are assembled in cell membrane in small lateral micro-domains, also called “lipid rafts”. These lipid rafts are believed to contain extensive amounts of lipid molecules, including cholesterol, sphingomyelin, and glycosphingolipids, as well as a few protein molecules that act as mediators in the information flow from the outer to the inner side of the cell. However, phospholipids are excluded from the lipid rafts. The lipid rafts may also contain several receptors, for example epidermal growth factor receptor, and insulin receptor, thus the signalling function of these receptors is influenced by glycolipids. The modulation of protein function by glycosphingolipids in the same plasma membrane is one of the two major categories into which their function can be placed. The second category includes the glycosphingolipids that mediate cell-cell interactions through binding to complementary structures on the neighbouring plasma membrane. This function plays important role in the development at the whole-organism level.[2]

2.1.1.4 Glycan-binding proteins

Many specific biological functions of glycans are mediated through recognition by glycan-binding proteins. These can be classified into two groups: lectins and glycosaminoglycan-binding proteins. Different groups of lectins are defined by their carbohydrate-recognition domains (CRDs) that interact and bind to specific glycans. The bacterial lectins can be used as an example of glycan-binding proteins[3]. The bacterial lectins fall into two categories: either they are lectins (adhesins) on the bacterial surface that assist in adhesion and colonization, or they are bacterial toxins. The PapG adhesin located on the tip of P-fimbriae of uropathogenic *Escherichia coli* binds to Galα4Gal-containing glycosphingolipids and thereby initiates the infection. The FimH adhesin part of the Type I fimbriae expressed by various *E. coli* strains, is also a glycan-binding protein, and binds to mannose. These are only a few examples of known bacterial carbohydrate-recognising adhesins. The adhesins presented on enterotoxigenic *E. coli* (ETEC) are called colonization factors[4], and more information about them will be provided in the following chapter.

2.2 *Escherichia coli*

Escherichia coli (E. coli) is a Gram-negative rod-shaped bacterium from the family *Enterobacteriaceae*. Its isolation and characterisation was first reported by Theodor Escherich in 1885 when he isolated it from infant stool (publication reprinted in English[5]). Over 130 years later, it is known as a harmless commensal of the gastrointestinal tract in human and many other animal species and is used as a useful tool in many laboratories worldwide. *E. coli* is able to grow aerobically and anaerobically at 37 °C and can be either motile (with flagella) or non-motile[6]. *E. coli* colonizes the human gastrointestinal tract within a few hours after birth. Despite its utility and frequent symbiotic relationship with humans, *E. coli* can through gene acquisition or loss become an adapted pathogen that causes disease in humans by colonizing the gastrointestinal tract or extraintestinal sites such as the urinary tract and may also enter both the blood stream and central nervous system[7].

2.2.1 Pathogenic *E. coli*

Commensal *E. coli* strains cause disease only in immune compromised hosts or in case that gastrointestinal barriers are interrupted. In the normal state these bacteria coexist with the host with mutual benefit for many years[6].

E. coli genomes are split between the core genome (shared, conserved genes) and the accessory genome which consist of a flexible pool of genes. The pathogenic ability of *E. coli* is obtained by the flexible gene pool. *E. coli* are highly diverse in their genome content, sharing less than half of their genes. Between 1,500 to 2,000 genes have been identified as the core genome, depending on the strains analysed. The virulence genes are part of the accessory genome and are often located on plasmids[8]. Plasmids, bacteriophages, insertion sequences, and transposons are mobile genetic elements (MGEs) by which genes can be transfer from cell to cell and either integrate into the host chromosome or self-replicate in the host cell. MGEs encode enzymes and other proteins that mediate the movement of DNA: transposases, site-specific recombinases, etc. There are three different mechanism by MGEs may be transferred: by transformation (the uptake of DNA from the environment by cells), transduction (transfer of the DNA by bacteriophages), and conjugation (cell-to-cell exchange of genetic information from donor bacterium carrying a conjugative plasmid or chromosomally integrated conjugative elements to recipient bacterium using pilus encoded by *tra* genes)[9]. Virulence and other genes can be moved through all these mechanisms, called horizontal gene transfer (HGT), and can result in strains with advantageous traits, for example new virulence factors or the ability to use a different energy source. The most successful combinations may result in bacteria that are capable of surviving in a particular environmental niche and causing disease. Generally, *E. coli* pathotypes can cause three different diseases: diarrhoeal disease, urinary tract infections and sepsis/meningitis[7].

Intestinal pathogens are divided into six categories: enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC), and diffusely adherent *E. coli* (DEAC) (Table 2 and Fig. 2). Extraintestinal infections can be caused either by uropathogenic *E. coli* (UPEC) or meningitis-associated *E. coli* (MNEC)[7].

Table 2: Overview of enteric pathotypes of *E. coli,* adapted from Croxen *et al.*[6].

Pathotype	Site of colonization	Disease	Pathogenesis
EAEC	Small intestine and colon	Travellers´ diarrhoea, persistent diarrhoea	Adhere to intestinal epithelia in a thick biofilm, secrete enterotoxins and cytotoxins
EHEC	Colon	Bloody diarrhoea, haemorrhagic colitis	Adhere to enterocytes, destroy microvillar architecture, inflammatory response, elaboration of Shiga toxin (Stx)
EIEC	Colon	Bacillary dysentery	Invade epithelial cell, lyse the phagosome and move through the cell
EPEC	Small intestine	Profuse watery diarrhoea, potent. fatal for infants	Adhere to enterocytes, destroy microvillar architecture, inflammatory response
ETEC	Small intestine	Travellers´ diarrhoea, persist. watery diarrhoea	Adhere to enterocytes through colonization factors (CFs), secrete heat-labile (LT) and heat-stable (ST) enterotoxins
DAEC	Small intestine	Persistent watery diarrhoea	Form a signal transduction effect in enterocytes that leads to the growth of long finger-like cellular projection

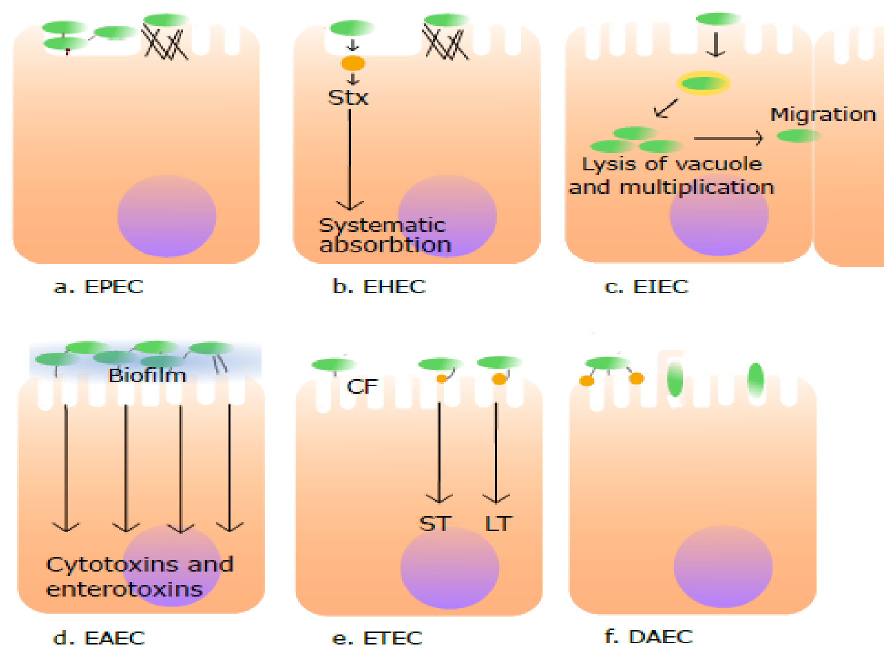

Figure 2: Enteric pathotypes of *E. coli* (adapted from Kaper *et al.*[7])**:**
a) Enteropathogenic *E. coli* (EPEC): After adherence to host enterocytes, it effaces microvilli and forms characteristic attaching and effacing lesions, cytoskeletal derangements and pedestal formation that are accompanied by inflammation response.
b) Enterohaemorrhagic *E. coli* (EHEC): After adherence it effaces microvilli in colon similarly to EPEC and produces LEE-positive Shiga toxin.
c) Enteroinvasive *E. coli* (EIEC): Intracellular pathogen penetrates the intestinal cells, induces lyses of phagosome and spreads laterally in colonocytes.
d) Enteroaggregative *E. coli* (EAEC): Bacteria form biofilm on the mucosa and secrete cytotoxins and enterotoxins.
e) Enterotoxigenic *E. coli* (ETEC): Adhere to surface using colonization factors and produce heat-labile and/or heat-stable enterotoxins.
f) Diffusely adherent *E. coli* (DAEC): Elicits signal transduction that leads to the growth of finger-like cellular projections.

2.2.2 ETEC

Enterotoxigenic *E. coli* is a major cause of diarrhoea in people in developing countries and travellers to endemic areas[10]. It is considered the most common diarrhoeagenic pathotype in developing countries of the six previously described intestinal pathotypes[11]. In addition, ETEC is the most common cause of *E. coli*-induced diarrhoea in farm animals, especially neonatal pigs, cattle, sheep, and dogs[12]. The main virulence factors of ETEC are the heat-stable (ST) or heat-labile (LT) enterotoxin and colonization factors (CFs).

2.2.2.1 Epidemiology

According to WHO, diarrhoea is often associated with malnutrition and is the second most common cause of the death of children in developing countries, estimated to cause 18% of childhood mortality[13,14]. There are between 280 to 400 million ETEC-induced diarrhoea cases per year, from which 325,000 children under the age of five die[11,15]. In addition, 46,6 million children under the age of four are asymptomatic carriers, who may increase the spreading of the infection[16]. ETEC is also the most common etiological agent related to travellers' diarrhoea in specific regions of Latin America, Africa, and Asia[17].

ETEC infection can be transmitted through the faecal-oral route, usually from contaminated water or food. The infectious dose is between 10^6 and 10^8 cells, which is a relatively high dose in comparison to other pathotypes, although it may be lower in the young, elderly or malnourished individuals[18]. There have been several outbreaks in Europe and North America during the last 20 years[19,20], however, ETEC is mainly endemic in developing countries where lack of clean water and proper sanitation allow it to spread easily.

2.2.2.2 Symptoms, onset and treatment

Although the human body has an effective defence system including peristalsis, gastric acid, mucus, bile salts, normal flora, and secretory IgA, ETEC can cause diarrhoea ranging from mild to severe, very similar to diarrhoea caused by *Vibrio cholerae*, with a high risk of dehydration[21]. Infection can be accompanied by headaches, abdominal pain, fever, and nausea or vomiting. The onset is very quick, with duration of symptoms usually about five days with a risk of prolongation for more than a week[10].

Complications that can follow ETEC-related diarrhoea are malnutrition and implication for growth in children under age of two in developing countries[22] or developing post-infectious irritable bowel syndrome[23].

The disease is usually self-limiting if rehydration and alimentation through diet is sufficient. In severe cases intravenous rehydration can be used[10]. Loperamide may reduce the number of stools in adults[24]. To shorten the duration of infection fluoroquinolones, rifaximin or azithromycin can be used. However, risk and benefit ratio should be considered for the possibility of increasing selection of antibiotic-resistant strains and developing secondary infections[24].

2.2.2.3 Diagnosis

Diagnosis of ETEC depends upon the recognition of the enterotoxins it produces. GM1 enzyme-linked immunosorbent assays (GM1-ELISA) using monoclonal antibodies against ST and LT are methods widely used nowadays. The ST GM1-ELISA method is based on the ability of STa to inhibit the binding of anti-ST antibody to solid-phase ST, which is obtained by binding of ST-CTB conjugate (ST covalently bonded to the cholera B subunit)[25]. The LT GM1-ELISA also uses GM1 ganglioside as a specific sorbent for enterotoxin. However, this method is based on the binding of LT to solid-phase GM1[26,27]. Detection of colonization factors includes genotypic methods, i.e. DNA probes (with either radioactive or nonradioactive detections) and PCR, or phenotypic methods, such as the dot blot assays with several anti-CF monoclonal antibodies binding to CFs expressed on the bacteria bonded to a nitrocellulose membrane[27,28].

The method of choice depends on investigator and laboratory. Due to the fact, that the virulence factors are encoded on plasmids that can be lost or become silent (due to the loss of regulatory genes)[27], the genotypic and phenotypic diagnostic methods combined are the best option. Unfortunately, there are no simple methods that can be used to identify ETEC in field sites in developing countries[25,27,29].

2.2.2.4 Pathogenesis

ETEC infections begin with adherence to the epithelial surface of the small intestine with bacterial outer membrane proteins called colonization factors. Adherence is mediated by bonds between CFs and different receptors on host cell, mostly glycoproteins and glycosphingolipids, for example fibronectin and sulfatide

(SO_3-3Galβ1ceramide)[30,31]. After ETEC binds to cells, it releases enterotoxins: heat-labile enterotoxins and heat-stable enterotoxins, which incidence can vary: some strains express only LT, some express only ST and others express both. Enterotoxins are responsible for the enzymatic activation of different pathways in the cell leading to the increase of Cl^- secretion from crypt cells resulting in diarrhoea[32].

2.2.2.5 Serotypes

Besides the enterotoxins and colonization factors, two other factors have been used to characterise ETEC: determination of the cell wall lipopolysaccharide O serogroups and H serogroups of flagella. Serotypes can combine from at least 78 O antigens and 34 H antigens which the most common are O6:H16, O8:H9, O25:H24, O78:H12[33]. A genotypic methods that use comparative genomics revealed nine new O genotypes that were previously unrecognised by the standard methods used for serotyping, suggesting more ETEC isolates may not be recognised and assigned to any known O serogroup[34].

2.2.2.6 Heat-labile and heat-stable enterotoxins

LT is structurally and functionally related to the cholera enterotoxin (CT), from another Gram-negative pathogen *V. cholerae,* with 80% amino acid identity. LT consist of one single A subunit and five identical B subunits. The function of the B subunits is to mediate a bond between toxin and cell surface, mostly through binding to ganglioside GM1 (Galβ3GalNAcβ4(NeuAcα3)Galβ4Glcβ1Cer). The A subunit has enzymatic activity and transfers ADP-ribosyl part from NAD to the α-subunit of a basolateral membrane G-protein with adenylate cyclase regulation activity. Activated adenylate cyclase increases the level of cAMP, which subsequently activates cAMP-dependent kinases and ultimately activates the chloride channel (CFTR) in epithelial cells, resulting in an electrolyte imbalance and efflux of water (Fig. 3a)[32].

ST is a single-peptide toxin, which is according to its different structure and function, divided into one of the classes: STa and STb. STa is 18–19 amino acids long and is associated with human infections. It can be divided into two subgroups: STh (identified in human) and STp (identified in porcine). It binds to the membrane-spanning guanylate cyclase, which activity increases the amount of cGMP leading to activation of cGMP–dependent and/or cAMP-dependent kinases and increased secretion of Cl^- into the intestinal lumen and indirectly inhibits

the re-absorption of sodium by sodium-hydrogen exchanger (Fig. 3b)[35]. STb is associated with animal disease. STb-positive strains have been isolated almost exclusively from pigs, but also sporadically from chickens, cattle, cats, dogs, ferrets and even humans[12]. This 48 amino acids long protein elevates the concentration of Ca^{2+} in cells via stimulating a GTP-binding regulatory protein which leads to activation of protein kinase C and phosphorylation of the CFTR. The increased Ca^{2+} levels also activate the phospholipases A2 and C, leading to the release of arachidonic acid and formation of prostaglandin E2 and serotonin. Both molecules mediate the secretion of electrolytes and water out of the enterocytes by an unknown mechanism[12,32].

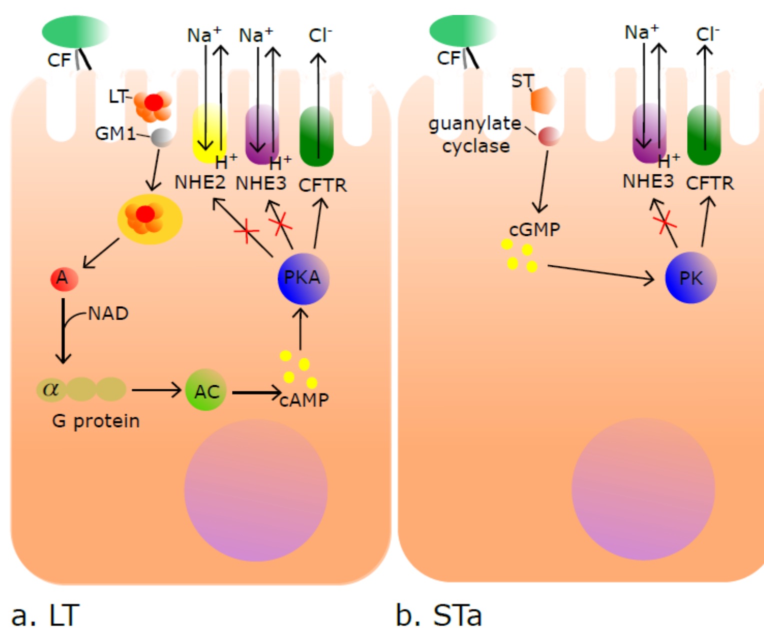

Figure 3: Pathogenesis of ETEC enterotoxins (adapted from Dubreuil *et al.*[12]):
a) Heat-labile enterotoxin (LT): LT binds to GM1 through the B subunits and the holotoxin is endocytosed. Subunit A is released and catalyses the transfer of ADP-ribosyl from NAD to the α-subunit of G-protein. Subsequently the G-protein activates adenylate cyclase (AC) which raises the level of cyclic adenosine monophosphate (cAMP) and the protein kinase A (PKA) is activated. The cystic fibrosis transmembrane regulator (CFTR) is phosphorylated by PKA and that results in the efflux of chloride ions and the uptake of sodium ions is blocked.
b) Heat-stable enterotoxin (ST): ST binds to guanylate cyclase C and activates its domain that results in elevated levels of cyclic guanosine monophosphate (cGMP) and activation of protein kinase II (PK), similarly to LT, CFTR is activated and uptake of Na^{+} is blocked leading to electrolyte imbalance and efflux of water.

2.2.2.7 Colonization factors

There are more than 25 colonization factors identified in human specific ETEC, and it is likely that there are additional CFs to be identified since 20–50% of all clinical ETEC isolates lack a known CF. There are three major groups into which CFs are divided: the CFA/I-like group (CFA/I, CS1, CS2, CS4, CS14, CS17, CS19 and PCFO71), the CS5-like group (CS5 and CS7), Class 1b group (CS12, CS18, CS20, and

CS30), and also one additional group of CFs not included in any other three groups[36,37]. The prevalence of different CFs differ from various geographical areas according to epidemiological studies, although CFA/I, CS1-CS7, CS14, and CS21 seem to be the most frequent[21]. There are also combinations of CFs and toxins in ETEC isolates that occur more frequently and often have specific geographic distribution, e.g. CS1 + CS3 (±CS21) LT+ STh, CS2+ CS3 (±CS21) LT+ STh, CS5 + CS6 LT + STh, CS6 STh, CFA/I (±CS21) STh and CS7 LT[10,21,38]. Structural subunits, transport, and assembly proteins for CFs are usually encoded in operons. In general, they consist of a major subunit, a periplasmic chaperone, an outer membrane usher protein and a minor subunit[36]. The gene expression may be influenced by different host environmental factors, e.g. sodium bicarbonate secreted from the pancreas into duodenum or bile produced in the liver as it is in the case of CS5 and CS6[39]. CFs may have fimbrial, fibrillar, helical or non-fimbrial structure. Fimbriae, for example CFA/I, are 5–10 nm in diameter, rigid, hair-like, filamentous organelles, whose structure consists of numerous copies of an identical subunit protein which determines its immunological characteristics. Fibrillae are flexible and thinner, 2–4 nm in diameter, their structure consists of fewer subunits[21]. The structure of CS5 and CS7 can be described as helical. Non-fimbrial adhesins (e.g. CS6) are a heterogeneous group sharing a few characteristics. They consist of non-covalently linked subunits and each subunit is expressing a binding epitope which makes the structure multivalent[40].

2.2.2.8 CS30

Although, intense efforts and hard work have been put into identification of different CFs, mostly with successful outcomes, more than 30% of all ETEC isolates lack a known CF (“CF negative”). In a recently published article by von Mentzer *et al.*, 94 whole genome sequenced “CF negative” ETEC isolates were used to search for novel CFs. By using nucleotide reference database of all published and known CFs, they identified four “CF negative” isolates containing seven genes with sequence homology and operon structure similar to known CFs (CS12, CS18, CS20 and the porcine CF 987P), and two isolates which also harboured CS13.

Table 3: ETEC colonization factors. Characteristic features of known colonization factors, adapted from von Mentzer *et al.*[41].

Colonization factor	Morphology	Size (kD)
CFA/-like group		
CFA/I	Fimbrial	25.0
CS1	Fimbrial	15.2
CS2	Fimbrial	15.4
CS4	Fimbrial	15.0
CS14	Fimbrial	15.0/15.5
CS17	Fimbrial	15.5
CS19	Fimbrial	15.0
PCFO71	Not determined	Not determined
CS5-like group		
CS5	Helical	18.6
CS7	Helical	18.7
Class Ib group		
CS12	Fimbrial	17.9
CS18	Fimbrial	18.5
CS20	Fimbrial	17.5
CS26	Not determined	Not determined
CS27A	Not determined	Not determined
CS27B	Not determined	Not determined
CS28A	Not determined	Not determined
CS28B	Not determined	Not determined
CS30	Fimbrial	18.5
Additional		
CS3	fibrillar	15.0
CS6	non-fimbrial	15.1/15.9
CS8	Fimbrial	25.3
CS21	Fimbrial	25.2
CS15	non-fimbrial	18.2
CS22	fibrillar	15.03
CS10	non-fimbrial	Not available
CS11	fibrillar	Not available
CS13	fibrillar	24.8
CS23	fibrillar / non-fimbrial	16.9

The novel CF, named CS30, has a major subunit with a predicted molecular weight of 18.5 kD. CS30 is encoded by a gene cluster (*csmA-G*) consisting of the major subunit gene, genes encoding the protein domains of chaperones and ushers, and genes

encoding a minor fimbrial subunit and an adhesin (Fig. 4). In addition, CS30 is thermo-regulated with protein expression at 37 °C as described by von Mentzer *et al.*, and maintains a rigid fimbrial structure of 7 nm in diameter.

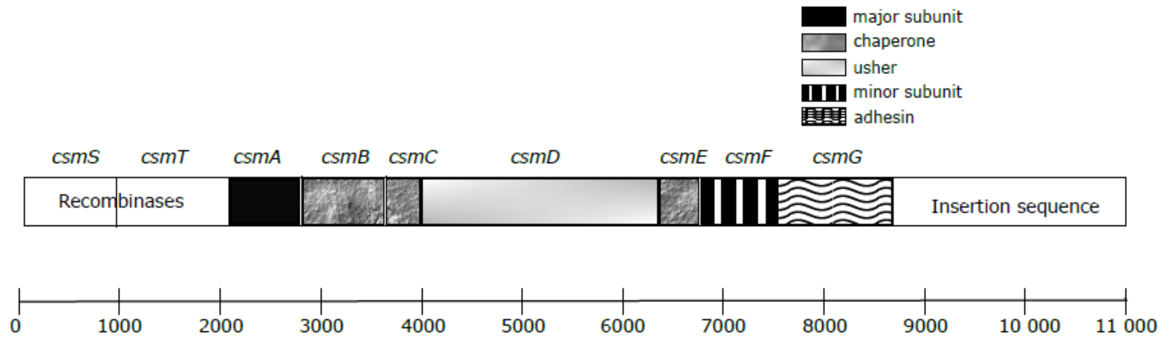

Figure 4: CS30 operon - characteristics in Table 4, adapted from von Mentzer *et al.*[41].

Table 4: CS30 operon. List of genes and their function, adapted from von Mentzer *et al.*[41].

Gene	Length (nt)	Function
csmA	606	Major subunit
csmB	708	Chaperone
csmC	432	Chaperone
csmD	2508	Usher
csmE	411	Chaperone
csmF	537	Minor subunit
csmG	1191	Adhesin

Genes encoding CS30 alongside with the ST and LT enterotoxins are encoded on an F-like plasmid with a FII replicon. The plasmid additionally contains five transfer associated genes (*traM, traY, traA, traL and traE*), replication genes, plasmid stability genes and DNA-binding regulator gene. Furthermore, the plasmid contains a disrupted gene *eatA* that encodes a protein belonging to a family of molecules called serine protease autotransporters of the *Enterobacteriaceae* (SPATE). EatA degrades the major mucin secreted by intestinal epithelium, MUC2, and enhances toxin access to epithelial cells, which suggests its role in the virulence of ETEC and other mucosal pathogens[41,42].

The ETEC strain E873 used in this project was identified as one of the "CF negative" strains carrying the CS30 operon[41].

2.2.2.9 Other virulence factors

In addition to CFs, there are also chromosomally located genes, e.g. *tia* and *tib*, which are involved in disease initiation and cell-to-cell interactions. The *tib* gene encodes a non-fimbrial 104 kDa glycosylated adhesin TibA. This structure not only

binds to a specific receptor on epithelial cells, but also promotes aggregation of bacterial cells and the formation of biofilm[43,44,45]. Another locus encodes the adhesin Tia, 25 kDa outer membrane protein binding to its specific receptor[46]. EAST1, a heat stable enterotoxin (ST) originally described in enteroaggregative *E. coli* can be found in some ETEC strains as well[47]. The previously mentioned protein EatA and the additional YghJ are two mucinases also involved in ETEC colonization[48]. The virulence two-partner secretion locus etpBAC encodes protein EtpA, which is situated on the tip of the flagella and mediates binding[49]. All listed virulence factors induce protective immunity thus are potential candidates in vaccine development.

2.3 Theoretical background about gene cloning methods

Gregor Mendel is considered a founder of genetics. Since he first described laws of the inheritance of biological characteristics, this scientific field has come a long way in search for understanding genes, their structure and function. A revolution in experimental biology started in 1970´s with developing a whole new methodology referred to as recombinant DNA technology or genetic engineering. The core of the technology lies in the process of gene cloning[50].

The basic steps in a gene cloning include insertion of a fragment DNA into a transfer DNA molecule called vector. This recombinant DNA molecule is then transported into a host cell, usually a bacterial strain, within which a vector can multiply and produce numerous identical copies of itself and of the gene it is carrying. After a large number of cell divisions, a colony or clone of identical cells is produced and each cell contains one or more copies of the recombinant DNA molecule[50].

In the following pages, a short theoretical background of every step in gene cloning will be provided with a description of methods or enzymes that are usually used.

2.3.1 Amplification of DNA fragment – Polymerase chain reaction (PCR)

PCR is an elegant enzymatic assay designed by Kary Mullis in 1983 enabling relatively fast amplification of specific DNA from a variety of tissues or organisms that is vastly used in many different scientific fields[51]. Each reaction requires the template DNA, specific primers, nucleotides, DNA polymerase and characteristic buffer to provide proper environment. The whole process occurs in three basic steps that repeat continuously several times until sufficient yield is reached. Firstly, a thermal block in a thermal cycler is heated above the melting point of the complementary DNA strands, usually 94–98 °C, to separate them in process called denaturation. In the next step called hybridization or annealing, the temperature is lowered to the rate enabling specific primers to bind to the complementary sequence of DNA. Raising the temperature again, the DNA polymerase is able to add nucleotides to the primers, thus elongation occurs.

For reaching reasonable outcomes, several conditions must be adjusted. Design of specific oligonucleotide primers is essential. Optimal length of PCR primers is 18–22 base pairs (bp), but even longer primers can be used. If the primers are too short, they might anneal non-specifically and give undesired products. If too long, hybridization might occur at a lower rate and primers simply wouldn´t have enough time to anneal. DNA-DNA hybridization is a temperature dependent process. For ideal hybridization between primers and DNA template the temperature must be low enough to enable the annealing, but high enough to prevent forming of mismatched hybrids. The ideal annealing temperature is 1–2 °C below the melting temperature (T_m) of primer-template hybrid, which can differ from 50–70 °C depending on GC content and can be calculated from the simple formula: $T_m = (4\times[G+C])+(2\times[A+T])$ °C. Primer secondary structures, runs, repeat and cross homology should be avoided[50]. For primer design number of design software can be used, i.e. Primer Premier, AlleleID, Geneious, etc.

Thermostability of DNA polymerase is one of the features that enables repetition of PCR cycles without a presence of the operator and reduces the amount of enzyme needed. The first enzyme used for PCR was *Taq* polymerase, isolated from thermophilic bacterium *Thermus aquaticus*[52]. The optimal activity demands temperature in range of 68–72 °C and divalent magnesium ions (1–5 mM). *Taq* has a 5´→ 3´exonuclease activity and no 3´→ 5´activity. Therefore, its fidelity is lower than

other recombinant enzymes with error rate approximately $1–2\times10^{-4}$ mutation frequency per base pair per duplication. A new strategy to enhance and improve the enzyme performance is to fuse it with a thermostable DNA binding protein[53]. Nowadays, there are many types of thermostable DNA polymerases differing in origin, optimal temperature, error rate and activity, for example *Pfu* polymerase, *Tgo* polymerase, *Tpe* polymerase, etc.

Likewise, the basic PCR procedure has many variations, e.g. reverse transcriptase PCR (RT-PCR) that uses cDNA strand synthesised from total RNA or mRNA transcript, nested PCR that uses two sets of primers for higher specificity and fidelity, inverse PCR (IPCR) that is used to amplify unknown regions, or several others.

2.3.2 Vectors

Vectors are DNA molecules used to transfer a foreign DNA into a host cell. There are several different types of vectors and the choice of vector depends on the host and the aim of the cloning experiment. Vectors for bacterial cells include plasmid vectors, bacteriophage vectors, cosmids, and phagemids[50].

2.3.2.1 Plasmid vectors

Plasmid vectors are one of the most studied from previously listed. Their origin comes from extra-chromosomal, self-replicating circular DNA, called plasmids. Plasmids are able to replicate autonomously, have mechanism for controlling their copy number and also encode addiction system which kills daughter cells that do not inherit the plasmid during cell division thus stable inheritance is ensured[54].

Plasmid vectors used for cloning are usually less than 5 kb and contain number of structural elements[50]:

1. Cloning sites, which are artificially designed recognition sites for restriction enzymes and serves for insertion of a foreign DNA.
2. Replication origin which aids the vector to utilise the bacterial host system.
3. Selectable markers, usually antibiotic resistance genes that facilitate the screening for transformed cells.
4. Expression vectors also contain a promoter upstream of the cloning site enabling expression of inserted DNA. The *lac* promoter is often used for vectors, because its expression is inducible by isopropyl-β-D-thiogalactopyranoside (IPTG).

Depending on the number of vectors, plasmids can be divided in two groups, high copy plasmids with 100 copies and more per cell and low copy plasmids with 1–25 plasmids per cell. Copy number depends on the origin of replication, size of the plasmid and associated inserts. For high yield of recombinant DNA, the high copy plasmid vectors are preferable. On the other hand, if the gene product may cause adverse effects to the host a low copy vectors may be desirable[50,55].

Another classification divides plasmids into conjugative and non-conjugative. The first can mediate their own transfer by process called conjugation. The transfer (*tra*) and mobilising (*mob*) regions on the plasmid are required for this function. Non-conjugative plasmids are not self-transferable, but if their mob region is functional a conjugation-proficient plasmid may enable transfer[50].

In 1971, Hedges and Datta suggested a classification based on plasmid incompatibility, which can be defined as the stability of plasmids during conjugation or inability of two related plasmids to be propagated in the same cell line. Each incompatibility group is manifested by the relatedness of plasmids that share common replication controls and can be identified by using cloned replication regions (replicons). Currently, 27 incompatibility groups are recognised in *Enterobacteriaceae*[54].

Plasmids can be classified based on the genes that code their main characteristic[50]:

1. F plasmids (fertility) carry *tra* genes and have the ability to transfer plasmids by conjugation.
2. R plasmids (resistance) carry genes for antibacterial resistance. These plasmids bring dangerous consequences in clinical microbiology, because their spread can exacerbate antibiotic resistance in human population.
3. Col plasmids code proteins that kill other bacteria, called colicins.
4. Degradative plasmids allow the host cell to metabolize unusual compounds, for example salicylic acid or toluene.
5. Virulence plasmids carry genes coding pathogenic features of the host bacteria.

2.3.3 Digestion of DNA-restriction enzymes

One of the most important enzymes used for gene cloning are the restriction enzymes. The original function of these enzymes is protection of host cell from

exogenous DNA as a part of the restriction-modification system[51]. The restriction enzyme only hydrolyses exogenous DNA while the host DNA is protected by methylation of particular bases in sequence. There are over 1200 characterised enzymes divided into three types (I, II, and III). Types I and III are complex and have only limited role in genetic engineering, while Type II is vastly used (Table 5). These endonucleases, also called molecular scissors, cut at an internal position in DNA strand with high specificity. Each enzyme recognises four, five or six base-pairs sequences and cut double-strand in two different ways. Either they make a simple double-strand cut in the middle of sequence which results in a blunt end or the DNA strands are cut in different position which results in two or four nucleotide single-stranded overhangs at each end, called sticky ends. Frequently, the agarose gel electrophoresis is used to analyse digested products.

Table 5: Restriction enzymes used in this project.

Enzyme	Organism	Recognition sequence	Cutting sites
AccI	*Xanthomonas maltophilia*	5'-GTMKAC-3'	5'-GT↓MKAC-3' 3'-CAKM↑TG-5'
EcoRI	*Escherichia coli*	5′-GAATTC-3′	5′-G↓AATTC-3′ 3′-CTTAA↑G-5′
HindIII	*Haemophilus influenzae Rd*	5′-AAGCTT-3′	5′-A↓AGCTT-3′ 3′-TTCGA↑A-5′
PstI	*Providencia stuartii*	5′-CTGCAG-3′	5′-CTGCA↓G-3′ 3′-G↑ACGTC-5′
SacI	*Streptomyces achromogenes*	5′GAGCTC-3′	5′-GAGCT↓C-3′ 3′-C↑TCGAG-5′
XmaI	*Xanthomonas malvacearum*	5′-CCCGGG-3′	5′-C↓CCGGG-3′ 3′- GGGCC↑C-5′

2.3.4 Ligation of inserted DNA and vector

DNA ligase is another enzyme that is important in genetic engineering. Its function is to repair defective phosphodiester bonds that may occur in DNA strands during, for example, DNA replication. In gene cloning it is used as the final step in creation of a recombinant DNA molecule ligating two molecules together, e.g. fragments and plasmids, by creating two phosphodiester bonds, working as molecular glue. The most commonly used ligase is T4 DNA ligase purified from *E. coli* strands infected with bacteriophage T4 that seals sticky ends as well as blunt ends under appropriate conditions. Ligation of complementary sticky ends is more efficient, because base pairs of these overhangs can form a relatively stable structure by hydrogen

bonding. In contrast, blunt ends have to wait for chance associations to ligate. A common situation is where the sticky ends are desirable, but not available on both DNA molecules or only one of them has sticky ends and the other has blunt ends. This can be solved by using linkers, adaptors or homopolymer tails[50].

2.3.5 Transport of recombinant DNA into the cell: transformation and electroporation

Transformation refers to the uptake of plasmid DNA, transfection describe uptake of phage DNA. The choice of appropriate method depends on the type of host/vector system. Transformation and transfection are some of the simplest methods, but still play critical part in genetic engineering, because they are not very efficient in that only a small percentage of cells become transformed. In nature, these are not major processes to acquire genetic material. Although, most bacterial species are able to take up DNA from medium, it is often degraded. To survive and replicate in host cell, the plasmid must carry an origin of replication recognised by the host. The cells meant for transformation need to be competent, which means that they can acquire plasmid DNA. This can be achieved by treating bacterial cells with an ice-cold calcium chloride solution for chemically competent cells or with and ice cold 10% glycerol for electrocompetent cells.

Chemically competent cells are mixed with plasmid DNA and incubated on ice for 20–30 minutes, then they undergo brief heat shock at 42 °C for 30–120 seconds and they are incubated in a nutrient broth at 37 °C for 1–2 hours. The cells are then plated out on selective media.

Electroporation is another option that can be used for insertion of recombinant DNA into host cell. This method is less frequently used in bacteria and more in yeasts, fungi, and plant cells. A brief electrical pulse (1.7–2.5 V) is used in electrocompetent cells, causing disorganization of the cell membrane and increase of its permeability. After adding a nutrient broth, the cells are incubated at 37 °C for 1–2 hours and plated out on selective media[50,51].

3 AIMS

The overall aim of this thesis is to identify the intestinal receptor(s) for the ETEC colonization factor CS30. Due to the fact that ETEC strains express several outer membrane proteins which may cause false positive results in binding assays, construction of recombinant strain expressing only CS30 is desirable to ensure binding specificity and accuracy.

I hypothesize that the ETEC CS30 can bind to both human and porcine specific intestinal receptors.

The specific aims are:

- To identify CS30 specific carbohydrate receptors by screening a mixture of glycosphingolipids isolated from human and animal small intestine using solid phase binding studies.
- To determine if CS30 binds to glycosphingolipids isolated from both human and pig intestine.
- To generate a recombinant TOP10 strains harbouring the CS30 operon to be used in the binding studies and for purifying CS30.

4 MATERIALS AND METHODS

4.1 CS30 binding to glycosphingolipids

4.1.1 Thin-layer chromatography

Thin-layer chromatography (TLC) is a useful method for a qualitative and quantitative separation of different substances. It is fast and easy to perform, thus vastly used in laboratories. A stationary phase is coated in a thin layer on a glass, aluminium or plastic plate. A small volume of different samples is applied at the bottom side of the plate and plate is put in a developing chamber with mobile phase. Due to capillary forces, the mobile phase moves up in the plate and the applied samples move up as well. Different molecules migrate through stationary phase with different speed depending on the size of the molecule and its polarity. In our systems, the polar compounds interact with the stationary phase more than hydrophobic and they are slowed down. The chromatographic mobility of a compound is usually compared to a reference compound.

For detection of GSLs, the universal staining can be done by anisaldehyde-sulfuric acid (4-methoxybenzaldehyde, glacial acetic acid and concentrated sulfuric acid; 1:98:2, v/v/v). GSLs are coloured green and contaminants are coloured blue or pink. Another staining can be done by resorcinol. This reagent is able to distinct sialic acid containing gangliosides, which produce a blue-violet colour.

For GSLs analysis glass-backed or aluminium-backed high performance thin-layer chromatography plates coated with silica gel 60 (Merck, Darmstadt, Germany) were used. A mixture of chloroform/methanol/water (65:25:4, v/v/v) was used for elution. The glycosphingolipids were applied in quantities of 2–8 μg on the bottom of the plate by a microsyringe and after elution and evaporation of redundant mobile phase, anisaldehyde was applied with a fine mist sprayer for detection. Sufficiently moistened plates were placed in the oven for sufficient time period until the bands were detectable.

4.1.2 Culture and labelling of bacteria

E. coli strain E873 expressing CS30 (wild type) was cultured on Luria-Bertani (LB) plates at 37 °C overnight. CFA broth with bile was inoculated with a colony from the LB plate and incubated in a shaker for 3 hours. Thereafter, 10 μl of ^{35}S-methionine

was added to 8 ml of CFA broth with bile and inoculated with 5 drops of the culture described earlier. The suspension was incubated for 3 hours in a shaker. The bacteria were harvested by centrifugation, washed three times with PBS, and then re-suspended in PBS/BSA/TWEEN (PBS with 2% (w/v) BSA, 0.1% (w/v) Tween 20 and 0.1% (w/v) NaN_3).

4.1.3 Chromatogram binding assay

One of the methods that GSLs are analysed with is the chromatogram binding assay. This highly sensitive method is based on the specific binding of ligands to the GSL(s), either antibodies, bacterial toxins, bacteria, viruses or lectins.

The aluminium-backed TLC plates prepared as described above are firstly dipped in diethylether/*n*-hexane (1:5, v/v) containing 0.5% (w/v) polyisobutyl-methacrylate for 1 minute and dried. Coating plates in polyisobutyl-methacrylate has two reasons. Firstly, it prevents flaking of silica gel from aluminium support during further incubation. Secondly, polyisobutyl-methacrylate mimics the cell membrane, thus the hydrophobic part of the GSLs is planted in the plastic and the glycan part is facing the outer environment (at least in theory). The chromatograms are then blocked with PBS/BSA/TWEEN to prevent unspecific binding. The plates are incubated with ^{35}S-labeled bacteria for 2 hours at room temperature, washed with PBS, dried and subjected to autoradiography (overnight exposure to BIOMAX MR Film, Carestream, USA).

4.1.4 Anti-sulfatide MAbs

Monoclonal antibodies directed towards SO_3-3Galβ that were used in binding assays were a kind gift from Dr. M. Blomqvist (Institute of Biomedicine, The Sahlgrenska Academy at University of Gothenburg, Sweden).

Firstly, the aluminium-backed TLC plates prepared and coated as described above are blocked with PBS with 2% (w/v) BSA and 0.1% (w/v) NaN_3 for 1.5 h at room temperature. Afterwards, the plates are incubated with solution of antibodies (25 μl in 5 ml of PBS/BSA) for 2 hours at room temperature and washed 5 times with PBS. This is followed by incubation with anti-mouse antibodies labelled with ^{125}I (100 μl in 5 ml of PBS/BSA) for 2 hours at room temperature. Thereafter the plates were washed 5 times with PBS, dried and subjected to autoradiography.

4.1.5 Separation of the acid glycosphingolipids fractions from human intestine

The total acid fraction from human small intestine, containing CS30^{+} *E. coli* binding fast-migrating compounds, was selected for isolation of the CS30 binding glycosphingolipids. First this fraction (43.2 mg) was separated on a 10 g Iatrobeads column (Iatron Laboratories Inc.; 6RS-8060) eluted with chloroform/methanol/water (60:35:8, v/v/v), 30 × 1 ml. The fractions obtained were analysed by thin-layer chromatography and anisaldehyde staining, and pooled into four subfractions according to mobility on the thin-layer chromatograms. The *E. coli* CS30 binding activity of these four fractions was assessed using the chromatogram binding assay. Thereby a one fraction (18.6 mg) containing CS30 binding fast-migrating compounds was obtained.

The fraction containing the CS30 binding compounds was further separated on a second 10 g Iatrobeads column eluted with chloroform/methanol/water (60:35:8, v/v/v), 40 × 0.5 ml. Again, the fractions obtained were analysed by thin-layer chromatography and anisaldehyde staining, and pooled into three subfractions according to mobility on the thin-layer chromatograms. The first and the third fraction obtained had distinctly different mobility on thin-layer chromatograms indicating different ceramide compositions. In order to isolate pure ceramide species to test for CS30 binding these two fractions (approximately 4 mg each) were further separated on 10 g Iatrobeads columns, eluted and pooled as described above.

4.1.6 Mass spectroscopy

The intact glycosphingolipids were analysed by LC-ESI/MS using Agilent 1100 Series HPLC (Agilent, USA). Separation was performed using the analytical column made of fused silica capillary (250 µm × 10 cm) (Scandinavian Genetec AB, Sweden) and filled with modified silica gel Polyamine II (ScantecLab AB, Sweden), that is usually used for hydrophilic interaction chromatography (HILIC). HILIC is a variant of normal phase liquid chromatography for separating polar compounds, in which the separation of samples is based on the interactions and retention of analytes by hydrogen bonding, ionic interactions, and dipole-dipole interactions. Elution is performed by binary eluent consisting of aqueous (contains salt or acid) and organic part. By beginning gradient elution with low-polarity organic solvent and increasing the polar aqueous content, better sensitivity and good on-column

retention for polar compounds can be achieved. A Finnigan LTQ mass spectrometer (Thermo Electron Corporation, USA) with ESI ionization in negative mode and linear iontrap mass analyser was used for mass spectrometry analysis of sample separated by HPLC.

The LC-ESI/MS system was washed with 50% buffer A (10 mM NH_4HCO_3) for 10 minutes (100–120 bar), followed by 100% B (100% AcCN) for 20 minutes (50–60 bar). The samples for analysis were prepared using 30 µg of the GSL solution that were dried in the heat block under N_2 atmosphere. Afterwards, they were dissolved in acetonitrile/methanol 3:1 (v/v) and put into the auto-sampler.

4.2 Recombinant strain TOP10-CS30

4.2.1 Construction of recombinant strain TOP10-CS30

DNA fragment *csmA-G*. To obtain CS30 operon containing genes *csmA-G*, ETEC strain E873 was cultured on Luria-Bertani agar plates and left overnight at 37 °C.

The rapid boil method was used to isolate DNA from bacterial culture. Bacterial colonies were diluted in genetic water (gH_2O), vortexed until homogenous suspension was reached, and boiled in water bath for 10 minutes. The suspension was afterwards centrifuged, and the supernatant containing the DNA was moved to a new tube. DNA concentration was measured on Nanodrop spectrophotometer (Nanodrop Technologies Inc, USA). The extracted DNA was used as a template for amplifying the *csmA-G* fragment by polymerase chain reaction.

Master mix for PCR reaction included specific 5X Phusion High Fidelity buffer (7.5 mM $MgCl_2$), dNTP (10 mM), primermix (10 pmol/µl), DNA template, gH_2O and Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific, USA). It has 5′→ 3′ DNA polymerase activity as well as 3′→ 5′ exonuclease activity generating blunt ends in the products and it is suitable for amplification of long amplicons with error rate determined to be 4.4×10^{-7}. PCR program was set as described in Table 6.

PCR products were analysed by gel electrophoresis. For separation of DNA, electrophoresis relies on the fact that nucleic acids are polyanionic at neutral pH and agarose gel can be used as a sieve. DNA fragments migrate through the gel towards the anode electrode. The bands are stained with different chemical compounds, *i.e.* ethidium bromide or Midori Green. For better interpretation of results markers or Gene ladders with defined lengths of DNA molecules are loaded in one (or more)

of the wells alongside the samples. PCR products were analysed in 1.5% agarose gel with Midori Green Advance DNA Strain (Nippon Genetics Europe GmbH, Germany) added into the gel and viewed under UV light. Additionally, the *csmA-G* was partially sequenced to confirm that the fragment had been amplified correctly.

Table 6: PCR program used for *csmA-G* amplification

Step	Temperature	Time	
Denaturation	94 °C	5 min	
Annealing	67 °C	1 min	
Elongation	72 °C	7 min	
Denaturation	94 °C	0.5 min	29×
Annealing	67 °C	0.5min	
Elongation	72 °C	7 min	
Extension	72 °C	15 min	

Afterwards, the GeneJet PCR Purification Kit (Thermo Fisher Scientific, USA) was used for purification of amplified DNA according to the manufacturer′s instructions, except for the Elution Buffer which we replaced with 25 µl of gH_2O. Concentration of prepared fragment *csmA-G* was measured on Nanodrop spectrophotometer.

Plasmid pMT-CTA. To obtain plasmid pMT-CTA (chloramphenicol resistant), *V. cholerae* strain 1275 was cultured in 5 ml LB broth with chloramphenicol (LB+CM) at 37 °C and 180 rpm overnight. Plasmid purification was performed with ZymeRapid Plasmid Miniprep (The Epigenetics Company, USA) according to manufacturer′s instructions. The plasmid concentration was measured on Nanodrop spectrophotometer.

Plasmid was amplified with PCR using master mix: specific 5X Phusion High Fidelity buffer (7.5 mM $MgCl_2$), dNTP (10 mM), primermix (10 pmol/µl), DNA template, gH_2O and Phusion High-Fidelity DNA polymerase. PCR program was set as in Table 7.

Table 7: PCR program used for pMT amplification

Step	Temperature	Time	
Denaturation	94 °C	5 min	
Annealing	68 °C	1 min	
Elongation	72 °C	4 min	
Denaturation	94 °C	0.5 min	29×
Annealing	68 °C	0.5 min	
Elongation	72 °C	4 min	
Extension	72 °C	8 min	

PCR product was analysed by gel electrophoresis and purified using GeneJet PCR purification Kit, following manufacturer's instructions. Concentration was measured on Nanodrop spectrophotometer.

Table 8: Primers (Eurofins Genomics, Germany) used in project for amplification of *csmA-G* and pMT

Primers	Sequence	T_m
For_csmA_SacI	5´-CCCGAG CTC ATG AAA AAG ACA ATT ATG TCT CTG GC-3´	70 °C
Rev_csmG_XmaI	5´-CCCCCC GGG TCA GCT TTC CCA TCG GGA AAT GAT-3´	70 °C
For_pMT_XmaI	5´-CCCCCC GGG AAG GAT GAA TTA TAA GCT TAG CCC G-3´	74 °C
Rev_pMT_SacI	5´- CCCGAG CTC CCT GTG TGA AAT TGT TAT CCG CTC-3´	70 °C

Digestion of pMT and *csmA-G*. Digestion of pMT and *csmA-G* was performed by two restricted endonucleases: SacI and XmaI (see Table 5). Master Mix containing 10 µl of DNA, 3 µl of XmaI specific buffer, 1 µl of XmaI, 2 µl of SacI (all restricted endonucleases are from Fermentas, USA) and 14 µl of gH_2O was incubated for 1 hour at 37 °C. Enzymes were then inactivated at 65 °C for 15 minutes. Solutions were purified with GeneJet PCR Purification Kit and eluted with 20 µl of gH_2O.

Dephosphorylation of pMT. To prevent self-ligation of pMT, 5´P terminus of sticky end is modified to 5´OH terminus thus unable to form a phosphodiester bridge between 5´OH and 3´OH ends of the same DNA molecule. 10 µl of pMT was mixed with 3 µl of FastAP (Thermosensitive Alkaline Phosphatase, Fermentas, USA), 2 µl of FastAP buffer, 5 µl of gH_2O and left at 37 °C for 10 minutes. Enzyme was inactivated at 75 °C for 5 minutes, and then the sample was purified with GeneJet PCR Purification Kit.

Ligation of pMT and *csmA-G*. The ligation mix consists of purified fragment *csmA-G* and plasmid pMT in a 1:5 concentration ratio, 10X T4 ligation buffer and gH_2O. For the ligation, 1 µl of T4 ligase (Fermentas, USA) was added to 20 µl of the ligation mix and 10 µl of mix without ligase was used as a control. Both samples were incubated at room temperature for 1 hour. The results were analysed by agarose gel electrophoresis.

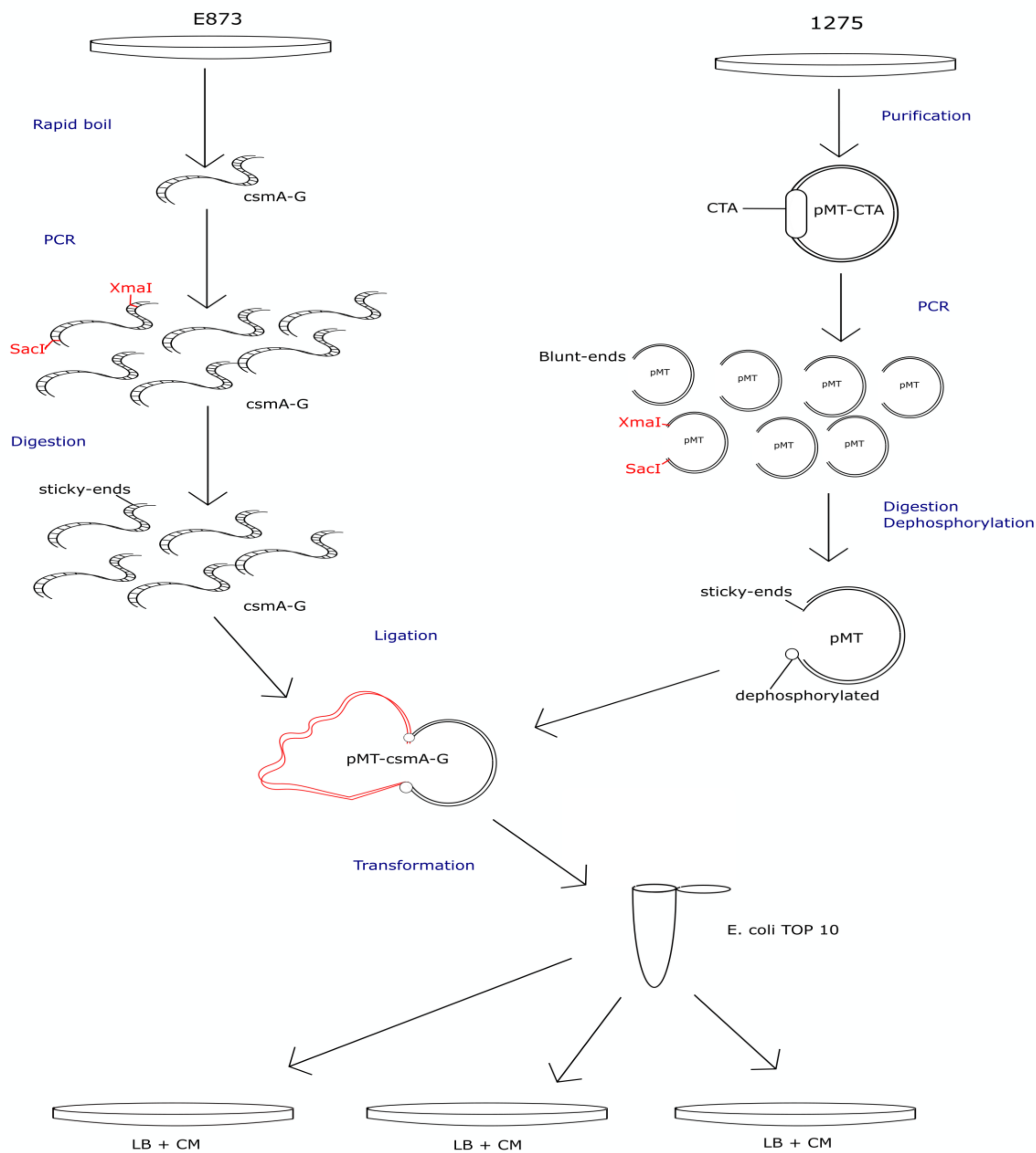

Figure 5: Scheme of the TOP10-CS30 cloning procedure. DNA molecules extracted from plated strains E873 and 1275 by rapid boil method or commercial plasmid purification kit were amplified using primers designed to produce sequences harbouring XmaI and SacI restriction sites. Both molecules were digested by appropriate restriction endonucleases and pMT molecule was dephosphorylated to prevent self-ligation. The *csmA-G* fragment and pMT vector (carrying chloramphenicol resistance genes) were ligated and transformed into commercial TOP10 strain. Transformed cells were plated on LB+CM plates using different volume of bacterial suspension and after incubation chloramphenicol resistant colonies were selected for further analysis.

Transformation of commercial TOP10 strain with pMT-*csmA-G*. Transformation was performed with both, chemically and electro competent cells.

Commercial chemically competent *E. coli* strain TOP10 was thawed on ice bath. Depending on the concentration, between 1–5 µl of pMT-*csmA-G* solution was added into the vial. The mixture was incubated in an ice bath for 30 minutes and then

put into a 42 °C water bath for 30 seconds. 250 µl of SOC medium was added and tubes were incubated at 37 °C, 225 rpm for 2 hours. Afterwards, the suspension was centrifuged, supernatant was discarded and 60 µl of PBS was added. Solution was plated on LB plates with chloramphenicol (20 and 40 µl) and left overnight at 37 °C.

Commercial electrocompetent TOP10 strain was thawed on ice. 1 µl of pMT-*csmA-G* was added into vial and gently mixed. Suspension was transferred into pre-cooled electroporation cuvette and electric pulse (2.5 V) was applied on cells. 250 µl of SOC medium was added into cuvette and mixture was transferred into tubes. Bacterial culture was incubated at 37 °C, 225 rpm for 2 hours and then plated on chloramphenicol LB plates, incubated overnight at 37 °C.

4.2.2 Fragment length and restriction enzyme size analyses

In screening for transformed cells, the fragment length and restriction enzyme size analyses were performed. Twenty colonies were cultured in LB broths with chloramphenicol overnight at 37 °C. Plasmid purification was performed with ZymeRapid Plasmid Miniprep according to manufacturer's instructions. Every plasmid sample was then digested with restriction endonuclease AccI. Master mix contained 16.5 µl gH_2O, 3 µl buffer B, 10 µl of purified plasmid and 0.5 µl of AccI. Samples were incubated at 37 °C for 1 hour and analysed by gel electrophoresis.

4.2.3 Expression of CS30

To analyse the right expression of genes *csmA-G* in recombinant strain TOP10-CS30, Reverse transcription PCR was used. After RNA extraction from TOP10-CS30 strain and cDNA synthesis, the specific primers for each gene (Table 10) were used to amplify them.

Strain culture and RNA extraction. Six colonies from TOP10-CS30 strain were inoculated in 25 ml of LB+CM broth and left at 37 °C, 180 rpm for 3 hours. 10^7 bacteria/ml from pre-culture was transferred to 25 ml LB+CM broth with added IPTG (100 mM) and cultured at 37 °C, 180 rpm for 7 hours until OD ~0.8–1.0 was reached. Volume corresponding to 10^9 bacteria was then collected, treated with RNA Protect®Bacteria Reagent (Qiagen, Germany), and stored at -80 °C. ReliaPrep™ RNA Cell Miniprep System (Promega Corporation, USA) was used for RNA extraction. Following GRAM-bacterial RNA Purification protocol, lysozyme (10 mg/ml) was

added to 10^9 bacteria prepared as described before and incubated for 5 minutes at room temperature. After adding 250 µl of BL+TG buffer and vortexing the sample until the cell pellet is dispersed, 85 µl of isopropanol was added and vortexed for 5 seconds. The lysate was then transferred to a mini-column and centrifuged for 30 seconds at the highest speed. The mini-column was washed by RNA Wash Solution and centrifuged for 30 seconds. Freshly prepared DNase I incubation mix (24 µl of Yellow Core Buffer, 3 µl of 0.09M $MnCl_2$ and 3 µl of DNase I) was applied to the mini-column membrane and incubated for 15 minutes at room temperature. The mini-column was then washed once with Column Wash Solution and twice with RNA Wash Solution. The sample was eluted with 50 µl of Nuclease-Free Water and RNA concentration was measured on Nanodrop.

cDNA synthesis. cDNA was synthesised using QuantiTect® Reverse Transcription Kit (Qiagen, Germany) and all reaction components were kept on ice during the preparation. Up to 600 ng of total RNA was diluted in RNase-Free Water for total volume 12 µl and treated with 2 µl of DNA wipe-out buffer (42 °C for 3 minutes). Afterwards, 6 µl of Master mix consisting of 1 µl of Reverse Transcriptase, 4 µl of RT buffer, and 1 µl of RT primermix was added and the mixture was incubated at 42 °C for 20 minutes. For enzyme inactivation, the sample was incubated at 95 °C for 3 minutes and then kept on ice until the next use.

Reverse transcription PCR (RT-PCR). For amplification of cDNA RT-PCR was used. Master mix contained 5 µl of cDNA (samples diluted in gH_2O; 1:10 v/v), 0.75 µl of primer mix (10 pmol/µl), 5 µl of specific Phusion High Fidelity buffer (7.5 mM $MgCl_2$), 0.5 µl of dNTP (10 mM), 13.5 µl of gH_2O and 0.25 µl of Phusion High-Fidelity DNA polymerase. PCR program was set as in Table 9. Afterwards all samples were analysed by agarose gel electrophoresis.

Additionally, the pMT-*csmA-G* was partially sequenced to confirm that the fragment had been ligated correctly.

Table 9: PCR program used for RT-PCR.

Step	Temperature	Time	
Denaturation	94 °C	5 min	
Annealing	60 °C	1 min	
Elongation	72 °C	1 min	
Denaturation	94 °C	0.5 min	29×
Annealing	60 °C	0.5min	
Elongation	72 °C	1 min	
Extension	72 °C	3 min	

Table 10: Primers (Eurofins Genomics, Germany) used for RT-PCR.

Primers	Sequence	T_m	Predicted products
F_csmA	5′-CCCGAA TTC ATG AAA AAG ACA ATT ATG TCT CTG GC-3′	70 °C	624 bp
R_csmA	5′-CCCAAG CTT TTA CGG AGT GTT TGC TTT GT-3′	56 °C	
For_csmB	5′-ATC CGT GTT CTC TGT TCG GG-3′	62 °C	230 bp
Rev_csmB	5′-ACC ATT CAA GGC TTT CGG GT-3′	60 °C	
For_csmC	5′-GTG CAA GAG TTA GGT GTT GCT G-3′	66 °C	208 bp
Rev_csmC	5′-GCG CTC GGC TTC TTT TCT TT-3′	60 °C	
For_csmD	5′-TAT TCG AGA GGC TGA CGG GA-3′	62 °C	879 bp
Rev_csmD	5′-TTA TCG TTC CCC CAA CTG CC-3′	62 °C	
For_csmE	5′-ACC CAG GAA GTT TGG TTT GGT-3′	62 °C	142 bp
Rev_csmE	5′-TCA GGA GTG CTT TTC GGG TA-3′	60 °C	
For_csmF	5′-AGT TAG CGA ACG GGG ATC AA-3′	60 °C	348 bp
Rev_csmF	5′-TAT CTG TCG GGA CGA CTT GC-3′	62 °C	
For_csmG	5′-TGC TAA TGA CGG CAC AGG AG-3′	62 °C	232 bp
Rev_csmG	5′-CAT GCG ATA ATA CGC CCC CT-3′	62 °C	
CS30_F	5′-AGT CAG CTC TTG CAG CCA GT-3′	62 °C	219 bp
CS30_R	5′-CCT TGG TAC CAT TGC TGG TT-3′	60 °C	

4.2.4 Negative staining of TOP10-CS30 and transmission electron microscopy (TEM)

Pre-culture of TOP10-CS30 was prepared by inoculating of approximately six colonies from cultured LB+CM plates into 25 ml LB+CM broth, and left in shaker at 37 °C, 190 rpm for 3 hours. 10^7 bacteria/ml from pre-culture was transferred to two flasks containing 25 ml LB+CM broth with added IPTG (100 mM) and the first flask was cultured at 37 °C, 190 rpm for 6 hours until OD ~0.8–1.0 was reached, while the second flask was cultured under the same conditions overnight. Volume corresponding to 10^9 bacteria was then collected, washed with PBS, and re-suspended in appropriate volume of PBS to adjust the concentration to 1.25×10^{10} bacteria/ml.

Glow-discharged grids were placed on a drop of prepared bacterial suspension and incubated for 2 minutes. Then, the bacteria were fixated in 2% glutaraldehyde for 5 minutes and rinsed twice in gH_2O. The grids were blocked with 0.1% BSA in PBS for 10 seconds and rinsing three times in the drop of gH_2O for 10 seconds, the grids are at last dipped in 1% phosphotungstic acid (PTA) (30 seconds), left on filter paper for 5 minutes, and stored at 4 °C until microscopy.

5 RESULTS

5.1. CS30 expressing *E. coli* binds in the sulfatide region in the acid glycosphingolipids fractions from human and porcine small intestine.

When the binding of CS30$^+$ *E. coli* to acid glycosphingolipids of human small intestine was tested a distinct binding to a fast-migrating compound was obtained (Fig. 6, lane 3). The bacteria also bound to reference sulfatide (lane 2), whereas no binding to reference cholesterol-sulfate (lane 1) was obtained.

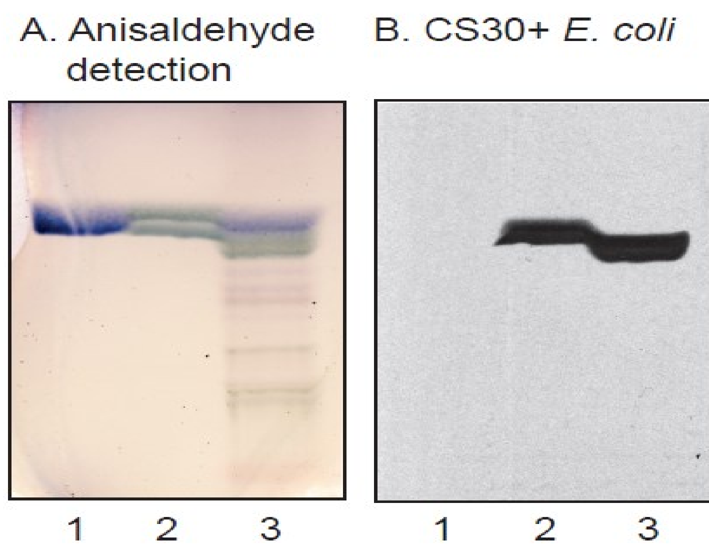

Figure 6: Binding of CS30-expressing *E. coli* to acid glycosphingolipid of human small intestine. Thin-layer chromatogram after detection with anisaldehyde (A), and autoradiogram obtained by binding of ^{35}S-labeled *E. coli* strain E873 (B). The glycosphingolipids were separated on aluminium-backed silica gel plates, using chloroform/methanol/water 60:35:8 (v/v/v) as solvent system, and the binding assays were performed as described in the Materials and methods section. Autoradiography was for 12 h. The lanes were: **Lane 1:** reference cholesterol-sulfate, 4 µg; **Lane 2:** reference sulfatide, 4 µg; **Lane 3:** total acid glycosphingolipids of human small intestine, 4 µg.

The binding of CS30$^+$ *E. coli* to glycosphingolipids of porcine small intestine was also tested. The bacteria bound in the sulfatide region in acid glycosphingolipid fraction from human small intestine (Fig. 7, lane 1), as well as in acid fraction from pig small intestine (Fig. 7, lane 3). No binding to non-acid glycosphingolipid fraction from pig small intestine (Fig. 7, lane 2) or human (not shown) was obtained.

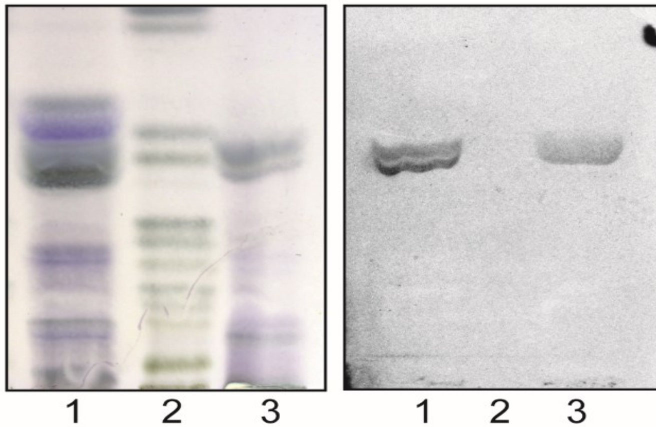

Figure 7: Screening for CS30 carbohydrate binding: CS30-expressing wild type ETEC binds in the sulfatide region in acid glycosphingolipid fractions from human and porcine small intestine. Thin-layer chromatogram after detection with anisaldehyde (A), and autoradiogram obtained by binding of ^{35}S-labeled *E. coli* strain E873 (B). The glycosphingolipids were separated on aluminium-backed silica gel plates, using chloroform/methanol/water 60:35:8 (v/v/v) as solvent system, and the binding assays were performed as described in the Materials and methods section. Autoradiography was for 12 h. The lanes were: **Lane 1:** acid glycosphingolipids of human small intestine, 40 µg; **Lane 2:** Non-acid glycosphingolipids from pig small intestine, 40 µg; **Lane 3:** Acid glycosphingolipids from pig small intestine, 40 µg.

5.2. Separation of the acid glycosphingolipids fractions from human intestine.

Isolation of the CS30 binding glycosphingolipids from the total acid fraction from human small intestine was done by a series of Iatrobeads column chromatographies. Fractions obtained were pooled according to mobility on thin-layer chromatograms, and binding of CS30$^+$ *E. coli*. After the first separation, one fraction (18.6 mg) containing the CS30 binding fast-migrating compounds was obtained (Fig. 8B, lane 2). This fraction was separated into a number of subfractions with different thin-layer chromatogram migration, indicating different ceramide composition.

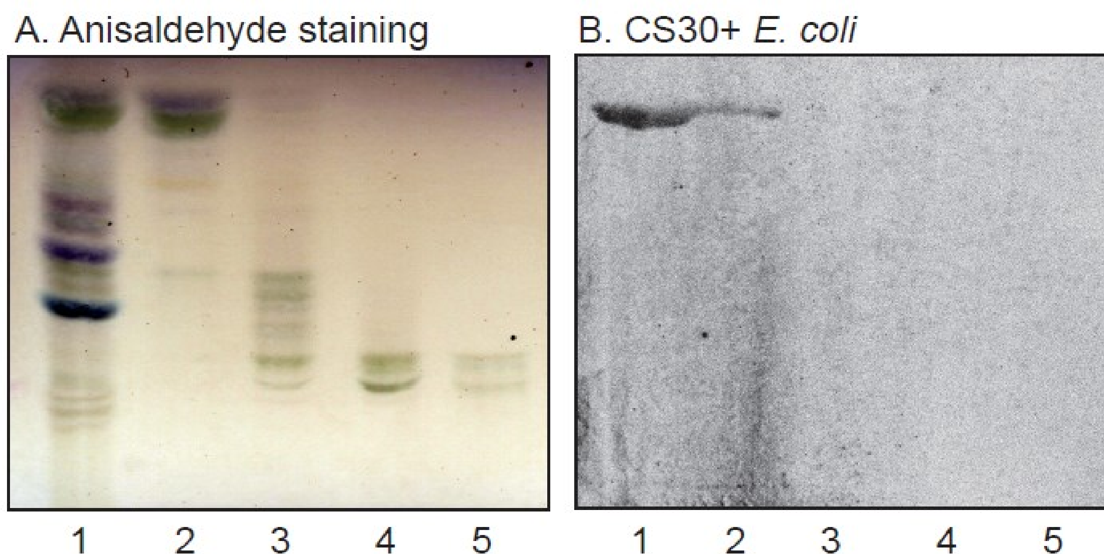

Figure 8: Binding of CS30-expressing *E. coli* to acid glycosphingolipid subfractions from human small intestine. Thin-layer chromatogram after detection with anisaldehyde (A), and autoradiogram obtained by binding of ^{35}S-labeled *E. coli* strain E873 (B). The glycosphingolipids were separated on aluminium-backed silica gel plates, using chloroform/methanol/water 60:35:8 (v/v/v) as solvent system, and the binding assays were performed as described in the Materials and methods section. Autoradiography was for 12 h. The lanes were: **Lane 1:** reference total acid glycosphingolipids of moose large intestine, 40 μg; **Lanes 2-5:** acid glycosphingolipid subfractions of human small intestine, 8 μg/lane.

5.3. Characterisation of CS30$^+$ *E. coli* binding acid glycosphingolipids fractions from human intestine.

The various subfractions obtained were all tested for binding of CS30 expressing *E. coli* and characterised by LC-ESI/MS. The mass spectrometry showed that the first CS30 binding fraction from the first separation (Fig. 8B, lane 2), and the subfractions isolated from this fraction, all contained sulfatide (SO_3-3Galβ1Cer), in most cases as mixtures of different ceramides. However, some sulfatide fractions from the later separations contained relatively pure ceramide species, as shown in Fig. 9.

The three base peak chromatograms had molecular ions at *m/z* 924, 906 and 794, corresponding to glycosphingolipids with a sulfated hexose (SO_3-Hex) and t18:0-h24:0, d18:1-h24:0 and d18:1-h16:0, respectively (Fig. 9, A-C). The three MS^2spectra (Fig. 9, D-F) all had a B_1 ion at *m/z* 241, and a C_1 ion at *m/z* 259, demonstrating a terminal SO_3-Hex. The ions at *m/z* 522, 540 and 568 are due to loss of the fatty acyl from the molecular ion[56].

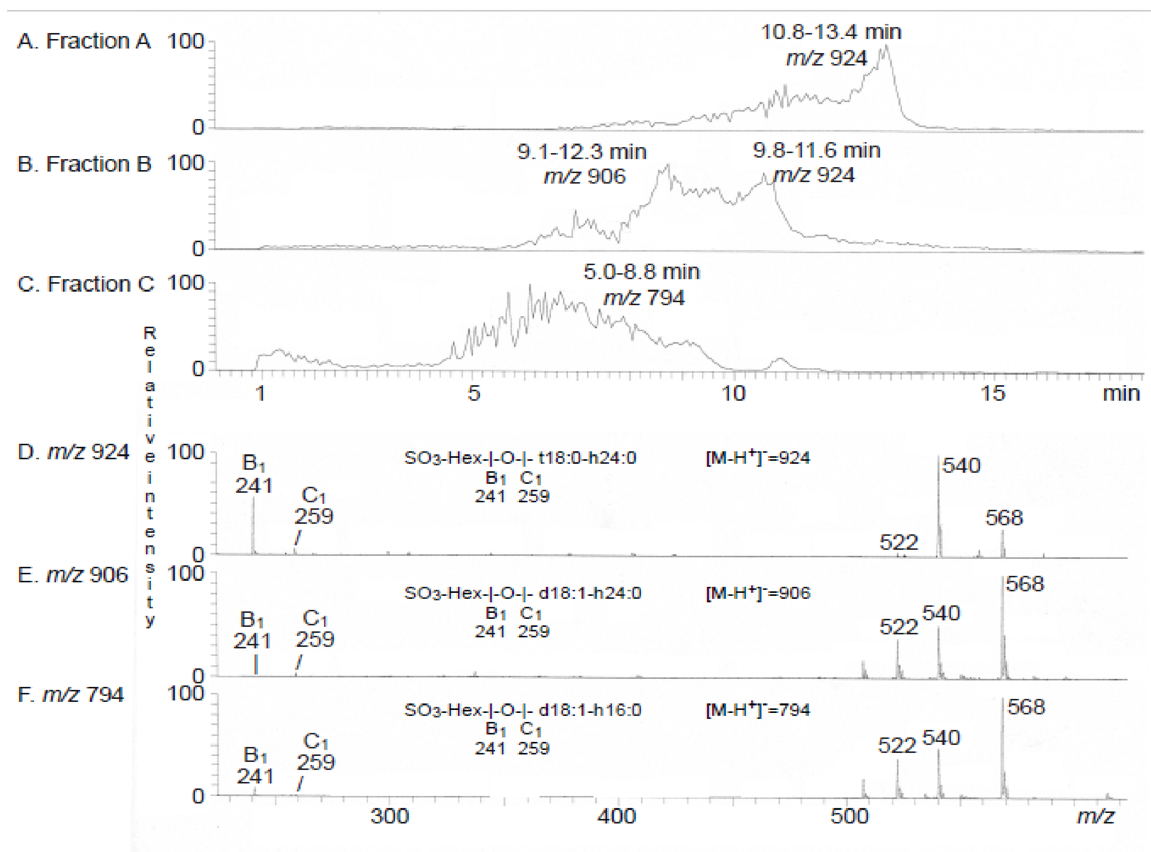

Figure 9: LC-ESI/MS of selected CS30 binding acid glycosphingolipid subfractions of human small intestine. (A) Base peak chromatogram from LC-ESI/MS of fraction A. **(B)** Base peak chromatogram from LC-ESI/MS of fraction B. **(C)** Base peak chromatogram from LC-ESI/MS of fraction C. **(D)** MS^2 of the ion at *m/z* 924 in (A) (retention time 12.6 min). **(E)** MS^2 of the ion at *m/z* 906 in (B) (retention time 8.8 min). **(F)** MS^2 of the ion at *m/z* 794 in (C) (retention time 6.5 min).

Next the binding of monoclonal antibodies directed towards SO_3-3Galβ to these three fractions (denoted fractions A, B, and C) was tested. Binding to all three fractions was obtained (Fig. 10B, lanes 4-6), confirming the sulfatide content deduced from mass spectrometry.

The results from the structural characterisation are summarized in Table 11.

Table 11: Summary of results from characterisation of the sulfatides isolated from human small intestine.

Fraction	m/z	LC-ESI/MS	Anti-SO_3-3Galβ
A=K13953-1	*m/z* 924	SO_3-3Hex-t18:0-h24:0[a]	+
B=K13950-1	*m/z* 906 *m/z* 924	SO_3-3Hex-d18:1-h24:0 SO_3-3Hex-t18:0-h24:0	+
C=K13950-3	*m/z* 794	SO_3-3Hex-d18:1-h16:0	+

[a]In the nomenclature for bases and fatty acids, the number before the colon refers to the length of carbon chain and the number after the colon refers to the total number of double bands in the molecule. Fatty acids with a 2-hydroxy group are denoted by the prefix h e.g. h24:0. For sphingoid bases, d denotes dihydroxy and t trihydroxy. Thus sphingosine (1,3-dihydroxy-2-aminooctadecene) is abbreviated as d18:1 and phytosphingosine (1,3,4-trihydroxy-2-aminooctadecene) is abbreviated as t18:0.

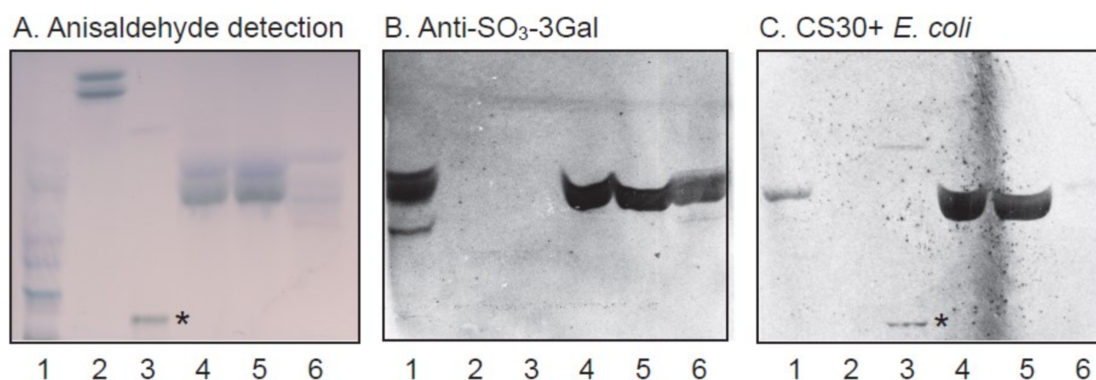

Figure 10: Binding of monoclonal antibodies directed towards SO_3-3Galβ, and of CS30-expressing *E. coli* to acid glycosphingolipid subfractions A, B, and C. Thin-layer chromatogram after detection with anisaldehyde (A), and autoradiograms obtained by binding of anti-SO_3-3Galβ antibodies (B), and CS30$^+$ *E. coli* strain E873 (C). The glycosphingolipids were separated on aluminium-backed silica gel plates, using chloroform/methanol/water 60:35:8 (v/v/v) as solvent system, and the binding assays were performed as described under "Materials and methods”. Autoradiography was for 12 h. The lanes were: **Lane 1:** reference total acid glycosphingolipids of moose large intestine, 40 μg; **Lane 2:** reference galactosylceramide (Galβ1Cer), 4 μg; **Lane 3:** reference sulf-gangliotetraosylceramide (SO_3-3Galβ3GalNAcβ4Galβ4Glcβ1Cer), 4 μg; **Lane 4:** fraction B (sulfatide with d18:1-h24:0 and t18:0-h24:0 ceramides), 4 μg; **Lane 5:** fraction A (sulfatide with t18:0-h24:0 ceramide), 4 μg; **Lane 6:** fraction C (sulfatide with d18:1-h16:0 ceramide), 4 μg.

5.4. Binding of CS30$^+$ ETEC to glycosphingolipids

The three sulfatide fractions (fractions A, B, and C) were selected for further definition of the detailed binding specificity of CS30 expressing *E. coli*. The first binding assays indicated that the binding to sulfatide with d18:1-h16:0 (Fig. 10, lane 6) was weaker than the binding to sulfatide with d18:1-h24:0 and t18:0-h24:0 ceramides (Fig. 10, lanes 4 and 5). This was repeated by binding to different concentrations of the three sulfatides, where again the binding of the CS30$^+$ bacteria to sulfatide with d18:1-h16:0 was weaker (Fig. 11, lanes 5 and 6).

It should be noted that anisaldehyde staining showed several of the subfractions obtained had a bluish band in addition to the green-coloured sulfatide band. This bluish contaminant is cholesterol-sulfate, which often co-migrates with sulfatide. Reference cholesterol-sulfate was therefore included in the chromatogram binding assays. No binding of CS30 expressing *E. coli* to this reference cholesterol-sulfate was however obtained (exemplified in Fig. 6, lane 1).

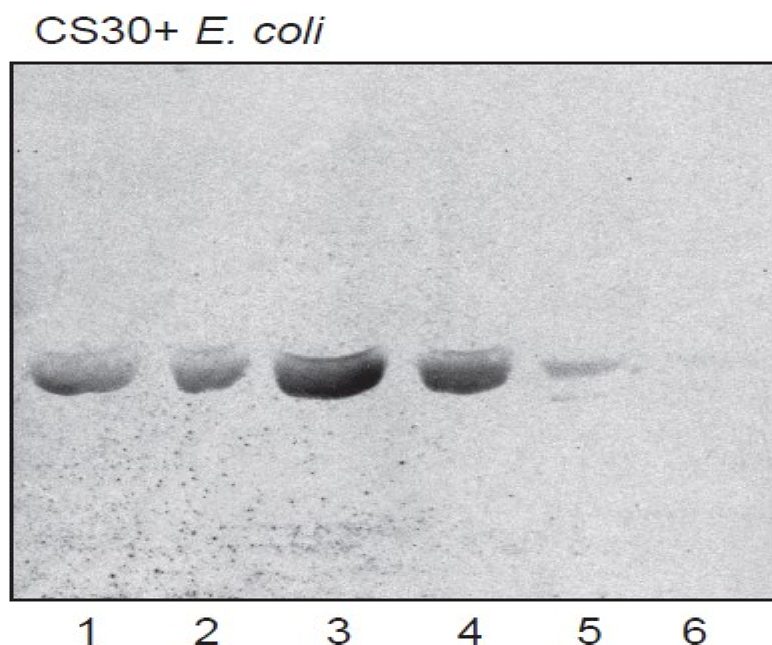

Figure 11: Binding of CS30-expressing *E. coli* to sulfatides with different ceramides. Autoradiogram obtained by binding of ^{35}S-labeled *E. coli* strain E873. The glycosphingolipids were separated on aluminium-backed silica gel plates, using chloroform/methanol/water 65:25:4 (v/v/v) as solvent system, and the binding assays were performed as described under "Materials and methods". Autoradiography was for 12 h. The lanes were: **Lane 1:** fraction B (sulfatide with d18:1-h24:0 and t18:0-h24:0 ceramides), 8 µg; **Lane 2:** fraction B, 4 µg; **Lane 3:** fraction A (sulfatide with t18:0-h24:0 ceramide), 8 µg; **Lane 4:** fraction A, 4 µg; **Lane 5:** fraction C (sulfatide with d18:1-h16:0 ceramide), 8 µg; **Lane 6:** fraction C, 4 µg.

Some glycosphingolipids related to sulfatide were also examined for binding of CS30 expressing *E. coli*. The bacteria did not bind to galactosylceramide (Galβ1Cer) (Fig. 10C, lane 2), which shows that the sulfate group of sulfatide is important for the interaction. On the other hand, there was a binding, albeit weak, to sulf-ganglioteraosylceramide (SO_3-3Galβ3GalNAcβ4Galβ4GlcCer) (Fig. 10C, lane 3).

The results from the chromatogram binding assays are summarized in Table 12. Taken together this demonstrates that CS30 binds to glycosphingolipids with a terminal SO_3-3Gal. In the case of sulfatides there is a tendency of preference for compounds with long fatty acids (C24), indicating that these ceramides give a better exposure of the terminal SO_3-3Gal group.

Table 12: Summary of results from binding of CS30+ *E. coli*

Trivial name	Structure	CS30 binding
Sulfatide t18:0-h24:0	SO_3-3Galβ1Cer	+
Sulfatide d18:1-h24:0	SO_3-3Galβ1Cer	+
Sulfatide d18:1-h16:0	SO_3-3Galβ1Cer	(+)
GalCer	Galβ1Cer	-
Sulf-gangliotetraosylCer	SO_3-3Galβ3GalNAcβ4Galβ4Glcβ1Cer	(+)
GM3	Neu5Acα3Galβ4Glcβ1Cer	-
GD3	Neu5Acα8Neu5Acα3Galβ4Glcβ1Cer	-
Cholesterol-sulfate	cholest-5-en-3β-olsulfate	-

5.5 Construction of recombinant E. coli strain expressing CS30

The CS30 operon containing *csmA-G* genes and plasmid pMT (carrying chloramphenicol resistance genes) were amplified by PCR. Template DNA for *csmA-G* was obtained from ETEC strain E873 by Rapid boil method (see section Materials and methods) and template DNA for pMT was obtained from *V. cholerae* strain 1275 by using plasmid purification kit. Primers used in this project were designed to contain restriction sites for XmaI on the one side and SacI on the other side of the amplified sequences, *csmA-G* and pMT (Fig. 12). Amplified PCR products were analysed by gel electrophoresis, showing one band between 6,000 and 8,000 bp for three samples of *csmA-G* (6,623 bp), and one band between 3,500 and 4,000 bp for three samples of pMT (3,878 bp) (Fig. 13A, lanes 1-6). Fragment *csmA-G* was Sanger sequenced. Obtained sequences mapped to *in silico* predicted sequence. The concentration of each product was measured on Nanodrop and the samples with the highest concentration were selected for the digestion with SacI and XmaI. Self-ligation of pMT was prevented by dephosphorylation.

The ligation was performed by T4 DNA ligase and the results were analysed by gel electrophoresis. A single band in ligation lane (Fig. 13B, lane 3) indicates that the ligation ran properly, as well as its place and shape.

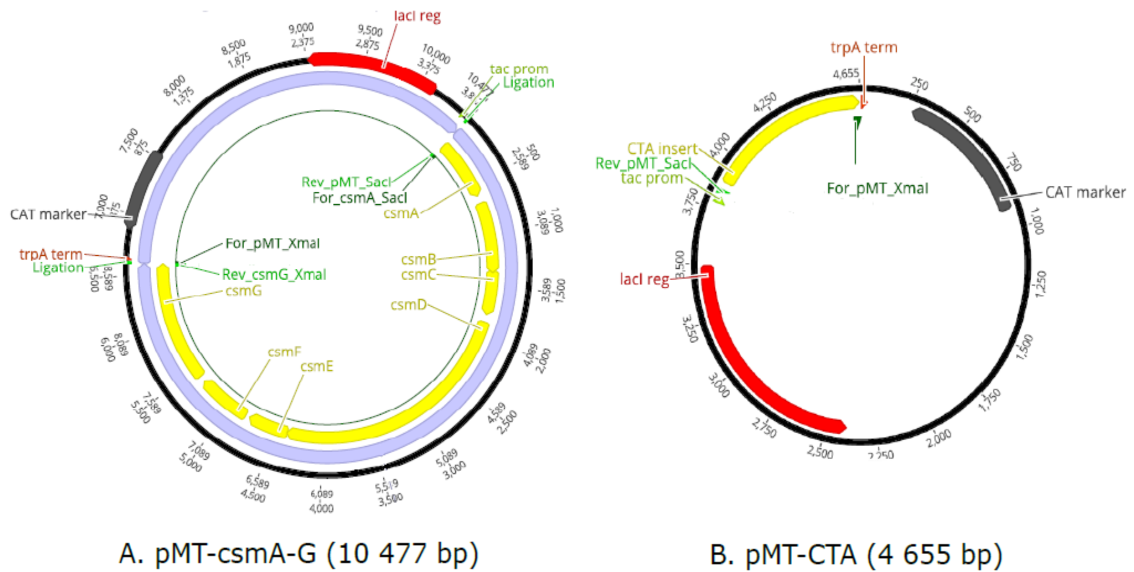

Figure 12: Ligated plasmid pMT-*csmA-G* and original plasmid pMT-CTA from cholera strain 1275.

Ligated DNA was transformed into chemically and electrocompetent *E. coli* TOP10. Chloramphenicol-resistant colonies were screened for the presence of pMT-*csmA-G* and further analysed by restriction analysis of isolated plasmids. Fifteen chemically competent and five electrocompetent transformed colonies were digested with AccI and analysed by gel electrophoresis. Lane 3 of colony number 12 (Fig. 13C) contained three bands (779, 3518, 6186) indicating pMT-*csmA-G* plasmid.

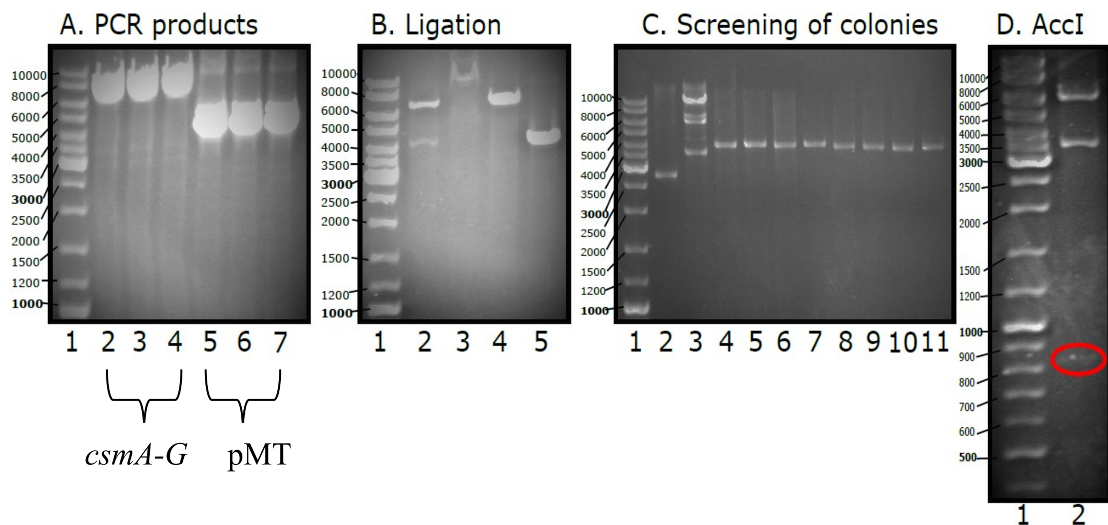

Figure 13: Construction of pMT-*csmA-G*. Agarose gel chromatography with Midori Green staining of amplified PCR products pMT and *csmA-G* **(A)**, ligation of prepared sequences with T4 DNA polymerase **(B)**, screening of transformed colonies by restriction analysis of isolated plasmids after 1 hour **(C)**, and restriction analyses of colony 12 with AccI after 2 hours **(D)**. The lanes for A-C are described in Table 13. The lanes for D were: **Lane 1:** Gene Ruler DNA Ladder, 4 µl; **Lane 2:** colony 12 cut with AccI.

Table 13: Description of lanes in Figure 14.

	Gel A	Gel B	Gel C
Lane 1	GeneRuler DNA Ladder	GeneRuler DNA Ladder	GeneRuler DNA Ladder
Lane 2	Amplified *csmA-G*	Ligation reference	Colony 11
Lane 3	Amplified *csmA-G*	Ligation of *csmA-G* and pMT	Colony 12
Lane 4	Amplified *csmA-G*	Fragment *csmA-G*	Colony 13
Lane 5	Amplified pMT	Vector pMT	Colony 14
Lane 6	Amplified pMT		Colony 15
Lane 7	Amplified pMT		Colony 16
Lane 8-11			Colony 17–20

The resulting 10,477 bp plasmid containing CS30 operon located downstream the IPTG-induced *tac* promoter was then tested for CS30 expression. Following RNA extraction and cDNA synthesis, reverse transcription PCR was performed using specific primers for each gene in CS30 operon. Gel electrophoresis of PCR products confirmed IPTG-induced expression of all seven genes (Fig. 14), when showing one band in every lane with a size as predicted *in silico* (Table 10). Isolated plasmid pMT-*csmA-G* was Sanger sequenced. Obtained sequences mapped to *in silico* predicted sequence.

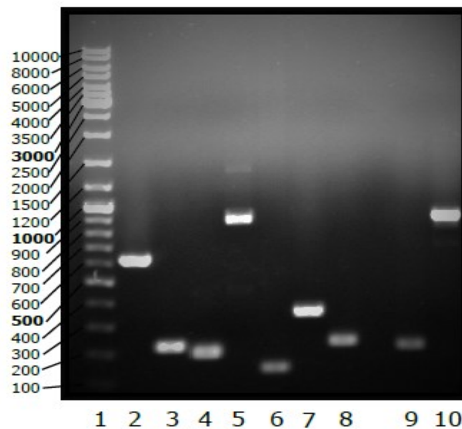

Figure 14: Expression of CS30. Agarose gel electrophoresis with Midori Green staining of RT-PCR products. Template cDNA was obtained by culturing colony 12 in LB+CM+IPTG that was followed by RNA extraction and cDNA synthesis (Lanes 2–9) or extracted from E873 (Lane 10). Used primers and *in silico* predicted products are listed in Table 10. The lanes were: **Lane 1:** GeneRuler DNA Ladder Mix, 4 µl; **Lane 2:** amplified *csmA*, 3 µl; **Lane 3:** amplified *csmB*, 3 µl; **Lane 4:** amplified *csmC*, 3 µl; **Lane 5:** amplified *csmD*, 3 µl; **Lane 6:** amplified *csmE*, 3 µl; **Lane 7**: amplified *csmF*, 3 µl; **Lane 8:** amplified *csmG*, 3 µl; **Lane 9:** *csmA* amplified using primers CS30_F and CS30_R, 3 µl; **Lane 10:** amplified *csmD* with DNA template extracted from E873, 3 µl.

Transmission electron microscopy imagining showed that reference strain E873 was associated with rigid fimbriae and that recombinant strain TOP10-CS30 lacked any visible fimbriae (Fig. 15).

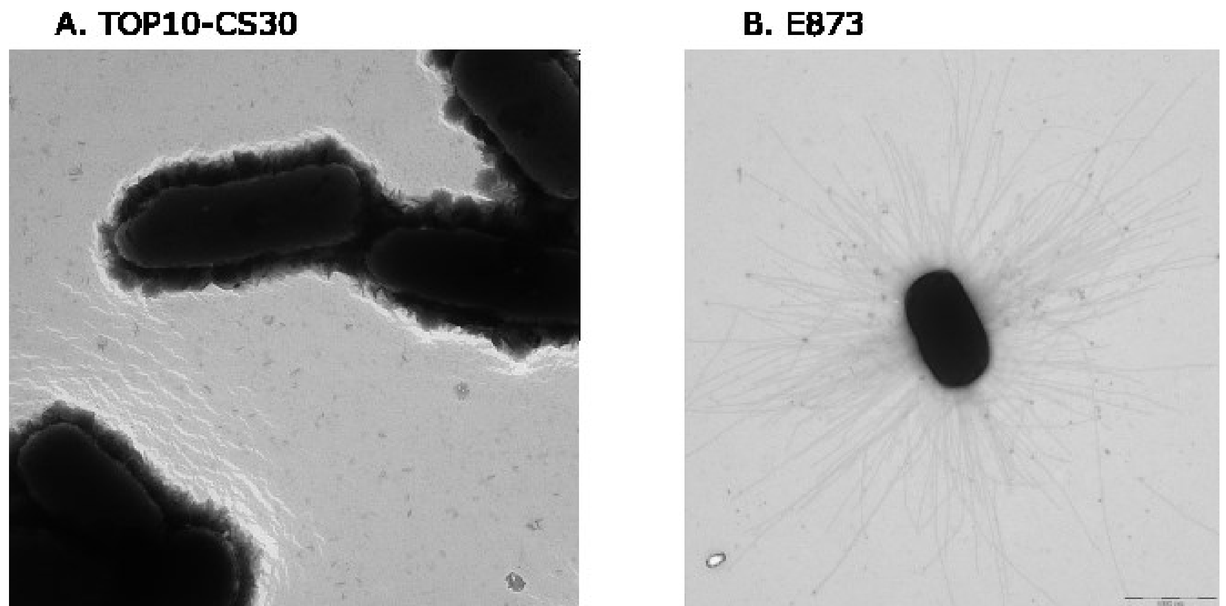

Figure 15: Transmission electron microscopy analysis of TOP10-CS30 (A) and reference strain E873 (B). Negative stain of the TOP10-CS30 **(A)** lacking visible fimbrial structures on the surface and of the wild type strain E873 **(B)** harbouring CS30 shows fimbriated bacteria.

6 DISCUSSION

The recently identified colonization factor CS30 belongs to the group of important virulence factors contributing to pathogenesis of ETEC-induced diarrhoea. Identification and detailed characterisation of CS30 receptors, which are presumably glycosphingolipids as in the case of other CFs, i.e. CS6 and CFA/I, is important for better understanding of ETEC pathogenesis as well as for developing treatments for ETEC infection and ETEC vaccine development.

Upon screening of ETEC CS30 glycosphingolipid recognition we obtained binding to a fast-migrating compound in the acid glycosphingolipid fraction from human small intestine. The mass spectrometry showed that this $CS30^{+}$ETEC strain binding glycosphingolipid is sulfatide (the major acid glycosphingolipid of the human enterocytes[57]), mostly as mixtures of different ceramide species. Binding assays with monoclonal antibodies directed towards SO_3-3Galβ confirmed the deduced sulfatide content in the tested samples. Further analysis by LC-ESI/MS defined the ceramide species as t18:0-h24:0, d18:1-h24:0 and d18:1-h16:0. The binding assays indicated a weaker binding of the $CS30^{+}$ bacteria to sulfatide with d18:1-h16:0 than to the other two ceramides, thus preference for compounds with long fatty acids may be assumed.

When some glycosphingolipids related to sulfatide were examined for binding, negative results of binding to galactosylceramide indicate importance of sulfate group for the interaction. The weak binding of CS30 expressing ETEC to the sulf-gangliotetraosylceramide (SO_3-3Galβ3GalNAcβ4Galβ4GlcCer) isolated from mouse intestine also highlights the importance of terminal SO_3-3Gal group. However, this glycosphingolipid has not been identified in human tissues, and may thus be excluded from potential CS30 binding receptors in human intestine. Interestingly, sulf-gangliotetraosylceramide is recognised by the CS30, but not by CS6, which is another sulfatide binding CF[30]. The difference in binding may be due to CS30 and CS6 different proteinous structure.

Testing of glycosphingolipids isolated from porcine small intestine confirmed binding of $CS30^{+}$ *E. coli* to the acid glycosphingolipid fraction, more specifically in the sulfatide region. Sulfatide is also present on the porcine intestinal surface[58]. On the other hand, non-acid glycosphingolipid fraction showed no binding. The function of the sulfatide in porcine ETEC pathogenesis gains importance due to

the fact that STb enterotoxin mainly occurring in animals is also recognised by this glycosphingolipid[12].

The binding studies performed in this project used the wild type $CS30^{+}$ *E. coli* strain E873. ETEC strains may have additional outer membrane proteins, in addition to CFs. To avoid false positive results of the binding studies, a recombinant strain expressing only CS30 was desirable. To achieve this, the CS30 operon, harbouring the seven genes *csmA-G,* was isolated from strain E873, inserted into vector plasmid pMT and transformed into the commercial strain TOP10, resulting into recombinant strain TOP10-CS30.

Firstly, the used method was performed as described by von Mentzer *et al*., i.e. the CS30 gene *csmA-G* was inserted into pMT-CTA plasmid and transformed into a TOP10 strain. Designed primers used for *csmA-G* amplification harbour restriction sites for EcoRI and HindIII that would form sticky ends. Plasmid pMT-CTA was isolated from cholera strain 1275 and digested by EcoRI and HindIII. The plasmid was cut in two sites to form sequence excluding CTA part, which was supposed to be replaced by *csmA-G*. However, these restriction endonucleases could not be used for digestion of sequence *csmA-G*, because it contains more restriction sites for HindIII, which would lead to wrong fragmentation. We decided to omit the digestion of fragment and use blunt ends instead. Despite several attempts, this approach was not successful. None of the transformed colonies harboured the recombinant DNA pMT-*csmA-G* for unknown reasons. The problem might be ligation with blunt ends, when there is a higher chance of self-ligation, loss of the material during purifying steps or toxicity of recombinant molecules for transformed cells.

Thereafter we designed and used a different approach in which we amplified both molecules, *csmA-G* and pMT, and designed primers in both cases harboured restriction sites for SacI and XmaI. Therefore, sticky ends were produced after digestion in both sequences. Correct ligation of fragment and plasmid was confirmed by gel electrophoresis. CS30 positive strain was selected from the samples of transformed colonies and tested by RT-PCR for presence of mRNA from *csmA-G*.

The reverse transcription PCR confirmed that the constructed recombinant strain (TOP10-CS30) harbours all seven genes *csmA-G* and that they are transcribed. However, the presence of the fimbriae on the bacterial surface could not be seen by transmission electron microscopy imagining for unknown reasons. The lack of CS30

surface expression could not be explained by a non-functioning *tac* promoter or by the operon being silent, because of the positive outcome from RT-PCR.

The loss of surface expression of the previously mentioned CS6 was described by Nicklasson *et al*. CS6 operon contains four genes *cssABCD*, encoding two structural subunits (CssA and CssB), a periplasmic chaperone (CssC), and a molecular usher (CssD). Phenotypic expression of CS6 starts by transportation of the subunits from cytosol to the periplasmic space mediated assumingly by the Sec machinery of the general secretory pathway. The chaperone subunit protects the structural subunits from proteolytic degradation and transports them across the periplasm to the outer membrane, while the usher translocates the CssA and CssB to the surface. In the paper, six out of eight studied CS6 positive strains that lack surface expression contained attenuating single-point mutations within the coding region of the periplasmic chaperone. This might negatively affect the assembly of the structural subunits in the periplasm[59]. Similarly, there might be single-point mutations within the coding region of CS30 operon causing defective assembly of the protein on the outer membrane. Additionally, the exact mechanism of expression has not been described yet, suggesting mutation in any part of the sequence might cause defective or none expression on the surface. The regulation of CS30 secretion merits further studies. There is a possibility that proper construction of CS30 fimbriae requires either native promoter or some other regulatory sequence from the native plasmid.

The mechanism of adherence mediated by CS30 merits further studies. Inhibition studies with sulfated oligosaccharides will be done to determine the role of sulfate group in CS30 binding. Due to the negative charge of sulfatide, testing of the binding with the point mutations in the CS30 protein (changing lysines) is proposed for more detailed characterisation of this interaction. For the expression of CS30 on the cell surface, adding 50 bp sequence in forward primer for initiation of secretion is planned to be done. In the case of negative outcome, strains with native promoter are planned to be constructed. CS30 positive colonies expressing protein on the outer membrane will be tested for binding to CaCo-2 cells. CS30 protein will be isolated from the cells, and tested by binding assays with glycosphingolipids, alongside with the bacterial culture. Additionally, mutant strain lacking *csmG* (encoding adhesin) will be constructed to determine importance of this adhesin in the mechanism of CS30 adherence and pathogenesis.

7 CONCLUSION

The main aims of the diploma thesis were to identify the intestinal receptor(s) for the ETEC colonization factor CS30 and to construct the recombinant strain expressing specifically CS30. We confirmed the hypothesis that the ETEC CS30 binds to the human and porcine specific intestinal receptors. In summary, the screening of a mixture of glycosphingolipids isolated from human and porcine small intestine revealed the fast-migrating compound that binds to CS30 and contains sulfatide (SO_3-3Galβ1ceramide) with different ceramide parts: SO_3-3Galβ1-t18:0-h24:0, SO_3-3Galβ1-d18:1-h24:0, and SO_3-3Galβ1-d18:1-h16:0. CS30 binds to glycosphingolipids with a terminal SO_3-3Gal. In the case of sulfatides there is a tendency of preference for compounds with long fatty acids (C24), indicating that these ceramides give a better exposure of the terminal SO_3-3Gal group.

The newly produced recombinant strain TOP10-CS30 is expressing all seven genes of the CS30 operon (*csmA-G),* but CS30 fimbriae cannot be viewed on the cell surface. Thus, further modifications of the used method will be performed in the near future to construct fully functional strain (in the means of adherence) and to isolate CS30 for further characterisation of receptor-CS30 interaction. There is a hope that the presented data and future studies will contribute information useful for constructing vaccines against ETEC.

8 REFERENCES

(1) Varki, A.; Sharon, N. Historical Background and Overview. In: Varki, A. (Ed.) *Essentials of Glycobiology*. 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY) 2009. ISBN 978-087-969-770-9

(2) Schnaar, R. L.; Suzuki, A.; Stanley, P. Glycosphingolipids. In: Varki, A. (Ed.) *Essentials of Glycobiology*. 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY) 2009. ISBN 978-087-969-770-9

(3) Varki, A.; Etzler, M. E.; Cummings, R. D.; Esko, J. D. Discovery and Classification of Glycan-Binding Proteins. In: Varki, A. (Ed.) *Essentials of Glycobiology*. 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY) 2009. ISBN 978-087-969-770-9

(4) Esko, J. D.; Sharon, N. Microbial Lectins: Hemagglutinins, Adhesins, and Toxins. In: Varki, A. (Ed.) *Essentials of Glycobiology*. 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY) 2009. ISBN 978-087-969-770-9

(5) Escherich, T. H. The Intestinal Bacteria of the Neonate and Breast-Fed Infant. *Rev. Infect. Dis.*, **1989**, *11*, 352–356.

(6) Croxen, M. A.; Law, R. J.; Scholz, R.; Keeney, K. M.; Wlodarska, M.; Finlay, B. B. Recent Advances in Understanding Enteric Pathogenic Escherichia Coli. *Clin. Microbiol. Rev.*, **2013**, *26*, 822–880.

(7) Kaper, J. B.; Nataro, J. P.; Mobley, H. L. T. Pathogenic Escherichia Coli. *Nat. Rev. Microbiol.*, **2004**, *2*, 123–140.

(8) Touchon, M.; Hoede, C.; Tenaillon, O.; Barbe, V.; Baeriswyl, S.; Bidet, P.; Bingen, E.; Bonacorsi, S.; Bouchier, C.; Bouvet, O.; et al. Organised Genome Dynamics in the Escherichia Coli Species Results in Highly Diverse Adaptive Paths. *PLoS Genet.*, **2009**, *5*, e1000344.

(9) Frost, L. S.; Leplae, R.; Summers, A. O.; Toussaint, A. Mobile Genetic Elements: The Agents of Open Source Evolution. *Nat. Rev. Microbiol.*, **2005**, *3*, 722–732.

(10) Qadri, F.; Svennerholm, A.-M.; Faruque, A. S. G.; Sack, R. B. Enterotoxigenic Escherichia Coli in Developing Countries: Epidemiology, Microbiology, Clinical Features, Treatment, and Prevention. *Clin. Microbiol. Rev.*, **2005**, *18*, 465–483.

(11) World Health Organization - WHO. Future Directions for Research on Enterotoxigenic Escherichia Coli Vaccines for Developing Countries. *Wkly. epidmiological Rec.*, **2006**, *11*, 97–104.

(12) Dubreuil, J. D.; Isaacson, R. E.; Schifferli, D. M. Animal Enterotoxigenic Escherichia Coli. *EcoSal Plus*, **2016**, *7*, 1–47.

(13) World Health Organization. WHO | Child mortality http://www.who.int/pmnch/media/press_materials/fs/fs_mdg4_childmortality/en/ (accessed Nov 14, 2017).

(14) World Health Organization. WHO | Diarrhoeal disease http://www.who.int/mediacentre/factsheets/fs330/en/ (accessed Oct 28, 2017).

(15) Gupta, S. K.; Keck, J.; Ram, P. K.; Crump, J. A.; Miller, M. A.; Mintz, E. D. Part III. Analysis of Data Gaps Pertaining to Enterotoxigenic Escherichia Coli Infections in Low and Medium Human Development Index Countries, 1984–2005. *Epidemiol. Infect.*, **2008**, *136*, 721–738.

(16) Wennerås, C.; Erling, V. Prevalence of Enterotoxigenic Escherichia Coli - Associated Diarrhoea and Carrier State in the Developing World. *J. Health. Popul. Nutr.*, **2004**, *22*, 370–382.

(17) Shah, N.; Ldupont, H.; Ramsey, D. J. Global Etiology of Travelers' Diarrhea: Systematic Review from 1973 to the Present. *Am. J. Trop. Med. Hyg.*, **2009**, *80*, 609–614.

(18) Levine, M. M.; Nalin, D. R.; Hoover, D. L.; Bergquist, E. J.; Hornick, R. B.; Young, C. R. Immunity to Enterotoxigenic Escherichia Coli. *Infect. Immun.*, **1979**, *23*, 729–736.

(19) Pakalniskiene, J.; Falkenhorst, G.; Lisby, M.; Madsen, S. B.; Olsen, K. E. P.; Nielsen, E. M.; Mygh, A.; Boel, J.; Mølbak, K. A Foodborne Outbreak of Enterotoxigenic E. Coli and Salmonella Anatum Infection after a High-School Dinner in Denmark, November 2006. *Epidemiol. Infect.*, **2009**, *137*, 396–401.

(20) Beatty, M. E.; Adcock, P. M.; Smith, S. W.; Quinlan, K.; Kamimoto, L. A.; Rowe, S. Y.; Scott, K.; Conover, C.; Varchmin, T.; Bopp, C. A.; et al. Epidemic Diarrhea due to Enterotoxigenic Escherichia Coli. *Clin. Infect. Dis.*, **2006**, *42*, 329–334.

(21) Gaastra, W.; Svennerholm, A. M. Colonization Factors of Human Enterotoxigenic Escherichia Coli (ETEC). *Trends Microbiol.*, **1996**, *4*, 444–452.

(22) Qadri, F.; Saha, A.; Ahmed, T.; Al Tarique, A.; Begum, Y. A.; Svennerholm, A. M. Disease Burden due to Enterotoxigenic Escherichia Coli in the First 2 Years of Life in an Urban Community in Bangladesh. *Infect. Immun.*, **2007**, *75*, 3961–3968.

(23) Okhuysen, P. C.; Jiang, Z. D.; Carlin, L.; Forbes, C.; DuPont, H. L. Post-Diarrhea Chronic Intestinal Symptoms and Irritable Bowel Syndrome in North American Travelers to Mexico. *Am. J. Gastroenterol.*, **2004**, *99*, 1774–1778.

(24) Turner, S. M.; Scott-Tucker, A.; Cooper, L. M.; Henderson, I. R. Weapons of Mass Destruction: Virulence Factors of the Global Killer Enterotoxigenic Escherichia Coli. *FEMS Microbiol. Lett.*, **2006**, *263*, 10–20.

(25) Svennerholm, A. M.; Wikstrom, M.; Lindblad, M.; Holmgren, J. Monoclonal Antibodies against Escherichia Coli Heat-Stable Toxin (STa) and Their Use in a Diagnostic ST Ganglioside GM1-Enzyme-Linked Immunosorbent Assay. *J. Clin. Microbiol.*, **1986**, *24*, 585–590.

(26) Svennerholm, A. M.; Wiklund, G. Rapid GM1-Enzyme-Linked Immunosorbent Assay with Visual Reading for Identification of Escherichia Coli Heat-Labile Enterotoxin. *J. Clin. Microbiol.*, **1983**, *17*, 596–600.

(27) Steinsland, H.; Valentiner-Branth, P.; Grewal, H. M. S.; Gaastra, W.; Mølbak, K. år.; Sommerfelt, H. Development and Evaluation of Genotypic Assays for the Detection and Characterization of Enterotoxigenic Escherichia Coli. *Diagn. Microbiol. Infect. Dis.*, **2003**, *45*, 97–105.

(28) Sjöling, A.; Wiklund, G.; Savarino, S. J.; Cohen, D. I.; Svennerholm, A.-M. Comparative Analyses of Phenotypic and Genotypic Methods for Detection of Enterotoxigenic Escherichia Coli Toxins and Colonization Factors. *J. Clin. Microbiol.*, **2007**, *45*, 3295–3301.

(29) Qadri, F.; Girón, J. A.; Helander, A.; Begum, Y. A.; Asaduzzaman, M.; Xicohténcatl-Cortes, J.; Negrete, E.; Albert, M. J. Human Antibody Response to Longus Type IV Pilus and Study of Its Prevalence among Enterotoxigenic Escherichia Coli in Bangladesh by Using Monoclonal Antibodies. *J. Infect. Dis.*, **2000**, *181*, 2071–2074.

(30) Jansson, L.; Tobias, J.; Jarefjäll, C.; Lebens, M.; Svennerholm, A. M.; Teneberg, S. Sulfatide Recognition by Colonization Factor Antigen CS6 from Enterotoxigenic Escherichia Coli. *PLoS One*, **2009**, *4*, 2–11.

(31) Jansson, L.; Tobias, J.; Lebens, M.; Svennerholm, A. M.; Teneberg, S. The Major Subunit, CfaB, of Colonization Factor Antigen I from Enterotoxigenic Escherichia Coli Is a Glycosphingolipid Binding Protein. *Infect. Immun.*, **2006**, *74*, 3488–3497.

(32) Sears, C. L.; Kaper, J. B. Enteric Bacterial Toxins: Mechanisms of Action and Linkage to Intestinal Secretion. *Microbiol. Rev.*, **1996**, *60*, 167–215.

(33) Merson, M. H.; Orskov, F.; Orskov, I.; Sack, R. B.; Huq, I.; Koster, F. T. Relationship Between Enterotoxin Production and Serotype in Enterotoxigenic Escherichia coli. *Infect. Immun.*, **1979**, *23*, 325–329.

(34) Iguchi, A.; Mentzer, A. von; Kikuchi, T.; Thomson, N. R. An Untypeable Enterotoxigenic Escherichia Coli Represents One of the Dominant Types Causing Human Disease. *Microb. Genomics*, **2017**, *3*, e000121.

(35) Weiglmeier, P. R.; Rösch, P.; Berkner, H. Cure and Curse: E. Coli Heat-Stable Enterotoxin and Its Receptor Guanylyl Cyclase C. *Toxins*, **2010**, *2*, 2213–2229.

(36) Anantha, R. P.; McVeigh, A. L.; Lee, L. H.; Agnew, M. K.; Cassels, F. J.; Scott, D. A.; Whittam, T. S.; Savarino, S. J. Evolutionary and Functional Relationships of Colonization Factor Antigen I and Other Class 5 Adhesive Fimbriae of Enterotoxigenic Escherichia Coli. *Infect. Immun.*, **2004**, *72*, 7190–7201.

(37) Nada, R. A.; Shaheen, H. I.; Khalil, S. B.; Mansour, A.; El-Sayed, N.; Touni, I.; Weiner, M.; Armstrong, A. W.; Klena, J. D. Discovery and Phylogenetic Analysis of Novel Members of Class B Enterotoxigenic Escherichia Coli Adhesive Fimbriae. *J. Clin. Microbiol.*, **2011**, *49*, 1403–1410.

(38) von Mentzer, A.; Connor, T. R.; Wieler, L. H.; Semmler, T.; Iguchi, A.; Thomson, N. R.; Rasko, D. A.; Joffre, E.; Corander, J.; Pickard, D.; et al. Identification of Enterotoxigenic Escherichia Coli (ETEC) Clades with Long-Term Global Distribution. *Nat. Genet.*, **2014**, *46*, 1321–1326.

(39) Nicklasson, M.; Sjöling, Å.; von Mentzer, A.; Qadri, F.; Svennerholm, A. M. Expression of Colonization Factor CS5 of Enterotoxigenic Escherichia Coli (ETEC) Is Enhanced in Vivo and by the Bile Component Na Glycocholate Hydrate. *PLoS One*, **2012**, *7*, 1–9.

(40) Schmidt, M. A. Non-Fimbrial Adhesins of Escherichia Coli. In: Klemm, P. (Ed.) *Fimbriae: Adhesins, Genetics, Biogenesis, and Vaccines*. 1st Ed., CRC Press, Boca Raton 1994. ISBN 978-084-934-894-5

(41) von Mentzer, A.; Tobias, J.; Wiklund, G.; Nordqvist, S.; Aslett, M.; Dougan, G.; Sjöling, Å.; Svennerholm, A.-M. Identification and Characterization of the Novel Colonization Factor CS30 Based on Whole Genome Sequencing in Enterotoxigenic Escherichia Coli (ETEC). *Sci. Rep.*, **2017**, *7*, 12514.

(42) Kumar, P.; Luo, Q.; Vickers, T. J.; Sheikh, A.; Lewis, W. G.; Fleckenstein, J. M. Eata, an Immunogenic Protective Antigen of Enterotoxigenic Escherichia Coli, Degrades Intestinal Mucin. *Infect. Immun.*, **2014**, *82*, 500–508.

(43) Elsinghorst, E. A.; Weitz, J. A. Epithelial Cell Invasion and Adherence Directed by the Enterotoxigenic Escherichia Coli Tib Locus Is Associated with a 104-Kilodalton Outer Membrane Protein. *Infect. Immun.*, **1994**, *62*, 3463–3471.

(44) Lindenthal, C.; Elsinghorst, E. A. Enterotoxigenic Escherichia Coli TibA Glycoprotein Adheres to Human Intestine Epithelial Cells. *Infect. Immun.*, **2001**, *69*, 52–57.

(45) Sherlock, O.; Vejborg, R. M.; Klemm, P. The TibA Adhesin/invasin from Enterotoxigenic Escherichia Coli Is Self Recognizing and Induces Bacterial Aggregation and Biofilm Formation. *Infect. Immun.*, **2005**, *73*, 1954–1963.

(46) Fleckenstein, J. M.; Kopecko, D. J.; Warren, R. L.; Elsinghorst, E. A. Molecular Characterization of the Tia Invasion Locus from Enterotoxigenic Escherichia Coli. *Infect. Immun.*, **1996**, *64*, 2256–2265.

(47) Gonzales, L.; Sanchez, S.; Zambrana, S.; Iñiguez, V.; Wiklund, G.; Svennerholm, A. M.; Sjölinga, A. Molecular Characterization of Enterotoxigenic Escherichia Coli Isolates Recovered from Children with Diarrhea during a 4-Year Period (2007 to 2010) in Bolivia. *J. Clin. Microbiol.*, **2013**, *51*, 1219–1225.

(48) Patel, S. K.; Dotson, J.; Allen, K. P.; Fleckenstein, J. M. Identification and Molecular Characterization of EatA, An Autotransporter Protein of Enterotoxigenic Escherichia Coli. *Infect. Immun.*, **2004**, *72*, 1786–1794.

(49) Fleckenstein, J. M.; Roy, K.; Fischer, J. F.; Burkitt, M. Identification of a Two-Partner Secretion Locus of Enterotoxigenic Escherichia Coli. *Infect. Immun.*, **2006**, *74*, 2245–2258.

(50) Brown, T. A. *Gene Cloning and Dna Analysis / T.A. Brown*. 7th Ed., John Wiley & Sons, Chichester 2016. ISBN 978-111-907-256-0

(51) Nicholl, D. S. T. *An Introduction to Genetic Engineering*. 2nd Ed., Cambridge University Press, Cambridge 2002. ISBN 052-180-867-7

(52) Kaledin, A. S.; Sliusarenko, A. G.; Gorodetskiĭ, S. I. Isolation and Properties of DNA Polymerase from Extreme Thermophylic Bacteria Thermus Aquaticus YT-1. *Biokhimiia*, **1980**, *45*, 644–651.

(53) Terpe, K. Overview of Thermostable DNA Polymerases for Classical PCR Applications: From Molecular and Biochemical Fundamentals to Commercial Systems. *Appl. Microbiol. Biotechnol.*, **2013**, *97*, 10243–10254.

(54) Carattoli, A. Resistance Plasmid Families in Enterobacteriaceae. *Antimicrob. Agents Chemother.*, **2009**, *53*, 2227–2238.

(55) Wong, D. W. S. *The ABCs of Gene Cloning*. 2nd Ed., Springer, New York 2006. ISBN 978-038-728-663-1

(56) Hsu, F. F.; Turk, J. Studies on Sulfatides by Quadrupole Ion-Trap Mass Spectrometry with Electrospray Ionization: Structural Characterization and the Fragmentation Processes That Include an Unusual Internal Galactose Residue Loss and the Classical Charge-Remote Fragmentation. *J. Am. Soc. Mass Spectrom.*, **2004**, *15*, 536–546.

(57) Falk, K. E.; Karlsson, K. A.; Leffler, H.; Samuelsson, B. E. Specific Pattern of Glycosphingolipids Enriched in a Mucosa Scraping of Human Small Intestine. *FEBS Lett.*, **1979**, *101*, 273–276.

(58) Suzuki, C.; Marita, A.; Yosizawa, Z. Glycolipids Isolated from Porcine Intestine. *Arch. Biochem. Biophys.*, **1968**, *127*, 140–149.

(59) Nicklasson, M.; Sjöling, Å.; Lebens, M.; Tobias, J.; Janzon, A.; Brive, L.; Svennerholm, A.-M. Mutations in the Periplasmic Chaperone Leading to Loss of Surface Expression of the Colonization Factor CS6 in Enterotoxigenic Escherichia Coli (ETEC) Clinical Isolates. *Microb. Pathog.*, **2008**, *44*, 246–254.