CRISPR-Cas system presents multiple transcriptional units including antisense RNAs that are expressed in minimal medium and upregulated by pH in Salmonella enterica serovar Typhi

Liliana Medina-Aparicio, Javier E. Rebollar-Flores, América A. Beltrán-Luviano, Alejandra Vázquez, Rosa M. Gutiérrez-Ríos, Leticia Olvera, Edmundo Calva and Ismael Hernández-Lucas*

Abstract

The CRISPR-Cas system is involved in bacterial immunity, virulence, gene regulation, biofilm formation and sporulation. In Salmonella enterica serovar Typhi, this system consists of five transcriptional units including antisense RNAs. It was determined that these genetic elements are expressed in minimal medium and are up-regulated by pH. In addition, a transcriptional characterization of cas3 and asce2-1 is included herein.

INTRODUCTION

Prokaryotes have developed strategies to survive phage invasion, including the CRISPR-Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated proteins) system that provides specific heritable immunity against these invaders [1, 2]. The expression of CRISPR-cas generates small RNAs (crRNAs) that, by homology, recognize genetic material of phage and facilitate cleavage of the invader by Cas proteins [3, 4]. These systems are classified according to the presence of signature Cas proteins [5]. The hallmark of the CRISPR-Cas Type I system is the presence of Cas3, which contains an N-terminal ssDNA nuclease and a C-terminal DExH helicase domains [6, 7]. This protein is involved in cleavage of exogenous target nucleic acids [7, 8]. The Type II system requires Cas9 and a trans-activating CRISPR RNA (tracrRNA) for DNA recognition and degradation [9]. Based on its few requirements, this system has been used for genomic engineering in prokaryotes and eukaryotes [10]. The Type III system uses a complex composed of the RAMP proteins and Cas10 nuclease to silence the invader [11, 12].

The CRISPR-Cas systems are widely distributed in bacteria [13] and are also involved in virulence [14, 15], fruiting body development [16], swarming motility, biofilm formation [17] and response to stress conditions such as the perturbation of the bacterial envelope [18–20] or exposure to γ-irradiation [21].

Salmonella enterica serovar Typhi (S. Typhi) IMSS-1, a clinical strain that causes typhoid fever, contains a Type I-E CRISPR-Cas cluster composed of cas3, cse1-cse2-cas7-cas5-cas6e-cas1-cas2, an 84 bp leader sequence, seven 29 bp repeats and six 32 bp spacers [22]. The conserved genetic organization of the cas genes in some Salmonella serovars is consistent with its having a biological function in these bacteria [23]. Interestingly, the expression of cas7 (STY3068) has been detected when S. Typhi infects macrophages [24]. Additionally, we previously reported the transcriptional expression of cse1 from this bacterium in N-minimal medium (N-MM) [22], which promotes the expression of the Salmonella pathogenicity island 2 genes [25]. Furthermore, the S. Typhi CRISPR-Cas locus is regulated by LeuO, H-NS and LRP [22, 26], which are involved in pathogenesis [27–32]. A role of H-NS and LeuO in the transcriptional regulation of the Type I-E CRISPR-Cas system from Escherichia coli K12 has also been reported [33, 34]. Since some evidence for the activity of the CRISPR-Cas system in virulence has been reported, we evaluated the expression of the S. Typhi IMSS-1 CRISPR-Cas locus in N-MM,
finding five transcriptional units. Two of them, the cse1-cse2-cas7-cas5-cas6e-cas1-cas2-CRISPR (cas-CRISPR operon) and cse2 (sense cse2 RNA), are located in the sense strand, whereas asc2-1 (antisense RNA of cse2 to cse1) and ascas2-1 (antisense RNA of cas2 to cas1) are present on the antisense strand. Additionally, the cas3 gene is transcribed as an independent unit divergent to the cas-CRISPR operon. Interestingly, these five genetic elements showed higher transcriptional expression at pH 7.5, a pH value characteristic of the human terminal ileum [35].

In this work, we also report that H-NS represses asc2e-1 and cas3 expression and that a 13 nt sequence (5’-TAATCCAGACAAA-3’) is involved in cas3 positive control.

METHODS

Bacterial strains, plasmids and culture conditions

The bacterial strains and plasmids used in this work are listed in Table S1 (available in the online Supplementary Material). S. Typhi [36] and E. coli strains were grown aerobically at 37°C in LB (10 g tryptone, 5 g yeast extract and 10 g NaCl per litre), MA (7 g nutrient broth, 1 g yeast extract, 2 ml glycerol, 3.75 g K2HPO4 and 1.3 g KH2PO4 per litre) [37] or N-MM (pH 7.5) medium. The pH 7.5 of regular N-MM containing 100 mM Tris/HCl decreases to 7.0 at OD600 of 1.3; therefore, for pH experiments, 200 mM Tris/HCl was used to avoid pH changes during bacterial growth. When required, the following antibiotics were added: kanamycin, 30 µg ml⁻¹; tetracycline, 12 µg ml⁻¹; and ampicillin, 200 µg ml⁻¹.

DNA and RNA manipulations

Plasmid and genomic DNA isolations were carried out according to published protocols [38]. Primers for PCR amplifications were provided by the Oligonucleotide Synthesis Facility at our institute (Table S2). Restriction enzymes, ligase, nucleotides and polymerases were acquired from New England Biolabs, Invitrogen or Thermo Scientific. Each mutation was further characterized by sequencing to verify the authenticity of the deletion.

Construction of transcriptional reporter constructs

For transcriptional cat constructs, oligonucleotides (see Table S2) were designed to amplify DNA fragments of different lengths. PCR products were double-digested with BamHI-KpnI, BamHI-HindIII or BamHI-XhoI and ligated into pKK232-8 or pKK232-9 (Table S1), which contains the promoterless cat gene. For transcriptional lacZ reporter constructs, PCR fragments were amplified, digested with BamHI and cloned into pRS415 plasmid [41] (Table S1). All constructs were sequenced to verify the correct DNA sequence of the PCR fragments.

Chloramphenicol acetyltransferase assays

To determine the genetic expression of the cat reporter gene mediated by the S. Typhi promoters, chloramphenicol acetyltransferase (CAT) assays were performed according to Martinez-Laguna et al. [42]. Briefly, S. Typhi strains harbouring the reporters were grown in N-MM to different optical densities. Cells were harvested, centrifuged, washed with 0.8 ml of TDTT buffer [50 mM Tris/HCl and 30 µM DL-DTT (pH 7.8)], resuspended in 0.5 ml of TDTT and sonicated on ice for 9.9 s intervals with 9.9 s rest periods until the extract was clear. The homogenate was centrifuged, and the supernatant was used for activity measurement. For CAT assays, 5 µl of each extract was added in duplicate to a 96-well ELISA plate, followed by the addition of 0.2 ml of a reaction mixture containing 1 mM DTNB [5,5′-dithiobis(2-nitrobenzoic acid)], 0.1 mM acetyl-CoA and 0.1 mM chloramphenicol in 0.1 M Tris/HCl (pH 7.8). The absorbance at 412 nm was measured every 5 s for 5 min using a Ceres 900 scanning autoreader and microplate workstation. The protein concentration of the cell extracts was obtained using the bicinchoninic acid protein assay reagent (Pierce). Protein values and the mean rate of product formation by CAT were used to determine CAT-specific activity as micromoles per minute per milligram of protein.

β-Galactosidase assays

To determine the genetic expression of the lacZ gene mediated by S. Typhi promoters, β-galactosidase activity was measured by the Miller method adapted for a microtitre plate format as described by Oropeza et al. [43]. Protein concentration was determined by the Lowry method. The results of transcriptional reporters presented are the mean of three independent experiments performed in duplicate.

Primer extension analysis

To identify transcriptional start sites, primer extension experiments were performed as follows. Forty microgram of total RNA was denatured at 65°C for 5 min and then annealed with the corresponding [γ-32P] ATP-labelled primers. Primers were extended with reverse transcriptase at 50°C for 30 min as indicated in the Maxima H Minus Strand cDNA Synthesis Kit (Thermo Scientific) protocol.
The extended products were precipitated and analysed by electrophoresis in 8% polyacrylamide/8 M urea gels alongside sequencing ladders [43].

RT-PCR

Experiments were performed according to Villarreal et al. [44]. Briefly, 20 µg of total RNA was treated with DNase I (Thermo Scientific), and 1 µg of RNA (DNA-free as assessed by PCR) was used for cDNA synthesis with specific primers as established in the RevertAid H Minus First Strand cDNA Synthesis kit protocol (Thermo Scientific). Finally, PCR reactions were performed using 2 µl of cDNA as a template, and the reaction products were analysed in 1.0% or 1.8% agarose gels. For all the RT-PCR experiments, a cDNA synthesis negative control reaction without reverse transcriptase was carried out.

Electrophoretic mobility shift assays

The H-NS protein was purified as described previously in De la Cruz et al. [45] and Medina-Aparicio et al. [22] and non-radioactive electrophoretic mobility shift assays (EMSA) were performed. The probes were obtained by PCR using primers described in Table S2. Each probe (100 ng) was mixed with increasing concentrations of purified protein in the presence of H-NS-binding buffer (40 mM HEPES, 8 mM MgCl₂, 50 mM KCl, 1 mM DTT, 0.05% NP40 and 0.1 mg ml⁻¹ BSA). The mixture was incubated for 20 min at room temperature and then separated by native 6% PAGE in 0.5× Tris/borate/EDTA buffer. The DNA bands were visualized by ethidium bromide staining.

RESULTS

Negative and positive genetic elements involved in S. Typhi cas3 transcriptional expression

Several studies regarding Cas3 structure, from other bacteria, and its enzymatic mechanism are available [46–48]. However, studies on its transcriptional regulation in Salmonella have not been performed. Therefore, a cas3 transcriptional characterization in S. Typhi is reported here.

Previously, we have determined that the cse1 promoter is repressed by the global regulators H-NS and LRP whereas LeuO participates in its positive regulation; also, it can be expressed independently of LeuO in N-MM [22]. In S. Typhi, cas3 is divergent to cse1, sharing a 5′ intergenic regulatory region of 357 bp (Fig. 1a). Therefore, we evaluated cas3 expression in N-MM. The pKK cas3 (−455/+112) transcriptional reporter construct, containing 455 bp upstream and 112 bp downstream of the cas3 translational start site, was evaluated in the S. Typhi IMSS-1 wild-type. cas3 expression was of 101 CAT units (Fig. 1a). This low transcriptional activity suggests that cas3 promoter is repressed in N-MM. Thus, to identify cas3 regions involved in its repression, transcriptional reporters encompassing different lengths of the cas3 regulatory region were analysed in S. Typhi grown in N-MM to OD₅₉₅ 1.3. Fig. 1(a) shows that constructs pKK cas3 (−241/+112), pKK cas3 (−180/+112) and pKK cas3 (−83/+112) had activities of 293, 934 and 1306 CAT units, respectively. These values were higher than those obtained with the longest reporter pKK cas3 (−455/+112) (101 CAT units), indicating that nucleotides −455 to −82 are involved in cas3 negative regulation. Additional constructs downstream of this cas3 repressing region, pKK cas3 (−83/+38), pKK cas3 (−83/−1) and pKK cas3 (−83/−25), produced 1553, 2393 and 4567 CAT units, respectively (Fig. 1a), showing that the DNA region between −24 and +112 is also implicated in cas3 repression. Thus, two regions (−455 to −82 and −24 to +112 relative to the cas3 translational start codon) contain DNA information involved in cas3 silencing, indicating that cas3 is negatively regulated downstream and upstream of its protein start codon.

In order to determine the transcriptional factors responsible for cas3 negative regulation, its transcriptional activity was assessed in individual mutants lacking the H-NS [49], LRP, CRP, CpxR, SsrB, IHF, PhoP, PmrA, FNR or OmpR regulators, particularly since some of them (H-NS or CRP) have been previously implicated in the regulation of other CRISPR-Cas systems [33, 50, 51] or in virulence of Salmonella (CpxR, SsrB, IHF, PhoP, PmrA, FNR or OmpR) [52–58]. We found that only H-NS represses cas3 expression (Fig. 1a), since an increase of CAT activity using pKK cas3 (−455/+112) (627 units) was observed in the hns-deficient strain (STYhns99) with respect to the wild-type (101 CAT units, Fig. 1a). To further characterize H-NS negative regulation, we evaluated the CAT activity of the cas3 transcriptional reporter constructs in the STYhns99 strain. The experiments showed higher activity with pKK cas3 (−241/+112) (697 CAT units) compared to the expression in the wild-type (293 CAT units). However, with plasmid pKK cas3 (−180/+112), similar CAT levels (780–930 units) were observed in the wild-type and the STYhns99 strain (Fig. 1a), indicating that H-NS represses cas3 expression by acting between nucleotides −455 and −181 relative to its ATG codon. These data are in agreement with the previous report that demonstrated the interaction of H-NS with this region in order to also repress the cas3-divergent cse1 gene [22]. The results obtained also showed that H-NS-independent regulation involves bases −180 to +112, since in the wild-type background, the CAT values of this construct were nine times greater than the pKK cas3 (−455/+112) reporter (Fig. 1a).

The data on cas3 repression suggest that nucleotides −241 to −181, upstream of the cas3 translational start site, are more relevant for negative regulation than those at −455 to −240, since CAT activity levels were higher, in the wild-type strain, when the sequence from −241 to −181 [pKK cas3 (−180/+112)], was removed (Fig. 1a). Thus, it was determined whether these nucleotides corresponded to an H-NS-binding site for cas3 repression. EMSA using purified H-NS protein and two cas3 DNA fragments, cas3-A (nt −241 to +112) and cas3-B (nt −180 to +112), were performed. As expected, H-NS had more affinity to cas3-A region than for cas3-B, validating that this nucleoid protein directly represses cas3 expression through binding at bases −241 to −181 (Fig. 1b).

Interestingly, the pKK cas3 (−83/−25) plasmid, which lacks negative regulatory regions upstream and downstream of...
**Fig. 1.** Genetic characterization of cas3. (a) The cas3 gene is divergent to cse1 and they share a 5' intergenic region of 357 bp (thin line). The bent arrow represents the cas3 transcriptional start sites located 28 bp upstream of the cas3 translational start codon. Nucleotides −83 to −71 (5'-TAATCCAGAAA-3') are involved in cas3 positive regulation. The hatched rectangle indicates the H-NS repression site located between nucleotides −241 and −181. Below the diagram, CAT transcriptional constructs of different length are shown. They were named (boldface) according to the nucleotides contained in each construct; the numbers below describe the coordinates upstream and downstream relative to the transcriptional start site. Primer extension experiments were performed with RNA from S. Typhi IMSS-1 grown in N-MM (100 mM Tris/HCl) to OD$_{699}$ 1.3. The values are the means±standard deviations of three independent experiments performed in duplicate. (b) Binding of H-NS to cas3. EMSA with increasing concentrations of purified H-NS-Myc-His6 and the cas3 positive regulation. The hatched rectangle indicates the H-NS protein complex. (c) Mapping of cas3 transcriptional start site. Primer extension experiments were performed with RNA from S. Typhi IMSS-1 harbouring pKK cas3 (−83/−25) grown in N-MM (100 mM Tris/HCl) to OD$_{699}$ 1.3. The oligonucleotide of the vector pKK232-8 (pKK-R) was used to synthesize the cDNA which was resolved in an 8 % polyacrylamide gel alongside sequencing ladders. The location of the transcriptional start site at 28 bp upstream of the cas3 ATG start codon is shown in the sequence at the bottom of the figure (bent arrow); the −10 and −35 sequences (boldface) are also indicated. (d) Transcriptional reporter constructs used to validate the cas3 promoter. The pKK cas3 (−70/−25) contains the transcriptional start site (bent arrow) and the −10 and −35 sequences (boldface), whereas pKK cas3 (−83/−46) only comprises the −35 box. Both plasmids were evaluated in S. Typhi IMSS-1 wild-type strain grown in N-MM (100 mM Tris/HCl) to OD$_{699}$ 1.3 and the results of the CAT assays are shown in the left columns. The values are the means±standard deviations for three independent experiments performed in duplicate. <dl (<detection limit) represents values between 0 and 10 CAT units.
the cas3 ATG codon (−455 to −82 and −24 to +112), gave activity values of 4567 CAT units in the wild-type strain, whereas in the STYhns99, the expression levels with pKK

\textit{cas3} (−455/+112) were only of 627 CAT units (Fig. 1a). This suggests that there is another negative regulator for cas3 besides H-NS.

As a part of the transcriptional characterization of \textit{cas3}, an evaluation of the \textit{cis} genetic elements involved in its positive expression was also performed. First, its transcriptional start site was determined by primer extension, using RNA from S. Typhi IMSS-1 harbouring plasmid pKK \textit{cas3} (−83/−25), and grown in N-MM to OD\textsubscript{950} 1.3. The results showed a transcriptional start site localized 28 bp upstream of the \textit{cas3} translational start codon, with consensus −10 and −35 sequences corresponding to the σ\textsubscript{70} factor (Fig. 1c). To validate the \textit{cas3} promoter region, we constructed pKK \textit{cas3} (−70/−25) containing the transcriptional start site and the −10 and −35 sequences and pKK \textit{cas3} (−83/−46) lacking the transcriptional start site and the −10 box. Both plasmids were evaluated in S. Typhi grown in N-MM. CAT activity was obtained only with pKK \textit{cas3} (−70/−25) (760 units, Fig. 1d), validating the presence of the \textit{cas3} promoter between nucleotides −70 and −25 from its translational start codon. Although this construct contains the elements necessary for \textit{cas3} activity, the CAT values observed (760 CAT units) were six times lower than the expression levels of plasmid pKK \textit{cas3} (−83/−25) (4567 CAT units, see Fig. 1a), indicating that nucleotides −83 to −71 (5′-TATCCGACAAAA-3′) are relevant for its transcriptional expression in N-MM (Fig. 1a). These 13 nucleotides are adjacent to the −35 sequence, a characteristic location for the UP element, which corresponds to the binding site for RNA polymerase α-subunit and contains a high percentage of A/T [59].

EMSA performed with the region involved in \textit{cas3} positive regulation and purified α-subunit showed no interaction between these two elements (data not shown), suggesting that the 5′-TATCCGACAAAA-3′ nucleotides could be a binding site for a transcriptional activator. Since LeuO positively regulates the \textit{cas3}-divergent \textit{ces1} gene, we evaluated the role of this LysR regulator in \textit{cas3} expression. The results showed that LeuO overexpression did not promote \textit{cas3} transcriptional activity (data not shown).

In summary, we have identified the presence of negative regulatory elements upstream and downstream of the \textit{cas3} translational start site. H-NS participates in its silencing by interacting with nucleotides −241 to −181 from the \textit{cas3} translational initiation site. Additionally, 13 nucleotides located from −83 to −71 are involved in \textit{cas3} positive regulation.

The \textit{sense} strand of the \textit{S. Typhi} CRISPR-Cas locus encodes two transcriptional units: the \textit{cse1-cse2-cas7-cas5-cas6e-cas1-cas2-CRISPR} operon and the \textit{scse2} RNA

Previously, we showed that there was minimal expression of the CRISPR-Cas system in rich medium and that LeuO overexpression induces transcriptional activity of the \textit{cse1-cse2-cas7-cas5-cas6e-cas1-cas2-CRISPR} operon (\textit{cas-CRISPR}) in this condition [22]. Here, we evaluated by RT-PCR whether the \textit{ces1} promoter drives the genetic expression of the complete \textit{cse1-cse2-cas7-cas5-cas6e-cas1-cas2-CRISPR} locus as a polycistronic RNA in N-MM. Total RNA, isolated from the wild-type \textit{S. Typhi} strain grown in this medium to OD\textsubscript{950} 1.3, was used for synthesising cDNA with a specific CRISPR oligonucleotide (CRISPR +503 Xhol-R) located at the 3′ end of the locus (see Table S2). Using this cDNA and specific primers, the fragments \textit{cse1}, \textit{cse2-cas7}, \textit{cas7-cas5}, \textit{cas5-cas6e}, \textit{cas1-cas2} and \textit{cas2-CRISPR} were amplified by PCR (Fig. 2a). The results demonstrated that these genes together with the CRISPR sequences comprise an operon in N-MM.

To further support these data, a transcriptional reporter containing the \textit{ces1} promoter, the \textit{cse1-cse2-cas7-cas5-cas6e-cas1-cas2} genes and the CRISPR sequences, together with independent constructs of each 5′ intergenic \textit{cas} regions were obtained and transformed into \textit{S. Typhi} IMSS-1. The results in N-MM at OD\textsubscript{950} 1.3 showed, as expected, that the construct encompassing the \textit{ces1} promoter and the \textit{cse1-cse2-cas7-cas5-cas6e-cas1-cas2-CRISPR} region (pKK9 \textit{ces1RR-CRISPR}) gave CAT activity of 173, indicating that this promoter is able to drive the expression of the downstream \textit{cas} genes together with the CRISPR sequences (Fig. 2b). Moreover, no expression with the independent \textit{ces1} coding region and with the independent \textit{cas7}, \textit{cas5}, \textit{cas6e}, \textit{cas1}, \textit{cas2} and CRISPR 5′ intergenic regions was obtained (Fig. 2b). Interestingly, with the \textit{cse2} reporter pKK \textit{cse2} (−450/+181), containing 450 bp upstream and 181 bp downstream of the \textit{cse2} ATG codon, CAT activity values of 619 were observed, suggesting the presence of a promoter within this region (Fig. 2b). Additionally, nucleotides −450 to +181 were also fused to the \textit{lacZ} reporter gene and transcriptional activity of 1549.3±128.5 β-galactosidase units was detected (data not shown), supporting the notion that this region contains a promoter.

To determine the size of the \textit{cse2} transcriptional unit, several transcriptional reporters of different lengths were obtained and transformed into \textit{S. Typhi} IMSS-1. The CAT activity results in N-MM demonstrated that the shorter constructs pKK \textit{cse2} (−62/+181), pKK \textit{cse2} (+134/+181) and pKK \textit{cse2} (+142/+181) gave expression levels similar to those of the longest plasmid pKK \textit{cse2} (−450/+181) (Fig. 3a), showing that a promoter is present in the \textit{cse2} intragenic region, between nucleotides +142 and +181 relative to the \textit{cse2} ATG start codon. Surprisingly, the pKK \textit{cse2} (+142/+202) and pKK \textit{cse2} (+142/+267) were not expressed (Fig. 3a), indicating the presence of a terminator sequence downstream of nucleotide +181, an observation that also supports the difference observed in Fig. 2 (b) between the pKK9 \textit{ces1RR-CRISPR} and pKK \textit{cse2} (−450/+181) reporters. Collectively, these data indicate that a promoter is present between nucleotides +142 and +181 relative to the \textit{cse2} translational start site. This short region contains the elements necessary to induce \textit{scse2} expression. Bioinformatic analyses using MEME-MAST [60] showed the presence of
Fig. 2. The cse1-cse2-cas7-cas5-cas6e-cas1-cas2-CRISPR locus is expressed as an operon in N-MM. (a) Chromosomal RT-PCRs were performed with total RNA from S. Typhi IMSS-1 grown in N-MM (100 mM Tris/HCl) to OD_{595} 1.3. PCR fragments of the cse1 intragenic region (704 bp), cse2-cas7 (756 bp, containing the last 454 bp of cse2), a cse2-cas7 intragenic region of 16 bp and the first 286 bp of cas7), cas7-cas5 (1104 bp, encompassing the last 815 bp of cas7, cas7-cas5 intragenic region of 9 bp and the first 280 bp of cas5), cas5-cas6e (511 bp, harbouring the last 291 bp of cas5 and the first 220 bp of cas6e), cas1-cas2 (389 bp that contains the last 210 bp of cas1, cas1-cas2 intragenic region of 2 bp and the first 177 bp of cas2) and cas2-CRISPR (516 bp, including the last 13 bp of cas2, the 84 bp leader sequence and 419 bp of the CRISPR array) were generated using specific primers and, as template, a cDNA synthesized from genomic DNA molecular weight markers (MW) are included. (b) Genetic organization of the S. Typhi cas-CRISPR operon. The cse1 5’ intragenic region of 357 bp (regulatory region, RR) contains the promoter of cas-CRISPR unit; the bent arrow indicates the cse1 transcriptional start site located 94 bp upstream of its ATG codon. This polycistronic RNA is composed of cse1-cse2-cas7-cas5-cas6e-cas1-cas2 genes (the first five genes are also called csaA, csaB, csaC, csaD and csaE [3]), a leader sequence of 84 bp (grey rectangle) and the CRISPR array containing seven repeats (white rectangles) and six spacers (black diamonds). Below CRISPR-Cas diagram, CAT transcriptional constructs of the CRISPR-Cas locus are shown. These were named (boldface) according to the corresponding cas gene evaluated; the numbers below each line represent the upstream and downstream coordinates relative to the ATG codon of the respective ORF. For pKK9 cse1RR-CRISPR, the coordinates upstream and downstream are from the cse1 ATG codon; in the case of the CRISPR construct, the name and coordinates upstream and downstream are according to the first nucleotide of the CRISPR leader sequence. The right columns represent the expression of the CRISPR-Cas reporters evaluated in the wild-type S. Typhi IMSS-1 strain. CAT-specific activities were measured in N-MM (100 mM Tris/HCl) at OD_{595} 1.3. The values are the means±standard deviations for three independent experiments performed in duplicate; <dl (<detection limit) represents values between 0 and 10 CAT units.

−35 (GTGAA) and −10 (CCGCAATAG) consensus sequences for E. coli σ32 [61, 62], between nucleotides +142 and +181 (Fig. 3b), with an e-value of 10^{-6}. A transcriptional reporter lacking the −35 σ32 box was constructed, pKK cse2 (+154/+181), and evaluated in S. Typhi IMSS-1 grown in N-MM. Its CAT activity was eight times lower (80 CAT units) than the pKK cse2 (+142/+181) (664 CAT units, Fig. 3b), validating the presence of cse2 promoter in this region. Notably,
The first one encompasses part of the transcriptional units in the antisense strand of the CRISPR-Cas genes, the question was whether there were other transcriptional units in the antisense strand of the CRISPR-Cas operon, namely pKK cas2-1 (–450/+181) and cas3 since an RNA in the intragenic region of cas2 was found and cas3 is divergent to the cse1-cas2-cas5-cas6e-cas2-cas1 genes, the question was whether there were other transcriptional units in the antisense strand of the CRISPR-Cas locus, and hence, two independent promoters were found. The first one encompasses part of the cas2 and cas1 antisense strand (ascas2-1) since a transcriptional reporter containing this region (nucleotides –6245 to –5556 relative to the cas3 ATG start codon), namely pKK ascas2-1 (–6245/–5556), had activity values of 308 CAT units in S. Typhi IMSS-1 wild-type strain grown in N-MM (100 mM Tris/HCl) at OD620 1.3. (b) The cse2 promoter is present in nucleotides +142 to +181. The pKK cse2 (+142/+181) contains the −10 and −35 sequences (boldface), whereas the pKK cse2 (+154/+181) lacks the −35 box. Both plasmids were evaluated in S. Typhi IMSS-1 wild-type strain grown in N-MM (100 mM Tris/HCl) to OD620 1.3, and the results of the CAT assays are shown in the right columns. For all the CAT-specific activities, the values are the means ± standard deviations of three independent experiments performed in duplicate; <dl (<detection limit) represents values between 0 and 10 CAT units.

Fig. 3. Identification and delimitation of the cse2 transcriptional unit. (a) The cse2 gene is composed of 603 bp and localized downstream of cse1 with a 5′ intergenic region of 16 bp. The wavy arrow represents the cse2 RNA encompassing cse2 intragenic nucleotides +142 to +181. Below the diagram, CAT transcriptional constructs of different length are shown. They were named (boldface) according to the nucleotides contained in each cse2 construct; the numbers describing each reporter represent the coordinates upstream and downstream relative to the cse2 ATG start codon. The right columns represent the expression of the cse2 plasmids evaluated in the wild-type S. Typhi IMSS-1 strain. CAT-specific activities were measured in N-MM (100 mM Tris/HCl) at OD620 1.3. The right columns represent the expression of the cse2 plasmids evaluated in the wild-type S. Typhi IMSS-1 strain. CAT-specific activities were measured in N-MM (100 mM Tris/HCl) at OD620 1.3. The right columns represent the expression of the cse2 plasmids evaluated in the wild-type S. Typhi IMSS-1 strain. CAT-specific activities were measured in N-MM (100 mM Tris/HCl) at OD620 1.3. (b) The cse2 promoter is present in nucleotides +142 to +181. The pKK cse2 (+142/+181) contains the −10 and −35 sequences (boldface), whereas the pKK cse2 (+154/+181) lacks the −35 box. Both plasmids were evaluated in S. Typhi IMSS-1 wild-type strain grown in N-MM (100 mM Tris/HCl) to OD620 1.3, and the results of the CAT assays are shown in the right columns. For all the CAT-specific activities, the values are the means ± standard deviations of three independent experiments performed in duplicate; <dl (<detection limit) represents values between 0 and 10 CAT units.

Thus, in the sense strand of the S. Typhi CRISPR-Cas locus, the cse1 promoter drives the genetic expression of the cse1-cse2-cas7-cas5-cas6e-cas2-cas1-CRISPR operon, whereas another promoter controls the activity of the cse2 RNA.

The antisense strand of the S. Typhi CRISPR-Cas locus encodes two transcriptional units: asce2-1 and ascas2-1

Since an RNA in the intragenic region of cse2 was found and cas3 is divergent to the cse1-cse2-cas7-cas5-cas6e-cas2-cas1 genes, the question was whether there were other transcriptional units in the antisense strand of the CRISPR-Cas locus, and hence, two independent promoters were found. The first one encompasses part of the cas2 and cas1 antisense strand (ascas2-1)
The second promoter named as cse2-1 was identified between the cas2 and cse1 antisense strand, at −2090 to −1460 since a plasmid harbouring this region, pRS asc2-1 (−2090/−1460), rendered expression levels of 245 β-galactosidase units when it was evaluated in S. Typhi grown in N-MM (Fig. 4a). To define the asc2-1 transcriptional unit, RT-PCR experiments were performed. Total RNA purified from S. Typhi (N-MM, OD$_{595}$ 1.3) and the asc1 −34 BamHI-R oligonucleotide was used to generate the cDNA. For PCR amplification, this cDNA together with primers asc1 −34 BamHI-R and asc2 −1958 KpnI-F rendered a fragment of 1925 bp corresponding to the antisense asc2-1 (Fig. 4c).

To further characterize asc2-1 and asc2-1, it was determined whether H-NS represses its expression since this nucleoid protein negatively regulates other S. Typhi CRISPR-Cas
transcriptional units such as cas3 and the cas-CRISPR operon. Thus, the pKK ascas21 (−6245/−5556) and pRS ascas2-I (−2090/−1460) were evaluated in STYhns99 strain grown in N-MM. The transcriptional profiles showed that H-NS is not implicated in the control of ascas2-I since the CAT units obtained in STYhns99 were similar to those observed in the IMSS-1 wild-type strain (Fig. S1). However, the activity levels of the pRS ascas2-I (−2090/−1460) were higher in the H-NS-deficient strain (46 β-galactosidase units) than in the wild-type (245 β-galactosidase units) (Fig. 4a), showing that H-NS silences the expression of the antisense ascas2-I RNA. EMSA using purified H-NS protein and an a DNA ascas2-I fragment (nucleotides −2090 to −1460, 630 bp) demonstrated that this global regulator directly represses ascas2-I since it interacted with the DNA region evaluated (Fig. 4d). Thus, two transcriptional units in the antisense strand of the CRISPR-Cas loci were identified: ascas2-I and ascas2-I. Moreover, ascas2-I is negatively regulated by H-NS.

**pH regulates the genetic expression of the five transcriptional units present in the S. Typhi CRISPR-Cas system**

The expression of the CRISPR-Cas system has been previously shown to depend on environmental conditions [18, 19]. In addition, as described herein, there are transcriptional units on the sense strand (cas-CRISPR operon, sces2) and on the antisense strand (ascas2-I, ascas2-I) of the S. Typhi CRISPR-Cas locus that, together with cas3, are expressed in N-MM. Hence, in order to identify the specific signal in N-MM responsible for CRISPR-Cas induction, the influence of the carbon source (glycerol) present in this medium was tested by replacing it with glucose. Upon evaluation of the cse1 (cas-CRISPR promoter), sces2 and cas3 promoters, their expression was found to be similar in both carbon sources (data not shown) indicating that glycerol is not the signal that promotes CRISPR-Cas expression. Several reports show the induction of Salmonella pathogenicity genes in acid pH [63–65], so the influence of pH on CRISPR-Cas transcription was tested. The expression of cse1, sces2, ascas2-1, ascas2-I and cas3 promoters was relatively low at acid pH 6.0–6.5. However, their activity was higher in N-MM pH 7.5 and 7.7 (Fig. 5a). To confirm that, in fact, pH determines this increase, the five transcriptional units were evaluated in rich medium A (MA) at pH 6.0, 6.5, 7.0, 7.5 and 7.7. Even though this medium is significantly different from N-MM, the expression of cse1, sces2, ascas2-I, ascas2-I and cas3 also increased at pH 7.5–7.7 (Fig. S2). In the case of cse1, sces2-1 and cas3, their transcriptional levels were higher in N-MM than those obtained in MA (Figs 5a and S2) illustrating a significant effect of N-MM and pH. On the other hand, the expression of the sces2 and ascas2-I promoters in N-MM and MA was similar, although the effect of pH was maintained (Figs 5a and S2).

Additionally, primer extension experiments to determine the transcriptional start site of the ascas2-I unit were performed, by using total RNA from S. Typhi IMSS-1 harbouring the pRS ascas2-I plasmid grown in N-MM at pH 6.5 and 7.5. For generating the cDNA, we used the ascse1 –1766 BamHI-R oligonucleotide, located at nucleotide –1766 relative to cas3 translational initiation site. The results showed an ascas2-I transcriptional start site at 2062 bp upstream of the cas3 ATG codon, whose transcription increased at pH 7.5 compared to pH 6.5 (Fig. 5b).

Since all the transcriptional units identified were up-regulated at pH 7.5, we evaluated their role under this condition. Growth rate experiments comparing a strain devoid of the entire CRISPR-Cas locus with the wild-type strain were performed in N-MM at pH 7.5 and 7.7. The results showed the same growth rate in exponential and stationary phase in both strains, thus indicating that the CRISPR-Cas system is not necessary for S. Typhi replication in N-MM at pH 7.5–7.7 (data not shown).

**DISCUSSION**

The CRISPR-Cas system has mainly been studied for its function in bacterial immunity, and recently, its role in virulence [14], biofilm formation [17] and gene regulation [15] has been documented. However, there are relatively few studies on its transcriptional regulation and the signals that promote its expression. We previously reported that H-NS and LRP silence the cse1 promoter whereas the LysR-type regulator LeuO induces its expression in rich medium [22, 26]. In this study, we performed a genetic characterization of five transcriptional units present in the S. Typhi CRISPR-Cas locus: the cas-cas-CRISPR operon, sces2-cas2-1, casas2-I and cas3. These genetic elements were expressed in N-MM and their activity increased from pH 7.5 to 7.7.

Our data support the view that the CRISPR-Cas system in S. Typhi is coded by a complex operon, since it harbours specific independent transcriptional units not previously identified. Furthermore, regulation studies of the CRISPR-Cas Type I-E system present in E. coli K12 support our data, since this locus presents diverse transcriptional units: one containing the cas3 gene; a second unit comprises the cse1-cse2-cas7-cas5-cas 6-cas1-cas2 genes; the CRISPR loci are transcribed from a promoter present in the leader sequence; and an antisense promoter (anti-Pcas) was also identified in the 414 bp intergenic region between cas3 and cse1 [33, 66, 67]. In Sulfolobus islandicus, it has also been demonstrated that the Type I-A CRISPR-Cas system is transcribed as two units: the cas1-cas1 and the acas (cas1-cas2-cas4) operons [68]. Thus, all these works support the notion that multiple transcriptional units on sense and antisense strands may be characteristic of the CRISPR-Cas Type I systems of bacteria and archaea.

The transcriptional characterization of the S. Typhi cas3 regulatory region is also presented here. This nuclease–helicase protein has an essential role in bacterial immunity mediated by functional CRISPR-Cas Type I systems, since its HD nuclease domain degrades the invader genetic element [7, 8] and recently it has been implicated in the acquisition of new spacer sequences [69, 70]. We have thus demonstrated that, in
S. Typhi, cas3 is negatively regulated by the global regulatory protein H-NS, which also silences cse1 [26] and the antisense as

case2-1. Interestingly, in the CRISPR-Cas system from E. coli K12, H-NS binds to Pcas, anti-Pcas and Pcrispr1 promoters, silencing their expression [33]. Thus, in E. coli and S. Typhi, H-NS regulates more than one CRISPR-Cas transcriptional unit, suggesting that, in these bacteria, the CRISPR-Cas is silenced in multiple steps.
Although the divergently transcribed cse1 and cas3 genes share a 5' intergenic region, we found that LRP and LeuO only control cse1 expression. These data showed that cas3 is transcriptionally regulated by other factors or signals that maintain its expression at low levels in N-MM.

An evaluation to determine transcriptional factors involved in cas3 regulation, as well as those in cas-CRISPR operon, sce2 and ascas2-1, led to the finding that the CpxR, SsrB, IHF, PhoP, PmrA, FNR, CRP, LRP and OmpR regulators are not involved in the transcriptional control of these transcriptional units (Fig. S1). The role of H-NS in the regulation of sce2 and ascas2-1 was also evaluated but no effect was observed (Fig. S1). The proteins evaluated include transcriptional factors implicated in virulence [52–58], or proteins involved in regulation of other CRISPR-Cas systems such as cAMP receptor protein (CRP), whose role in positive regulation of Thermus thermophilus and Pectobacterium atrosepticum CRISPR-Cas systems [50, 71] or in the repression of E. coli CRISPR-Cas locus [51] has been established. Therefore, the lack of CRP participation in the control of S. Typhi CRISPR-Cas transcriptional units illustrates the differences in the transcriptional regulation of this system [22, 71].

The expression of the transcriptional units defined in this work increases at pH 7.5, a pH value characteristic of the distal ileum [35], a compartment that is invaded and colonized by S. Typhi. Thus, it is relevant to determine whether these genetic elements have a role in the response to the environment present in the gastrointestinal tract or directly participate in the invasion of this compartment. Although it is possible that the CRISPR-Cas systems would offer a mechanism for adaptation to environment changes, experiments of phage resistant or plasmid acquisition in conditions where the S. Typhi CRISPR-Cas locus is expressed are necessary to determine whether it participates in bacterial immunity.

Finally, the S. Typhi CRISPR-Cas system could be involved in other aspects of the pathogenesis and is a matter of future studies. For instance, the Type II CRISPR-Cas system of the human pathogen Francisella novicida downregulates the expression of bacterial lipoprotein, promoting the evasion of the host immune system [15]. In Campylobacter jejuni and Neisseria meningitidis, the Cas9 protein, a component of the CRISPR-Cas system, is necessary for intracellular survival of these bacteria in human epithelial cells from intestine and lung, respectively [14, 15].

In summary, we have characterized five transcriptional units present in the S. Typhi CRISPR-Cas system. In Fig. 6, all the known elements implicated in the transcriptional control of S. Typhi CRISPR-Cas locus are shown, providing an overview of its regulation.

**Fig. 6.** Transcriptional organization and regulation of the S. Typhi CRISPR-Cas system. Wavy grey arrows represent mRNAs of cas-CRISPR and cas3, whereas the RNA sce2, the antisense ascse2-1 and ascas2-1 are shown as wavy black arrows. We present all the elements implicated in the positive (+) and negative (−) regulation of each transcriptional unit from the S. Typhi CRISPR-Cas locus that have been identified in this work, as well as by Hernández-Lucas *et al.* [26] and Medina-Aparicio *et al.* [22] including the LysR-type regulator LeuO, LRP (leucine-response protein), H-NS (heat-stable nucleoid-structuring protein), N-MM (N-minimal medium) and pH.

**Funding information**

This work was supported by grants from Dirección General de Asuntos de Personal Académico, DGAPA/UNAM IN203312 and IN203215 to I. H.-L.; IN201513 to E. C.) and Consejo Nacional de Ciencia y Tecnología, CONACYT, México (89337 and 127298 to I. H.-L.; 179946 to E. C.)

**Acknowledgements**

We would like to thank N. Becerra-Lobato, M. L. Zavala-García, V. E. Osio-Becerro, C. Hernández, M. F. Mora, F. J. Santana, P. Gaytan, E. Bustos, S. Becerra and J. Yañez (IBT-UNAM) for technical help and J. M. Villarreal, A. Medrano-López, G. Gosset (IBT-UNAM); J.
Miranda (IIB-UNAM), and M. Dunn (CCG-UNAM) for stimulating discussions and critical reading.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References


Edited by: R. Lan and A. Achouak