Genetic diversity of the heat labile (LT) and heat stable (ST) toxins of human enterotoxigenic *Escherichia coli* (ETEC)

New insights into polymorphism, regulation, and gene transcription

Enrique Joffré

Department of Microbiology & Immunology

Institute of Biomedicine

Sahlgrenska Academy at University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2015

Cover illustration: Enrique Joffré

Genetic diversity of the heat labile (LT) and heat stable (ST) toxins of human enterotoxigenic *Escherichia coli* (ETEC): New insights into polymorphism, regulation, and gene transcription

enrique.joffre@gu.se

ISBN 978-91-628-9508-2

Printed in Gothenburg, Sweden 2015

INEKO AB

To my lovely family

Desyi, Henrry and Adri

"*Knowing is not enough; we must apply.*
Willing is not enough; we must do."
—Goethe

ABSTRACT

Infection with enterotoxigenic *Escherichia coli* (ETEC) is a leading cause of diarrhea in children in developing countries and travelers to endemic regions. ETEC is a diverse pathogen, with a wide range of virulence factors including enterotoxins and more than 25 identified colonization factors (CFs). ETEC infection causes varying symptoms (mild to profuse, watery, cholera-like diarrhea) as a result of the colonization of the small intestine via CFs, secretion of heat labile (LT) and/or heat stable enterotoxins (STp and STh).

To expand the knowledge about the complexity of ETEC pathogenesis we studied the genetic diversity of the LT and ST toxins, using a clinical ETEC strains collection isolated worldwide during three decades. By genomic sequencing we found high diversity in the toxin amino acid sequences, especially in LT where 20 amino acid variants were identified. The LTA subunit was highly polymorphic while the LTB subunit was more conserved. The most common LT variants were LT1 and LT2. ST was less heterogeneous, including 3 ST alleles found in STp and 3 in STh. Phylogenetic analysis of the toxins revealed worldwide distribution of the different variants, and an association with specific CF profiles. The most frequent toxin variants belonged to ETEC linages that have disseminated globally over decades. We also found that main variants differed in ability to produce and secrete the toxins. The STp variant STa5 was linked to disease in adults while the STh variant STa3/4 was associated with disease in children.

The gene expression levels of LT (*eltAB)*, and ST (*estA*) were analyzed by qPCR. We found significantly lower levels of *eltAB* in presence of glucose in LT1 strains. No polymorphisms were found at the CRP binding sites at *eltAB* promoter. ST alleles were also significantly downregulated by glucose while bile supplementation favored STp expression.

Finally, we performed an RNA-transcriptome study, which showed a dramatic change in global gene expression at the onset of stationary phase. During a specific transient phase we observed up- and down-regulation of genes involved in mechanisms related to virulence, such as biofilm formation, indole induction, iron uptake, fucose catabolism, and the putrescine pathway. The expression levels of the toxins and CFs remained high during this phase.

Altogether, this study highlights the diversity within the ETEC population and its virulence factors. We propose that certain combinations of virulence genes influence strain specific responses to host factors that may impact the pathogenesis and severity of ETEC infections.

ISBN 978-91-628-9508-2

LIST OF PAPERS

The thesis is based on the following studies, referred to in the text by their Roman numerals (I-IV).

I. **Joffré E,** von Mentzer A, Abd El Ghany M, Oezguen N, Savidge T, Dougan G, Svennerholm AM, and Sjöling Å. Allele variants of enterotoxigenic *Escherichia coli* heat-labile toxin are globally transmitted and associated with colonization factors. J Bacteriol 2015;197:392-403

II. **Joffré E,** von Mentzer A, Wiklund G, Iniguez V, Svennerholm AM, and Sjöling Å. Identification of new heat-stable (STa) enterotoxin allele variants produced by human enterotoxigenic *Escherichia coli* (ETEC).

Manuscript

III. **Joffré E** and Sjöling Å. The LT1 and LT2 variants of enterotoxigenic *Escherichia coli* (ETEC) heat labile toxin (LT) are associated with major ETEC linages

Submitted for publication

IV. Xiao X, **Joffré E**, Nookaew I, Wang Z, Klena J, Zhu B, and Sjöling Å. RNA-seq transcriptome, transcription factor, and metabolome analysis of enterotoxigenic *Escherichia coli* (ETEC) indicate a transient transcription phase during early stationary phase.

Manuscript

Papers not included in the thesis

- Gonzales L, **Joffré E**, Rivera R, Sjöling Å, Svennerholm AM, and Iniguez V. Prevalence, seasonality and severity of disease caused by the pathogenic *Escherichia coli* in children with diarrhea in Bolivia. J Med Micriobiol 2013; 62:1697-706.

- von Mentzer A, Connor TR, Wieler LH, Semmler T, Iguchi A, Thomson NR, Rasko DA, **Joffré E**, Corander J, Pickard D, Wiklund G, Svennerholm AM, Sjöling Å, Dougan G. Identification of enterotoxigenic *Escherichia coli* (ETEC) clades with long-term global distribution. Nat Gent 2014; 46:1321-6.

TABLE OF CONTENTS

ABBREVIATIONS

AcfD	Accessory colonization factor
CFA/I	Colonization factor antigens
cAMP	Cyclic adenosine monophosphate
CFTR	Cystic fibrosis transmembrane conductance regulator
CS	Coli Surface
Cya	Adenyl cyclase
CRP	Cyclic AMP receptor protein
DsbA	Bacterial disulfide oxidoreductase A
EAST1	Enteroaggregative *Escherichia coli* heat-stable enterotoxin 1
ER	Reticle endoplasmatic
GC-C	Guanylate cyclase C
GM1	Monosialotetrahexosylganglioside
Gsp	General secretion pathway
HGT	Horizontal gene transfer
LPS	Lipopolysaccharides
MUC2	Mucin 2
OMV	Outer membrane vesicles
PAIs	Pathogenicity islands
PDZ	Post synaptic/Drosophila/zonula occludens-1 protein
Sec	Secretory pathway
T2SS	Type 2 secretion system
TF	Transcription factor
TFBSs	Transcription factors binding site

INTRODUCTION

Commensal and pathogenic *Escherichia coli*

In humans, the gastrointestinal tract is home to an extraordinary diversity of bacterial species including *Escherichia coli* (*E. coli*). *E. coli* was first thought to be the predominant commensal bacterial species in human gut, however it was later confirmed that it only comprises 0,1% of the total gut flora, which is dominated by obligate anaerobic bacteria [1]. *E. coli* is a facultative anaerobic gram-negative bacterium that is usually harmless but it can cause disease by acquisition of mobile elements such as pathogenicity islands (PAIs) and plasmids as well as bacteriophages and transposons integrated into either the bacterial chromosome or plasmid [2]. As such, every pathogenic form of *E. coli* (pathovar or pathotype) share and display a set of common virulence factors to cause a common disease. However, single strains of each pathotypes can different sets of virulence markers that define the severity of the disease[2, 3].

Different pathotypes of *E. coli* have found to cause a wide range of human diseases by colonizing the gastrointestinal tract (diarrheagenic *E. coli,* DEC) or disseminated along the urinary tract, bloodstream and central nervous system (extraintestinal *E. coli*, ExPEC). DEC comprises six different pathotypes: entero-pathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), enterohemorrhagic *E. coli* (EHEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC), and diffusely-adhering *E. coli* (DAEC) [2, 3].

Worldwide, DEC is the major cause of diarrheal disease, which remains a leading cause of morbidity and mortality in children under 5 years of age [2]. DEC strains are also associated to numerous outbreaks of diarrheal cases among travelers. Some pathotypes have a major impact on the global health burden of diarrhea disease, especially in developing countries [4].

Enterotoxigenic *Escherichia coli* (ETEC)

Among DEC pathotypes, ETEC alone accounts for millions of diarrheal episodes and it is one of the major agents of moderate to severe infantile diarrhea in developing countries [4].

ETEC is estimated to be the second biggest cause of diarrheal disease in children in developing countries, who experience a median of 3,2 episodes of diarrhea/child-year. ETEC infections are also responsible for fatal cases of diarrhea. Annually, approximately 300,000-500,000 deaths in children in endemic regions are reported. ETEC also commonly cause diarrhea in travelers to endemic regions in Africa, Asia and Latin America [5, 6].

ETEC is also considered as an important emerging cause of food-borne [7] and water-borne disease [8] with negative health and economic consequences in both developed and developing countries [7]. ETEC is transmitted through the fecal-oral route by ingestion of contaminated food and drinking water exposed to human sewage. Ingested ETEC can reach the human intestinal tract, colonize and then cause diarrhea by deregulation of ion channels in the epithelium [2]. The clinical manifestation of an ETEC infection ranges from mild diarrhea without dehydration to severe cholera-like disease [9]. The infection dose is considered to be approximately 10^8 colony-forming units (CFUs) with an incubation period between 15-50 h after ingestion of the bacteria [10]. ETEC infections have been found to be directly related to delayed growth and malnutrition that increase the susceptibility to acquire another ETEC infection during the first years of life [11]. The disease is usually self-limited in 1 to 5 days and in few cases prolong beyond 10 days [11, 12].

Virulence and Pathogenesis of ETEC

Classical and non-classical virulence genes

ETEC is a complex and heterogeneous pathogen with a genome size of approximately 4,8 to 5,2 Mbp including several plasmids [13, 14]. It harbors virulence genes and putative virulence genes involved in different mechanisms of pathogenicity leading to diarrhea. In order to cause diarrhea, ETEC express and produce either one or both of two well characterized plasmid-encoded enterotoxins, the heat labile enterotoxin (LT) and the heat stable enterotoxins (ST), and up to 25 antigenically different, mainly also plasmid-encoded colonization factors (CFs) [15].

Some ETEC virulence-related plasmid and chromosomal-encoded genes have also been involved in the outcome of infection. These non-classical virulence determinates have been found to be located on the chromosome, pathogenicity islands and in plasmids [16, 17]. TibA is an example of chromosomally encoded adhesin that mediates adhesion to human cells and subsequently induces invasion [18]. Another chromosomal gene, *clyA* encodes a

pore-forming hemolytic protein and its expression leads to cytotoxic effects on mammalian cells [19]. The *tia* and *leoA* genes, on the contrary, are harbored on pathogenicity islands. Tia is an outer membrane protein acting as an adhesin [20], while LeoA is involved in toxin secretion by membrane vesicles [21]. The plasmid gene *eatA*, encodes a serine protease that degrades the major protein on the mucosa layer of the small intestine and facilitate translocation of toxins [22]. The EtpA extracellular adhesin is also a plasmid-encoded protein that is located on the tip of the flagella and is required for optimal delivery of LT to epithelial cells [23]. In addition, a large diversity of more than 100 different O antigens is attributed to ETEC strains collected from different countries around the world. In early studies, serotyping was a used to identify and characterize ETEC strains until molecular techniques using virulence markers as targets were applied to improve diagnostic sensitivity [11].

Thus, ETEC´s heterogeneity is made up not only of different combinations of toxins and CFs but also by the expression of other distinct virulence determinants located on the bacterial chromosome or extra chromosomal DNA, and generally transmitted by horizontal gene transfer (HGT). It is evident that any individual ETEC strain only harbor a subset of the virulence genes and putative virulence markers mentioned above which might affect virulence. The hallmark of ETEC is however the expression of one or both of the two major toxins LT and ST.

Enterotoxins: heat labile (LT) toxin and heat stable (ST) toxin

ETEC is though to have evolved when nonpathogenic *E. coli* acquired virulence plasmid (s) [14, 24]. The transition from commensal bacteria to a pathogenic form causing millions of deaths of people was driven by acquisition of the two LT and ST enterotoxins [25]. The LT and ST genes were probably acquired by *E. coli* from *Vibrio cholerae* in the remote past (~ 130 millions year ago), long before the origin of the genus *Homo* suggesting that in the beginning the niche of ETEC was the environment and that the toxins initially had other purposes [26]. New ETEC clones have however evolved independently at several occasions [27]. The current pandemic ETEC clones seem to have emerged rather recently 50-150 years ago [24].

The toxins are ligands to receptors expressed in the human small intestine. Once enterotoxins are liberated into the intestinal tract, they specifically interact with the gastrointestinal mucosa by binding to their receptors and interfere with signal transduction pathways leading to imbalance in intracellular homeostasis [28]. The LT toxin comes in different variants, LTI and LTII where LT-I is associated with disease in both human and animals while LT-II has been

associated only with animal disease. Several variants of the heat stable toxin do exist in pathogenic *E. coli* infection in both animals and humans and are termed STa or STI and STb or STII. ETEC strains infecting human produce the STa and LT-1 variants, which are the only toxin variants thought to play an important role in human disease [11] and the focus of this thesis.

Human heat-labile toxin (LT-I)

In 1991, the crystal structure of LT was elucidated and the multimeric AB_5 conformation of the toxin was revealed, typically found in members of the AB_5 toxin family [29]. Other bacterial toxins are also belong to this large family, including Shiga toxin, Pertussis toxin, Anthrax toxin, Ricin, and Cholera toxin (CT) [30]. LT is closely related to CT and shares similar physiology, structure and antigenic properties [9].

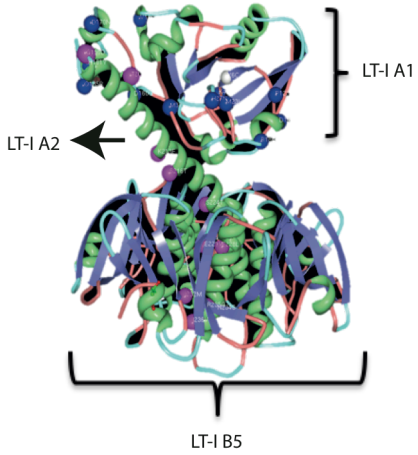

Figure 1. The crystal structure of the human heat labile toxin (LT).

LT is encoded by the *eltAB* operon composed by two genes, *eltA* (LTA) and *eltB* (LTB), which have a 4 nucleotide overlap in the nucleotide sequence. The *eltAB* sequence has an identity of approximately 80% with the *ctxAB* sequence encoding CT and both toxins have totally 240 amino acids [31, 32].

As is shown in Figure 1, the holotoxin structure of LT is organized in two subunits: LTA and LTB. The single LTA subunit has a catalytic domain with ADP-ribosylation activity and binds to a pentamer of non-toxic B subunits (LTB). The LTB pentamer binds to glycosphingolipid receptors on the surface of eukaryotic cells (*e.g.,* GM1 gangloside). The LTA subunit is basically divided into a large A1 domain with the enzyme-active site, which is linked to a small A2 domain responsible for embedding the LTA1 subunit into the center of the LTB pentamer. A single trypsin-sensitive loop and a long α-helix join the two subunits. The molecular masses for subunit A and B is 27,200 and 11,800 Da,

respectively [9, 32, 33]. Several steps are necessary for the toxin production and uptake to be able to trigger the diarrhea in the host [31, 34, 35] (Figure 2 and 3).

Diversity of heat labile toxin

In early studies the molecular heterogeneity of LT was assessed by analysis of the electrophoresis properties and immunological studies. Honda *et al.* [36] and Tsuji *et al.* [37] first desribed that the LT-1 toxins found in porcine infections (LTp) and human LT (LTh) were similar but not identical. These observations raised the question of differences between LTs at the sequence level. After sequencing of LT gene derived from ETEC from human and porcine origins it was possible to show that they share 95% identity but have some polymorphic amino acids in their sequences of the A subunit (K4R, K213K and N238D) and the B subunit (S4T, A46E, and E102K) [38, 39]. By applying discriminatory techniques *i.e.* RFLP it was possible to distinguish LTp and LTh through differences in a single *HhaI* restriction site [40]. This technique was used to test several ETEC isolates from different sources and to characterize LT types when DNA sequencing was not accessible and affordable. However, during the last decade the dramatic reduction in cost has made sequencing more accessible for all labs and therefore it has been possible to include more ETEC LT strains and analyze the natural diversity of this toxin. As a result of the sequencing of the LT-I gene from human derived-ETEC strains isolated from a restricted geographic region (Brazil), 16 LT variants were found. This finding provided a new perspective about the heterogeneity of LT [41]. In our recent study (**Paper I**) that examined amino acid polymorphisms from a geographic and temporal diverse set of 192 LT human-ETEC strains, 20 different LT variants were found, including 8 previously described in Lasaro´s study and 12 novel variants [42]. Altogether, these studies provided new insights about the remarkable diversity harbored in human derived-ETEC and its enterotoxins and also indicate a link with the phenotype heterogeneity of the disease.

Heat stable toxin (STa)

ETEC isolates can express two distinct heat-stable toxin families, STa, ST_A, STI, or ST1 and STb, ST_B, STII, or ST2 with significant differences in structure, function, antigenic cross-reactivity, methanol solubility and activity in infant mouse [10]. STa found in human isolates is a small cysteine rich enterotoxin of 18-19 amino acids. Three cysteine-based disulfide bonds link the peptide into a small molecule with a molecular mass of ca. 2 kDa. STa is encoded on a plasmid by a transposon associated *estA* gene. Within STa, two variants associated with human disease have been described, STh and STp, originally found in human

and pigs, respectively. STp and STh are synthetized as 72-amino acid residues including a pre -or signal peptide, a pro region and a mature ST region. ST is active even after 60 min of heating at 95°C. STa also has to go through different steps until it reaches the target and cause diarrhea (Figure 2 and 3)[10, 28, 33, 43-45].

Diarrhea induced by STa is of the secretory type with no signs of inflammation or colon involvement [45]. It also probably induces more severe disease than LT among children in developing countries [46]. A recent study demonstrated that ST is responsible for the rapid onset and shorter duration of ST-induced diarrhea, while if a LT+/ST+ ETEC strain is causing the diarrhea episode, a second phase with longer duration is due to LT-induced diarrhea [47].

Diversity of heat stable toxin

ST comprises a family of small cysteine-rich peptides that cause diarrhea in human and animals. Peptides with a high homology to *E. coli* STa have been found in other bacterial pathogens such as *Yersinia enterolitica*, *Citrobacter freudii*, cholera toxin positive *Vibrio cholerae* O1 and *Klebsiella pneumonia*. Within the DEC group, the enteroaggregative heat-stable toxin EAST1 encoded by the gene *astA* in EAEC strains has a 117-bp-long DNA sequence and belongs to a subfamily of heat stable toxins and it is genetically and immunologically distinct from ETEC STa [48].

Also three endogenous peptides display functions similar to ST: uroguanylin, guanylin and lymphoguanylin which has 16, 15 and 15 amino acids, respectively. The function of these peptides is to maintain normal fluid and electrolyte homeostasis in the kidneys and intestine. This explains how STa can deregulate fluid homeostasis in the human gut since STa and the endogenous molecules share the same receptor, the guanylate cyclase C (GCC) [44].

STb is encoded by *estB* gene and it was mostly associated to porcine strains than can also harbor the STa gene. STb has shown little heterogeneity and one natural variant has been reported. The STb gene is present in combination with STa and teracycle resistance gene, possibly in the same plasmid [49].

In contrast to STb, STa is a family that is more heterogeneous since several natural variants were identified. In early studies, the first variant identified designated *estA1* was the porcine type of ST (STp) with an 18 amino acid length sequence in comparison with the STh amino acid sequence of 19 aa [50]. Follow-up studies demonstrated the presence of three additional STa variants identified as *estA2* (STa2), *estA3* (STa3) and *estA4* (STa4)[50-54].

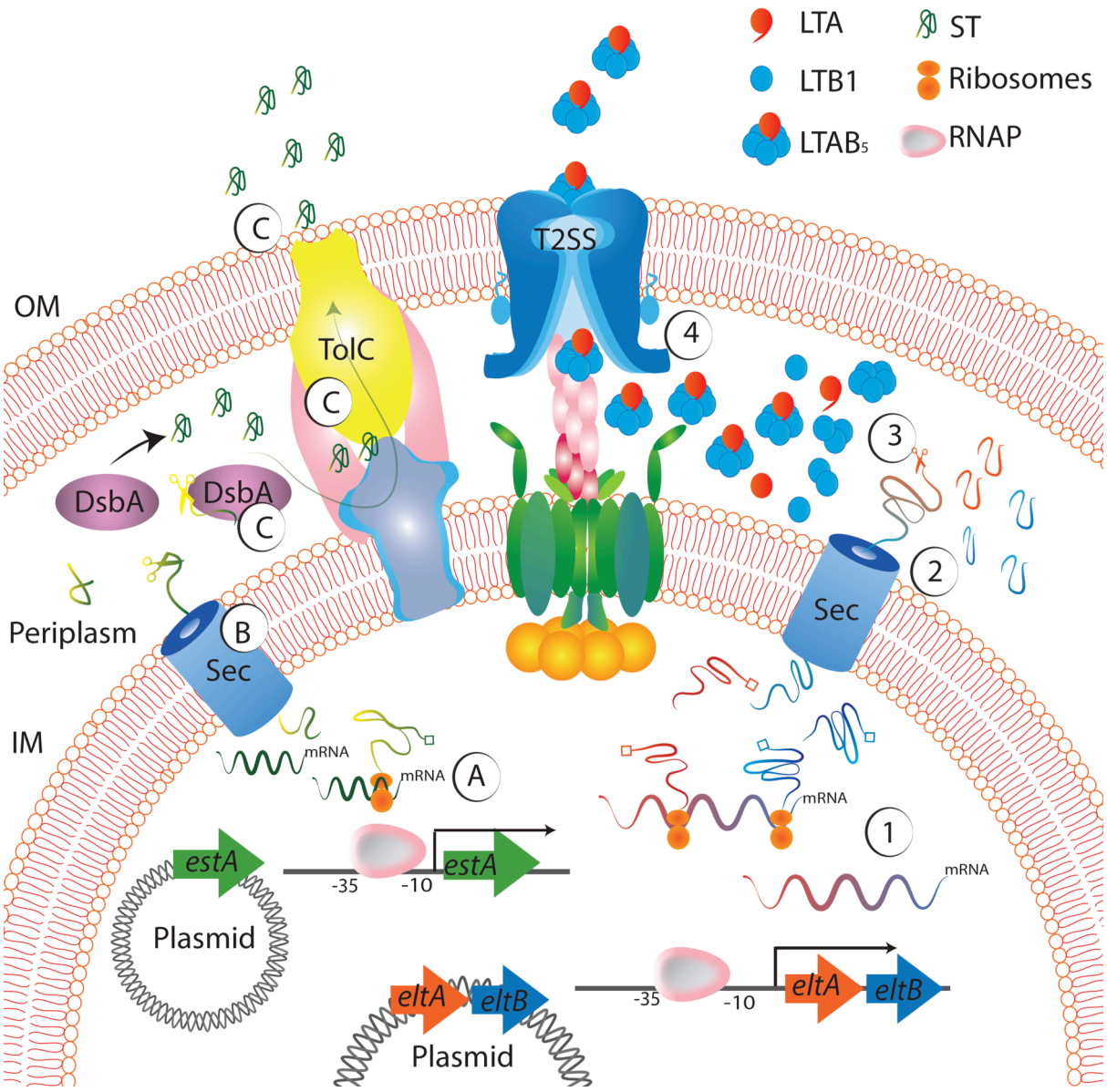

Figure 2. Mechanisms of synthesis and secretion of LT and ST.

LT: 1. Synthesis of LTA and LTB as precursors with signal peptides, which are transported across the cytoplasmic membrane to the periplasm. **2.** Periplasmatic proteolysis of the signal peptides of mature subunits. **3**. By releasing mature subunits into the periplasm, the holotoxin assembly is mediated by the pentamerization of LTB in a circular conformation, which embeds the C-terminal tail of the LTA2 domain. Pentamerization also can occur in either absence of LTA or preformed LTB pentamers. However, in presence of LTA the assembly process is three times faster promoting stability. **4**. Secretion of the assembled subunits by the type II secretion (T2SS) apparatus.

ST: A. Synthesis of STa as an intracellular pre-pro-STa with subsequent cleavage of the 19 amino acid signal sequence by a signal peptidase during or after translocation across the inner membrane. **B**. Translocation to the periplasm and formation of intracellular disulfide bonds by a SecA-dependent export pathway establishing the three-dimensional structure of the peptide. **C.** Inside the periplasm DsbA cleaves the 53-amino acid pro-STa leaving the mature STa to be secreted through the TolC channel.

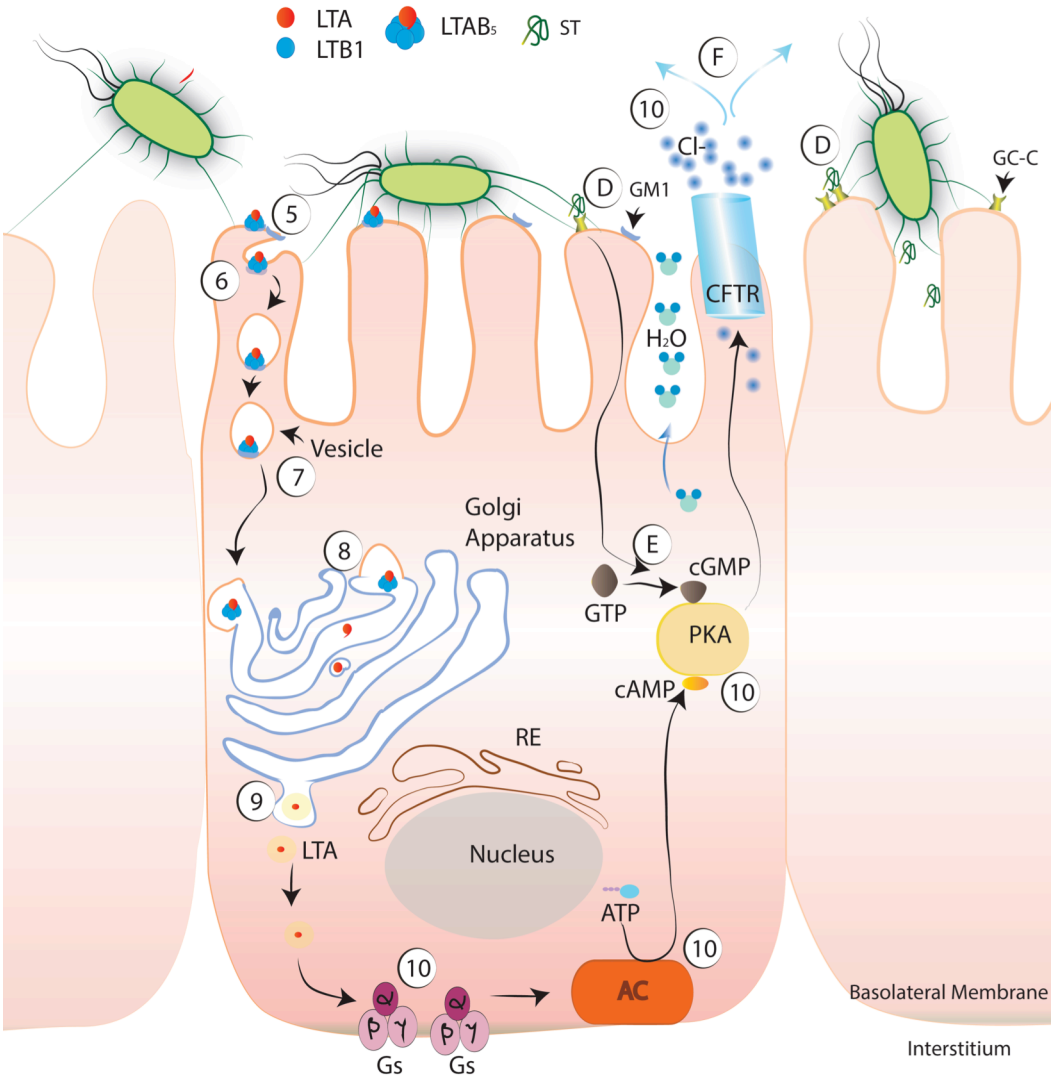

Figure 3. Pathogenesis of ETEC infection.

LT: 5. Recognition and binding of the pentameric complex of LTB monomers to receptors GM1 on the epithelial cell membrane. **6.** Internalization of toxin prior activation. **7.** Association of GM1 receptors with lipid rafts inserted in the cell membrane mediates toxin endocytosis. **8**. Toxin trafficking via retrograde manner to the endoplasmic reticulum (ER) and translocation of LTA1 subunit to the cytoplasm. **9**. Toxin activation is undertaken by two-step process: nicking and reduction. The photolytic processing using trypsin-like enzyme between amino acids 192 and 193 leads to activation and formation of activated LTA1 domain. The LTA2 domain, at the same time, still keeps linked to a B pentamers. **10.** Activation of G protein Gsα by LTA1 stimulates adenylate cyclase to produce cAMP resulting in a dramatic efflux of ions and water from the host leafing to watery diarrhea. AC: Adenylate cyclase. PKA: protein kinase. **ST**: The biological active STa which mimics the hormone guanylin and binds to the extracellular domain of guanylyl cyclase C (GC-C) receptors widely present in the brush border membranes of the intestinal epithelium. **5.** Consequently, the catalytic domain of GC-C is activated which leads to increased levels of intracellular cyclic GMP, stimulating chloride secretion through cystic fibrosis transmembrane receptor (CFTR) and/or preventing NaCl absorption. **6.** The result is net fluid accumulation into the intestinal lumen and secretory diarrhea.

The three new variants that belongs to the STh type were found to be >90% identical to each other at the amino acid sequence level with all of the polymorphic sites located at the pro region of ST. In a more recent study [51], by resequencing of *estA3* and *estA4* genes, the STa4 variant was discarded as a new variants, since it was found to be identical to *estA3*. We have analyzed additional STa positive ETEC strains and found three novel alleles, which are discussed in this thesis (Paper II).

Adherence mediated by expression of colonization factors (CFs)

In order to perform an effective delivery of its enterotoxins, ETEC colonize the small intestine as an initial step of its pathogenesis by means of plasmid-encoded fimbrial colonization factors (CFs) [55, 56]. CFs include a variety of pilus (fimbrial) or pilus-related adhesins and up to 25 different CFs have been described and putative new CFs are repeatedly discovered [57, 58]. In early studies, an ambiguous nomenclature was used to designate the different CFs; years later it was improved and standardized [55] giving a "CS" (Coli Surface antigen) designation, followed by an Arabic numeral, excepting CFA/I. CFs are pili with polymeric structures and conformed by either single (homolymeric) or more than one structure subunit (heteropolymeric) [58]. Based on the morphology, four main types were described including the well-described CFs: fimbrial (pilus) (CFA/I, CS1, CS2, CS4, CS8, CS12, CS14, CS17-21, CS26), fibrillar (CS3, CS11, CS13, CS22) helical (CS5, CS7) and afimbrial (CS6, CS10, CS15, CS23) [11, 58, 59]. Although epidemiologic studies have reported CFA/I, CS1-7, CS14, CS17 and CS21 to be most common in ETEC globally, almost 30-50% of ETEC strains are lacking of any characterized CF [11, 60].

The CF genes are genetically organized in operons, including all genes needed for the assembly of functional CFs. Thus, due to their plasmid localization, it is suggested that ETEC acquired the whole operons by horizontal gene transfer. For instance, CFA/I and CS1 are harbored by pCS1 [61] and pCoo [62] conjugative plasmids, respectively. Remains of insertion sequences flanking the pilus operons indicate mobilization of these genes via transposition [58].

CFs differ in receptor-binding specificity even though the natural intestinal receptor molecules for ETEC CFs are still largely unknown. They are able to hemagglutinate and attach to the intestine through binding to specific receptors, such as glycoproteins and glycosphingolipids [58]. Also, Lewis blood group "a" antigen have been associated to symptomatic infection with ETEC strains expressing a variety of CFs, particularly the Lewis blood group Lt (a^+b^-) was strongly associated with infection by ETEC expressing CFA/I [63].

Type II secretion system (T2SS) and its role in the secretion of LT

Enteropathogens have evolved and acquired specific mechanisms that enable them to colonize and proliferate by producing damage to the host in the process causing disease. ETEC delivers enterotoxins and proteases to the intestinal lumen by secreting them through a complex secretion system widely present in gram-negative species, the type II secretion system (T2SS) [64]. It is considered as a virulence factor because of its role in the secretion of LT [65]. Also other human pathogens have shown to harbor one or more T2SSs including *Vibrio cholerae* [66], EHEC, *Klebsiella* spp and *Legionella pneumophila* [64].

The type II secretion system is a sophisticated multiprotein machinery formed by 12-16 proteins that spans the inner and outer membrane leading to the controlled liberation of specific folded proteins and virulence factors directed to the periplasm through the Sec machinery [65, 67, 68]. The genetic structure of T2SS is arranged in a major operon composed of genes *gspC, -D, -E. –F, -G, h-H, -I, -J, -K, -L, -M, -N* and *–O* (*gspC-O*) and in some cases a minor operon containing *gspA* and *gspB* that codes for surface protein and a large serine-rich glycoprotein, respectively or an independently encoded *gspS* [64, 68, 69].

Genomic sequencing of the ETEC H10407 lab strain allowed identification of the presence of two T2SS operons encoded in the bacterial chromosome homologous to that used by *V. cholerae* to secrete CT [67]. The two distinct gsp-operons were designated alpha (T2SSα) and beta (T2SSβ). While T2SSβ is assembled and functionally active in LT secretion into the culture supernatant under standard laboratory conditions, T2SSα under the exactly same conditions is not assembled, probably due to the repression of the *gspABa* and *gspC-O* promoters by a global regulator H-NS [70]. Most likely T2SSα required specific *in vivo* or environmental conditions to be expressed. In contrast of H10407 and TW10598 ETEC strains that contains both secretion system, ETEC strains such as E24377A and B74 are lacking T2SSα, indicating that T2SSα is not involved in the secretion of LT in all ETEC strains [69]. In addition, sequencing data indicated that T2SSα is not conserved among other enteropathogens whereas T2SSβ is prevalent among them (ETEC, AIEC, EPEC, EAEC, UPEC, APEC and ExPEC) [71].

The structure of the T2SS suggests that this multiprotein complex consists of four parts: an inner membrane platform, a periplasmic pseudopilus, an outer membrane complex and the cytoplasmic secretion ATPase as is represented in Figure 4. The inner membrane platform consists of proteins GspC, F, L and M, forming the core of the system that interact with the pseudopilus (major pseudopilus GspG and minor pseudopilins GspH, I, J and K) the secretin (GspD)

and the cytoplasmic ATPase (GspE). The inner-membrane complex might have a key role in converting conformational changes in the ATPase into extension of the pseudopilus, which possibly acts as a piston and that pushes exoproteins through the outer-membrane channel [64, 72].

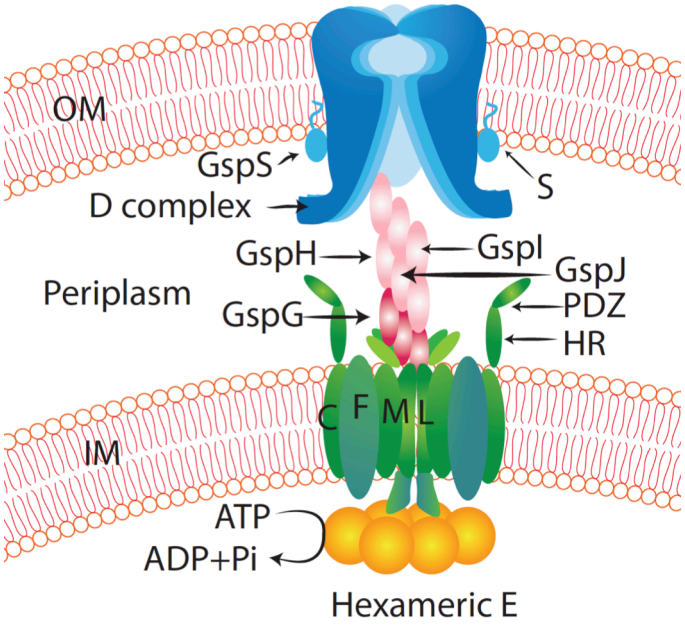

Figure 4. The type 2 secretion system (T2SS) in ETEC. The system is formed by the ATPase in orange, the inner-membrane proteins with a transmembrane helix colored in purple, the outer-membrane proteins are colored in blue and the pili in pink. OM, outer membrane; IM, inner membrane

GspC may regulate T2SS substrate specify though its homology region (HR) domain and PDZ (post synaptic/Drosophila/zonula occludens-1 protein) domain. Also, GspC acts as a tether between the outer membrane complex composed of GspD and the inner membrane complex [73]. GspE is a Zn-containing secretion ATPase, which probably forms hexamers at the interface with the inner membrane. Although the mechanism in which GspE powers the T2SS is unknown, it might couple energy derived form ATP hydrolysis to drive assembly/disassembly of the pseudopilus since GspE interacts with GspL which in turn interacts with the major pseudopilin GspG [74]. In absence of the ATPase, T2SS is not functional; therefore GspG has an essential function for T2SS [75].

The periplasmic pseudopilus (GspG, H, I, J and K) is designated as a pseudo-pilus due to the sequence identity with pilins of the Type I. GspK together with GspI and GspJ may form the "arrow head" of the pseudopilus whereby GspK may interact with secretin or substrates of the T2SS [76].

The outer membrane part of the complex contains for 12-14 subunits of the GspD, which is termed the secretin. It belongs to the outer membrane secretin transporters and its function is to act as the outer membrane pore through which proteins are translocated [69].

In later studies on the T2SSβ operon, three atypical genes (*yghJ, pppA,* and *yghG*) upstream of *gspC* were found [70]. The *yghJ* gene encodes a putative lipoprotein, homologous to the accessory colonization factor (AcfD) of *V. cholerae,* and also essential for colonization in mouse. It is not a structural protein of the T2SS since it is secreted to the culture supernatant. YghJ is also identified as a mucinase that cleaves MUC2 and MUC3 present in the small intestine and is also known as SslE [77]. The *pppA* gene encodes prepilin peptidases required for processing of pseudopilin subunits GspG-K. Deletion of this gene does not affect assembly of the secretin multimer and does not prevent secretion of LT [69, 70]. The third gene *yghG* located directly upstream of *gspC* encodes an outer membrane lipoprotein. Deletion of *yghG* prevents assembly of GspD secretin, which results in a nonfunctional T2SS and inability to secrete LT [69, 70].

Vesicles secrete LT

Although both LT and CT are secreted through the outer membrane via the general secretion pathway, some authors have found that a portion of LT remains associated to the outer membrane of ETEC whereas CT is completely secreted from *V. cholerae*. Comparison between LPS between ETEC and *V. cholerae* revealed that two unphosphorylated D-manno-octulosomic acid (Kdo) bound to lipid A was present in ETEC LPS while in *V. cholerae* LPS contained a single phosphorylated Kdo. Later, it was demonstrated that LT is only able to bind to Kdo molecule that are not phosphorylated [78]. It was described that when LT is bound to outer membrane vesicles (OMV), it acts as an adhesin by mediating the internalization of ETEC vesicles into the intestinal epithelial cells. The molecular delivery of LT begins when the LT-bound vesicles either bind to a receptor in a lipid raft such as caveolin prior internalization and retention and finally LT is trafficked to the Golgi and ER [79].

Similar mechanism of virulence factors delivered by pathogen-derived vesicles also has been described in other pathogens, for instance Shiga toxin was found associated with vesicles from *E. coli* O157:H7 strain [80]. *H.pylori* is another example where the vacuolating toxin VacA is associated to vacuoles and transported through lipid raft [81]. In *Vibrio cholerae,* a pore forming toxin called *V. cholerae* cytolysin (VCC) is also translocated to the eukaryotic cell by OMVs [82]. Based on these observation and additional studies the mechanisms of secretion of LT by OMVs produced by ETEC might be an alternative route where LT can be delivered into the host and might play a role in virulence of ETEC.

Transcriptional Regulation in *Escherichia coli*

Bacteria are microorganisms with a very complex but at the same time efficient mechanism to respond to external stimuli by modifying their genome expression pattern. A major step of regulation of the gene expression is the transcription initiation [83-85].

The principal component during transcription is RNA polymerase (RNAP), which is a holoenzyme, comprised of a multi-subunit core enzyme with subunit composition α2ββ'ω, and one of the seven known sigma factor σ subunits with promoter recognition activity. The recognition of promoters by RNAP holoenzyme is determined by the type of associated sigma (σ) factor. The order of transcription level is determined by the strength of the promoter and it is significantly affected by the presence of the upstream (-35 to -65) region of the promoter, which encompass the UP element, a binding site for the C-terminal domain of the α-subunit of RNA [83, 86, 87]. However, the promoter strength undergoes modification by the second set of regulatory protein, so-called transcriptional factors. They modulate at transcriptional level from the promoter by a direct interaction with the target DNA, located close to the promoter. The transcriptional apparatus is formed once the DNA-binding transcription factors interact with DNA-bound RNA polymerase subunits. This interaction also can involve changes in the DNA curvature. The distribution, concentration and activity of each transcriptional factor are influenced by external signals and internal metabolic states [83, 87].

Transcription factors

The classification of the TFs is based on at least two domains, which allow them to act as regulatory switches and divided in several families. The two-domain structure contains a signal sensor domain and a responsive domain. The signal sensor is characterized by a ligand-binding or protein-protein interaction. More often the ligand is a metabolite or a physical or chemical signal that channels the information, which is either endogenous or environmental. The responsive domain directly interacts with the target DNA sequence or transcription factors-binding sites (TFBSs). In some cases TFBSs exist as direct repeats or palindromes and are located at various positions, from far upstream to inside or downstream of the promoter, depending the canonical -35 and -10 promoter sequences [88, 89]. *In E. coli* the domain more representatively found is the helix-turn-helix domain [90].

The role of the TF can be summarized by repression or activation of the transcription. The repression mechanism is characterized by binding to the promoter and consequently interfering with RNA polymerase. Three

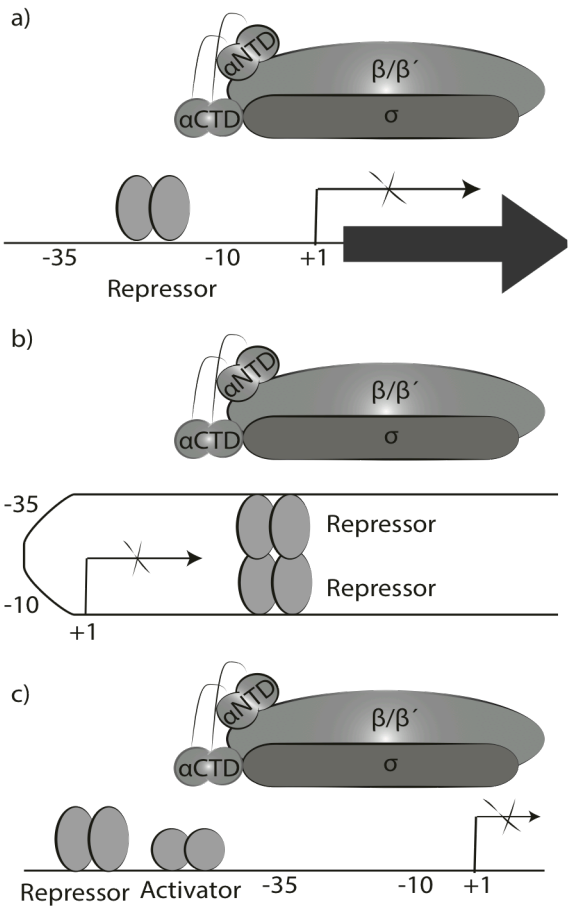

Figure 4. Different mechanisms of repression in prokaryotes

a) *Repression steric hindrance*. The repressor-binding site overlaps core promoter elements and blocks recognition of the promoter by the RNA polymerase holoenzyme.

b) *Repression by looping*. Repressor binds to distal sites and interact by looping, repressing the intervening promoter.

c) Repression by modulation of an activator. The repressor binds to an activator and prevents the activator from acting by blocking promoter recognition by the RNA polymerase holoenzyme.

mechanisms of repression have been described [88, 91] in Figure 4.

In contrast, some others act as positive regulators and bind to the region upstream the promoter, helping in the recruitment of the polymerase to star the transcription. Similar to the repression mechanism, three mechanisms for a simple activation have been proposed [88, 91] in Figure 5.

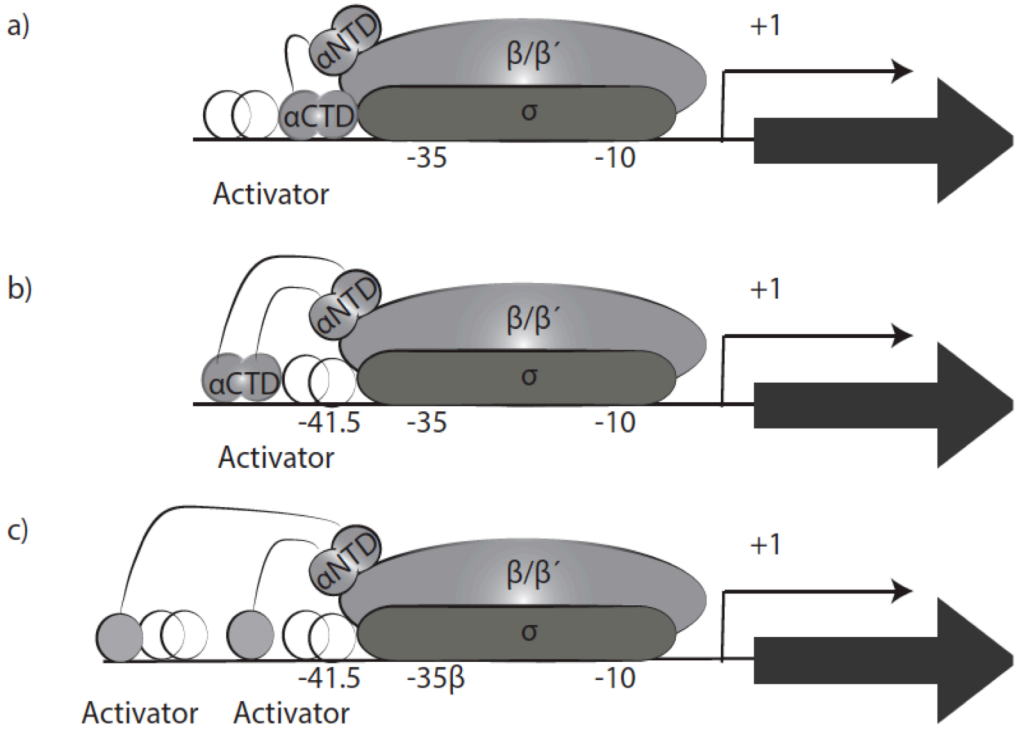

Figure 5. Different mechanisms of activation in prokaryotes

a) *Class I activation*. The activator binds to a target at -35 element to recruit RNA polymerase to the promoter trough direct interaction with the RNA polymerase αCTD.

b) *Class II activation*. The activator binds to the target, which overlaps the promoter -35 element to contact the domain 4 of the RNA polymerase, leading the recruitment of RNA polymerase into the promoter.

c) Activation by conformation change. It is mediated by alteration of the conformation of the target promoter to help the interaction of RNA polymerase with the promoter -10 and/or -35 elements.

Global regulators are characterized for displaying pleiotropic phenotypes and their ability to regulate operons involve in different metabolic pathways [92]. Interestingly, seven regulatory proteins (CRP, FNR, IHF, FIS, ArcA, NarL and Lrp) are able to modulate the gene expression of more than 50% of genes in *E. coli* [91] and an overview of their main features is described as follows.

In the first level of hierarchy the cAMP repressor protein, CRP, acts as a master regulatory protein. It is in charge of sensing the energy available for the metabolism by cAMP levels. CRP is by far the TF that regulate the most TFs, including itself. This global regulator will be described in more detail below. FNR (fumarate and nitrate reductase) and ArcA (aerobic respiration control protein) are responsible of the direct regulation of energy production by modulating respiratory modes. FNR is a oxygen level sensor through an iron-sulphur cluster at the N-terminus of the protein and synchronize the transcriptional response to oxygen limitation [87]. ArcA is a member of the two-complement regulatory system for regulation of expression of genes encoding enzymes involved in mainly anaerobic catabolic pathways [93]. The leucine-responsive protein (Lrp) is involved in the activation of anabolism and repression of other catabolic pathways; helping to the bacterium to adapt to changes in the nutritional environment [94]. Fis (factor for inversion stimulation), IHF (integration host factor) and Hns (histone-like nucleotide structuring protein) are DNA-binding proteins and act as sensors of cellular energy levels by modulating the DNA topology. They are nucleoid-associated proteins (NAPs) and are believed to be the bacterial equivalent of eukaryotic histones [95]. Individually, Fis is found in high concentration during the exponential phase due to activation of rRNA operons to accelerate fast growth. Its role involves response to a range of nutritional environments [96]. Furthermore, IHF is a sequence-specific DNA-binding protein that bends the DNA by over 160° [97]. It is one of the most abundant NAP during the early throughout late stationary phase. Lastly, H-NS regulates a variety of physiological functions such as metabolism, fimbriae expression, virulence flagella synthesis, and proper function [84]

Auto-regulation is a common mechanism of regulation among global regulators. Lrp, FIS, IHF and FNR have a negative auto-regulation, frequently found in TFs with complex connectivity and crucial importance in regulatory network because of homeostatic properties [98].

Sigma factors

Sigma factors are multi-domain subunits of the bacterial RNA polymerase (RNAP) and play an important role in transcription initiation [99]. They enable binding of RNA polymerase to DNA to initiate formation of the open complex and the initiation of the transcription. Bacteria are capable to response to a broad variety of environmental signals by switching on the transcription through a large number of sigma factors. Sigma factors are able to alter the gene expression to be induced or repressed by competition of different σ factors to bind to the core RNA-polymerase or changes in their synthesis [85, 98, 100].

In *E. coli,* seven sigma factors has been identified and classified into two families based on the homologies to two σ factors: the primary factor σ^{70}, which recognize most of the housekeeping gene promoters and is in charge for the bulk of transcription during growth and the structurally unrelated σ^{54} that leads transcription in response to environmental signals and recognize promoters of specific regulons involved in nitrogen regulation [99, 101]. The σ^{70} has ben divided into four groups. Group 1, including σ^{70} itself, composed by sigma factors essential for cell growth while group 2 groups sigma factors (σ^{s} or σ^{38}, also called RpoS), which is closely related to σ^{70} but not essential for bacterial growth. Group 3 and 4 includes sigma factors that control heat shock response (σ^{32}, RpoH), flagellar biosynthesis and sporulation (σ^{28}, RpoF), and extra-cellular stress (RpoE), respectively [84, 85, 100].

Growth phases and transcriptional regulation

When bacterial cells are inoculated into a fresh medium, they go through a growth cycle composed of four phases. As it is shown in Figure 1, when cells enter a new habitat and face different nutritional conditions without an increase of bacterial cell number is called *lag phase*. At this phase the cells experience a reprograming of the metabolic system to allow the adaptation required for bacterial cells to begin to explore new environmental conditions. Even thought this stage has not been studied extensively, it has been described that during this adaptation in the fresh LB medium. bacterial cells upregulate approximately 900 genes, encoding processes such as transcription, translation, iron-sulfur protein assembly, nucleotide metabolism, LPS biosynthesis and aerobic respiration while transcription of genes related to osmotolerance, acid resistance, oxidative stress and adaptation to other stresses was downregulated [102]. Curiously, at the earliest stage of growth, there is a transient sensitivity to oxidative damage due to metal accumulation [103]. Promoters of the genes that are regulated during the lag phase exhibit a strong σ^{70} binding motifs [103, 104]. FIS is the TF remarkably expressed during the lag phase in order to activate promoters of ribosomal genes and it is concentrated in chromosomal zones of actively expressed genes [105]. Also, a dual transcriptional activator SoxS is significantly expressed and involved in the removal of superoxide and nitric oxide for protection of *E. coli* cell against superoxide–generating agents. The SoxS regulon controls 25 operons; all are involved in the production of metabolic energy for restart of cell growth from resting state [106].

The next stage known as *exponential phase* or *log phase* is when the bacterial cells divide asexually by binary fission maintaining a constant rate. At this phase of growth, cell physiology and metabolic activity alters dramatically and this leads to changes in physical and chemical properties of cell components. The

growth rate depends on the richness of the medium. For instance *E. coli* growing at 37°C in a rich medium divide every 20 min [102]. Bacterial cell in the exponential phase express only one-quarter to one-third of the genes on its genome, while the rest of the silent genes are only expressed during adaptation and survival when the bacteria encounter stressful conditions [106].

During the exponential phase, the levels of housekeeping sigma factor σ^{70} reaches its peak, followed by σ^{28} and σ^{54} whereas σ^{38} has been reported to be almost undetected [85]. In early exponential phase, Fis is an abundant transcriptional regulator, which upregulate a large portion of genes involved in translation, flagellar biosynthesis and motility, nutrient transport, carbon compound metabolism, and energy metabolism. There is a growth phase-dependent Fis expression, which gradually shifts its the gene expression towards downregulation as the cells enter stationary phase while an progressive increase of the CRP levels of which are inversely proportional to glucose concentration in the medium is observed [96]. The StpA protein is an analogous nucleoid protein H-NS and varies with growth phase; it is controlling the levels of σ^{38} at mid-exponential phase by preventing its activation during rapid bacterial growth. In contrast, StpA activates the CRP-cAMP regulon during late exponential phase [107]. At the onset of the exponential growth, there is a significant increase in the rate of H_2O_2 production and therefore increased OxyR-dependent transcription to cope with the endogenous oxidative stress [108]. RpoS a major regulator required for adaptation to stationary phase in *E. coli* is also present during the exponential phase and participates in the regulation of genes responsible for carbon source transport, protein folding and iron acquisition [109].

On the other hand, H-NS and HU (Histone-Like) have been found maximally expressed during this stage. H-NS in *E. coli* K-12 binds to the intrinsically curved DNA associates with genes that are thought to have been acquired horizontally [110]. H-NS has a negative influence on components of the growth-arrested regulatory machinery by maintaining GadX at its lowest levels and consequently preventing activation of *rpoS*. In addition, SoxS, MarA and Rob (homologous of AraC family stress response) are highly expressed [105].

As a consequence of the high cell density, the concentration of nutrients depletes and waste is accumulated inducing the bacterial cells to enter to *stationary phase* (carbon starved phase), where bacteria stop dividing. Transcriptional ability of σ^{70} is diminished in a reversible manner; favoring alternative sigma factors *i.e.* σ^{S} (encoded by r*poS* gene) σ^{38} and σ^{32} (encoded by *rpoH*) [102, 106]. This switchover of transcription during starvation is carried out by (p)ppGpp, DksA, Rsda and 6S RNA. The action of these effectors is facilitating the transcription of genes involved in the maintenance of cell functions [85]. Changes in the culture conditions trigger activation of RpoS, the master

regulator of the stationary phase or stress-induced genes and involved in the resistance to various stress condition (*e.g.* oxidative stress, heat shock, osmotic stress, near-UV irradiation or pH changes), metabolic processes and virulence [102].

Aside of stress response, RpoS regulates expression of DNA repair enzymes, genes involved in the cell morphology and genes encoding transport and binding proteins [84]. *rpoS* gene transcription is controlled by the cAMP receptor protein as well as ppGpp signaling. The transcription of *rpoS* increases as growth rate decreases while high cell density, high osmolality, phosphorus starvation, low temperatures and pH induce the synthesis of already present *rpoS* mRNA. Increased expression of several TFs was identified at this stage, such as HdfR (flagellar master regulator), McbR (sensor of quorum sensing) and NadR (transport and *de novo* synthesis of NAD) suggesting an altered metabolic system for energy by entry into the stationary phase [106]. The global transcriptional regulator Lrp plays a key role during the transition to stationary phase by activating proteins involved in the mobilization of internal nutrient supplies and to metabolize fermentation products [111]. Another regulator that contributes the regulation of genes at onset of the stationary phase is IHF (histone-like protein) is growth phase-dependent concentration and regulate genes, such as curli-producing genes. Curli fimbriae are an essential for cell-cell contacts within biofilms. When the IHF levels are increased the silencing effect of H-NS is stopped [112]. FadR regulon is also increased during entry to stationary phase by controlling modulation of long-chain fatty acid pathway in order to provide carbon energy out of endogenous membrane digestion [113]. Changes in the catabolic activity were observed during this phase by regulatory response of ArcB/ArcA/RssB regulon [84]. Also, aerobic metabolism is repressed to prevent waste of energy and also as defense mechanism to avoid formation of reactive oxygen species by the respiratory chain [102]. While bacterial cells redirect metabolic circuits to scavenge nutrients and cope with the stress, other pathways *i.e.* DNA repair controlled by RpoS is downregulated due to a large amount of required energy leading to an increased generation of mutation [114].

Finally, as a consequence of accumulation of damaged molecules in starved cells or under certain unfavorable conditions, cells begin to program their own death in some cases mediated by toxin-antitoxin (TA) molecules. TA biological function is still in debate, but it cause the death of a part of the population, allowing survivors to feed with debris released from the dead cells [102].

Virulence gene regulation in ETEC

As any other pathogen, ETEC is capable to sense different environmental stimuli and modulate the gene expression of its diverse set of virulence genes. Although little is known about the mechanisms behind the transcriptional regulation of the enterotoxins and colonization factors of ETEC, some studies [115-119] have identified the regulatory role of some global regulators such as CRP and H-NS in the modulation of LT and ST expression in response to molecules that may be found in the small intestine *i.e.* glucose and bile salts.

Transcriptional regulation mediated by CRP

Gene expression in bacteria has been very extensively studied showing the existence of global regulators, where a regulatory element controls the expression of many targets involved in complex cellular pathways [98]. The cAMP receptor protein termed CRP or catabolite activator protein (CAP) is a good example of global regulator in bacteria due to its control of a minimum of 378 promoters and perhaps more than 500 genes in *E. coli* [120, 121]. In addition, CRP plays a role as 'master' regulator for 70 'slave' transcription factors. Thus, the CRP manages catabolic pathways, usually in response to environmental conditions and specifically transports the substrates, glycolysis, the Krebs cycle, anaerobic respiration and also virulence [122]. CRP regulation involves promoters from four different σ factors and it is capable to be auto-regulated by itself in a positive or negative fashion [98].

CRP has a sole effector cAMP, which is formed from the catalysis of ATP by a Class I adenyl cyclase (Cya) whose activity is controlled by glucose availability. It is known that when there is availability of glucose in the media, it is transported into the cell by a glucose phosphotransferase system (PTS), which converts glucose into glucose-6-phosphate during transport to the cytoplasm. The cell detects the phosphorylation state of the PTS in order to sense the abundance of available glucose – lower phosphorylation state of the PTS indicates saturation of glucose transporter, while accumulation of phosphorylated PTS proteins occurs when glucose is absent [122]. By phosphorylated PTS interaction with Cya, the adenylyl cyclase activity is enhanced and cAMP concentration increased; however it is also believed that cAMP concentration increase as a consequence of low ATP, promoting catabolism and turning off the anabolism [122].

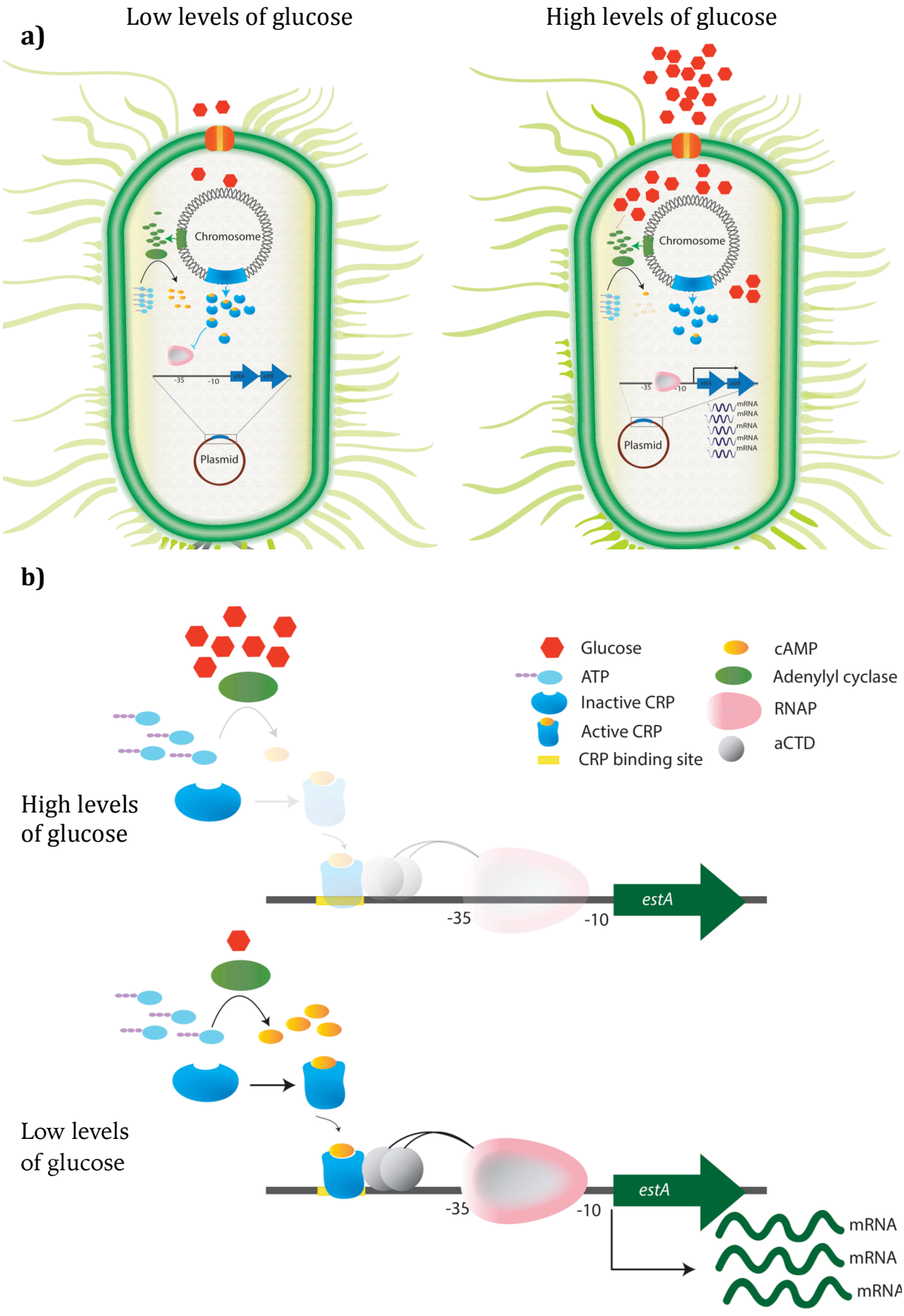

Figure 6. Effect of the glucose in the gene expression of LT (a) and ST (b) genes.

The CRP-cAMP complex is capable to activate transcription by binding to specific DNA sequences in the promoter, often upstream of the core promoter (-10 and -35 elements), interacting with the RNA polymerase. On the contrary, it can repress the expression when the binding site overlaps with or is located downstream of the core promoter. The consensus-binding site for CRP-cAMP is TGTGA-N6-TCACA and any variation in the consensus sequence affects the affinity of CRP-cAMP to bind to different sites. Depending on the location of the binding, CRP regulation can be divided in two groups: Class I and II. Class I activation takes place upstream of the DNA site for RNAP, allowing interaction only with αCTD, which facilitate binding of RNAP to promote to form the RNAP-promoter closed complex [123]. At class II promoters, there is an overlapping site -35 element between the CRP binding site and DNA site for RNAP. At these promoters there are interaction protein-protein interaction between CRP and the RNAPα subunit either C-terminal or N-terminal domain that assist isomerization of the RNAP-promoter closed complex to the RNA-promoter open complex [124]. Both mechanism allows synergistic transcription activation and permits "anti-activation" by negative regulators [121].

CRP is a virulence-required regulator of several bacterial pathogens [125], including in ETEC [115, 116, 118]. This global regulator is involved in modulating the transcription of many genes in ETEC, such as colonization factors antigens [126] and heat-labile and heat-stable enterotoxin genes. CRP has been considered as a pleiotropic regulator of ETEC enterotoxins transcription [118].

Briefly, since glucose is downregulating CRP and adelylate cyclase (the enzyme producing cAMP), several studies have aimed to determine the role of CRP and/or glucose on ETEC virulence. Early studies suggested that *eltAB* expression is inhibited by glucose [127], although addition of this carbohydrate to the medium supported increased LT production. Based on this controversy, Bodero and Munson [115] showed that in the ETEC type strain H10407, CRP repressed the *eltAB* promoter by binding to three DNA target sequences within the promoter region. Later, Kansal and colleagues [117] used a transcriptome *in vivo* study to suggest possible inter-strain transcriptional variation. They observed opposite CRP modulation of *eltAB* virulence genes expression using two different strains using two different strains H10407 and E234377A. Sahl & Rasko [118] and subsequently Haycocks and colleagues [116] with a more descriptive and robust data found an indirect repression of CRP on the *eltAB* and a differential regulation of *estA1* and *estA2*, which might be due to occupancy of H-NS at target promoter sites. All this studies affirm the central role of CRP and the cAMP in the regulation of the enterotoxins but the complete picture of ETEC toxin regulation by CRP is still highly elusive and probably differs

between ETEC strains. The mechanisms of regulation for both toxins are described in Figure 6.

Transcriptional regulation mediated by H-NS

The histone-like nucleoid structuring protein (H-NS) is a transcriptional repressor and an abundant protein in *E. coli* (approximately $2*10^4$ molecules per cell). H-NS belongs to the family of small nucleoid associated proteins together with FIS and IHF [87]. The structure of H-NS is formed by a N-terminal oligomerization domain connected to a carboxyl-terminal nucleic-acid-biding domain via a flexible linker [110]. For the biological activity of H-NS, the oligomeric state of the protein is crucial. The transcriptional repression of H-NS is mediated by preferentially binding to promoters exhibiting AT-rich and highly curved DNA region [128, 129]. Also, H-NS is able to bind to different parts of the same molecule of DNA or even form complexes between different DNA molecules such as DNA-H-NS-DNA [110, 128, 129]. Depending on the motif context H-NS can effect local repression or act more globally altering the packaging of DNA thereby silencing the packaged genes [130].

As the majority of the global regulators, H-NS is auto-repressed by itself and also repressed by the chromosomally encoded H-NS paralogue StpA while Fis activates *hns* transcription. In *E. coli*, H-NS levels are associated to the bacterial cell cycle and therefore maintain a proportional ratio of H-NS protein to chromosomal DNA[110].

As CRP, H-NS has been incorporated into the virulence gene regulatory network and its repressor role has been investigated in bacterial pathogens such as EPEC [131], ETEC [132], EIEC [133], *Shigella flexneri* [134] and *Vibrio cholerae* [135]. It has been suggested that H-NS silences horizontally transferred genes to avoid a competitive disadvantage and unwelcome effect on the physiology of the bacterial host. However, if the new host benefits from the new genetic information, H-NS repression is relieved [110] as consequence of response to environmental cues such as temperature and osmolality by activating the expression of other regulators with overlapping binding sites that of H-NS [136]

In ETEC, Yang and colleagues [137] elucidated the molecular mechanism of H-NS repression of *eltAB* by demonstrating the presence of H-NS-binding regions located downstream of the *eltAB* promoter (+31 and +110, and +460 and +556), which were occupied by H-NS protein at 22°C. The presence of two binding sites indicates DNA loop formation by cooperative interaction between H-NS proteins bound at the two sites. Thus, RNA polymerase is excluded from the nucleoprotein complex formed by H-NS and DNA. Affinity of H-NS has

shown to be increased at 22°C indicating temperature-dependent gene regulation [137].

Another study described that not only LT gene but also STa genes (*estA1* and *estA2*) were subject to repression when H-NS bound to both *estA1* and *estA2* promoter regions. This repression was relieved under increased osmolality. The mechanism suggested by Haycocks *et al.* [116] indicate that *estA1* and *estA2* promoters are target of CRP and H-NS regulation. H-NS represses *estAs* expression probably by occluding the binding sites of CRP, the binding of RNAP or trapping RNA at promoter. Although was demonstrated that binding sites of H-NS are located within the coding sequence of the gene, oligomerization of H-NS in surrounding DNA is necessary to prevent CRP binding.

Bile salts

Extracellular signal such as low pH, elevated temperature and osmolarity can stimulate regulator protein and promote the desire gene expression. During bacterial translocation through the gastrointestinal (GI) tract, they are exposed to a number of different potentially toxic compounds such as bile salts.

Bile is a yellow-green aqueous solution produced by the liver and secreted into the upper duodenum (upper small intestine) from the bile duct. Bile is mainly constituted by bile salts with a concentration in the small intestine that ranges from 0,2 – 2% (wt/vol), depending upon the individual and the type and amount of food ingested. The main purpose of bile secretion is to emulsify and dissolve ingested fat, but a significant bactericidal effect is also achieved due to the detergent-like properties[138, 139].

Several studies described the role of bile in gene expression modulation. In Classical *Vibrio cholerae*, decreased expression of CT and TcaA in about 80% was reported in presence of crude bile extract while motility was favored by about 150% [140]. In transcriptome studies carried out in *S. typhimurium*, bile had a great impact on the gene expression of flagella biosynthesis [141].

In a study using *E. coli* O157:H7, where the transcriptome response to bile was assessed, a significant upregulation of genes associated to the flagella hook-basal body structure was found, in addition of increased levels of mRNA for genes associated with iron scavenging [142]. The ETEC strain E23477A was also subjected to a transcriptional profiling in presence of bile, showing upregulation of many ETEC virulence factors, including *estA* and *eltA* genes while downregulation was observed for the CFs CS1 and CS3 gene expression [118]. On the contrary our group have described that CF CS5 is upregulated in response to bile [143]. The same effect was found for CS7, CS17 and CS19 [144]. Thus, these

studies suggest that enteropathogens have evolved to be capable to sense and modulate gene expression in response of environmental signals such as bile.

Genomic and phylogenetic relationship of ETEC

In early studies using sequence-based PCR analysis of ETEC strains with restricted geographic isolation showed that strains with same toxin-CF profile were closely related, which provided some insights of clonal groups, which share same virulence genes [145]. Years later, another study based on MSLT data from more than thousand human ETEC isolates from different countries provided information about 42 different ETEC linages, which probably came from well-established and wide-spread ETETC linages with evidence of extensive exchange of enterotoxin and colonization factor genes between lineages [146].

With the arrival of the next-generation sequencing technology it became feasible to study hundreds of strains to help to understand the evolutionary process acting in ETEC populations at the whole-genome. Initially, by sequencing and comparing sequenced genomes of single ETEC strains (H10407, E24377A, B7A and clinical isolates) it was possible to identify a conserved genomic pathovar core for ETEC but also confirm the variability on virulence and antigenically dominants genes, indicating that such variability extends beyond the virulence genes [14, 147, 148].

Using the whole-genome sequencing approach we have identify signatures of ETEC linages from a representative collection of ETEC strains with global and long-term distribution [24]. The phylogenetic structure of our ETEC collection consisting of 363 strains placed ETEC throughout the context of the *E. coli* species (phylogroups A, B1, B2, D/E) highlighting the high genetic diversity of this pathovar. However, some linages of ETEC were very discrete including strains with similar virulence and plasmid profiles. Therefore 21 (L1-L22) ETEC linages were identified of which 5 appeared to be the major linages L1-L5, which have emerged in modern time [24]. ETEC strains from major linages expressed the most prevalent virulence profiles (CFA/I, CS1+CS3, CS2+CS3, CS5+CS6 and CS6) according to previous studies. Additional analysis of the major linages demonstrated virulence profile pattern and this finding was also seen in the rest of the linages. In this sense, this study provide a framework of the structure of global ETEC populations based on the acquirement of plasmid encoded virulence factors followed by clonal spreading [24].

In a recent study, the variation of the ETEC population during infection in patients was investigated by whole genome sequencing of multiple distinct ETEC isolates from individual patients. The identification of multiple distinct ETEC isolates with even heterogeneity in virulence profiles during infection

suggest another level of complexity where subpopulation of genomically diverse ETEC co-exist and causes the disease in one individual [149].

AIMS OF THE THESIS

The general aim of this thesis was to study polymorphisms, expression and regulation of LT and ST produced by enterotoxigenic *Escherichia coli* (ETEC).

Specific aims of this thesis

- To identify single nucleotide polymorphisms (SNPs) variants in the genes encoding LT and ST among ETEC strains by DNA sequencing.
- To compare the identified polymorphic sequences with phenotypic production of produced, and secreted toxins, and clinical characteristics and to evaluate whether LT and ST variants belong to different clonal groups or geographic origins.
- To study the impact of host factors such glucose and bile on toxin expression and regulation.
- To expand the knowledge of ETEC transcriptional gene regulation during bacterial growth.

METHODOLOGY

Bacterial strains

The University of Gothenburg has a large bacterial strain collection that comprises approximately 3500 enterotoxigenic *Escherichia coli* (ETEC) strains isolated worldwide, during the period of 1980 – 2014. The ETEC strains were collected from all age groups (children <5 years old, and adults), including diarrheal and asymptomatic patients, as well as from outbreak cases, adult travellers and soldiers visiting endemic areas. A representative selection of 362 ETEC strains (appr. 10% of the collection) were subjected to whole-genome sequencing at Wellcome Trust Sanger Institute for further studies (More details about the ETEC strain collection are found in von Mentzer´s publication [24]). The strain selection criterion was based on a representative proportion of ETEC strains with a diverse virulence profile of toxin (only-LT, LT/ST and only-ST) and colonization factors and relevant geographic and temporal distribution including all groups of age and type of patients.

As is represented in the Figure 7, 2 datasets were generated from the 362 whole-genome sequenced ETEC strains based on the toxin profile. The first dataset is discussed in Paper I and included 186 ETEC strains expressing LT either alone or combined with ST. Additionally, 6 LT-expressing ETEC strains isolated from Bolivia during 2002-2011 were included. The second dataset included in Paper II contained 108 ETEC strains expressing only-ST and LT/ST. The toxin profiles were characterized by phenotype (inhibition ELISA for ST and GM1-ELISA for LT) and genotype assays (multiplex PCR for LT and ST genes) [144]. The CF characterization was performed by dot blot and multiplex PCR. In order to confirm the presence of toxin (*eltAB* and *estA1*), and CF genes, nBLAST gene screening was used on the ETEC genome sequences [24].

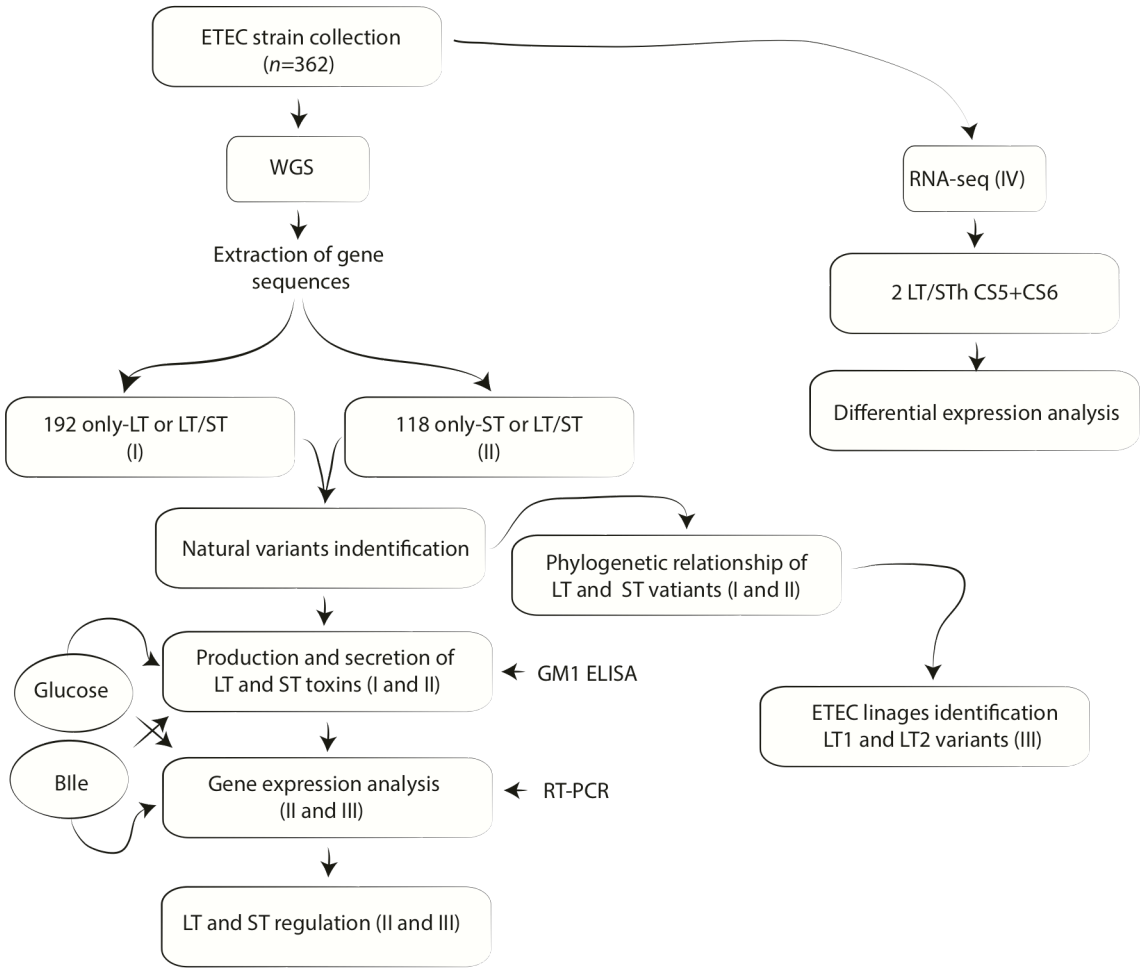

Figure 7. Scheme of the methodology applied in the present thesis

In Paper IV two ETEC strains that were selected fro RNA sequencing were selected from the ETEC strain collection of the University of Gothenburg: E1777 and E2265 (LT STh/CS5+CS6). Both strains were previously subjected to whole-genome sequencing: [13]1777 and E2265 ETEC strains were isolated from adult patient with diarrhea in Dhaka, Bangladesh in 2005 and 2006, respectively. The both belong to the global lineage 5 discussed in von Mentzer et al. [150]

Genomic sequencing (Paper I and II)

All the selected ETEC strains were grown on horse blood plates overnight at 37°C. Later, pure ETEC cultures were used to extract the DNA from each strain following the instructions in the Wizard Genomic DNA kit (Promega). The genomic library preparation and DNA sequencing performed on the Illumina HiSeq 2000 platform have been described by von Mentzer *et al.* [24].

Extraction of gene sequences

To obtain the DNA extraction of the encoding genes *eltAB* (LT-I), *estA1* (STp) and *estA2* (STh) and type 2 Secretion System (T2SS) operon (*gspC-M*, *pppA* and *yghG* genes) from the whole genome sequencing data of each isolate, we used nBLAST using the respective GenBank accessory numbers are mentioned in Paper I, II and III.

LT and ST variants identification and phylogenetic analysis

For the identification of natural polymorphism at amino acid sequence level, we translated the extracted DNA sequence of each gene to the corresponding amino acid sequence using the respective reference sequence. For LT (Paper I) variants analysis, the signal peptides of *eltA* (LTA) and *eltB* (LTB) were extracted and the overlapping sequence between both genes was corrected resulting in a total length of 1035 nucleotides or 344 amino acids. For ST (Paper II) the encoding DNA sequence of 219 nucleotides was translated into a 72 amino acids-length sequences, which formed the pre (1-19 aa) – pro (20-54 aa) – mature (55-72 aa) region. The multialignment analysis was performed by the MEGA 6,06 program separately for each enterotoxin and the sequences were compared to the corresponding reference sequence together with the amino acid sequences of previously reported LT and ST variants (accessory numbers see in Paper I and II). We defined a "natural variant" to when the translated target amino acid sequence differed in at least one amino acid sequence from the reference amino acid sequence.

To establish the phylogenetic relationships of LT and ST variants a Neighbor-Joining (NJ) method was applied to construct the phylogenetic trees. NJ produces a unique final parsimonious tree minimizing the total branches length at each stage of clustering of OTUs (Operational Taxonomic Units) starting with a star-like tree. This method applies the principle of minimum evolution with a reliable estimation of branches lengths, and uses a relatively constant rate of evolution, which is suitable for analyzes of large dataset [151]. A bootstrap analysis (1000 replicates) was performed on the trees obtained to provide a confidence assessment for each clade of an observed tree, based on the proportion of bootstrap trees showing that same clade [152]. Phylogenetic and bootstrap analysis was accomplished by using MEGA 6,06 program.

Growth conditions and culture media (Paper I - IV)

Bacterial strains were growth onto blood horse blood plates overnight at 37°C and then 1 colony was inoculated in 20 ml of either Luria broth (LB) for

LT-expressing ETEC strains or CFA medium for ST-expressing ETEC strains at 37°C, 180 rpm during 3 h (bacterial exponential phase). The bacterial culture density of each strain was measured at OD_{600} and the adjusted volume (considering that 0,8 OD_{600} is approximately 10^9 bacteria) for the next inoculum was inoculated into 15 ml of fresh medium LB (pH=7,5) (Paper I) or CFA (pH=7,4) or CFA supplemented with either glucose (0,2% w/v) or bile (0,15%) (Paper II). Thus, the concentration of the starting cultures were always 10^7 bacteria per ml medium. The cultures were subsequently incubated at 37°C, 180 rpm for 4 h. Finally, the OD values were measured and 1 ml of bacterial cultures was collected to perform the GM1-ELISA and inhibition ELISA.

For gene expression analysis, the bacterial strains were growth as previously described but with some modification. After inoculation of the adjusted volume into fresh media of either LB (only LB or supplemented with glucose) or CFA (only CFA, CFA supplemented with glucose or bile) (Paper II, III)the bacterial culture was incubated until it reached an OD_{600} of 0,3. A sample of the culture media equivalent of 10^8 bacteria/ml per was incubated in RNAprotect (Qiagen, Germany) for 5 min. The samples were centrifuged at 10000 rpm for 10 min and the pellet was frozen at -70°C for RNA isolation.

In Paper IV, a 10 ml LB starting culture from 10 inoculated colonies were diluted hundred-fold in 250 ml Erlenmeyer flask containing 20 ml of LB medium and incubated with aeration at 150 rpm at 37°C. After 3, 4 and 5 hours, 500 ul of bacterial culture was collected and mixed with as RNAprotect as described previously was described.

GM1-ELISA for LT and ST toxin quantification (Paper I and II)

First, in order to assess the produced LT among the LT variant strains we set up a modified and quicker GM1-ELISA using the Svennerholm & Wiklund [153] protocol for testing a large number of ETEC strains (Paper I). In brief, 1 ml of bacterial culture in LB was subjected to bacterial cell disruption using ultrasonication on ice and immediately transferred into the GM1-coated ELISA plates. The assay was performed using triplicates of each sample. The development and the plate reading were performed as described previously [153]. The data analysis was based on the extrapolation of the mean of the OD values from the plates read, thus OD values >0.3, 0.3 to <0.5 and ≥0.5 were considered as low, medium and high levels of LT produced respectively.

Later, for the analysis of the total production and secretion of LT and ST, quantitative GM1-ELISA and GM1-inhibitory ELISA, respectively, were performed according to the previously described protocol [153]. The culture samples were centrifuged for 1 min at 10000 rpm and the supernatant fraction

was transferred to a new tube while the pellet was dissolved in 1 ml of PBS and the bacterial suspension was sonicated for cell disruption. Both fractions were analyzed for LT activity. In case of ST (Paper II), the sonication step was omitted and only supernatant fraction was tested for ST activity, due to that ETEC secretes most of the detected mature ST toxin into the supernatant. The concentration of LT and ST was calculated using the purified r-CTB (0.3 ug/ml) and ST-ref 881108 (0.3nmol/ml), respectively. The total production of LT was calculated by combing the concentration of LT in the pellet and supernatant while secretion rate (%) was obtained by dividing the concentration of LT in the supernatant and total production and multiplied 100.

Another modified GM1-ELISA from Svennerholm & Wiklund [153] for LT detection was set up to determine the amount of free B subunit in relation with the AB holotoxin (Paper I). As is described in [42] two sets of monoclonal antibodies (mAb) were used in two different GM1-coated microtiter plates; in the first plate we used an in-house MAb for the detection of LTB and CTB (LT39:1:1) subunits whereas in the second plates we added the antibody LT17 which detects only the A subunit of the assembled holotoxin. The next steps were performed according to the original protocol [153]. We based the analysis on the assumption that LT17 targeting anti-CTA detects the AB_5 holotoxin, bound through the B_5 subunit-mediated binding to the GM1-coated plates, while anti-CTB detects both holotoxin and dissociated B5 subunits i.e., the total amount of B5 formed. The assays were performed in duplicated by measurement of the amounts of product in the ELISA targeting anti-CTA and anti-CTB, respectively.

RNA extraction and cDNA preparation (Paper II, III and IV)

The RNA isolation was performed from stored bacterial pellets in RNA protect RNAprotect (Qiagen, Germany) at -70°C using the RNeasy Kit (Qiagen, Germany). The resulting RNA was measured at 260/280 nm and by gel electrophoresis in 1% agarose gel. The RNA samples were stored at -70°C until used. The cDNA preparation was carried out using 600 ng of RNA from each sample using the QuantiTect cDNA kit (Qiagen, Germany). In Paper IV, RNA samples were precipitated carefully and RNA pellets were stored under 99,5% ETOH during the shipment to Beijing, China. The quality and integrity of the samples were re-tested upon arrival to the sequencing facility. The RNA Integrity Number (RIN) assessment of RNA samples gave values above 9,8 for all of them. Values above 8 are optimal for transcriptome sequencing.

Gene expression quantification of LT-I and STa by RT-PCR (Paper II and III)

We measured the gene expression of *eltAB* and *estA* genes in LT and ST variant strains respectively under different conditions by quantitative real time PCR (qPCR). All reactions were run in 96-well plates using the standard amplification conditions as is described for ABI 7500 qPCR SYBR green system (AppliedBiosystems Foster city, CA, USA). Specific primers for STp and STh as well as LT-I were used [144]. The reactions were run in duplicates, and the control experiments (only-LB for LT and only-CFA for ST) were used as the reference expression levels in all experiments. The difference between the control experiment and the treatment experiment (LB and CFA+Glucose or CFA+Bile) (ΔCT) was determined and the relative expression was calculated using the formula $2^{\Delta CT}$. The values were normalized against expression in the control experiment. A negative control (NTC, non template control) was included in each run.

RNA-seq and data analysis (Paper IV)

The RNA samples extracted from ETEC strains E1777 and E2265 grown LB collected during exponential (3h), late exponential (4h) and stationary phases (5h) were subject to RNA-seq analysis. The rRNA was removed using a Ribo-Zero kit according to the manufacturer´s specification. The library construction was performed using the TruSeq protocol following the instructions of the manufacturer. Hundred base-pair reads were obtained from each library using standard operating procedures for the Illumina HiSeq 2000 at the Beijing Genome Institute (BGI), Shenzhen, China. SOAPdenovo was used for assembly of the short-reads and the CAP sequence assembly program (CAP3) was used to assembly larger contigs to create a set of non-redundant unigenes. Blast analysis against protein database and functional annotation of unigenes are described in more detail in Paper IV. Differential expression of each unigene was calculated using Reads Per Kilobase of transcript per Million (RPKM), which represents the relative expression of a transcript. The RPKM value of each gene was compared between the samples and only significant changes in RPKM values of more than 2-fold ($Log2>1$, $p<0,05$) and more than 4-fold ($Log2>2$, $p<0,05$) were considered significant. A Poisson distribution in R was used to calculate the significant changes. The hierarchical clustering and heat map visualization were performed using R as is described in paper IV as well as additional metabolic profiling.

Statistical analysis (Paper I-IV)

One–way analysis of variance (ANOVA) and Mann-Whitney U test were applied in Paper I. Nonparametric test, Wilcoxon matched pair test was used in Paper II and III. P-values<0.05 were considered statistically significant. All tests were performed using GraphPad Prism, version 6.

RESULTS AND DISCUSION

In this section of the thesis, the main results of the four scientific articles are highlighted and summarized based on similar topics and additional data not published are incorporated into the context to support the discussion and provide a more complete view of the study.

Genetic diversity of ETEC enterotoxins

To better understand the extent of ETEC enterotoxin gene diversity, the DNA sequences of the corresponding encoding genes of the heat labile (*eltAB*) and heat stable toxin (*estA*) expressed by human-derived ETEC strains isolated from children and adults with worldwide distribution and over three decades were studied. Originally from a whole genome sequencing data set of 362 ETEC strains [24] and 4 additional strains from Bolivia, 192 and 118 gene sequences of LT-1 (Paper I) and STa (Paper II), respectively, were selected and translated into amino acid sequence to be analyzed.

It is noteworthy to mention that these are the first studies about the genetic diversity of the enterotoxins secreted by human derived-ETEC isolates isolated globally and over a longer time period. Firstly, in Paper I (also discussed in Paper III) the genetic variability of the heat labile toxin (LT) gene was studied by identifying the amino acid changes within the DNA sequence translated into the LTA and LTB amino acid sequence. A previous study had demonstrated 16 polymorphic LT in a set of 51 Brazilian LT ETEC strains [41]. In Paper I we found 12 new LT variants (LT17-LT28) from ETEC strains isolated from children and adults with worldwide distribution and over 31 years. In addition to the 12 new variants, 8 out of 16 previously described LT variants (LT1-3, LT7, LT8, LT11-13) already reported were also found in out dataset. In total 20 LT variants were found among the 192 *eltAB* operons using multi-alignment analysis with LT1 (expressed by the type strain H10407) as reference. Two LT variants stood out for being frequently identified in the LT ETEC collection, LT1 (40,6%) and LT2 (25,8%).

The genetic diversity of ST (Paper II) was studied to provide new insights about the natural heterogeneity of this toxin in the same ETEC strain collection. Characterization of the two types of ST, called STp and STh were performed using previously reported nucleotide sequences for each type (STa1 for STp and STa2 for STh). Ninety STh and eighteen STp sequences were identified and six

STa variants were discovered (STh: STa2, ST34 and a novel variant designated STa7; STp: STa1, and two novel variants STa5 and STa6).

The paper by Lasaro [41] and our publication (Paper I) are the only studies about the genetic diversity of human LT gene. However, Zhang and Zhang [154] decided to analyze the diversity of LT porcine gene (LTp) in 52 porcine ETEC strains. LTh and LTp are up to 95% identical in the amino acid sequence and the data reveled little heterogeneity of LTp (1 variant). On the contrary, the related LT toxin, LT-II, which is not associated to human disease, is a much more diverse group and more divergent with 57-59% identity compared to LT-I in the amino acid sequence. LT-II is located in the bacterial chromosome and includes three variants (LT-IIa, LT-IIb and LT-IIc). A study including 50 uncharacterized type II ETEC strains from various sources and isolation dates displayed a potential variability within the LT-IIc family reporting a total of 6 variants (LT-IIc1 – LT-IIc6) [155]. Among other AB_5 toxins, such as the Shiga toxins (Stx1 and Stx2) several variants of Stx1 (Stx1c and Stx1d) and Stx2 (Stx2c, Stx2d, Stx2e, Stx2f, and Sxt2g) have been described [156, 157]. *Vibrio cholerae* is reported to express either classical type CT or biotype El Tor specific CT. Only 2 amino acids of the *ctxB,* which encodes the B subunit of CT, differ between CTBs from CTX phages varieties [158].

Regarding STa variants, only one study analyzed the *estA* gene from 33 STa-porcine ETEC strains and found identical sequences [154]. Another study based on the sequencing of the porcine heat stable toxin gene B gene (*estB*, STb) in 100 porcine ETEC strains isolated over than 20 years identified the presence of only one variant [49]. Hence the STa and STb toxins in porcine ETEC seem to be more stable that STa derived from human.

Findings of Paper I and Paper II highlight the existence of a large genetic diversity of human ETEC enterotoxins, which was previously reported but extended by our studies (Table 1). There is a large difference in number of variants harbored by LT (20 variants) comparing with ST (6 variants). This might be due to differences in terms of size, complexity, structure, function and immunogenicity. However, all natural variants seem to have retained virulence potential. When the distribution of the polymorphisms along the amino acid sequences of both enterotoxins were analyzed, sequences of critical regions such as the ADP-ribosylation active site (between 47-56 aa) in LTA and the mature STa toxin (18-19 aa) were highly conserved and intact, while regions located on the A2 helix domain of LTA and, pre- and pro- segments of ST were most variable.

Table 1. Compilation of the most prevalent virulence gene profiles of ETEC with worldwide distribution that resulted from the variant identification analysis in Paper I and II.

LT variant	ST variant	CF profile	No.
LT1	STa3/4	CS1+CS3 (+CS21)	15
LT1	STa3/4	CS2+CS3 (+CS21)	12
LT1	-	CS7	4
LT1	-	CS17	5
LT1	-	-	10
LT2	STa3/4	CS5+CS6	16
LT2	STa2	CFA/I (+CS21)	5
LT2	-	-	19
LT8	-	-	3
LT11	-	-	7
LT11	-	-	7
LT13	-	CS6	11
LT18	STa6	CS12	3
-	STa2	CFA/I (+CS21)	14
-	STa3/4	CS1+CS3 (+CS21)	11
-	STa3/4	CS5+CS6	8
-	STa3/4	CS14	4
-	STa5	CS6	23

We hypothesized that the amino acid variation in LT might cause instability during the assembly of the holotoxin, but our data using ELISA with MAbs targeting at LTB and the intact holotoxin did not show any difference between the LT2 variant that harbors 4 amino acid changes (S190L, G196D, K213E, S224T) comparing with LT1. Additional protein modeling of the LT1 and LT2 structures mapped onto the crystal structure of 1LTS (LTp) did not provide new insights of the impact of this polymorphism on the function and structure. Nevertheless, the *in silico* modeling reveled interaction sites with potential protein-protein interaction (antibody-antigen) at, or nearby, polymorphic sites. Although these polymorphisms could potentially indicate possible advantages for the toxin to escape immune response others have shown that the immune response to LT1 and LT2 is identical [159].

Phylogenetic relationship of ETEC based on diversity of enterotoxins

After the identification of several variants of the both enterotoxins, especially LT, we wanted to assess how these variants distribute among ETEC isolates. We were interested to analyze clustering with respect to virulence profiles, geographic isolation and phylogenetic relationship.

The phylogenetic analysis of each group of LT (Paper I and Paper III) and ST (Paper II) sequences showed that the variants were strongly associated to the most prevalent CFs. The most prevalent variants of LT, LT1 and LT2 were found in ETEC strains expressing also CS1, CS2, CS3, and CS5+CS6, CFA/I, respectively. These strains also belonged to some of the most successful ETEC linages with global distribution [24](Paper III). Among the ST variants, all STa5 variant strains co-expressed CS6, STa2 expressing strains co-expressed CFA/I, CFA/I+CS21 and CS6, STa3/4 was the most prevalent STh variant and co-expressed 12 different combination of CFs. The complexity of the LT phylogenetic three confirms the level of genetic variability of the LT gene in contrast to the ST gene. The most frequently virulence profiles found among our collection of ETEC strains including the enterotoxin variant are summarized in Table 1 and Paper III. The data indicate that ETEC with specific and successful combinations of toxin variants and CFs are circulating in endemic areas causing a large portion of diarrheal cases. Although, we did not include epidemiological data of the patients with diarrhea caused by LT ETEC, we could find an association between STh and STp alleles (STa3/4 and STa5) with diarrheal disease in two different age groups of patients (children and adults, respectively). The expression of STa5 was associated to disease in adults while STa34 was associated to infant diarrhea (>5 years of age). This finding might indicate that some variants are prompt to cause diarrhea in specific hosts or that the intestinal environment in adults and children favors different toxin variants and/or CFs to be more successful.

Data of the ETEC genomic linages of LT1 and LT2 ETEC strains [24] (discussed in Paper III) have revealed that LT variants, in particular LT1 and LT2, belonged to major and stable ETEC linages. For instance, a big portion of LT1 and LT2 strains falls into L1-L5 major linages. Altogether this data suggest that horizontal gene transfer is the principal vehicle of transmission and dispersion of successful combination of LT variants and CFs genes. The virulence factor combinations are probably located in compatible plasmids, which have advantageous traits in a specific chromosome background. Hence we hypothesized that both plasmids traits and chromosomal background needs to act in concert to generate stable and successful ETEC linages.

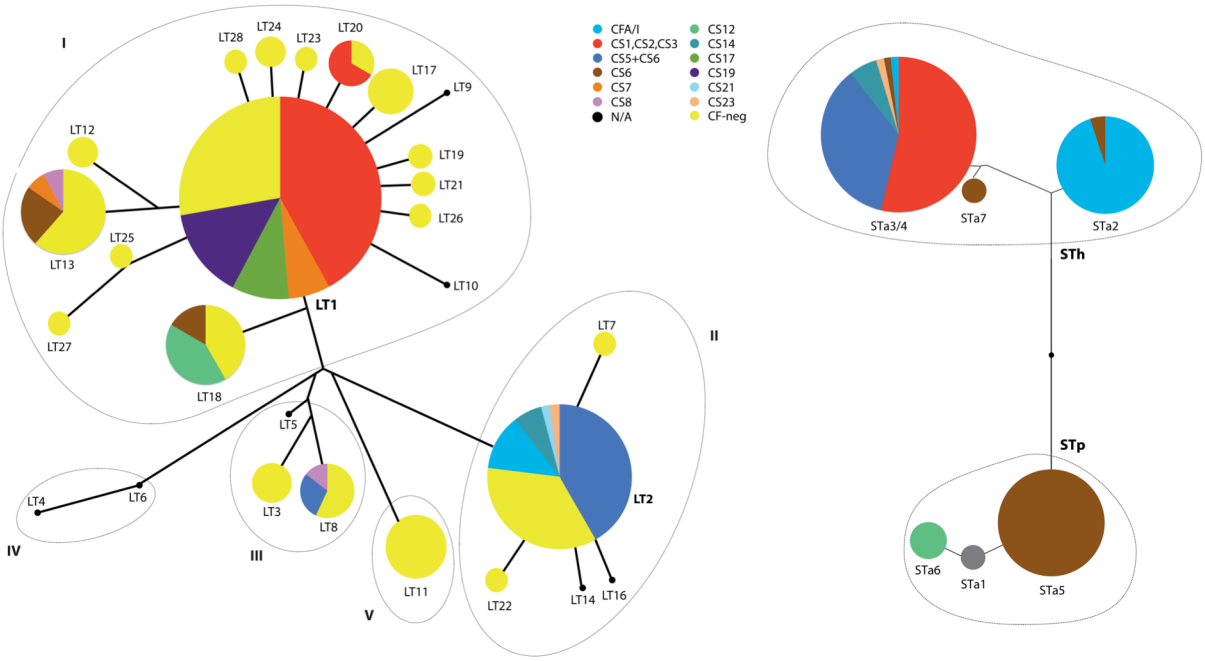

Figure 8. Diversity of CF profiles associated to LT and ST variants. The phylogenetic relationship representation was performed based on the amino acid sequences of the encoding gene of LT **(a)** and ST **(b)** using NJ method. A color pattern was used to identify the CF profiles and the proportion of each CF profile identified in either LT or ST variants is represented in pies. The area of each pie is proportional to the number of strains expressing the respective toxin variant.

Production and secretion of ETEC enterotoxins variants

In light of new findings we were interested in investigating the effects of genetic polymorphism on the phenotype by assessing the production and secretion of both toxin variants. As it is described in Paper I, from a screening of the total amount of produced LT performed in 155 ETEC strains expressing the main LT variants, significant differences in the levels of produced LT were identified, particularly between the LT1 (low producers) and LT2 (high producers) expressing ETEC strains. A further quantitative GM1-ELISA of LT1 and LT2 confirmed our previous finding reporting 5-fold increased production of LT in LT2 variant strains compared with strains expressing LT1 variant.

Based on our previous results, we wanted to investigate the transcriptional levels (mRNA) of the *eltAB* gene encoded by LT1 and LT2 strains (Paper III) and correlate with their respective phenotype (toxin production) (Paper I). Our data revealed slightly higher but not significantly different expression of LT2 gene variant compared to the LT1 gene variant. Also, variations in the amount of mRNA transcripts within LT variant strains were observed. One explanation of

these differences could be due to the levels of sensibility of both techniques and the sampling method. Real time PCR is a more sensitive technique than ELISA since fluorescent primers are used to quantify the DNA or cDNA molecules. An unequal sampling timing during bacterial growth in both techniques also could have an affect on the mRNA and protein yields. GM1-ELISA was performed with samples collected with an optic density of 0.8 (10^8 bacteria/ml) (Paper I, see methodology) while for Real Time PCR; the samples were taken once the optic density reached 0.3 (Paper III).

The secretion levels of LT1 and LT2 were further studied. Initially the data of Paper I indicated comparable percentage of secreted toxin between LT1 and LT2 strains. By looking into the percent of LT secreted per strain, the variation ranged from 0 to 100%. Reports have shown fluctuation in LT secretion with 50-fold differences among human ETEC strains grown in CAYE (supplemented with 0,25% glucose at pH 8,5). CAYE is a medium suggested to yield the highest secretion of LT in the supernatant [160, 161]. We therefore tested growth in LB, CAYE and CFA media on a selected set of LT1 and LT2 strains. Toxin production and secretion after 4 h and over night incubation (data not shown) indicated comparable values between CAYE and LB, while CFA did not promote secretion. Hence our data on LB is comparable to previous publications using CAYE.

We next compared the secretion capacity of LT1 and LT2 strains that belonged to major ETEC lineages 1,2,3 and 5 [24]. We found that ETEC strains in Linage 1 and 2, that express the LT1 variant, were impaired both in production and secretion od the LT toxin compared to the other linages that express LT2 (Paper III).

Based on these findings in Paper III, we extended the analysis of the secretion capacity of LT1 and LT2 strains by including genomic data of the Type 2 secretion system (T2SS) genes in order to correlate such variation in the secretion capacity with possible polymorphism found in genes encoding the T2SS (Figure 9). Once the LT toxin is assembled in the periplasm it can be either retained in the periplasm or secreted through the T2SS [65], or membrane transport machinery via vesicles to the extracellular milieu [162]. Type II secretion pathway is present in several proteobacteria and it is demonstrated to be involved in pathogenesis and to be critical for secretion of many other proteins. In ETEC H10407 genome two different T2SS were identified (so called T2SSα and T2SSβ) and they differ because alfa is not assembled in ETEC under standard laboratory conditions [69]. We could identify the T2SSβ in all LT1 and LT2 strains. A subsequent phylogenetic analysis of the genes forming the *gsp* operon was carried out, and including additional data of the secretion levels of LT per strain, revealed a clustering trend based on similar levels of secretion and

an obvious disaggregation of LT1 with respect to LT2 *gsp* sequences. For instance, the *gsp* operon of LT1 expressing strains branched off from the rest and clustered together in a cluster characterized by low levels of secretion (<34%) while hypersecreting (>50%) LT1 and LT2 strains grouped in different clusters. High values in the bootstrap test supports the sequence cluster generated [163], hence this data might indicate a differential secretion capacity as a consequence of the impact of the amino acid substitution in the Gsp proteins. Protein modeling and amino acid sequence analysis of LT have not reveled polymorphisms at the N-terminal α1 of LTB, associated to the recognition of secretory apparatus. So our discovery supports a previous hypothesis of variation of T2SS might result in different capacities of secretion.

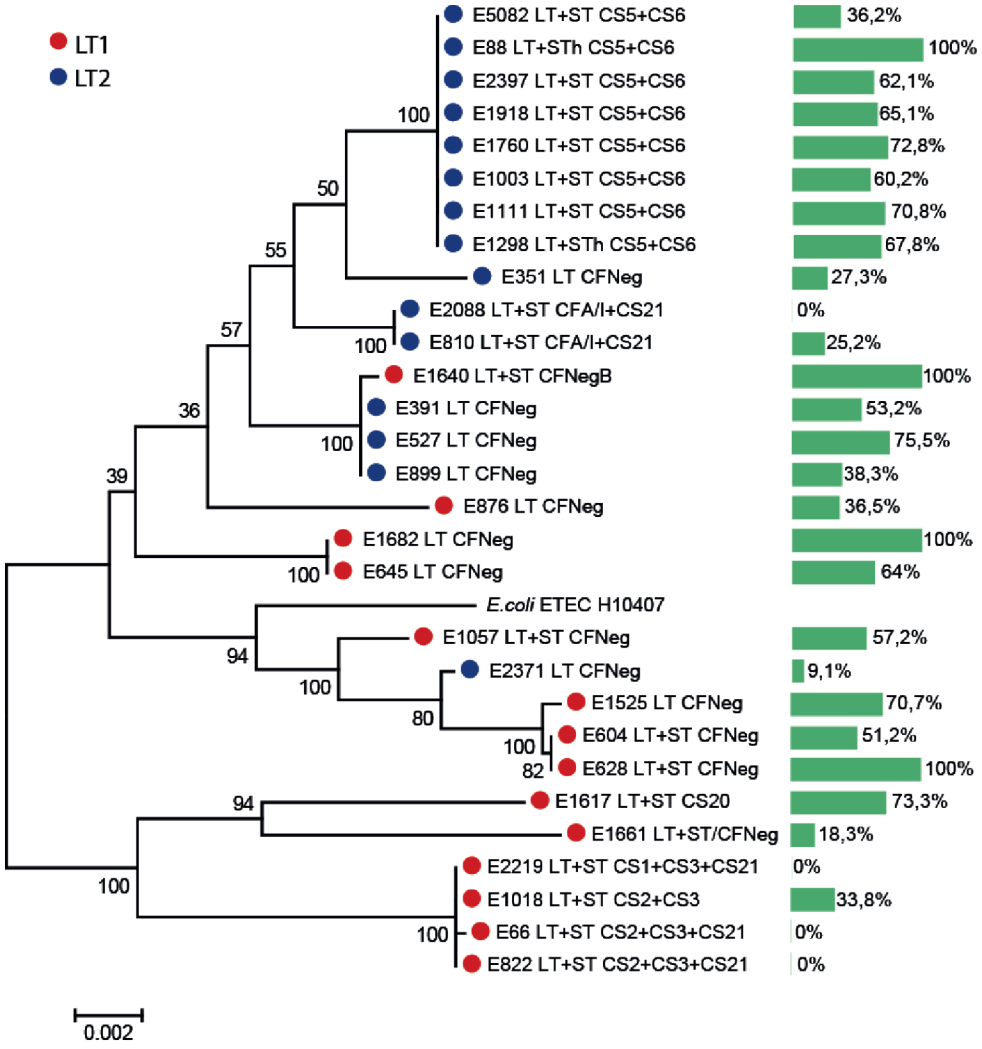

Figure 9. Maximum-likelihood tree of LT1 and LT2 strains based on the concatenated amino acid sequence of the Type 2 secretion system (T2SS) in ETEC, which includes the complete operon *gspC-M*, and *pppA* and *yghG* genes located upstream *gspC*. Each strain designation is followed by the toxin and CF profile. Numbers below-left of the nodes are bootstrap percentages of 1000 replicates. The corresponding levels of secretion (%) of each isolate were represented in green bars. The M-L analysis was performed by MEGA6.06.

Along the Gsp operon, the lowest levels of polymorphisms were found in GspE, which is an ATPase that is thought to be essential for secretion. Interestingly, LT1 strains with impaired secretion had more polymorphisms in GspE. Also, GspC, which forms the multimeric pore for translocation of secreted protein, showed the highest polymorphism suggesting that GspC variants affect secretion levels. Obviously further studies are required for confirmation of our preliminary data, but altogether, this implies that polymorphisms in the T2SS is linked to the capability to secrete the LT toxin.

Similar to the above studies of toxin LT production, we evaluated the production of the ST among the most prevalent ST variants (STa2, STa3/4 and STa5) strains (Paper II). First, comparison of ST production between STh (STa2 and STa3/4) and STp (STa1 and STa5) showed significant higher production of ST in STh strains.

Next, when we examined the ST variants and we found significantly higher ST production in STa2 expressing strains, compared to STa3/4 and STa5 variant strains. Changed amino acid residues of the pre and pro regions might play a role in the translocation of the mature protein across the inner membrane and alter the efficacy of toxin translation [164]. Studies have shown that amino acid substitutions at position 29-31 and 37-38 within the pro region significantly affected the maturation pathway of STp and altered the enterotoxicity [164]. Our data showed that in particular, the region between position 27-35 and 37-39 were conserved between STp and STh sequences; however amino acid changes at position 36 (between sequences of STh and STp), 42, 45 and 49 (only among STh variants) might have minor effect in the efficacy translocation of intermediates ST leading to conformational changes and decreasing the recognition of processing enzymes in the periplasm prior secretion. Also, alteration of amino acid with negative charge near to position 30 was shown to be involved in facilitating translocation of ST through the inner membrane as a consequence of the membrane electrical potential [165]. In our strains the charge near this domain was not changed.

Until now we provided some insides about the link between the natural polymorphism found in LT and ST sequences with the variable capacity of ETEC to produce and secrete its toxin in laboratory conditions, but we believe that ETEC responds to a more complex scenario where fine tuning of regulation of the virulence genes might cause the differences of production rather that the mere amino acid polymorphism.

Transcriptional regulation of virulence in response of host factors

As a result of coevolution between bacteria and their host and an intimate relation between them, bacterial pathogens are capable to employ different strategies to cope host defense mechanisms, to survive and multiply resulting in disease. ETEC has acquired a variable array of virulence factors required for the pathogenesis, which are interconnected to regulation systems of chromosomal and metabolic regulons not directly linked to pathogenesis.

In Paper II and III we addressed questions about the role of gastrointestinal-related environmental signals such as glucose (for LT and ST) and bile (for ST) and their modulation of the toxin expression and production in ETEC strains that belong to different ETEC linages, in order to provide some hints about the influence of the host environment to the virulence of ETEC.

Several studies have described that CRP and H-NS besides their roles of controlling gene expression in a global scale; also are linked to virulence [32, 115, 116, 119]. Based on previous studies [115, 116] using H10407 as prototype to study the regulation of ETEC, the regulatory mechanism is described as follow: H-NS directly represses expression of both genes *eltAB* and *estA* by blocking the binding of RNAP or by sequestering RNAP at promoters. This repression is reverted by increasing the osmolarity and/or the temperature. CRP, on the other hand, acts as dual-regulator, activating transcription of *estA* directly and indirectly repressing the *eltAB* promoter [116]. When glucose is present, repression of CRP is abolished by decreased levels of intracellular cAMP and consequently prevented cAMP-CRP complex formation [122]; thus glucose favors *eltAB* transcription and also prevents *estA* expression by catabolic repression [115, 166].

Contrary to what was mentioned in the literature, our data of LT gene provides a new approach about the role of cAMP-CRP by performing *in vitro* analysis of several ETEC strains in presence of the glucose.

Gene regulation and LT variants

Initially, in Paper III we grew the LT1 (*eltAB1*) and LT2 (*eltAB2*) strains in presence and absence of glucose to compare whether gene regulation between LT1 and LT2 is similar or different as was observed at phenotype level. Although our observation showed lowered expression of LT1 and LT2 by glucose, only downregulation of LT1 gene was significant. Next, we analyzed the 270-nucleotide promoter region of *eltAB* (*peltAB*) gene of LT1 and LT2 expressing strains in order to identify polymorphism associated to CRP binding sites. Non-polymorphic sites were found at CRP binding sites sequences in LT1 and LT2

peltAB, although a SNP located at C6T within the RNAP binding position (-10) was frequent among LT1 strains, which showed lower gene expression. Deletion of CRP binding sites sequences in the *peltAB* promoter has shown no effect on CRP regulation, therefore and in agreement with Haycocks *et al.,* [116] we also suggest that CRP regulation on *peltAB* is mediated by an intermediate unknown factor.

Out data of glucose-induced expression of *eltA* from H10407 are in agreement with previous studies using the same strain. Other studies [116, 118] that used the ETEC model strain E24377A ETEC strain found no effect of glucose on the *eltA* expression. On the contrary, *eltA* from a large portion of our LT1 and LT2 strains were downregulated by adding glucose in the media. Thus, based on our results first we support Sahl and Rasko´s observations regarding the need to not extrapolate the findings using only one strain since ETEC is a quite diverse pathovar. Moreover, our findings postulate the role of glucose as negative signal for transcription of *eltA*, and differential regulation might be explained by a strain-specific regulation depending on the genomic background.

The *estA* gene regulation (ST) displayed a contradictory scenario than was observed in *eltA* regulation. In Paper II we analyzed the effect of bile and glucose on *estA* gene expression, since bile is an abundant compound found in the gastrointestinal tract and has been involved in modulating the virulence expression [142]. By assessing the secretion and expression of STp, STh and variants in presence of glucose and bile, we found that glucose has common pattern among ST variants characterized by decreased secretion and downregulation of the *estA* gene. Bile, on the other hand, displayed a variable effect, favoring the secretion of STp variant (STa5) and not affecting secretion in STh alleles, while downregulation of the gene expression was a general trend of all ST variants.

Transcriptional regulation of ETEC related-virulence mechanism

In Paper IV we extended the scope of ETEC regulation from assessing the role of specific transcription factors (Paper II and Paper III) to a global transcriptome scale to identify underlying mechanisms that might be related to the virulence during different growth phases. When the bacterium is inside of the host, it experiences a myriad of different stress factors *e.g.* nutrient starvation, oxidative stress and decreased growth. Some of these factors are also present during *in vitro* growth cultures when bacteria face growth transition, specifically from exponential to early stationary phase. We performed the LB transcriptome analysis during growth from mid-exponential phase to stationary

phase to study the global transcriptome changes and gain addiction additional insights into ETEC gene regulation.

As was expected, the transcriptome data of two ETEC isolates grown for 3, 4 and 5 hours in LB medium showed a massive variation in gene regulation of essential metabolic pathways and physiologic processes frequently observed when bacteria face reduction of nutrients in the medium. Interestingly the dataset that stand out also reveled a portion of genes with a peculiar pattern of gene regulation characterized by a significant up or down regulation at 4 hours with opposite trend at 3 and 5 hours of growth, described in Paper IV. This peculiar gene expression pattern included mechanisms involved in biofilm formation, indole induction, iron uptake, fucose catabolism, and the putrescine pathway.

Our transcriptome dataset indicated that during the period when ETEC is about to enter stationary phase, mechanism such as biofilm formation and iron uptake are remarkably repressed, while pathways of alternative carbon (fucose) and nitrogen (putrescine) sources as well as signaling through indole induction were highly active. In a biological context, when ETEC copes with a highly stressful environment, this pathogen is capable to drastically regulate the gene expression for a certain short period of time in response to environmental perturbations.

One interesting example is that the ETEC strains we analyzed in Paper IV (from ETEC lineage 5), seems to favor the assimilation of fucose as a carbon source during the transition to stationary phase. Fucose can be found in high concentrations in the intestinal lumen. This feature seems to reflect an adaptation among enteropathogens to obtain alternative carbon sources when the primary carbohydrates (glucose or fructose) are scarce [167, 168]. Fucose was also implicated in virulence by modulating virulence genes [169] and promoting colonization [170]. Similarly to fucose, putrescine is an alternative energy source abundantly found in the small intestine where the intestinal microflora are a major source of putrescine produced by *Enterobacteriaceae* spp, [171] but also it can be found in food contaminated by putrescine-producing strains [172]. Thus, this indicates that ETEC has the ability to take advantage of the commensal metabolism to successfully colonize the small intestine and also to be benefited by the diet.

Another interesting feature in ETEC was an apparent pulse-induction of indole, which plays a role as a signaling molecule directly related with suppression of biofilm formation. The indole activation was mediated through regulation of TnaA (tryptophanase) that converts tryptophan to indole. The ETEC strains also turned off gene expression of several genes involved in biofilm

formation and promoted chemotaxis and motility during the transition to stationary phase. It suggests that biofilm formation is more likely to occur in later stages of bacterial growth, since the transition to stationary phase, ETEC seems to be more actively motile. Moreover, we observed a repression of genes involved in iron assimilation. Bacterial pathogens have evolved to take up and store iron from the host for their own survival. This mechanism was linked to the virulence in many other enteropathogens, but not described in ETEC yet. However, temporary reduction in iron uptake upon entry to stationary phase might indicate that ETEC uses its iron reserves to face this transition.

CONCLUDING REMARKS

In the present thesis we have provided new insights of the diversity of ETEC with focus on the enterotoxins as major virulence factors causing diarrhea in humans. The approaches used in this study included the analysis of different levels of biological complexity of ETEC.

At amino acid sequence level: We demonstrated that the LT peptides are polymorphic, with several natural variants, which are described for first time in this study. The ST peptide is also polymorphic but to a lesser extent than LT.

Phenotype, production and secretion of enterotoxins: Quantification of the amount of produced and secreted LT and ST showed that ETEC strains that express different toxins variants have differential capacity of production and secretion of their toxins and this difference could be associated to their natural polymorphism.

Gene regulation of enterotoxins: By analyzing the role of CRP as a major regulator and the effect of glucose and bile as host factors in the transcription of LT and ST variants genes in several ETEC strains, we could show that CRP regulation on LT and ST transcription might differ between strains and toxin variants. A general trend showed that glucose downregulated both transcription and secretion of LT and ST, but this is not consistent with other reports on LT regulation in the literature. Bile downregulated STh expression but promoted STp secretion. These data indicate that there is a strain specific variation in response to external cues that might modulate virulence differentially.

Genetic relationship of ETEC strains: Analysis of the relatedness of ETEC strains indicated that clonally related ETEC lineages harbor successful combinations of enterotoxins variants (LT1, LT2, STa3/4 and STa5) and colonization factors (CFA/I, CS1+CS3, CS2+CS3 and CS5+CS6), which allow them to persist over time and spread globally.

Global transcriptome analysis of ETEC strains during growth: A framework study of the *in vitro* transcriptome was performed that provided us with new insights on ETEC transcriptional modulation during growth phases. ETEC in early stationary phase regulate distinct pathways that could have a competitive advantage during infection of the host.

The disease: The data generated in this thesis gives a better understanding of the complexity of the underlying factors that influence the severity of the disease. Our results also provide a framework for further studies about the influence of host factors present in the human gut on ETEC virulence.

ACKNOWLEDGEMENTS

Now that I am here about to finish what I call “the most enriching experience of my life”, good stuff like voices, laughs, faces, memories, places and emotions are running through my head. Thanks to all of you who made it happen. This new experience of living abroad not only helped me a lot to learn about myself but also to learn from others. I have been always fascinated by new cultures and lifestyles and Sweden was for me a new niche, which I found very exciting and full of challenges. One of my biggest challenges was to do the PhD, because now I know that “It always seems impossible until it is done”. This dream would not have become true without people who has accompanied, supported, motivated and encouraged me during the last years, whom I would now like to thanks.

Gracias Dios mio porque fuiste y eres el amigo que nunca me falla.

First, thanks to the best supervisor **Åsa Sjöling,** I am sure that you heard that quite often but it is totally true. I am very grateful for having you as my supervisor and for sharing all your vast knowledge, skills and time with me. Second, for being my mentor, although I have never asked you to be, you also played that role very well; all your comprehension and friendly attitude made me feel very confortable since the first time I arrived to Sweden. I will be endlessly grateful for all your trust, support, and freedom and especially for having given me the opportunity to come to Stockholm, no regrets ☺. During the last six months I enjoyed a lot to work with you side by side. You were very supportive and caring during the thesis writing. It is a great feeling to know that I can count on you. Thanks Åsa!. Many thanks for the inspiring chats and for listening my crazy ideas that sometimes I don’t even understand :P. Thanks for creating the best work atmosphere for everybody. Thanks to you today I am a better scientist than yesterday. I am so looking forward to working with you!

Volga Iñiguez, dear Doc, you were that person who inspired the most when I was doing my Licenciatura. You showed me how to work with passion and dedication, which always pay off. I could not have asked for better supervisor during my undergraduate studies and for guided me during my PhD. I found you an endless source of enthusiasm, knowledge and energy. At the same time I am very proud of you for all your work done in Bolivia. Finally, thanks a lot for taking care of me and for being always there. I admire your strength and determination.

I would like to thanks to **Ann-Mari Svennerholm** for being my co-supervisor. Thanks for your advices, scientific discussion and taking care of my education. Also thank you for give me the opportunity to visit Dhaka, Bangladesh.

Gudrun Wiklund, I feel privileged to work with you. You are so caring, kind and nice with me since the first day. Thank you for all your expertise and help not only in

the lab, also in the everyday life. We are very lucky to have you with us in the group. And a super big thanks for helping me with the ELISA :D.

Cristina and **Maria**, Changas! I was so lucky to have two little sisters working in the same department. I lost count of all the fun and exciting moment I spend with you traveling, having fikas, lunches, dinners, hangouts, long experiments (Cris, besides Åsa and I, you knew almost everything about my experiments, thanks for your extra help), gym (Maria so much fun to go to the gym with you, remember how was in the beginning lol) and so on. Las quiero mucho amigas y les debo mis mejores momentos en Göteborg. I miss you!.

My Bolivian friends and colleagues**: Claudia Rodas**, thanks for your true friendship, all your help and the great moments together here in Sweden and in Bolivia bolicheando (clubbing). **Lucia,** my little sister, thanks a lot for your help, your friendship and the good times we spend inside and outside the lab. **Jessica,** my old classmate, friend, colleague, officemate, roommate; you are everywhere;) lol. I am so lucky to have you close to me. It is so much fun living and spending time with you. Thanks for taking care of me like a mom and feed me with your delicious chicken soup. Good luck in your PhD studies. **Olga** thanks to you for your sincere friendship, nice talks, fikas and support ☺ here in Stockholm. **Ceci Curi,** thanks for visiting me not only in Göteborg, also here in Stockholm. It is always a lot of fun to have you around. Good luck in your PhD.

Kaisa, my officemate in Göteborg and now at MTC, thanks so much for your kindness, enthusiasm and friendship, I am glad you are with us ☺ and thanks a lot for your help!. My BFF **Rahil**, I am very glad that we got the chance to share the office because I found a very good friend, who is very caring, funny and crazy ;). Thanks for the funny moments together and but also all you support. **Jenni,** my former officemate, thank you for having fikas, lunches, hangouts and walks together. It is always fun to meet you. **Joanna**, **Madde** and **Annelie** thanks for your friendship, big smile, nice chats and afterworks. **Anna Lundgren** for your kindness.

To all my awesome colleagues at **IBMB** in Bolivia, my home. Thanks for being so nice and caring every time I have been in Bolivia. It so great to work with you guys.

All my school friends**,** thank you a lot for the good times and make my stay in Bolivia very enjoyable. Every time that I was with you guys I felt like time did not pass by.

Stephan, for your friendship and kind help in the lab or with computers also thanks for the ice cream fikas and parties. **Paulina** for invite me to your weeding, your place for cake, Melodifestivalen and fikas ☺. **Mike** for the interesting discussions and help with the plasmid extraction protocol. **Josh** for your help in the lab, your kindness and interesting chats.

Astrid, for your very valuable bioinformatics help, nice discussions, and journal club.

To the **ETEC group**, an especial thanks to all of you for the friendly environment, great discussions, kickoffs, interaction and help. Erik and Matilda for your friendship, your warm welcome and your help in the lab.

To my dear friends in Göteborg: **Elisa, Xandre, Nacho, Jairo, Pamela, Ana** and **Elena** for all the fun we have together at parties, clubs, trips, dinners and basically every single weekend ☺

To our group at MTC, KI **Åsa, Jessica, Kaisa, Patricia, Daeho, Daniela** and **Maike** for the nice environment, support and encourage during this writing process. It is a lot of fun to work with all of you.

Also, a special thanks to my colleagues at **MTC, KI**, who gave me a warm welcome and made me feel like home. It is a great place

A special thank to everybody in the **Department of Microbiology and Immunology** who offered help and support for the experiments and my everyday. I really enjoyed my time there where I met so many great people always ready to help others. Also for the inspiring lectures and scientific discussion. To all of the PhD students at the department, for your friendship, fikas, discussion, pubs and afterworks.

Lars Storm thanks you all so much for your care, help and support during my thesis. I have really good memories with you (the warmest Swedish summer, BBQs, your visit to Bolivia, Copenhagen, and Lyrestad). It was a pleasure to meet your family. Thank you so much for making me feel so welcome.

Gergö, I want to thank you for the great moments we had spend together. I had so much fun going to the "illegal parties", but what I enjoyed the most was your company; you were always there for me, even in the lab helping me to measure the OD. We shared many things in common. I wish you all the best for this new start. As I told you before, all your sacrifices are paying off.

Last but not least, I am truly and deeply blessed to have the support of my family. Adri, my lovely sister, thanks you for being always there for me, caring and asking about me all the time. **Papa and Mama**. I love you so much. I owe all my success to you. Ustedes saben cuanto significan para mi y no me alcanzará la vida para agradecerles por todo lo que hace e hicieron por mi. Welcome to Sweden!!.

***

This work is part of the UMSA-IBMB "Diarrheal Disease Project" supported by the Swedish Agency for Research Economic Cooperation (SIDA). I would like to thanks to ISP for allowances. The studies were supported by grants from the Swedish Foundation for Strategic research (SSF) and the Swedish Research Council.

REFERENCES

1. Tenaillon, O., et al., *The population genetics of commensal Escherichia coli.* Nat Rev Micro, 2010. **8**(3): p. 207-217.
2. Croxen, M.A., et al., *Recent advances in understanding enteric pathogenic Escherichia coli.* Clin Microbiol Rev, 2013. **26**(4): p. 822-80.
3. Kaper, J.B., J.P. Nataro, and H.L. Mobley, *Pathogenic Escherichia coli.* Nat Rev Microbiol, 2004. **2**(2): p. 123-40.
4. Levine, M.M., et al., *The Global Enteric Multicenter Study (GEMS): Impetus, Rationale, and Genesis.* Clinical Infectious Diseases, 2012. **55**(suppl 4): p. S215-S224.
5. Black, R.E., et al., *Global, regional, and national causes of child mortality in 2008: a systematic analysis.* Lancet, 2010. **375**(9730): p. 1969-87.
6. WHO, *Future directions for research on enterotoxigenic Escherichia coli vaccines for developing countries.* Wkly Epidemiol Rec, 2006. **81**(11): p. 97-104.
7. Diseases, W.I.t.E.t.G.B.o.F., *First formal meeting of the Foodborne Disease Burden Epidemiology Reference Group (FERG.* 2008.
8. Rooney, R.M., et al., *A review of outbreaks of waterborne disease associated with ships: evidence for risk management.* Public Health Rep, 2004. **119**(4): p. 435-42.
9. Sanchez, J. and J. Holmgren, *Virulence factors, pathogenesis and vaccine protection in cholera and ETEC diarrhea.* Curr Opin Immunol, 2005. **17**(4): p. 388-98.
10. Sizemore, D.R., K.L. Roland, and U.S. Ryan, *Enterotoxigenic Escherichia coli virulence factors and vaccine approaches.* Expert Review of Vaccines, 2004. **3**(5): p. 585-595.
11. Qadri, F., et al., *Enterotoxigenic Escherichia coli in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention.* Clin Microbiol Rev, 2005. **18**(3): p. 465-83.
12. Mondal, D., et al., *Contribution of Enteric Infection, Altered Intestinal Barrier Function, and Maternal Malnutrition to Infant Malnutrition in Bangladesh.* Clinical Infectious Diseases, 2012. **54**(2): p. 185-192.
13. Liu, F., et al., *Draft genomes of four enterotoxigenic Escherichia coli (ETEC) clinical isolates from China and Bangladesh.* Gut Pathog, 2015. **7**: p. 10.
14. Crossman, L.C., et al., *A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic Escherichia coli strain H10407.* J Bacteriol, 2010. **192**(21): p. 5822-31.
15. Sjoling, A., A. von Mentzer, and A.M. Svennerholm, *Implications of enterotoxigenic Escherichia coli genomics for vaccine development.* Expert Rev Vaccines, 2015. **14**(4): p. 551-60.
16. Fleckenstein, J.M., et al., *Identification of a gene within a pathogenicity island of enterotoxigenic Escherichia coli H10407 required for maximal secretion of the heat-labile enterotoxin.* Infect Immun, 2000. **68**(5): p. 2766-74.
17. Del Canto, F., et al., *Distribution of classical and nonclassical virulence genes in enterotoxigenic Escherichia coli isolates from Chilean children and tRNA gene screening for putative insertion sites for genomic islands.* J Clin Microbiol, 2011. **49**(9): p. 3198-203.
18. Sherlock, O., R.M. Vejborg, and P. Klemm, *The TibA Adhesin/Invasin from Enterotoxigenic Escherichia coli Is Self Recognizing and Induces Bacterial Aggregation and Biofilm Formation.* Infection and Immunity, 2005. **73**(4): p. 1954-1963.

19. Ludwig, A., et al., *Molecular Analysis of Cytolysin A (ClyA) in Pathogenic Escherichia coli Strains.* Journal of Bacteriology, 2004. **186**(16): p. 5311-5320.
20. Mammarappallil, J.G. and E.A. Elsinghorst, *Epithelial cell adherence mediated by the enterotoxigenic Escherichia coli tia protein.* Infect Immun, 2000. **68**(12): p. 6595-601.
21. Michie, K.A., et al., *LeoA, B and C from enterotoxigenic Escherichia coli (ETEC) are bacterial dynamins.* PLoS One, 2014. **9**(9): p. e107211.
22. Kumar, P., et al., *EatA, an immunogenic protective antigen of enterotoxigenic Escherichia coli, degrades intestinal mucin.* Infect Immun, 2014. **82**(2): p. 500-8.
23. Roy, K., D.J. Hamilton, and J.M. Fleckenstein, *Cooperative Role of Antibodies against Heat-Labile Toxin and the EtpA Adhesin in Preventing Toxin Delivery and Intestinal Colonization by Enterotoxigenic Escherichia coli.* Clinical and Vaccine Immunology, 2012. **19**(10): p. 1603-1608.
24. von Mentzer, A., et al., *Identification of enterotoxigenic Escherichia coli (ETEC) clades with long-term global distribution.* Nat Genet, 2014. **46**(12): p. 1321-1326.
25. Pupo, G.M., et al., *Evolutionary relationships among pathogenic and nonpathogenic Escherichia coli strains inferred from multilocus enzyme electrophoresis and mdh sequence studies.* Infect Immun, 1997. **65**(7): p. 2685-92.
26. Yamamoto, T., T. Gojobori, and T. Yokota, *Evolutionary origin of pathogenic determinants in enterotoxigenic Escherichia coli and Vibrio cholerae O1.* J Bacteriol, 1987. **169**(3): p. 1352-7.
27. Turner, S.M., et al., *Phylogenetic comparisons reveal multiple acquisitions of the toxin genes by enterotoxigenic Escherichia coli strains of different evolutionary lineages.* J Clin Microbiol, 2006. **44**(12): p. 4528-36.
28. Popoff, M.R., *Multifaceted interactions of bacterial toxins with the gastrointestinal mucosa.* Future Microbiol, 2011. **6**(7): p. 763-97.
29. Sixma, T.K., et al., *Crystal structure of a cholera toxin-related heat-labile enterotoxin from E. coli.* Nature, 1991. **351**(6325): p. 371-7.
30. Odumosu, O., et al., *AB toxins: a paradigm switch from deadly to desirable.* Toxins (Basel), 2010. **2**(7): p. 1612-45.
31. Fan, E., et al., *Structural biology and structure-based inhibitor design of cholera toxin and heat-labile enterotoxin.* Int J Med Microbiol, 2004. **294**(4): p. 217-23.
32. Mudrak, B. and M.J. Kuehn, *Heat-labile enterotoxin: beyond G(m1) binding.* Toxins (Basel), 2010. **2**(6): p. 1445-70.
33. Sears, C.L. and J.B. Kaper, *Enteric bacterial toxins: mechanisms of action and linkage to intestinal secretion.* Microbiol Rev, 1996. **60**(1): p. 167-215.
34. Streatfield, S.J., et al., *Intermolecular interactions between the A and B subunits of heat-labile enterotoxin from Escherichia coli promote holotoxin assembly and stability in vivo.* Proc Natl Acad Sci U S A, 1992. **89**(24): p. 12140-4.
35. Mekalanos, J.J., R.J. Collier, and W.R. Romig, *Enzymic activity of cholera toxin. I. New method of assay and the mechanism of ADP-ribosyl transfer.* J Biol Chem, 1979. **254**(13): p. 5849-54.
36. Honda, T., et al., *Immunological nonidentity of heat-labile enterotoxins from human and porcine enterotoxigenic Escherichia coli.* Infection and Immunity, 1981. **34**(2): p. 337-340.
37. Tsuji, T., et al., *Molecular heterogeneity of heat-labile enterotoxins from human and porcine enterotoxigenic Escherichia coli.* Infect Immun, 1982. **38**(2): p. 444-8.
38. Yamamoto, T., T. Tamura, and T. Yokota, *Primary structure of heat-labile enterotoxin produced by Escherichia coli pathogenic for humans.* J Biol Chem, 1984. **259**(8): p. 5037-44.

39. Leong, J., A.C. Vinal, and W.S. Dallas, *Nucleotide sequence comparison between heat-labile toxin B-subunit cistrons from Escherichia coli of human and porcine origin.* Infect Immun, 1985. **48**(1): p. 73-7.
40. Vinal, A.C. and W.S. Dallas, *Partition of heat-labile-enterotoxin genes between human and animal Escherichia coli isolates.* Infection and Immunity, 1987. **55**(5): p. 1329-1331.
41. Lasaro, M.A., et al., *Genetic diversity of heat-labile toxin expressed by enterotoxigenic Escherichia coli strains isolated from humans.* J Bacteriol, 2008. **190**(7): p. 2400-10.
42. Joffre, E., et al., *Allele variants of enterotoxigenic Escherichia coli heat-labile toxin are globally transmitted and associated with colonization factors.* J Bacteriol, 2015. **197**(2): p. 392-403.
43. Lima, A.A. and M.C. Fonteles, *From Escherichia coli heat-stable enterotoxin to mammalian endogenous guanylin hormones.* Braz J Med Biol Res, 2014. **47**(3): p. 179-91.
44. Nair, G.B. and Y. Takeda, *The heat-stable enterotoxins.* Microb Pathog, 1998. **24**(2): p. 123-31.
45. Weiglmeier, P.R., P. Rosch, and H. Berkner, *Cure and curse: E. coli heat-stable enterotoxin and its receptor guanylyl cyclase C.* Toxins (Basel), 2010. **2**(9): p. 2213-29.
46. Wolf, M.K., *Occurrence, distribution, and associations of O and H serogroups, colonization factor antigens, and toxins of enterotoxigenic Escherichia coli.* Clin Microbiol Rev, 1997. **10**(4): p. 569-84.
47. Read, L.T., et al., *Simultaneous Exposure to Escherichia coli Heat-Labile and Heat-Stable Enterotoxins Increases Fluid Secretion and Alters Cyclic Nucleotide and Cytokine Production by Intestinal Epithelial Cells.* Infection and Immunity, 2014. **82**(12): p. 5308-5316.
48. Savarino, S.J., et al., *Enteroaggregative Escherichia coli heat-stable enterotoxin 1 represents another subfamily of E. coli heat-stable toxin.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(7): p. 3093-3097.
49. Taillon, C., et al., *Heterogeneity of Escherichia coli STb enterotoxin isolated from diseased pigs.* J Med Microbiol, 2008. **57**(Pt 7): p. 887-90.
50. Takao, T., et al., *Amino acid sequence of a heat-stable enterotoxin isolated from enterotoxigenic Escherichia coli strain 18D.* FEBS Lett, 1983. **152**(1): p. 1-5.
51. Guzman-Verduzco, L.M. and Y.M. Kupersztoch, *Rectification of two Escherichia coli heat-stable enterotoxin allele sequences and lack of biological effect of changing the carboxy-terminal tyrosine to histidine.* Infect Immun, 1989. **57**(2): p. 645-8.
52. Thompson, M.R. and R.A. Giannella, *Revised amino acid sequence for a heat-stable enterotoxin produced by an Escherichia coli strain (18D) that is pathogenic for humans.* Infect Immun, 1985. **47**(3): p. 834-6.
53. Moseley, S.L., et al., *Isolation and nucleotide sequence determination of a gene encoding a heat-stable enterotoxin of Escherichia coli.* Infect Immun, 1983. **39**(3): p. 1167-74.
54. Chan, S.K. and R.A. Giannella, *Amino acid sequence of heat-stable enterotoxin produced by Escherichia coli pathogenic for man.* J Biol Chem, 1981. **256**(15): p. 7744-6.
55. Gaastra, W. and A.M. Svennerholm, *Colonization factors of human enterotoxigenic Escherichia coli (ETEC).* Trends Microbiol, 1996. **4**(11): p. 444-52.
56. Fleckenstein, J., A. Sheikh, and F. Qadri, *Novel antigens for enterotoxigenic Escherichia coli vaccines.* Expert Rev Vaccines, 2014. **13**(5): p. 631-9.

57. Nada, R.A., et al., *Discovery and Phylogenetic Analysis of Novel Members of Class b Enterotoxigenic Escherichia coli Adhesive Fimbriae.* Journal of Clinical Microbiology, 2011. **49**(4): p. 1403-1410.
58. Madhavan, T.P. and H. Sakellaris, *Colonization factors of enterotoxigenic Escherichia coli.* Adv Appl Microbiol, 2015. **90**: p. 155-97.
59. Anantha, R.P., et al., *Evolutionary and Functional Relationships of Colonization Factor Antigen I and Other Class 5 Adhesive Fimbriae of Enterotoxigenic Escherichia coli.* Infection and Immunity, 2004. **72**(12): p. 7190-7201.
60. Isidean, S.D., et al., *A systematic review of ETEC epidemiology focusing on colonization factor and toxin expression.* Vaccine, 2011. **29**(37): p. 6167-78.
61. Forest, K.T., *Enterotoxigenic Escherichia coli CS1 pilus: not one structure but several.* J Bacteriol, 2013. **195**(7): p. 1357-9.
62. Froehlich, B., et al., *The pCoo plasmid of enterotoxigenic Escherichia coli is a mosaic cointegrate.* J Bacteriol, 2005. **187**(18): p. 6509-16.
63. Ahmed, T., et al., *Children with the Le(a+b-) blood group have increased susceptibility to diarrhea caused by enterotoxigenic Escherichia coli expressing colonization factor I group Fimbriae.* Infection and Immunity, 2009. **77**(5): p. 2059-2064.
64. Korotkov, K.V., M. Sandkvist, and W.G. Hol, *The type II secretion system: biogenesis, molecular architecture and mechanism.* Nat Rev Microbiol, 2012. **10**(5): p. 336-51.
65. Mudrak, B. and M.J. Kuehn, *Specificity of the Type II Secretion Systems of Enterotoxigenic Escherichia coli and Vibrio cholerae for Heat-Labile Enterotoxin and Cholera Toxin.* Journal of Bacteriology, 2010. **192**(7): p. 1902-1911.
66. Sandkvist, M., et al., *General secretion pathway (eps) genes required for toxin secretion and outer membrane biogenesis in Vibrio cholerae.* J Bacteriol, 1997. **179**(22): p. 6994-7003.
67. Tauschek, M., et al., *Identification of a protein secretory pathway for the secretion of heat-labile enterotoxin by an enterotoxigenic strain of Escherichia coli.* Proc Natl Acad Sci U S A, 2002. **99**(10): p. 7066-71.
68. Sandkvist, M., *Biology of type II secretion.* Molecular Microbiology, 2001. **40**(2): p. 271-283.
69. Strozen, T.G., G. Li, and S.P. Howard, *YghG (GspS β) Is a Novel Pilot Protein Required for Localization of the GspS β Type II Secretion System Secretin of Enterotoxigenic Escherichia coli.* Infection and Immunity, 2012. **80**(8): p. 2608-2622.
70. Yang, J., et al., *Transcriptional Regulation of the yghJ-pppA-yghG-gspCDEFGHIJKLM Cluster, Encoding the Type II Secretion Pathway in Enterotoxigenic Escherichia coli.* Journal of Bacteriology, 2007. **189**(1): p. 142-150.
71. Genin, S. and C.A. Boucher, *A superfamily of proteins involved in different secretion pathways in gram-negative bacteria: modular structure and specificity of the N-terminal domain.* Mol Gen Genet, 1994. **243**(1): p. 112-8.
72. Douzi, B., A. Filloux, and R. Voulhoux, *On the path to uncover the bacterial type II secretion system.* Philos Trans R Soc Lond B Biol Sci, 2012. **367**(1592): p. 1059-72.
73. Gerard-Vincent, M., et al., *Identification of XcpP domains that confer functionality and specificity to the Pseudomonas aeruginosa type II secretion apparatus.* Mol Microbiol, 2002. **44**(6): p. 1651-65.
74. Gray, M.D., et al., *In vivo cross-linking of EpsG to EpsL suggests a role for EpsL as an ATPase-pseudopilin coupling protein in the Type II secretion system of Vibrio cholerae.* Mol Microbiol, 2011. **79**(3): p. 786-98.
75. Possot, O.M., et al., *Multiple interactions between pullulanase secreton components involved in stabilization and cytoplasmic membrane association of PulE.* J Bacteriol, 2000. **182**(8): p. 2142-52.

76. Forest, K.T., *The type II secretion arrowhead: the structure of GspI-GspJ-GspK.* Nat Struct Mol Biol, 2008. **15**(5): p. 428-30.
77. Luo, Q., et al., *Enterotoxigenic Escherichia coli secretes a highly conserved mucin-degrading metalloprotease to effectively engage intestinal epithelial cells.* Infect Immun, 2014. **82**(2): p. 509-21.
78. Horstman, A.L., S.J. Bauman, and M.J. Kuehn, *Lipopolysaccharide 3-deoxy-D-manno-octulosonic acid (Kdo) core determines bacterial association of secreted toxins.* J Biol Chem, 2004. **279**(9): p. 8070-5.
79. Kesty, N.C., et al., *Enterotoxigenic Escherichia coli vesicles target toxin delivery into mammalian cells*. Vol. 23. 2004. 4538-4549.
80. Kolling, G.L. and K.R. Matthews, *Export of Virulence Genes and Shiga Toxin by Membrane Vesicles of Escherichia coli O157:H7.* Applied and Environmental Microbiology, 1999. **65**(5): p. 1843-1848.
81. Fiocca, R., et al., *Release of Helicobacter pylori vacuolating cytotoxin by both a specific secretion pathway and budding of outer membrane vesicles. Uptake of released toxin and vesicles by gastric epithelium.* J Pathol, 1999. **188**(2): p. 220-6.
82. Elluri, S., et al., *Outer membrane vesicles mediate transport of biologically active Vibrio cholerae cytolysin (VCC) from V. cholerae strains.* PLoS One, 2014. **9**(9): p. e106731.
83. Ishihama, A., *Prokaryotic genome regulation: a revolutionary paradigm.* Proc Jpn Acad Ser B Phys Biol Sci, 2012. **88**(9): p. 485-508.
84. Shimizu, K., *Regulation Systems of Bacteria such as Escherichia coli in Response to Nutrient Limitation and Environmental Stresses.* Metabolites, 2013. **4**(1): p. 1-35.
85. Sharma, U.K. and D. Chatterji, *Transcriptional switching in Escherichia coli during stress and starvation by modulation of sigma activity.* FEMS Microbiol Rev, 2010. **34**(5): p. 646-57.
86. Mutalik, V.K., et al., *Promoter Strength Properties of the Complete Sigma E Regulon of Escherichia coli and Salmonella enterica.* Journal of Bacteriology, 2009. **191**(23): p. 7279-7287.
87. Grainger, D.C. and S.J.W. Busby, *Chapter 4 - Global Regulators of Transcription in Escherichia coli: Mechanisms of Action and Methods for Study*, in *Advances in Applied Microbiology*, S.S. Allen I. Laskin and M.G. Geoffrey, Editors. 2008, Academic Press. p. 93-113.
88. van Hijum, S.A.F.T., M.H. Medema, and O.P. Kuipers, *Mechanisms and Evolution of Control Logic in Prokaryotic Transcriptional Regulation.* Microbiology and Molecular Biology Reviews, 2009. **73**(3): p. 481-509.
89. Balleza, E., et al., *Regulation by transcription factors in bacteria: beyond description.* Fems Microbiology Reviews, 2009. **33**(1): p. 133-151.
90. Perez-Rueda, E., et al., *The functional landscape bound to the transcription factors of Escherichia coli K-12.* Comput Biol Chem, 2015. **58**: p. 93-103.
91. Browning, D.F. and S.J. Busby, *The regulation of bacterial transcription initiation.* Nat Rev Microbiol, 2004. **2**(1): p. 57-65.
92. Gottesman, S., *Bacterial regulation: global regulatory networks.* Annu Rev Genet, 1984. **18**: p. 415-41.
93. Alexeeva, S., K.J. Hellingwerf, and M.J. Teixeira de Mattos, *Requirement of ArcA for Redox Regulation in Escherichia coli under Microaerobic but Not Anaerobic or Aerobic Conditions.* Journal of Bacteriology, 2003. **185**(1): p. 204-209.
94. Landgraf, J.R., J. Wu, and J.M. Calvo, *Effects of nutrition and growth rate on Lrp levels in Escherichia coli.* Journal of Bacteriology, 1996. **178**(23): p. 6930-6936.
95. Hatfield, G.W. and C.J. Benham, *DNA topology-mediated control of global gene expression in Escherichia coli.* Annu Rev Genet, 2002. **36**: p. 175-203.

96. Bradley, M.D., et al., *Effects of Fis on Escherichia coli gene expression during different growth stages.* Microbiology, 2007. **153**(Pt 9): p. 2922-40.
97. Prieto, A.I., et al., *Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in Escherichia coli K12.* Nucleic Acids Res, 2012. **40**(8): p. 3524-37.
98. Martinez-Antonio, A. and J. Collado-Vides, *Identifying global regulators in transcriptional regulatory networks in bacteria.* Curr Opin Microbiol, 2003. **6**(5): p. 482-9.
99. Paget, M.S., *Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution.* Biomolecules, 2015. **5**(3): p. 1245-65.
100. Paget, M.S. and J.D. Helmann, *The sigma70 family of sigma factors.* Genome Biol, 2003. **4**(1): p. 203.
101. Cook, H. and D.W. Ussery, *Sigma factors in a thousand E. coli genomes.* Environ Microbiol, 2013. **15**(12): p. 3121-9.
102. Navarro Llorens, J.M., A. Tormo, and E. Martinez-Garcia, *Stationary phase in gram-negative bacteria.* FEMS Microbiol Rev, 2010. **34**(4): p. 476-95.
103. Rolfe, M.D., et al., *Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation.* J Bacteriol, 2012. **194**(3): p. 686-701.
104. Pin, C., et al., *Network analysis of the transcriptional pattern of young and old cells of Escherichia coli during lag phase.* BMC Syst Biol, 2009. **3**: p. 108.
105. Martinez-Antonio, A., et al., *Regulatory design governing progression of population growth phases in bacteria.* PLoS One, 2012. **7**(2): p. e30654.
106. Yamamoto, K., H. Watanabe, and A. Ishihama, *Expression levels of transcription factors in Escherichia coli: growth phase- and growth condition-dependent variation of 90 regulators from six families.* Microbiology, 2014. **160**(Pt 9): p. 1903-13.
107. Lucchini, S., et al., *The H-NS-like protein StpA represses the RpoS (sigma 38) regulon during exponential growth of Salmonella Typhimurium.* Mol Microbiol, 2009. **74**(5): p. 1169-86.
108. González-Flecha, B. and B. Demple, *Transcriptional regulation of the Escherichia coli oxyR gene as a function of cell growth.* Journal of Bacteriology, 1997. **179**(19): p. 6181-6186.
109. Dong, T., M.G. Kirchhof, and H.E. Schellhorn, *RpoS regulation of gene expression during exponential growth of Escherichia coli K12.* Mol Genet Genomics, 2008. **279**(3): p. 267-77.
110. Dorman, C.J., *H-NS, the genome sentinel.* Nat Rev Micro, 2007. **5**(2): p. 157-161.
111. Tani, T.H., et al., *Adaptation to famine: a family of stationary-phase genes revealed by microarray analysis.* Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13471-6.
112. Ogasawara, H., et al., *Regulation of the Escherichia coli csgD promoter: interplay between five transcription factors.* Microbiology, 2010. **156**(Pt 8): p. 2470-83.
113. Farewell, A., et al., *Role of the Escherichia coli FadR regulator in stasis survival and growth phase-dependent expression of the uspA, fad, and fab genes.* J Bacteriol, 1996. **178**(22): p. 6443-50.
114. Saint-Ruf, C., et al., *Causes and consequences of DNA repair activity modulation during stationary phase in Escherichia coli.* Crit Rev Biochem Mol Biol, 2007. **42**(4): p. 259-70.
115. Bodero, M.D. and G.P. Munson, *Cyclic AMP Receptor Protein-Dependent Repression of Heat-Labile Enterotoxin.* Infection and Immunity, 2009. **77**(2): p. 791-798.
116. Haycocks, J.R.J., et al., *The Molecular Basis for Control of ETEC Enterotoxin Expression in Response to Environment and Host.* PLoS Pathogens, 2015. **11**(1): p. e1004605.

117. Kansal, R., et al., *Transcriptional modulation of enterotoxigenic Escherichia coli virulence genes in response to epithelial cell interactions.* Infect Immun, 2013. **81**(1): p. 259-70.
118. Sahl, J.W. and D.A. Rasko, *Analysis of global transcriptional profiles of enterotoxigenic Escherichia coli isolate E24377A.* Infect Immun, 2012. **80**(3): p. 1232-42.
119. Gonzales, L., et al., *Alkaline pH Is a signal for optimal production and secretion of the heat labile toxin, LT in enterotoxigenic Escherichia coli (ETEC).* PLoS One, 2013. **8**(9): p. e74069.
120. Harman, J.G., *Allosteric regulation of the cAMP receptor protein.* Biochim Biophys Acta, 2001. **1547**(1): p. 1-17.
121. Busby, S. and R.H. Ebright, *Transcription activation by catabolite activator protein (CAP).* J Mol Biol, 1999. **293**(2): p. 199-213.
122. Green, J., et al., *Cyclic-AMP and bacterial cyclic-AMP receptor proteins revisited: adaptation for different ecological niches.* Curr Opin Microbiol, 2014. **18**: p. 1-7.
123. Ebright, R.H., *Transcription activation at Class I CAP-dependent promoters.* Mol Microbiol, 1993. **8**(5): p. 797-802.
124. Busby, S. and R.H. Ebright, *Transcription activation at class II CAP-dependent promoters.* Mol Microbiol, 1997. **23**(5): p. 853-9.
125. Martínez-Antonio, A. and J. Collado-Vides, *Identifying global regulators in transcriptional regulatory networks in bacteria.* Current Opinion in Microbiology, 2003. **6**(5): p. 482-489.
126. Edwards, R.A. and D.M. Schifferli, *Differential regulation of fasA and fasH expression of Escherichia coli 987P fimbriae by environmental cues.* Mol Microbiol, 1997. **25**(4): p. 797-809.
127. Gibert, I., V. Villegas, and J. Barbé, *Expression of heat-labile enterotoxin genes is under cyclic AMP control inEscherichia coli.* Current Microbiology, 1990. **20**(2): p. 83-90.
128. Lang, B., et al., *High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes.* Nucleic Acids Res, 2007. **35**(18): p. 6330-7.
129. Owen-Hughes, T.A., et al., *The chromatin-associated protein H-NS interacts with curved DNA to influence DNA topology and gene expression.* Cell, 1992. **71**(2): p. 255-65.
130. Lang, B., et al., *High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes.* Nucleic Acids Research, 2007. **35**(18): p. 6330-6337.
131. Bustamante, V.H., et al., *Transcriptional regulation of type III secretion genes in enteropathogenic Escherichia coli: Ler antagonizes H-NS-dependent repression.* Mol Microbiol, 2001. **39**(3): p. 664-78.
132. Murphree, D., B. Froehlich, and J.R. Scott, *Transcriptional control of genes encoding CS1 pili: negative regulation by a silencer and positive regulation by Rns.* J Bacteriol, 1997. **179**(18): p. 5736-43.
133. Falconi, M., et al., *Involvement of FIS in the H-NS-mediated regulation of virF gene of Shigella and enteroinvasive Escherichia coli.* Mol Microbiol, 2001. **42**(2): p. 439-52.
134. Buchrieser, C., et al., *The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of Shigella flexneri.* Mol Microbiol, 2000. **38**(4): p. 760-71.
135. Nye, M.B., et al., *Vibrio cholerae H-NS silences virulence gene expression at multiple steps in the ToxR regulatory cascade.* J Bacteriol, 2000. **182**(15): p. 4295-303.
136. Dorman, C.J. and K.A. Kane, *DNA bridging and antibridging: a role for bacterial nucleoid-associated proteins in regulating the expression of laterally acquired genes.* FEMS Microbiol Rev, 2009. **33**(3): p. 587-92.

137. Yang, J., et al., *The H-NS protein represses transcription of the eltAB operon, which encodes heat-labile enterotoxin in enterotoxigenic Escherichia coli, by binding to regions downstream of the promoter.* Microbiology, 2005. **151**(Pt 4): p. 1199-208.
138. Maillette de Buy Wenniger, L. and U. Beuers, *Bile salts and cholestasis.* Dig Liver Dis, 2010. **42**(6): p. 409-18.
139. Alrefai, W.A. and R.K. Gill, *Bile acid transporters: structure, function, regulation and pathophysiological implications.* Pharm Res, 2007. **24**(10): p. 1803-23.
140. Gupta, S. and R. Chowdhury, *Bile affects production of virulence factors and motility of Vibrio cholerae.* Infect Immun, 1997. **65**(3): p. 1131-4.
141. Hamner, S., et al., *Bile Salts Affect Expression of <italic>Escherichia coli</italic> O157:H7 Genes for Virulence and Iron Acquisition, and Promote Growth under Iron Limiting Conditions.* PLoS ONE, 2013. **8**(9): p. e74647.
142. Hamner, S., et al., *Bile salts affect expression of Escherichia coli O157:H7 genes for virulence and iron acquisition, and promote growth under iron limiting conditions.* PLoS One, 2013. **8**(9): p. e74647.
143. Nicklasson, M., et al., *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**(4): p. e35827.
144. Sjoling, A., et al., *Comparative analyses of phenotypic and genotypic methods for detection of enterotoxigenic Escherichia coli toxins and colonization factors.* J Clin Microbiol, 2007. **45**(10): p. 3295-301.
145. Steinsland, H., et al., *Clonal relatedness of enterotoxigenic Escherichia coli strains isolated from a cohort of young children in Guinea-Bissau.* J Clin Microbiol, 2004. **42**(7): p. 3100-7.
146. Steinsland, H., et al., *Ancestral lineages of human enterotoxigenic Escherichia coli.* J Clin Microbiol, 2010. **48**(8): p. 2916-24.
147. Sahl, J.W., et al., *A comparative genomic analysis of diverse clonal types of enterotoxigenic Escherichia coli reveals pathovar-specific conservation.* Infect Immun, 2011. **79**(2): p. 950-60.
148. Rasko, D.A., et al., *The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates.* J Bacteriol, 2008. **190**(20): p. 6881-93.
149. Sahl, J.W., et al., *Examination of the Enterotoxigenic Escherichia coli Population Structure during Human Infection.* MBio, 2015. **6**(3): p. e00501.
150. von Mentzer, A., et al., *Identification of enterotoxigenic Escherichia coli (ETEC) clades with long-term global distribution.* Nat Genet, 2014. **46**(12): p. 1321-6.
151. Saitou, N. and M. Nei, *The neighbor-joining method: a new method for reconstructing phylogenetic trees.* Molecular Biology and Evolution, 1987. **4**(4): p. 406-425.
152. Efron, B., E. Halloran, and S. Holmes, *Bootstrap confidence levels for phylogenetic trees.* Proceedings of the National Academy of Sciences, 1996. **93**(23): p. 13429.
153. Svennerholm, A.M. and G. Wiklund, *Rapid GM1-enzyme-linked immunosorbent assay with visual reading for identification of Escherichia coli heat-labile enterotoxin.* J Clin Microbiol, 1983. **17**(4): p. 596-600.
154. Zhang, C., D. Rausch, and W. Zhang, *Little heterogeneity among genes encoding heat-labile and heat-stable toxins of enterotoxigenic Escherichia coli strains isolated from diarrheal pigs.* Appl Environ Microbiol, 2009. **75**(19): p. 6402-5.
155. Jobling, M.G. and R.K. Holmes, *Type II heat-labile enterotoxins from 50 diverse Escherichia coli isolates belong almost exclusively to the LT-IIc family and may be prophage encoded.* PLoS One, 2012. **7**(1): p. e29898.

156. Burk, C., et al., *Identification and characterization of a new variant of Shiga toxin 1 in Escherichia coli ONT:H19 of bovine origin.* J Clin Microbiol, 2003. **41**(5): p. 2106-12.
157. Friedrich, A.W., et al., *Escherichia coli Harboring Shiga Toxin 2 Gene Variants: Frequency and Association with Clinical Symptoms.* Journal of Infectious Diseases, 2002. **185**(1): p. 74-84.
158. Kim, E.J., et al., *Molecular insights into the evolutionary pathway of Vibrio cholerae O1 atypical El Tor variants.* PLoS Pathog, 2014. **10**(9): p. e1004384.
159. Lasaro, M.A., et al., *Functional and immunological characterization of a natural polymorphic variant of a heat-labile toxin (LT-I) produced by enterotoxigenic Escherichia coli (ETEC).* FEMS Immunol Med Microbiol, 2009. **55**(1): p. 93-9.
160. Lasaro, M.A.S., et al., *Production and release of heat-labile toxin by wild-type human-derived enterotoxigenic Escherichia coli.* 2006. **48**(1): p. 123-131.
161. Wijemanne, P. and R.A. Moxley, *Glucose significantly enhances enterotoxigenic Escherichia coli adherence to intestinal epithelial cells through its effects on heat-labile enterotoxin production.* PLoS One, 2014. **9**(11): p. e113230.
162. Horstman, A.L. and M.J. Kuehn, *Enterotoxigenic Escherichia coli Secretes Active Heat-labile Enterotoxin via Outer Membrane Vesicles.* Journal of Biological Chemistry, 2000. **275**(17): p. 12489-12496.
163. Sitnikova, T., A. Rzhetsky, and M. Nei, *Interior-branch and bootstrap tests of phylogenetic trees.* Mol Biol Evol, 1995. **12**(2): p. 319-33.
164. Yamanaka, H. and K. Okamoto, *Amino acid residues in the pro region of Escherichia coli heat-stable enterotoxin I that affect efficiency of translocation across the inner membrane.* Infect Immun, 1996. **64**(7): p. 2700-8.
165. Geller, B., et al., *Charged residues render pro-OmpA potential dependent for initiation of membrane translocation.* Journal of Biological Chemistry, 1993. **268**(13): p. 9442-7.
166. Busque, P., et al., *Production of Escherichia coli STb enterotoxin is subject to catabolite repression.* Microbiology, 1995. **141 (Pt 7)**: p. 1621-7.
167. Eisenreich, W., et al., *Metabolic host responses to infection by intracellular bacterial pathogens.* Frontiers in Cellular and Infection Microbiology, 2013. **3**: p. 24.
168. Staib, L. and T.M. Fuchs, *From food to cell: nutrient exploitation strategies of enteropathogens.* Microbiology, 2014. **160**(Pt 6): p. 1020-39.
169. Pickard, J.M. and A.V. Chervonsky, *Intestinal fucose as a mediator of host-microbe symbiosis.* J Immunol, 2015. **194**(12): p. 5588-93.
170. Pacheco, A.R., et al., *Fucose sensing regulates bacterial intestinal colonization.* Nature, 2012. **492**(7427): p. 113-7.
171. Wunderlichová, L., et al., *Formation, Degradation, and Detoxification of Putrescine by Foodborne Bacteria: A Review.* Comprehensive Reviews in Food Science and Food Safety, 2014. **13**(5): p. 1012-1030.
172. Bover-Cid, S., et al., *Relationships between microbial population dynamics and putrescine and cadaverine accumulation during dry fermented sausage ripening.* Journal of Applied Microbiology, 2009. **106**(4): p. 1397-1407.