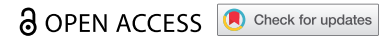



RESEARCH PAPER



Sequence variations in the ETEC CS6 operon affect transcript and protein expression

Jonathan Moon and Eileen M. Barry 

Center for Vaccine Development, University of Maryland School of Medicine, Baltimore, MD, USA

ABSTRACT

Enterotoxigenic *Escherichia coli* (ETEC) is a leading cause of diarrheal disease in developing nations where it accounts for a significant disease burden in children between the ages of 0 to 59 months. It is also the number one bacterial causative agent of traveler's diarrhea. ETEC infects hosts through the fecal-oral route and utilizes colonization factors (CF) to adhere within the small intestine. Over 25 CFs have been identified; 7 are considered major CFs and a vaccine targeting these is predicted to provide protection against up to 66% of ETEC associated disease. Coli Surface Antigen 6 (CS6) is a major CF and is associated with disease-causing ETEC isolates. Analysis of the CS6 operon sequence led to the identification of two regions of variability among clinical isolates which we predicted exert effects on CS6 transcript and protein expression. A total of 7 recombinant *E. coli* strains were engineered to encode the CS6 operon in wild-type, hybrid, and mutant configurations. Western blot analysis and RT-qPCR provided evidence to support the importance of an intergenic hairpin structure on CS6 expression. Our results reveal the significance of CS6 sequence selection regarding ETEC vaccine development and present novel information regarding CS6 sequence variation in WT ETEC strains.

ARTICLE HISTORY

Received 30 June 2021
Revised 25 August 2021
Accepted 8 September 2021

KEYWORDS

ETEC; CS6; expression; regulation; vaccine


Introduction

Enterotoxigenic *Escherichia coli* (ETEC) is a noninvasive enteric pathogen that is a significant cause of moderate-to-severe diarrheal disease (MSD) in children under five years of age in low resource regions. It is also a significant cause of traveler's diarrhea [1–4]. ETEC infects hosts through the fecal-oral route and colonizes the small intestine by utilizing colonization factors (CF) to adhere to intestinal epithelial cells. Adherence within the gut is required for successful ETEC infection and it has been shown that CF-specific antibodies can inhibit colonization and prevent disease [5,6]. Subsequent delivery of heat-labile (LT) and/or heat-stable (ST) toxin increases the production of second messenger cyclic nucleotides within intestinal epithelial cells resulting in ion efflux [7–9]. This shift in the osmotic gradient results in excretion of water into the intestinal lumen and leads to the common clinical presentation of watery diarrhea [10]. Over 25 CFs have been identified; seven are classified as major CFs including CFA/I, and CS1-CS6; it has been estimated that a vaccine targeting the major CFs could provide protection against up to 66% of ETEC-associated disease [11]. Although multiple ETEC

vaccine development strategies are ongoing, no FDA-approved ETEC vaccine currently exists [11–14].

Coli Surface Antigen 6 (CS6) is categorized as a major CF and multiple ETEC vaccine strategies have incorporated or targeted CS6 through a variety of approaches [15–17]. The CS6 operon is composed of four genes: *cssA* and *cssB* encode two structural subunits of 15 kDa and 16 kDa, respectively. *cssC* and *cssD* encode a chaperone and usher protein, respectively [18]. The crystal structure of both subunits was resolved elucidating that the two subunits which comprise the CS6 antigen are present in a 1:1 stoichiometric ratio linked by donor strand complementation [19]. Previous characterization of the CS6 antigen implicated fibronectin and sulfatide as host receptors within the gut [19,20]. Evidence supports the presence of multiple alleles of *cssA* and *cssB* in wild-type (WT) isolates and that SNPs in the sequences of these genes impact binding efficiency to Caco-2 cells [21]. Little information is available regarding other variations within the sequence of the CS6 operon and the resulting effects on CS6 expression although it has been suggested that CS6 gene expression may be regulated by a negative feedback loop [22].

CONTACT Eileen M. Barry  embarry@som.umaryland.edu; embarry@som.umaryland.edu  Center for Vaccine Development, University of Maryland School of Medicine, Baltimore, MD, USA

 Supplemental data for this article can be accessed [here](#)

Our analysis of the CS6 operon led to the identification of two regions which display heterogeneity in WT isolates. The first region occurs between the genes *cssB* and *cssC* where a stem-loop structure is predicted to form. The second variation is in the length of the *cssD* gene which may be truncated at the 3' terminus resulting in the loss of another stem-loop structure. Stem-loop structures form as a result of intramolecular base pairing that result in secondary structures that have been shown to play a role in regulation [23]. Stem-loops acting as regulators of CF expression through enhancement of mRNA stability has previously been described [24], however secondary structures encoded within CF operons are largely uncharacterized. Based on preliminary CS6 expression data in ETEC isolates, we hypothesized that these regions of variability may contribute to differential CS6 expression. Recombinant plasmids encoding WT, hybrid, and mutant CS6 operons were engineered and confirmed the importance of these regions on CS6 expression. Ultimately, we endeavor to identify factors that affect CS6 expression and their contribution to optimization of CS6 expression in the context of a live-attenuated multivalent vaccine [25].

Materials and methods

Bacterial strains and culture conditions

Table 1 lists the strains and plasmids used in this study. Recombinant CS6 operons were cloned in DH5 α and grown at 30°C on LB agar or in LB broth culture in the presence of 100 μ g/mL carbenicillin or kanamycin. ETEC strains were grown on CFA agar (2% agar) or in CFA broth (1% Casamino acids, 0.15% Yeast Extract, 0.005% MgSO₄, 0.0005% MnCl₂) at 37°C [26,27].

Sequencing of ETEC isolates

The primers CS601F and CS601R were used with the Roche Expand Long Template PCR kit (RocheLife Sciences, Mannheim, Germany) to amplify a 2195bp PCR product containing both the *cssB* and *cssC* genes. The primer CS6JF1 was used with either the primer *cssD_long_R1* or *cssD_short_R1* to amplify a 340bp or 328bp product, respectively, from 23 CS6 encoding ETEC isolates (Supplementary Figure 1). These PCR products were then sequenced to enable categorization of strains. Nineteen published CS6 sequences were also assessed and categorized [28].

Molecular genetic techniques and construction of strains

The primers used to amplify the CS6 operon were generated by Integrated DNA Technologies (Coralville, IA) and are listed in Table 2. PCR reagents were purchased from Roche Life Sciences (Mannheim, Germany) and restriction enzymes were purchased from New England Biolabs (Ipswich, MA).

The CS6 operon was amplified by PCR as a 4294bp fragment from the ETEC strain E17018A. The amplified product was ligated into the TOPO-XL cloning vector. The CS6 insert was then excised with XhoI/SpeI and ligated into pZA112 (formerly named pGA2) [29] to create the plasmid pZA112-mLpp-CS6. The operon was cloned under the control of a mutated Lpp promoter (PmLpp) to enhance transcription [30]. To develop the recombinant CS6 operons, the genes *cssA* and *cssB* were amplified from pZA112-mLpp-CS6 as a single 986bp fragment using the Roche Expand Long Template PCR kit (Roche Life Sciences, Mannheim, Germany) with XhoI and AscI ends. This PCR product was digested using XhoI and AscI; T4

Table 1. Strains used in this study.

Strain Name	Strain Derivation	Description
CS6WT1	DH5 α (pGRG-mLpp-CS6.1)	Small intergenic hairpin; Short <i>cssD</i>
CS6WT2	DH5 α (pGRG-mLpp-CS6.2)	Large intergenic hairpin; Long <i>cssD</i>
CS6H1	DH5 α (pGRG-mLpp-CS6.3)	Small intergenic hairpin; Long <i>cssD</i>
CS6H2	DH5 α (pGRG-mLpp-CS6.4)	Large intergenic hairpin; Short <i>cssD</i>
CS6M1	DH5 α (pGRG-mLpp-CS6.5)	Altered palindromic sequence of intergenic hairpin; no stem-loop formation
CS6M2	DH5 α (pGRG-mLpp-CS6.6)	No intergenic region between <i>cssB</i> and <i>cssC</i>
CS6M3	DH5 α (pGRG-mLpp-CS6.7)	PCR mutation in palindromic sequence of stem-loop; deletion of two bases
-	DH5 α (pZA112-mLpp-CS6)	CS6 vector cloned from ETEC strain E17018A
ETEC 214-4	-	Wild-type ETEC (Group 1 CS6)
ETEC E17018A	-	Wild-type ETEC (Group 2 CS6)
-	DH5 α (pGRG-mLpp)	Vector; CS6 Negative Control

Table 2. Primers used in this study.

Primer	Description	Sequence (5' – 3')
cssA-F	<i>cssA</i> forward primer	TAACTCGAGATGAAGAAAACAATTGGTTTAATTCTAATTCCTG
cssB-R	<i>cssB</i> reverse primer	TAAGGCGCGCCTTAATTGCTGTAAAATGATACAGTCAAATTCCTG
cssC-F1	<i>cssC</i> forward – no intergenic region	TAAGGCGCGCCATGAAATCAAAGTTAATTATACTATTGACGTTAGTGCCATTTTC
cssC-F2	<i>cssC</i> forward – small stem- loop	TAAGGCGCGCCAAAAGGCCGATTATTGCGGCCATTGACGATACTGCTAGGCCAAAAT
cssC-F3	<i>cssC</i> forward – altered stem- loop	TAAGGCGCGCCAAAAGTCATGATTATTGATTGCGGCCATTGACGATACTGCCAG
cssC-F4	<i>cssC</i> forward – large stem- loop	TAAGGCGCGCCAAAAGGCCGATTATTGATTGCGGCCATTGACGATACTGCCAGGCCAA
cssD-R1	<i>cssD</i> reverse – short <i>cssD</i>	GGCGCGCCTAACATTGTTTATTACAACAGATAATTGTTTGCTAG
cssD-R2	<i>cssD</i> reverse – long <i>cssD</i>	TTAGGCGCGCCTTAGTCTCCAGAATTTTCGGGGCG
cssA -F-qPCR	<i>cssA</i> forward RT-qPCR	GGACGACTCGTAAATACCGCT
cssA-R-qPCR	<i>cssA</i> reverse RT-qPCR	TTAGGCGTAACCTCTGCACC
cssB-F-qPCR	<i>cssB</i> forward RT-qPCR	TCTGGACAGCAGATCGGAAAG
cssB-R-qPCR	<i>cssB</i> reverse RT-qPCR	TGCCCTGCCATAAACTTACCA
cssC-F-qPCR	<i>cssC</i> forward RT-qPCR	TGAATCAATGCCACCAACAGA
cssC-R-qPCR	<i>cssC</i> reverse RT-qPCR	AATGCATCCCCGAATGCTGA
cssD-F-qPCR	<i>cssD</i> forward RT-qPCR	CGGCAACCAGTTCTGTAGGT
cssD-R-qPCR	<i>cssD</i> reverse RT-qPCR	TACGGGTCTGTTCTGTTCTGC
rpoD-F	<i>rpoD</i> forward RT-qPCR	GAGCAAGGCTATCTGACCTATG
rpoD-R	<i>rpoD</i> reverse RT-qPCR	GCCCATGTCTGTTGATTTG
CS601F	<i>cssB</i> - <i>cssC</i> forward primer	TGGTGCAGAGGTTACACCTAATC
CS601R	<i>cssB</i> - <i>cssC</i> reverse primer	GAGAGTCTGAATCAGCCACTCCATG
cssD_long_R1	Longer <i>cssD</i> reverse primer	CTCTTTCTCAGGAAGTTTAGTCTCCAGAATTTTCGG
cssD_short_R1	Shorter <i>cssD</i> reverse primer	CACATGTCTACTAATTGGATGCACTACCTAAC

DNA Ligase (New England Biolabs, Ipswich, MA)) was used to ligate the fragment into the vector pGRG-mLpp to yield pGRG-mLpp-*cssA*-*cssB* [31]. The *cssC* and *cssD* genes, including the non-coding region between *cssB* and *cssC*, were subsequently amplified from pZA112-mLpp-CS6 as either a 3,139bp or 3,167bp product with flanking *AscI* ends. The 3,139bp product encodes the smaller stem-loop structure and the shorter *cssD* gene and was amplified using the primers *cssC*-F2 and *cssD*-R1 to create the construct CS6WT1. The construct CS6WT2 was created using the 3,167bp product, which encodes the longer versions of each region, and was amplified with the primers *cssC*-F4 and *cssD*-R2. Two hybrid CS6 constructs were also engineered one of which, CS6H1, contained the smaller stem-loop and longer *cssD* (3,163bp), amplified with primers *cssC*-F2 and *cssD*-R2, and the other, CS6H2, containing the larger stem-loop and shorter *cssD* (3,143bp) was amplified with primers *cssC*-F4 and *cssD*-R1. Alternative CS6 operon constructs were created in which the stem-loop between *cssB* and *cssC* was mutated to not form or was deleted entirely. The alternative construct, CS6M1, was engineered by changing the nucleotides in one half of the stem-loop sequence such that the structure would no longer form while leaving the entire intergenic region intact. This was accomplished during PCR amplification of the *cssC* and *cssD* genes using the forward primer *cssC*-F3, in which one half of the

palindromic stem sequence was altered from GGCCGC to GTCATG, and the reverse primer *cssD*-R1. In the alternative construct CS6M2, the entire intergenic region was deleted by amplifying the *cssC* and *cssD* genes with the forward primer *cssC*-F1 which binds at the *cssC* start codon to exclude the intergenic sequence completely and the reverse primer *cssD*-R1. Finally, the construct CS6M3 acquired a mutation (G and C deletion from one of two palindromic sequences) within the stem-loop during PCR with the primers *cssC*-F2 and *cssD*-R1 that increased the predicted Gibbs Free Energy (ΔG) of the structure thus reducing the likeliness of its random formation. All *cssC*-*cssD* PCR products were digested and ligated into the vector pGRG-mLpp-*cssA*-*cssB*. Correct cloning was confirmed by colony PCR using the primers CS601F and CS601R and positive transformants were isolated and stored in LB stocks with 20% glycerol at -80°C . Plasmids were extracted using the GeneJet Plasmid Miniprep Kit (Thermo Fisher Scientific, Waltham, MA), and sequenced for confirmation.

RT-qPCR

Each recombinant CS6 strain was grown from frozen stocks in LB broth culture (Becton Dickinson, Franklin Lakes, NJ)). Subcultures were grown to an OD_{600} of 0.5.

10 mL of each culture was sampled for total RNA isolation. Bacterial pellets were resuspended in TriPure Isolation Reagent (Roche Life Sciences, Mannheim, Germany), extracted with chloroform and precipitated with isopropanol. Each RNA pellet was resuspended in nuclease-free water (Thermo Fisher Scientific, Waltham, MA) plus 1 μ L of RNase Out (Thermo Fisher Scientific, Waltham, MA). Contaminating DNA was removed from the total RNA sample using the Turbo DNA-free kit (Thermo Fisher Scientific, Waltham, MA) and the absence of contaminating DNA was confirmed by endpoint PCR. cDNA was generated using the qScript cDNA Synthesis Kit (Quantabio, Beverly, MA)

qPCR primer sequences (Table 2) were generated using the NCBI primer design tool. Each gene was analyzed in triplicate from two biological replicate samples using FastStart Universal SYBR Green Master Mix (Rox) (Roche Life Sciences, Mannheim, Germany)). The gene *rpoD* was used to normalize across samples. The fold change of each transcript was calculated using a $\Delta\Delta C_t$ analysis: the C_t value recorded for each CS6 sample was subtracted from the average C_t value of the *rpoD* replicates to give the normalized ΔC_t value. CS6WT2 was used as the control sample and the triplicate ΔC_t values from each individual CS6 gene detected from the Group 2 construct were averaged to give an average control value. Then, the average control value was subtracted from the ΔC_t value of each replicate sample from each other construct variation to give the $\Delta\Delta C_t$ value for each replicate sample. The fold change in each sample was then calculated as $2^{-\Delta\Delta C_t}$ and the average fold change from triplicate wells was then calculated from these values which allowed for comparison of relative transcript quantity.

A primer efficiency curve was also carried out using a 1:10 serial dilution of CS6WT1 cDNA to a final dilution of 1:10,000,000. Reactions were prepared as previously stated. The average C_t of the replicate wells analyzing each CS6 gene were then plotted linearly against the respective dilution factor.

Western blot analysis

Whole cell bacterial lysates were prepared for western blot analysis as follows. Bacteria were harvested from overnight growth on agar plates into 1xDPBS (Corning, Corning, NY) and the OD_{600} was used to normalize to an OD of 3.0. The normalized suspensions were then mixed 1:1 with 2x Laemmli Sample Buffer (0.125 M Tris pH 6.8, 4% SDS, 20% Glycerol, 0.2% Bromophenol Blue, 10% β -mercaptoethanol). Proteins were separated on a 12% Mini-Protean TGX Precast Gel (Bio-Rad,

Hercules, CA) before being transferred to a PVDF membrane. The membrane was blocked in a 10%(w/v) nonfat milk buffer in 1xDPBS and then incubated with polyclonal Rabbit α -CS6 sera (Rockland Immunochemicals, Limerick, PA) in 10% (w/v) nonfat milk buffer overnight at 4°C on an orbital shaker. Following washing, the membrane was incubated with secondary antibody Goat α -Rabbit 680 nm (Invitrogen, Carlsbad, CA) and proteins visualized using the 700 nm wavelength setting on an LI-COR Odyssey Laser Scanner.

Results

Sequence analysis of ETEC and recombinant strains encoding CS6

Preliminary characterization of a subset of CS6 encoding ETEC isolates demonstrated differential CS6 expression under identical growth conditions; ETEC strain 214-4 expressed a higher level of CS6 than strains E17018A, 203740 and 204576 (Figure 1). Sequence analysis of ETEC isolates 214-4 and E17018A revealed two regions of variation within the CS6 operon (Figure 2). The first region occurs in a noncoding sequence between the genes *cssB* and *cssC*. In 214-4, this noncoding sequence is comprised of 48bp while in E17018A the sequence totals 52bp. Importantly, the difference of 4bp corresponds to the formation of a stem-loop structure containing either 4 or 8 bases within the loop (Figure 2(b)). This structural variation of the stem-loop may impact its formation and stability as indicated by the Gibbs Free Energy (ΔG) which was calculated to be -10.3 and -9.5 in WT1 strain 214-4 and WT2 strain E17018A, respectively.

The second variable region occurs at the 3' terminus of the *cssD* gene. In 214-4, the gene totals 2436bp while in E17018A it measures 2460bp (Figure 2(c)). The difference of 24bp translates to an additional 8 amino acids at the carboxyl terminus of *CssD* in E17018A. Also, within this extra 24bp sequence another stem-loop structure is predicted to form in E17018A that is absent in the 214-4 sequence.

We analyzed the CS6 sequences of a total of 42 unique ETEC isolates from 8 geographically distinct locations to assess the distribution of the motifs we observed in 214-4 and E17018A (Table 3). Of the 42 strains analyzed, 9 encoded the CS6 operon with the smaller stem-loop and shorter *cssD*, consistent with ETEC strain 214-4, and were classified as Group 1 strains. The remaining 33 strains were classified as Group 2 strains and encoded the CS6 operon

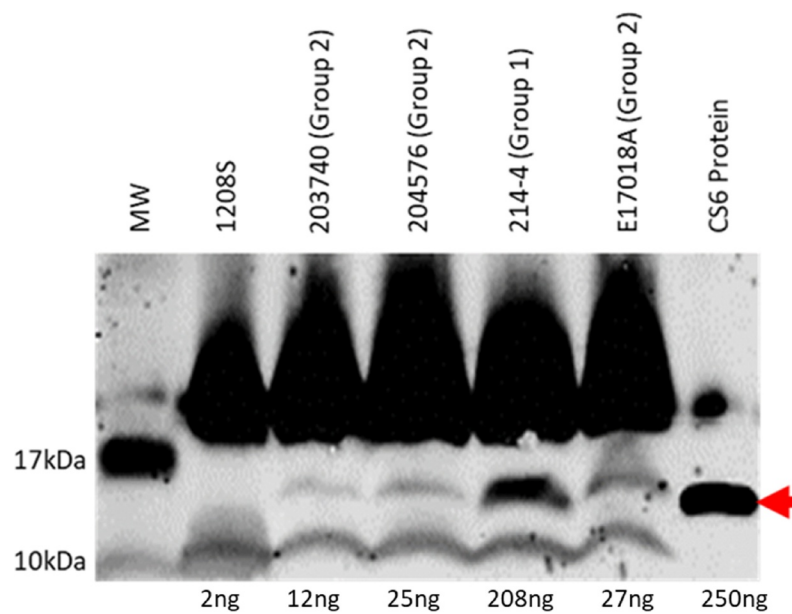


Figure 1. Western blot analysis of four different CS6⁺ ETEC strains. Expression of CS6 was determined by western blot analysis of whole cell lysates. Strains 203740, 204576, and E17018A were categorized as Group 2 ETEC strains and share the same CS6 operon structure. 214-4 was categorized as Group 1 and is distinct in sequence from Group 2 strains. Protein (CS6) amounts were quantified by densitometric analysis based on the single CS6 purified protein standard (lane 7) and the negative control 1208S (lane 2) which is an attenuated *Shigella flexneri* strain.

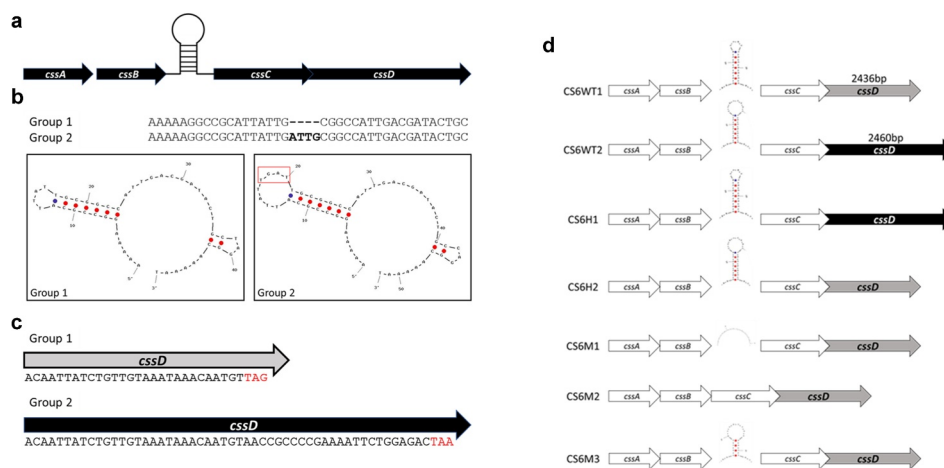


Figure 2. CS6 operon structures. a) Organization of the CS6 operon demonstrating the stem-loop structure the intergenic region between *cssB* and *cssC*. b) The sequence of the stem-loop structure denoting 4-bp smaller loop structure in Group 1 CS6 compared to Group 2 CS6. c) In Group 1 CS6, the final gene in the operon, *cssD*, is truncated by 24bp compared to Group 2 CS6 and exhibits no identifiable downstream sequence homology. d) Schematic display of the seven recombinant CS6 operons. CS6WT1 and CS6WT2 are found in WT-ETEC strains. CS6H1 and CS6H2 were engineered as hybrids and were not observed in any WT strains. In CS6M1, one half of the palindromic sequence was mutated such that the hairpin structure does not form. CS6M2 does not contain any intergenic region between *cssB* and *cssC*. CS6M3 acquired two deletions in the palindromic sequence during PCR amplification and forms a shorter stem comprised of 4 bases rather than the typical 6–7.

containing the larger stem-loop and longer *cssD* as observed in ETEC strains E17018A, 203740, and 204576. All strains examined in this subset conformed

to either the Group 1 or Group 2 classification. Further evaluation revealed that strains encoding the STh and LT toxins were present in both groups, but strains

Table 3. Distribution of CS6 operon motifs in geographically diverse CS6 ETEC isolates.

CS6 Operon Group	Location	Colonization Factors	Toxin Profile
Group 1			
201446	Mali	CS4, CS6	STp, LT
302054	Mozambique	CS6, CS21	STp, LT
401061	Kenya	CS4, CS6, CS21	STp, LT
503046	India	CS4, CS6, CS21	STh
503458	India	CS4, CS6, CS21	STh
510016	India	CS4, CS6, CS21	STh
520873	India	CS4, CS6, CS21	STh
214-4	Mexico	CS6, CS21	STp
E8775	Mexico	CS4, CS6	ST, LT
Group 2			
100137	Gambia	CS5, CS6	LT
100345	Gambia	CS5, CS6	STh
100491	Gambia	CS6	STh, LT
103605	Gambia	CS5, CS6	STh
120899	Gambia	CS5, CS6	STh
200006	Mali	CS6, CS21	LT
200065	Mali	CS5, CS6	STh
203740	Mali	CS6	STh
204446	Mali	CS5, CS6	STh, LT
204576	Mali	CS5, CS6	STh
300006	Mozambique	CS5, CS6	STh, LT
300007	Mozambique	CS5, CS6	STh, LT
300239	Mozambique	CS6, CS14, CS21	LT
400588	Kenya	CS5, CS6, CS14	STh, LT
400650	Kenya	CS6	STh
401068	Kenya	CS6, CS14, CS21	LT
500465	India	CS6, CS21	STh
500819	India	CS5, CS6	LT
503025	India	CS6, CS21	STh
503440	India	CS6, CS21	STh
503663	India	CS6	STh, LT
503829	India	CS6, CS21	STh
504211	India	CS6, CS21	STh
504237	India	CS5, CS6	STh
600489	Bangladesh	CS5, CS6	STh, LT
602354	Bangladesh	CS5, CS6	STh
604113	Bangladesh	CS6, CS21	LT
E10703	Bangladesh	CS5, CS6	ST
E17018A	Bangladesh	CS5, CS6	ST
700006	Pakistan	CS5, CS6	STh, LT
700173	Pakistan	CS5, CS6, CS14	STh, LT
700360	Pakistan	CS6	LT
703322	Pakistan	CS5, CS6	LT

encoding STp were limited to Group 1. Strains in both groups encoded operons for multiple CFs; we noted that no Group 1 strains contained genes that encode CS5 and that no Group 2 strains encoded CS4. These trends were consistent within the subset of strains analyzed.

Seven recombinant constructs were engineered in an identical plasmid-based system to assess the potential effects that these sequence variations confer on CS6 expression (Figure 2(d)). The operons were cloned

behind a constitutive promoter to remove any impact of upstream regulation. CS6WT1 and CS6WT2 correspond to the Group 1 and Group 2 WT operons, respectively. CS6H1 and CS6H2 are hybrid operons which contain the two combinations of Group 1 and Group 2 stem-loop and *cssD* motifs. CS6M1, CS6M2, and CS6M3 all include the Group 1 *cssD* motif (2436bp) but are mutated in the stem-loop sequence. In CS6M1 the nucleotides in one half of the *cssB-csC* stem-loop were altered such that the structure would no longer form. In CS6M2, the entire noncoding region between *cssB* and *cssC* was deleted. CS6M3 contains two deletion mutations in the stem-loop's palindromic sequence which results in formation of a 4bp stem as opposed to 7bp in the WT and hybrid configurations and an intergenic sequence composed of 46bp.

CS6 protein expression

Expression of the CS6 antigen in WT and recombinant strains was assessed by western blot analysis. Bands were quantitatively assessed using densitometric analysis based on a single CS6 protein standard. Initial analysis of WT ETEC strains demonstrated a 7.7-fold increase in CS6 expression in the Group 1 strain 214-4 compared to the Group 2 strain E17018A (Figure 1).

Expression of CS6 was compared among the recombinant CS6 constructs (Figure 3). CS6H1, which encodes the Group 1 stem-loop and Group 2 *cssD*, exhibited the highest levels of CS6 expression; expression was 40% greater than CS6WT1 and 110% greater than CS6WT2. CS6H2 which encodes the Group 2 stem-loop and Group 1 *cssD* expressed ~50% of the level of CS6 in CS6H1 and was intermediate between CS6WT1 and CS6WT2. CS6 expression in all three mutant constructs, CS6M1, CS6M2, and CS6M3 was very low and was approximately comparable between all three mutants.

CS6 transcript expression

The observed range of CS6 antigen expression in the recombinant CS6 constructs prompted investigation into CS6 mRNA expression as stem-loop structures have been implicated in the regulation of both transcription and translation [24,32]. We quantified transcripts of all four CS6 genes using RT-qPCR and calculated the relative amount of transcript expressed through a $\Delta\Delta C_t$ analysis. C_t values determined for each construct were generated relative to those values calculated for CS6WT2 (Figure 4). Overall, expression of all four *css* genes was higher in all engineered constructs

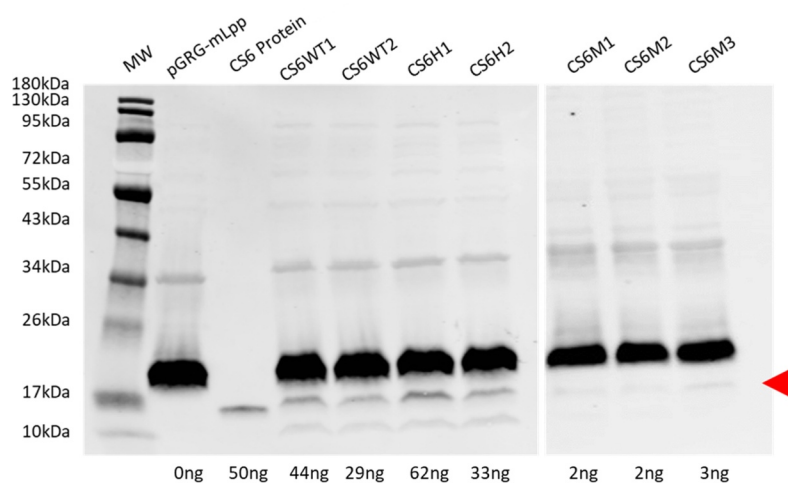


Figure 3. CS6 expression. Expression of CS6 from the recombinant operons expressed in DH5 α was determined by western blot analysis of whole cell lysates. Proteins (CS6) were quantified by densitometry based on the single CS6 protein standard (lane 3) and the negative control pGRG-mLpp. CS6M1, CS6M2, and CS6M3 were run on a separate blot with a single CS6 protein standard used for densitometric analysis.

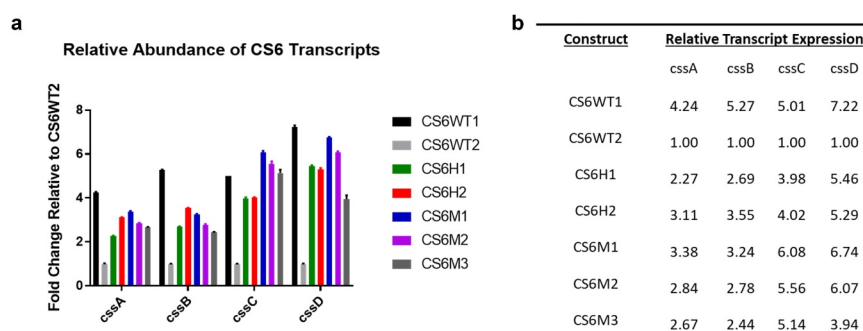


Figure 4. A) RT-qPCR analysis of CS6 transcript abundance compared to CS6WT2. Strains were grown to OD₆₀₀ 0.5 in LB broth and harvested to isolate total RNA. b) Relative fold change values of CS6 transcripts in each recombinant CS6 strain.

relative to CS6WT2. These data showed that, on average, CS6WT1 expressed 4.24-fold more *cssA* transcript, 5.27-fold more *cssB* transcript, 5.01-fold more *cssC* transcript, and 7.22-fold more *cssD* transcript than CS6WT2. CS6H1 and CS6H2 expressed 2.27 and 3.11-fold more *cssA* transcript, 2.69 and 3.55-fold more *cssB* transcript, 3.98 and 4.02-fold more *cssC* transcript, and 5.46 and 5.29-fold more *cssD* transcript, respectively. The mutants that abolished the *cssB*-*cssC* intergenic stem-loop structure (CS6M1 and CS6M2) or reduced the stem length (CS6M3) had higher levels of all transcripts than CS6WT2. CS6M1, CS6M2, and CS6M3 also expressed 3.38, 2.84, and 2.67-fold greater amounts of *cssA*, respectively. *cssB* transcripts in the mutant constructs were elevated by 3.24, 2.78, and 2.44-fold, respectively. *cssC* transcripts were increased by 6.08, 5.56, and 5.14-fold, respectively. Finally, *cssD*

transcripts were greater by 6.74, 6.07, and 3.94-fold, respectively when compared to CS6WT2 transcript quantities (Figure 4(a)).

Primer efficiency was calculated to be 102.46%, 99.15%, 91.97%, 100.37%, and 92.10% for the genes *cssA*, *cssB*, *cssC*, *cssD* and *rpoD*, respectively (Supplementary Figure 2).

Discussion

The regulation of ETEC CF expression is affected by environmental cues via upstream sequences and regulators including *rms* [33–35]. Additionally, regulation at the transcriptional level is impacted by intergenic sequences including those that form stem-loop structures [24,32,35]. Stem-loops have also been described in

the sequences of the genes *eltA* and *eltB* which encode subunits of the LT toxin and are suggested to have a regulatory role in expression providing some precedence for stem-loop structures as regulators in ETEC [36]. Understanding this regulation lends insight into the role of CFs during disease and provides information for optimizing their expression in vaccine constructs. The Center for Vaccine Development and Global Health has an ongoing program to develop a multivalent *Shigella*-ETEC vaccine that is composed of live attenuated *Shigella* strains expressing immunogenic ETEC antigens including major CFs [37,38]. This strategy requires optimization of heterologous antigen expression. CS6 is a major CF and is a critical component of a broadly protective vaccine against ETEC [11].

We initially observed variable CS6 expression in WT ETEC strains which led us to examine the sequences of the CS6 operons. Several SNPs were present in coding sequences, but interestingly we identified two regions within the operon that exhibited more pronounced divergence. Sequence analysis of a selection of clinical ETEC isolates [28] revealed that isolates could be categorized based on their CS6 operon profile as defined by the variable regions. We found that the smaller stem-loop was always present in CS6 operons that also had the shorter *cssD* gene (Group 1) and that the larger stem-loop was always present with the longer *cssD* gene (Group 2). This association occurred regardless of the geographical location from which the strain was isolated or the strains' toxin profile. We noted that the total CF profile of the strains did not associate with the CS6 operon variation (Group 1 or Group 2) although none of the Group 1 strains we analyzed encoded CS5 and none of the Group 2 strains encoded CS4. These data suggest that at least two variations of the CS6 operon exist in WT strains and are not geographically or genotypically restricted.

Western blot analysis of recombinant CS6 strains provided evidence to support a contribution of the stem-loop, which forms between the genes *cssB* and *cssC*, to regulation of expression. CS6WT1 and CS6H1 strains encode the smaller version of this structure, which contains four bases within the loop. These strains expressed more CS6 protein than constructs encoding the larger stem-loop containing eight bases in the loop (CS6WT2 and CS6H2). When the stem-loop sequence was altered or deleted, we observed a dramatic decrease in CS6 protein expression. These results suggest that there is an optimal configuration of the stem-loop which intrinsically promotes greater CS6 protein expression. This could be due to differences in how the mRNA folds and fits into the ribosome; further characterization of the mRNA's structure is necessary to evaluate this.

Although we detected differential protein expression across the recombinant strains, RT-qPCR analysis provided evidence that all CS6 transcripts were elevated to similar levels in recombinant CS6 strains when compared to the Group 2 recombinant strain, CS6WT2. We were surprised to find that CS6M1, CS6M2, and CS6M3 also expressed more CS6 transcripts. We hypothesize that when this structure is absent entirely there is a lack of inhibitory regulation of CS6 transcription, potentially due to the absence of physical interference by the secondary structure. The increase in *css* transcription by the mutant operons in conjunction with negligible protein expression, supports the requirement for the stem-loop structure for RNA stability and/or protein translation. The decoupling of transcript quantity and protein expression suggests another potential regulatory role of the two variable regions that we identified in the CS6 operon. Multiple post-transcriptional effects have been attributed to stem-loop structures including regulation of gene expression through transcript stability [24,35] and it has been shown that mutations resulting in the loss of specific stem-loops can impact protein expression [32]. Furthermore, secondary structures such as stem-loops and cruciforms are known to form in DNA and can exert a variety of effects, including sterically impeding binding of RNA polymerases or other transcription factors resulting in reduced transcriptional activity [39].

The effect of the truncated *cssD* gene is not entirely clear. However, the increase in transcription in CS6H2 which encodes the same *cssB*-*cssC* stem-loop as CS6WT2, but encodes the shorter *cssD*, supports its role in regulation. We posit that the lack of the 3' terminal stem-loop alters the folding of the molecule such that negative regulation of transcription is reduced.

Our data suggest a role in regulation of CS6 expression by two secondary structures we identified in this study potentially due to alterations in the conformation of the CS6 mRNA.

CS6 is an important antigen in the development of an ETEC vaccine as it is one of the most common CFs associated with ETEC disease. There have been various ETEC vaccine candidates developed which incorporate CS6 [15,16,40], however no FDA licensed vaccine currently exists. We showed that the sequence of the stem-loop and of the *cssD* gene impact both CS6 transcript and protein expression in recombinant *E. coli* strains which share identical genetic backgrounds. Our results provide support for the implementation of one version of the CS6 operon for use in vaccine design for optimization of high level CS6 expression. Further

characterization of the CS6 operon mRNA and visualization of recombinant strains by immunogold electron microscopy may yield more insight into differences in antigen export and surface expression which may provide further support for identification of a superior CS6 operon allele for implementation in ETEC vaccine design.

Acknowledgments

We would like to thank Drs David Rasko and Sharron Tennant at the University of Maryland, Baltimore for supplying ETEC sequences and isolates used in this study. This work was funded by NIH U19AI109776 and NIH U19AI142725.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Institutes of Health [U19AI142725]; National Institutes of Health [U19AI109776].

Data availability statement

Supporting sequence data of ETEC strains utilized in this study are available online through GenBank at https://www.ncbi.nlm.nih.gov/nuccore/?linkname=pubmed_nuccore&from_uid=30944874 [28].

ORCID

Eileen M. Barry  <http://orcid.org/0000-0002-5326-0047>

References

- [1] Kotloff KL, Nataro JP, Blackwelder WC, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet*. 2013;382(9888):209–222. Epub 2013/05/18. PubMed PMID: 23680352
- [2] Liu J, Platts-Mills JA, Juma J, et al. Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case-control study. *Lancet*. 2016;388(10051):1291–1301. Epub 2016/09/28. PubMed PMID: 27673470; PubMed Central PMCID: PMC45471845
- [3] Porter CK, Olson S, Hall A, et al. Travelers' Diarrhea: an update on the incidence, etiology, and risk in military deployments and similar travel populations. *Mil Med*. 2017;182(S2):4–10. Epub 2017/09/09. PubMed PMID: 28885918;
- [4] Hyams KC, Bourgeois AL, Merrell BR, et al. Diarrheal disease during operation desert shield. *N Engl J Med*. 1991;325(20):1423–1428. Epub 1991/11/14. PubMed PMID: 1656260
- [5] Freedman DJ, Tacket CO, Delehanty A, et al. Milk immunoglobulin with specific activity against purified colonization factor antigens can protect against oral challenge with enterotoxigenic *Escherichia coli*. *J Infect Dis*. 1998;177(3):662–667. Epub 1998/03/14. PubMed PMID: 9498445;
- [6] Savarino SJ, McKenzie R, Tribble DR, et al. Hyperimmune bovine colostrum anti-CS17 antibodies protect against enterotoxigenic *Escherichia coli* diarrhea in a randomized, doubled-blind, placebo-controlled human infection model. *J Infect Dis*. 2019;220(3):505–513. Epub 2019/03/22. PubMed PMID: 30897198
- [7] Rao MC. Toxins which activate guanylate cyclase: heat-stable enterotoxins. *Ciba Found Symp*. 1985;112:74–93. Epub 1985/01/01. PubMed PMID: 2861070.
- [8] Gill DM, Richardson SH. Adenosine diphosphate-ribosylation of adenylate cyclase catalyzed by heat-labile enterotoxin of *Escherichia coli*: comparison with cholera toxin. *J Infect Dis*. 1980;141(1):64–70. Epub 1980/01/01. PubMed PMID: 6988518;
- [9] Qadri F, Svennerholm AM, Faruque AS, et al. Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clin Microbiol Rev*. 2005;18(3):465–483; Epub 2005/07/16. PubMed PMID: 16020685; PubMed Central PMCID: PMC1195967;
- [10] Levine MM. *Escherichia coli* that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. *J Infect Dis*. 1987;155(3):377–389; Epub 1987/03/01. PubMed PMID: 3543152;
- [11] Vidal RM, Muhsen K, Tennant SM, et al. Colonization factors among enterotoxigenic *Escherichia coli* isolates from children with moderate-to-severe diarrhea and from matched controls in the Global Enteric Multicenter Study (GEMS). *PLoS Negl Trop Dis*. 2019;13(1):e0007037. Epub 2019/01/05. PubMed PMID: 30608930; PubMed Central PMCID: PMC6343939 following competing interests. We note that the following authors are involved in the development of vaccines to prevent diarrhea caused by enterotoxigenic *Escherichia coli*: James P. Nataro: Has been named in patents for technology that may have relevance for ETEC vaccine technology; JPN is also a co-investigator on grants from non-profit agencies that support ETEC vaccine development. Halvor Sommerfelt: Has been named in patents for technology that may have relevance for ETEC vaccine technology including: INT Application Number PCT/IB2014/000267; INT Publication Number WO2014128555 A2 “ETEC Vaccine”. European Patent Application No. 14711297.3 United States Patent Application No. 14/769,342. China Patent Application No. 201480022329.6. Puntervoll P, Sommerfelt H, Clements J, Nataro JP, Zhang W, Tacket A. HS is also a co-investigator on grants from non-profit agencies that support ETEC vaccine development Ann-Mari Svennerholm: Has shares in the University of Goteborg spin-out biotech company Gotovax AB which is entitled to royalty from Scandinavian Biopharma on

- sales in travelers of the ETEC vaccine Etvax if it becomes a commercial product. A-MS is a co-inventor on ETEC vaccine patent application owned by Scandinavian Biopharma and has a research grant from the Swedish Foundation for Strategic Research for infection biology research. Myron M. Levine: Is a co-inventor of patents for ETEC vaccine technology including US patent #6,902,736 B2. "Isolation and characterization of the CSA operon (ETEC-CS4 pili) and methods of using same." He is a member of the Scientific Advisory Board of the PaxVax Corporation. MML is also the recipient for grants relevant to ETEC vaccine development including the Enteric Center of Excellence for Translational Research (Enteric-CETR) "Immunoprophylactic Strategies to Control Emerging Enteric Infections" (NIAID U19AI109776; MML, PI)
- [12] Seo H, Zhang W. Development of effective vaccines for enterotoxigenic *Escherichia coli*. *Lancet Infect Dis.* **2020**;20(2):150–152; Epub 2019/11/24. PubMed PMID: 31757775;
 - [13] Bourgeois AL, Wierzba TF, Walker RI. Status of vaccine research and development for enterotoxigenic *Escherichia coli*. *Vaccine.* **2016**;34(26):2880–2886; Epub 2016/03/19. PubMed PMID: 26988259
 - [14] Khalil I, Walker R, Porter CK, et al. Enterotoxigenic *Escherichia coli* (ETEC) vaccines: priority activities to enable product development, licensure, and global access. *Vaccine.* **2021**;39(31):4266–4277. Epub 2021/05/10. PubMed PMID: 33965254
 - [15] Poole ST, Maciel M Jr., Dinadayala P, et al. Biochemical and immunological evaluation of recombinant CS6-derived subunit enterotoxigenic *Escherichia coli* vaccine candidates. *Infect Immun.* **2019**;87(3):Epub 2019/01/04. PubMed PMID: 30602504; PubMed Central PMCID: PMC6386543.
 - [16] Akhtar M, Chowdhury MI, Bhuiyan TR, et al. Evaluation of the safety and immunogenicity of the oral inactivated multivalent enterotoxigenic *Escherichia coli* vaccine ETVAX in Bangladeshi adults in a double-blind, randomized, placebo-controlled Phase I trial using electrochemiluminescence and ELISA assays for immunogenicity analyses. *Vaccine.* **2019**;37(37):5645–5656. Epub 2018/11/27. PubMed PMID: 30473185; PubMed Central PMCID: PMC6717083
 - [17] Laird RM, Ma Z, Dorabawila N, et al. Evaluation of a conjugate vaccine platform against enterotoxigenic *Escherichia coli* (ETEC), *Campylobacter jejuni* and *Shigella*. *Vaccine.* **2018**;36(45):6695–6702. Epub 2018/10/03. PubMed PMID: 30269917; PubMed Central PMCID: PMC6223012
 - [18] Ghosal A, Bhowmick R, Banerjee R, et al. Characterization and studies of the cellular interaction of native colonization factor CS6 purified from a clinical isolate of enterotoxigenic *Escherichia coli*. *Infect Immun.* **2009**;77(5):2125–2135. Epub 2009/02/25. PubMed PMID: 19237522; PubMed Central PMCID: PMC62681727
 - [19] Roy SP, Rahman MM, Yu XD, et al. Crystal structure of enterotoxigenic *Escherichia coli* colonization factor CS6 reveals a novel type of functional assembly. *Mol Microbiol.* **2012**;86(5):1100–1115; Epub 2012/10/11. PubMed PMID: 23046340;
 - [20] Jansson L, Tobias J, Jarefjall C, et al. Sulfatide recognition by colonization factor antigen CS6 from enterotoxigenic *Escherichia coli*. *PLoS One.* **2009**;4(2):e4487; Epub 2009/02/27. PubMed PMID: 19242561; PubMed Central PMCID: PMC62647841;
 - [21] Sabui S, Ghosal A, Dutta S, et al. Allelic variation in colonization factor CS6 of enterotoxigenic *Escherichia coli* isolated from patients with acute diarrhoea and controls. *J Med Microbiol.* **2010**;59(Pt 7):770–779. Epub 2010/03/20. PubMed PMID: 20299505
 - [22] Wajima T, Sabui S, Fukumoto M, et al. Enterotoxigenic *Escherichia coli* CS6 gene products and their roles in CS6 structural protein assembly and cellular adherence. *Microb Pathog.* **2011**;51(4):243–249. Epub 2011/07/07. PubMed PMID: 21729748
 - [23] Svoboda P, Di Cara A. Hairpin RNA: a secondary structure of primary importance. *Cell Mol Life Sci.* **2006**;63(7–8):901–908. Epub 2006/03/29. PubMed PMID: 16568238
 - [24] Jordi BJ, op den Camp IE, de Haan LA, et al. Differential decay of RNA of the CFA/I fimbrial operon and control of relative gene expression. *J Bacteriol.* **1993**;175(24):7976–7981. Epub 1993/12/01. PubMed PMID: 7504669; PubMed Central PMCID: PMC6206977;
 - [25] Barry EM, Levine MM. A tale of two bacterial enteropathogens and one multivalent vaccine. *Cell Microbiol.* **2019**;21(11):e13067. Epub 2019/06/14. PubMed PMID: 31194282; PubMed Central PMCID: PMC6842089;
 - [26] Evans DG, Evans DJ Jr., DuPont HL. Virulence factors of enterotoxigenic *Escherichia coli*. *J Infect Dis.* **1977**;136(Suppl):S118–23. Epub 1977/08/01. PubMed PMID: 330772;
 - [27] Evans DG, Evans DJ Jr., Tjoa W. Hemagglutination of human group A erythrocytes by enterotoxigenic *Escherichia coli* isolated from adults with diarrhea: correlation with colonization factor. *Infect Immun.* **1977**;18(2):330–337; Epub 1977/11/01. PubMed PMID: 336541; PubMed Central PMCID: PMC621235;
 - [28] Hazen TH, Nagaraj S, Sen S, et al. Genome and functional characterization of colonization factor antigen I- and CS6-encoding heat-stable enterotoxin-only enterotoxigenic *Escherichia coli* reveals lineage and geographic variation. *mSystems.* **2019**;4(1):doi: 10.1128/mSystems.00329-18. PubMed PMID: 30944874; PubMed Central PMCID: PMC6446980. Epub 2019/04/05.
 - [29] Barry EM, Altboum Z, Losonsky G, et al. Immune responses elicited against multiple enterotoxigenic *Escherichia coli* fimbriae and mutant LT expressed in attenuated *Shigella* vaccine strains. *Vaccine.* **2003**;21(5–6):333–340. Epub 2003/01/18. PubMed PMID: 12531629
 - [30] Inouye S, Inouye M. Up-promoter mutations in the *lpp* gene of *Escherichia coli*. *Nucleic Acids Res.* **1985**;13(9):3101–3110. Epub 1985/05/10. PubMed PMID: 3923441; PubMed Central PMCID: PMC62341223;
 - [31] McKenzie GJ, Craig NL. Fast, easy and efficient: site-specific insertion of transgenes into enterobacterial

- chromosomes using Tn7 without need for selection of the insertion event. *BMC Microbiol.* **2006**;6(1):39. Epub 2006/05/02. PubMed PMID: 16646962; PubMed Central PMCID: PMCPMC1475584;
- [32] Loh JT, Lin AS, Beckett AC, et al. Role of a stem-Loop structure in helicobacter pylori cagA transcript stability. *Infect Immun.* **2019**;87(2). doi: [10.1128/IAI.00692-18](https://doi.org/10.1128/IAI.00692-18). Epub 2018/12/05. PubMed PMID: 30510104; PubMed Central PMCID: PMCPMC6346133.
- [33] Caron J, Scott JR. A rns-like regulatory gene for colonization factor antigen I (CFA/I) that controls expression of CFA/I pilin. *Infect Immun.* **1990**;58(4):874–878. Epub 1990/04/01. PubMed PMID: 1969396; PubMed Central PMCID: PMCPMC258554;
- [34] Bodero Md, Munson GP. The virulence regulator Rns activates the expression of CS14 Pili. *Genes (Basel)*. **2016**;7(12):120. Epub 2016/12/13. PubMed PMID: 27941642; PubMed Central PMCID: PMCPMC5192496;
- [35] Emory SA, Bouvet P, Belasco JG. A 5'-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*. *Genes Dev.* **1992**;6(1):135–148. Epub 1992/01/01. PubMed PMID: 1370426;
- [36] Schlor S, Riedl S, Blass J, et al. Genetic rearrangements of the regions adjacent to genes encoding heat-labile enterotoxins (eltAB) of enterotoxigenic *Escherichia coli* strains. *Appl Environ Microbiol.* **2000**;66(1):352–358. Epub 2000/01/05. PubMed PMID: 10618247; PubMed Central PMCID: PMCPMC91829;
- [37] Medeiros P, Bolick DT, Ledwaba SE, et al. A bivalent vaccine confers immunogenicity and protection against *Shigella flexneri* and enterotoxigenic *Escherichia coli* infections in mice. *NPJ Vaccines.* **2020**;5(1):30. Epub 2020/04/08. PubMed PMID: 32257392; PubMed Central PMCID: PMCPMC7101394
- [38] Altboum Z, Levine MM, Galen JE, et al. Genetic characterization and immunogenicity of coli surface antigen 4 from enterotoxigenic *Escherichia coli* when it is expressed in a *Shigella* live-vector strain. *Infect Immun.* **2003**;71(3):1352–1360. Epub 2003/02/22. PubMed PMID: 12595452; PubMed Central PMCID: PMCPMC148885;
- [39] Bikard D, Loot C, Baharoglu Z, et al. Folded DNA in action: hairpin formation and biological functions in prokaryotes. *Microbiol Mol Biol Rev.* **2010**;74(4):570–588. Epub 2010/12/02. PubMed PMID: 21119018; PubMed Central PMCID: PMCPMC3008174;
- [40] Qadri F, Akhtar M, Bhuiyan TR, et al. Safety and immunogenicity of the oral, inactivated, enterotoxigenic *Escherichia coli* vaccine ETVAX in Bangladeshi children and infants: a double-blind, randomised, placebo-controlled phase 1/2 trial. *Lancet Infect Dis.* **2020**;20(2):208–219. Epub 2019/11/24. PubMed PMID: 31757774; PubMed Central PMCID: PMCPMC6990395