UNIVERSITY OF BIRMINGHAM
University of Birmingham Research at Birmingham

Organization of the LEE1 operon regulatory region of enterohaemorrhagic Escherichia coli O157:H7 and activation by GrlA.

Islam, Md; Bingle, LE; Pallen, Mark; Busby, Stephen

DOI:
10.1111/j.1365-2958.2010.07460.x

Citation for published version (Harvard):
Islam, M, Bingle, LE, Pallen, M & Busby, S 2011, 'Organization of the LEE1 operon regulatory region of enterohaemorrhagic Escherichia coli O157:H7 and activation by GrlA.', *Molecular Microbiology*, vol. 79, no. 2, pp. 468-83. https://doi.org/10.1111/j.1365-2958.2010.07460.x

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

• Users may freely distribute the URL that is used to identify this publication.
• Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
• User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
• Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 19. May. 2021

Molecular Microbiology vol 79 pp468-483 (2011)

Organisation of the *LEE1* operon regulatory region of enterohaemorrhagic *Escherichia coli* O157:H7 and activation by GrlA

Md. Shahidul Islam, Lewis E. H. Bingle, Mark J. Pallen and Stephen J. W. Busby*

School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

*for correspondence s.j.w.busby@bham.ac.uk

Keywords: Enterohaemorrhagic *E. coli*, Locus of enterocyte effacement, *LEE1* operon regulatory region, GrlA transcription factor, Transcription activation, MerR family

Running title: Action of GrlA at the *LEE1* regulatory region

Summary

Expression of the genes in the locus of enterocyte effacement (LEE) in enterohaemorrhagic *Escherichia coli* is primarily coordinated by expression of the *LEE1* operon. GrlA is a LEE-encoded transcription regulator that has been proposed to be involved in the regulation of expression of the *LEE1* operon. We describe a simple plasmid-based system to investigate the *LEE1* operon regulatory region and to study GrlA-dependent effects. We report that GrlA can activate transcription initiation at the *LEE1* P1 promoter by binding to a target located within the 18 base pair spacer between the promoter -10 and -35 elements, which were defined by mutational analysis. Shortening this spacer to 17 base pairs increases P1 promoter activity and short-circuits GrlA-dependent activation. Hence, at the P1 promoter, the action of GrlA resembles that of many MerR family transcription activators at their target promoters.

Introduction

Enterohaemorrhagic *Escherichia coli* (EHEC) serotype O157:H7 causes haemorrhagic colitis and haemolytic uremic syndrome in humans, exploiting virulence determinants that include a type III secretion system. Effector molecules are translocated through the type III secretion system, and this causes the formation of attaching and effacing lesions in intestinal epithelial cells (Nataro and Kaper, 1998; Bhavsar *et al*., 2007). The genes responsible for this phenotype are contained in a pathogenicity island termed the locus of enterocyte effacement (LEE) (McDaniel *et al*., 1995; Perna *et al*., 1998). The LEE genes are shared by many bacterial strains including enteropathogenic *E. coli* (Elliott *et al*., 1998) and *Citrobacter rodentium* (Schauer and Falkow, 1993). The LEE contains approximately 40 protein coding sequences that are transcribed in five major polycistronic operons (designated *LEE1*-*5*) and several smaller transcriptional units (Mellies *et al*., 1999). LEE gene expression is controlled by a complex network of signals and transcription factors (reviewed by Mellies *et al.*, 2007; Tree *et al*., 2009). The 'master' regulator of the system is the Ler protein (the LEE-encoded regulator encoded by the *ler* gene) that acts as the central activator for expression of most of the LEE genes (Mellies *et al*., 1999; Bustamante *et al*., 2001; Umanski *et al*., 2002; Haack *et al.*, 2003; Stoebel *et al.*, 2008).

The *ler* gene is the first cistron of the *LEE1* operon, and the *LEE1* regulatory region, located immediately upstream, is crucial in controlling expression of the LEE. In EHEC, this region

has been reported to contain two promoters, P1 and P2, respectively distal and proximal with respect to the *ler* translation start site (Figure 1) (Sperandio *et al*., 2002). Expression from the *LEE1* regulatory region is controlled directly or indirectly by a multitude of regulators, including IHF, Fis, H-NS, QseA, BipA GadX, Pch, DksA, Hha, EtrA, EivF and small RNAs together with Ler itself (Mellies *et al.*, 2007; Tree *et al*., 2009; Hansen and Kaper, 2009; a recent overview summary can be found in the introduction to Kendall *et al*., 2010). In addition, GrlR and GrlA, encoded by a bicistronic operon *grlRA* located within the LEE, between the *LEE1* and the *LEE2* operons (Figure 1), appear to play a role (Deng *et al*., 2004; Barba *et al*., 2005; Jimenez *et al*., 2010). GrlA (global regulator of LEE activator) is a transcription activator, whilst GrlR (global regulator of LEE repressor) acts as a repressor by binding to GrlA and preventing its activity (Jobichen *et al*., 2007). GrlA is homologous to the enterobacterial CaiF proteins (Mellies *et al*., 2007), which positively regulate expression of the carnitine pathway (Buchet *et al*., 1999), and a helix-turn-helix DNA binding motif has been predicted (Deng *et al*., 2004).

Many aspects of the action of GrlA at the EHEC *LEE1* regulatory region are still open to question. For example, both Huang and Syu (2008) and Jimenez *et al*. (2010) reported that purified GrlA fusion proteins could bind to the *LEE1* regulatory region but did not define its binding target, whilst Russell *et al.* (2007) suggested that GrlA may modulate *LEE1* expression via the action of another factor. Hence our aim in this work was to improve our understanding of the action of GrlA at the EHEC *LEE1* regulatory region. In our previous work with very complex bacterial promoters we had noted that individual regulatory components were often able to function in isolation (Barnard *et al*., 2004) and, thus, we chose to use a laboratory strain of *E. coli* K-12 as a 'test-tube' to reconstruct GrlA-dependent activation at the *LEE1* regulatory region. We have exploited this background to investigate the activity of the P1 and P2 promoters, to find the target for GrlA-dependent activation at the P1 promoter, and to identify functional determinants in GrlA.

At most *E. coli* promoters, the key sequences required for activity are the -10 and -35 hexamer elements, which are recognised by different domains of the RNA polymerase σ factor, and the optimal length of the spacer region between these elements is 17 base pairs (McClure, 1985). Most transcription activators function by binding to targets located upstream from or overlapping with the -35 element (Browning and Busby, 2004). Here we report that the functional -10 and -35 elements at the *LEE1* regulatory region P1 promoter are

separated by a non-optimal 18 base pair spacer and that GrlA activates expression from P1 by binding to a target within this spacer.

Results

Activity of the LEE1 *promoters and nested deletion analysis*

The regulatory region of the *E. coli* O157:H7 *LEE1* operon is located upstream from the *ler* gene. Thus, the start point of this work was to create a construct in which *lacZ* expression is controlled by the LEE10-568 fragment, which carries the base sequence from position -568 to position -19 upstream of the functional ATG start codon of the *ler* gene (Figure 1). Figure 2A illustrates a series of eight nested deletions of this fragment, showing the location of the upstream end of each fragment with respect to the two proposed *LEE1* promoters, P1 and P2. Each fragment was cloned into the *lac* expression vector, pRW224, and the resulting recombinants were transformed into the *Δlac E. coli* K-12 laboratory strain, M182, and into two EHEC serotype O157:H7 strains (Sakai and EDL933). Measurements of β-galactosidase expression illustrated in Figure 2B show the effects of the deletions on the activity of the *LEE1* promoters. The overall patterns of activity in all three *E. coli* strains resemble each other in both LB medium and DMEM medium (see supplementary material, Figure S2). Expression increases ~2-4 fold as upstream sequences are removed and optimal expression is found with the LE10-275 and LEE10-215 fragments, that carry DNA sequence up to positions -275 and -215 upstream from the *ler* start codon. Longer deletions to positions -203, -195 and -155 result in a large decrease in expression. Since the region from positions -215 to -155 contains the proposed sequence elements for the P1 promoter (Sperandio *et al*., 2002), we conclude that, in our conditions, P1 is the major functional promoter for the *LEE1* operon. With the LEE10-155 and LEE10-115 fragments, which lack all P1 elements, expression is 10-15% of the maximum. This expression is most likely due to the P2 promoter, since the longer deletion in the LEE10-75 fragment, which cuts into P2, causes a reduction in expression to basal levels.

In a complementary experiment, the LEE10-568, LEE10-315, LEE10-275 and LEE10-203 fragments were shortened to move the downstream HindIII site to position -158, rather than position -19, thereby deleting the downstream P2 promoter, and generating the LEE20-568, LEE20-315, LEE20-275 and LEE20-203 fragments (Figure 2A). After cloning these fragments into pRW224, the observed patterns of β-galactosidase expression in both the O157:H7 and M182 strains were similar to with the LEE10 fragment series (Figure 2B). The

data are consistent with the conclusion that the P2 promoter makes but a minor contribution to the activity measured here. Note that the increased expression seen with the LEE20 fragments is probably due to the shorter untranslated leader sequence upstream of the reporter *lacZ* gene.

Identification of functional elements at the LEE1 *regulatory region P1 promoter*

In the experiment illustrated in Figure 2, the shortest fragment that shows maximum promoter activity is the LEE20-275 fragment. This fragment carries the DNA sequence from positions -275 to -158 upstream of the *ler* start codon and includes the P1 promoter. The complete sequence of this fragment is shown in Figure 3, where the sequence has been renumbered 1-118, starting immediately downstream of the EcoRI linker. We adopted this simpler numbering system to describe the ensuing mutational analysis of the LEE20-275 fragment and the shorter LEE20-203 fragment.

To identify elements essential for promoter activity, error-prone PCR was used to generate ten independent preparations of the LEE20-275 fragment carrying random point mutations. The fragments were then cloned into the pRW224 *lac* reporter plasmid, the resulting recombinant plasmids were transformed into *E. coli* strain M182, and transformants were grown on MacConkey indicator plates. As expected, the majority of colonies scored as Lac^+ (red), but, after screening over 100,000 transformants, we identified ~100 Lac^- (pale pink) colonies. Sequence analysis showed that 43 of these carried single mutations in the LEE20-275 fragment cloned in pRW224.

The locations of the different single point mutations that reduced *lac* expression from the LEE20-275 promoter fragment are illustrated in Figure 3. Strikingly, 25 of the 43 point mutations fall in the TTGACA motif at positions 73-77 that was predicted by Sperandio *et al*. (2002) to be the *LEE1* P1 promoter -35 hexamer, and which corresponds exactly to the consensus hexamer -35 element for *E. coli* promoters. The effects of the different substitutions in this element were quantified and the data are shown in Table 1. The results are consistent with this being the functional -35 element controlling expression from the LEE20-275 fragment. Another of the point mutations (71A) falls just upstream of the -35 element, and its effects are consistent with the lower promoter activity of the LEE20-203 fragment compared to the LEE20-275 fragment (Figure 2B).

A second cluster accounts for 15 of the 43 point mutations and appears to identify the promoter -10 hexamer (Figure 3). Recall that base pairs at positions 1 and 2 of promoter -10 hexamers in *E. coli* are most crucial for promoter activity (McClure, 1985), so this suggests that the motif TACACA at positions 97-102 is the likely functional -10 hexamer element. Since the consensus -10 element for *E. coli* promoters is TATAAT, we used site-directed mutagenesis to create complementary mutations to check this suggestion. Data presented in Table 1 show that the 97C, 98C, 98G, 98T and 100G substitutions cause >90% reductions in expression, whilst the 99A, 99G, 101G, 102C and 102G substitutions have lesser effects. These results are consistent with our assignation of TACACA as the -10 element. However, this would mean that the spacer between the proposed -35 and -10 hexamer elements is 18 base pairs rather than the optimal 17 base pairs. We therefore constructed a LEE20-275 fragment derivative in which a T in the spacer was deleted (Δ94T), thus shortening the spacer to the optimal 17 base pairs. Data presented in Table 1 show that the measured promoter activity is doubled by the Δ94T mutation. Note that one of the Lac^- mutants, selected after random mutagenesis, carried the LEE20-275 fragment with an extra T between position 93 and 94. This insertion extended the spacer to 19 base pairs, and caused a sharp reduction in *lac* expression (Table 1), suggesting that P1 promoter activity tails off as the spacer length increases from the optimal 17 base pairs.

To confirm the location of the functional promoter in the LEE20-275 fragment, we studied complexes between purified DNA fragments and purified *E. coli* RNA polymerase holoenzyme. In these experiments, we compared the starting 'wild-type' DNA fragment with a corresponding fragment carrying the 98C substitution at position 2 of the -10 hexamer element. An experiment in which binary RNA polymerase-DNA complexes were probed with potassium permanganate, a reagent that modifies T residues in the single stranded 'bubble' produced after local unwinding of promoter DNA around the transcription start, is illustrated in Figure 4 (Savery *et al*., 1996; Browning *et al*., 2009). The experiment reveals extensive unwinding just downstream of the promoter -10 hexamer and, crucially, this unwinding is greatly reduced with the 98C mutant. We then analysed the labelled transcripts formed after labelled nucleoside triphosphates were added to the binary complexes. In this experiment, RNA polymerase runs to a terminator in the DNA fragment synthesising labelled RNA molecules that are detected by gel electrophoresis (Figure 4B). The major transcript starting in the LEE20-275 fragment is ~102 bases and this corresponds to a transcript starting at position

107A, 5 bases downstream from the proposed -10 element (Figure 3). This RNA species is completely absent when the DNA fragment carrying the 98C substitution is used (Figure 4B).

Activation by GrlA

A major aim of this study was to investigate the action of GrlA at the *LEE1* operon regulatory region. To do this, we used a series of pACYC184 derivatives carrying the *grlRA* operon (pSI01), *grlA* alone (pSI02), *grlR* alone (pSI03) or empty vector (Figure 1), pRW224 derivatives carrying different *LEE1* regulatory region fragments (Figure 2), and the *E. coli* K-12 M182 *Δlac* strain as a 'test tube'. In a preliminary experiment, M182 was co-transformed with pRW224 carrying the LEE10-568 fragment and the different pACYC184 derivatives, and β-galactosidase expression was measured. The results show ~6-fold activation in the presence of GrlA (Figure 5). As expected (Jobichen *et al*., 2007), the activation found with GrlA alone (pSI02) was suppressed by the presence of GrlR (pSI01).

GrlA-dependent activation at the *LEE1* regulatory region carried by fragments with the two series of nested deletions described earlier (Figure 2) was measured. Activation is observed with the LEE20 series of fragments that lack the P2 promoter, but is lost with the LEE10-195, LEE10-155, LEE10-115 and LEE10-75 fragments in which segments of the P1 promoter are deleted (Figure 6). From this, we deduce that the *LEE1* P1 promoter is the target for activation by GrlA, and this can be measured even with the 52 base pair LEE20-203 fragment. Note that GrlA-dependent activation at the *LEE1* P1 promoter carried by LEE20-203 fragment can also be measured in the Sakai EHEC serotype O157:H7 strain (see supplementary material, Figure S3).

Jimenez *et al*. (2010) recently reported that GrlA primarily activates *LEE1* expression by counteracting repression by the global repressor, H-NS, but also reported GrlA-dependent activation of *LEE1* in the absence of H-NS. To investigate this, we measured expression from the *LEE1* regulatory region and activation by GrlA in *E. coli* K-12 strain MG1655 and its isogenic Δ*hns* derivative. With the full-length LEE10-568 fragment, *LEE1* promoter activity is clearly repressed by H-NS, and, as expected, the repression is overcome by GrlA which activates to higher levels in both the *hns*$^{+}$ and Δ*hns* backgrounds (Figure 7). In contrast, with the short LEE20-203 fragment, *LEE1* promoter activity is not repressed by H-NS and a similar level of GrlA-dependent activation is observed in both genetic backgrounds. Hence, GrlA can activate the *LEE1* P1 promoter independently of H-NS.

Evidence that GrlA binds to the spacer region at the LEE1 *P1 promoter*

Most bacterial transcription activators function by binding to specific sites upstream of the -35 hexamer at target promoters (Browning and Busby, 2004). However, this cannot be the case for GrlA at the *LEE1* P1 promoter since GrlA-dependent activation was observed with both the LEE10-203 and LEE20-203 fragments, which begin immediately upstream of the -35 TTGACA hexamer element (Figure 8A). To identify essential promoter determinants for activation, we constructed a series of mutations throughout the LEE20-203 fragment and measured GrlA-dependent activation. The effects of 23 point mutations throughout this fragment, which carries only 46 base pairs of the *LEE1* regulatory region sequence, were measured. The results (Figure 8) show that GrlA-dependent activation is completely suppressed by the 89G, 90C, 91C, 92A and 98C substitutions (note that the numbering is as in Figure 3 and Table 1). As noted, the 98C substitution falls at the second position of the -10 hexamer and completely knocks out P1 activity (Figure 4). In contrast, positions 89-92 fall within the 18 base pair promoter spacer region. A simple explanation for the effects of changes at these positions would be that they fall within the binding site for GrlA and hence suppress GrlA binding and ensuing activation.

In a complementary set of experiments, effects of altering the length of the spacer length on GrlA-dependent activation at the *LEE1* P1 promoter were measured, using derivatives of the LEE20-203 fragment carrying insertions or deletions in the spacer either upstream or downstream of, but not within positions 89-92. The results, presented in Table 2, show that GrlA-independent expression from the LEE20-203 fragment is increased by single base deletions at positions 79 or 94, but decreased by two-base deletions at positions 79 and 80, or by single base insertions between positions 78 and 79 or between positions 93 and 94. Hence, as expected, in the absence of GrlA, optimal *LEE1* P1 activity is found with a 17 base pair spacer. However, GrlA-dependent activation of the *LEE1* P1 promoter is found only with the starting 18 base pair non-optimal spacer, and it is suppressed by all of the deletions and insertions that were tested. Note that, in this experiment, we selected locations for insertions and deletions that did not appear to be involved in GrlA binding.

We sought to use the well-established *in vitro* band shift and footprinting techniques (Minchin and Busby, 2009) to identify the binding target for GrlA at the *LEE1* P1 promoter. However our efforts were unsuccessful as we failed to purify soluble functional GrlA. Thus, we turned

to genetic approaches and, first, used suppression genetics. Error-prone PCR was used to create random mutations in the *grlA* coding sequence within pSI02. Mutated pSI02 DNA was then transformed into *E. coli* strain M182 containing pRW224 carrying the LEE20-203 fragment with the 92A substitution that reduces *lac* expression (Figure 8), resulting in Lac^- colonies on MacConkey indicator plates. After screening over 75,000 transformants, from 10 independent preparations of mutated pSI02, we identified 11 Lac^+ colonies containing pSI02 with single-base substitutions leading to the RK53, IT59 or IV59 substitutions in GrlA. These substitutions fall in the predicted helix-turn-helix motif of GrlA (Figure 9A). We found that the RK53, IT59 and IV59 substitutions partially restore GrlA-dependent activation with the LEE20-203 fragment carrying the 92A mutation but not with other mutations (Figure 9B). This implies that, when bound at the *LEE1* P1 promoter, residues 53 and 59 may be sufficiently close to make contact with the base pair at position 92 and suggests that the predicted helix-turn-helix in GrlA is functional in binding the target operator site. To investigate this further, a set of pSI02 derivatives encoding GrlA with alanine substitutions at sequential residues from L52 to S63 was constructed and GrlA-dependent activation of expression from the starting and mutant LEE20-203 fragments was measured. We found that the RA54 substitution causes the largest activation defect (Figure 10A). Taken together with the suppression genetics experiment, this suggests that the R54 side chain makes a contact with the GrlA binding target somewhere in its target from positions 89 to 92, and that the RK53 substitution in the neighbouring side chain makes a contact that compensates for the 92A substitution.

In a complementary experiment, we investigated whether any of the alanine substitutions could relieve the 'down' effects of specific base substitutions in the putative GrlA operator target in the LEE20-203 promoter fragment. Results in panels B and C of Figure 10 show that the FA57 substitution partially reverses the effects of the 89G and 90C substitutions. The simplest explanation is that residue F57 clashes with the base pairs at positions 89-90 in the mutated LEE20-203 89G or 90C fragments, probably because residue F57 is close to positions 89 and 90 in the GrlA-target DNA complex. A similar relationship was found between residue R53 and position 91 (Figure 10D).

Discussion

In pathogenic *E. coli* and related bacteria, the LEE plays a key role during infection and the expression of the different LEE genes must be carefully orchestrated in time and space. This

is likely to account for the complexity of *LEE1* regulation. Here, we have focussed on one single aspect, namely activation by GrlA, and studied it in the context of a simplified system, using a non-pathogenic *E. coli* K-12 strain as a host.

The first issue that we addressed was the relative activity of the two promoters, P1 and P2. Our conclusion, that P1 is the major promoter is in agreement with Porter *et al*. (2005). Other groups have reported a bigger role for P2 (e.g., Sperandio *et al*., 2002; Sharp and Sperandio, 2007) but differences are likely due to the precise fusions, genetic backgrounds and conditions used for the measurements. These differences underscore the point that any set of results applies only to the conditions in which the experiments are performed. Our experimental setup allowed us to make a detailed mutational analysis of the key P1 elements (Figure 3 and Table 1) and this identified TTGACA and TACACA as the functional -35 and -10 hexamer elements, respectively. The -35 element, which was predicted by Sperandio *et al*. (2002) and other groups, corresponds exactly to the established consensus for *E. coli* promoters (McClure, 1985). The -10 element was originally identified by Sperandio *et al.* (2002) as TACACA but, due to uncertainty in the location of the P1 transcript start, was later reassigned to a sequence 2 base pairs upstream, TTTACA (Porter *et al*., 2005; Russell *et al*., 2007; Sharp and Sperandio, 2007). Our mutational analysis argues unambiguously for TACACA, as originally assigned, and a key consequence of this is that the spacer region between the -10 and -35 elements is 18 base pairs, one above the optimal 17 base pairs for promoter activity (McClure, 1985). An important point to note is that there is considerable flexibility in the location of the 5' end of transcripts with respect to -10 hexamer regions at bacterial promoters and the structural basis for this is now understood (Darst, 2009). Indeed, the AT-rich nature downstream of the -10 element may confer unusual flexibility at the *LEE1* P1 promoter and this may be reflected in the greater than usual unwinding (>20 base pairs) seen in the open complex probed by permanganate (Figure 4A; compare with Browning *et al*., 2009). The consequence of this is that the different start points reported in different experiments could all be correct. For many bacterial promoters, the upstream end of the -10 hexamer provides the major anchoring point for RNA polymerase and the location of the first templated base is not fixed (Sclavi, 2009). For this reason, rather than referencing *LEE1* P1 promoter locations with the transcript start as +1, we used an arbitrary numbering system from an upstream fixed point to describe our experiments (Figure 3).

Most summaries of transcriptional regulation in the LEE (e.g. Kendall *et al*., 2010) show GrlA as a positive regulator of expression from the *LEE1* regulatory region and yet the current literature is not completely unambiguous. We were able to show significant activation at *LEE1* by GrlA in our K-12 system (Figure 5) and this activation was suppressed by GrlR, which suggests that our set-up is reporting physiologically relevant effects. Our most striking observations were GrlA-dependent activation from a 52-base pair promoter fragment (LEE20-203) that lacks LEE1 sequence upstream of the P1 -35 hexamer, and the localisation of the DNA target for GrlA to the spacer region between the P1 -10 and -35 hexamer elements. In agreement with Jimenez *et al*. (2010), our results (Figure 7) confirm that GrlA plays a dual role at the *LEE1* regulatory region by counteracting repression by the global repressor, H-NS, and activating the P1 promoter. Our observation that expression from the 52-base pair LEE20-203 promoter fragment is activated by GrlA, but not subject to repression by H-NS, argues that GrlA can function as a true transcription activator, most likely by directly accelerating one of the transactions between RNA polymerase holoenzyme and the *LEE1* P1 promoter.

To date, it has proven impossible to reconstitute GrlA-dependent transcription activation at the *LEE1* regulatory region *in vitro*. Indeed, we were unable to purify functional GrlA, in order to reproduce the *in vitro* binding reported by Huang and Syu (2008) and Jimenez *et al.* (2010). However, we were able to exploit different genetic approaches to confirm the proposed DNA-binding helix-turn-helix motif in GrlA, and to identify its target in the spacer region. Our conclusion that GrlA binds to the *LEE1* P1 promoter spacer region was surprising, since most bacterial transcription activators bind sites that are upstream of or overlapping the -35 element of the target promoter (Browning and Busby, 2004). However several members of the MerR family of transcription factors provide interesting exceptions to this rule since they also bind to sites located between the -10 and -35 hexamers at target promoters (reviewed by Brown *et al*., 2003). Furthermore, they activate transcription at target promoters where the spacing between the -10 and -35 hexamer elements is greater than the optimum 17 base pairs. Current models for activation by MerR family members suggest that their binding in the spacer bends or twists the target DNA, resulting in juxtaposition of the target promoter -10 and -35 elements that facilitates RNA polymerase binding (reviewed by Brown *et al*., 2003). The existence of an 18 base pair spacer in the *LEE1* P1 promoter, which, when changed to 17 base pairs, increases promoter activity and stops GrlA-dependent activation, argues that a similar activation mechanism may be occurring here (Figure 11), even though

there is no significant sequence similarity between GrlA and any member of the MerR family. Note that our conclusions are consistent with Jimenez *et al* (2010) who reported that the DNA site for GrlA is downstream of position -54 at its target promoter in the *LEE1* regulatory region of enteropathogenic *E. coli*.

In conclusion, it is clear that the interaction between GrlA and the *LEE1* P1 promoter represents an intriguing case study in the complexity of bacterial gene regulation, of interest, not just to researchers studying bacterial pathogenesis, but also to those interested in promoter organisation. Further insights into the action of GrlA may emerge from studying its other targets and its interactions with other proteins. To date, the full extent of the GrlA regulon is unknown, though it has been reported to activate the expression of the Ehx enterohemolysin (Saitoh *et al.,* 2008) and repress genes responsible for the formation of flagella (Iyoda *et al.,* 2006).

Experimental procedures

Strains, plasmids and oligonucleotides

In this work we used *E. coli* K-12 strain M182 (Casadaban, 1976), MG1655 and a Δ*hns* derivative (Grainger *et al*., 2008), and two EHEC O157:H7 strains lacking the *stx* toxin, Sakai 813, obtained from Chihiro Sasakawa and EDL933, given by Arthur Donohue-Rolfe. The vector plasmids used were pRW224 (Lodge *et al*., 1992), pSR (Kolb *et al*., 1995) and pACYC184 (Chang and Cohen, 1978). Plasmid pRW224 is an RK2-based low copy number *lac* expression vector, encoding resistance to tetracycline, designed to facilitate the cloning of EcoRI-HindIII fragments carrying a promoter directed towards the HindIII end of the fragment but without a translation start. Thus, a translation initiation signal for *lacZ* is located in the vector immediately adjacent to the HindIII site as described by Bingham and Busby (1987) and supplementary Figure S1. Plasmid pSR is a *colE1*-based general cloning vector, encoding resistance to ampicillin. Promoters that are cloned into pSR on EcoRI-HindIII fragments run into the bacteriophage λ *oop* terminator downstream of the HindIII site (Browning *et al*., 2009). Plasmid pACYC184 is a P15A-based general cloning vector, encoding resistance to tetracycline and chloroamphenicol. We routinely cloned DNA segments so as to disrupt resistance to tetracycline. Hence, for use as an empty vector, we constructed the pACYC184ΔHN derivative by deleting intervening sequences between the HindIII and NruI sites. This derivative encodes resistance to chloramphenicol but not to tetracycline. Standard techniques for recombinant DNA manipulations were used throughout

this work using PCR with synthetic oligodeoxyribonucleotide primers made by Alta Biosciences (University of Birmingham) that are listed in the supplementary material (Table S1). All cloned sequences were checked using the University of Birmingham Functional Genomics Facility (http://www.genomics.bham.ac.uk/sequencing.htm).

Cloning of LEE1 *regulatory region fragments and mutational analysis*

PCR was used to amplify the LEE10-568 fragment using the D61221 and D61222 oligos and genomic DNA from the O157:H7 Sakai 813 strain. The fragment was designed with flanking EcoRI and HindIII sites respectively located upstream and downstream of the *LEE1* regulatory region, with the HindIII site positioned such that the fragment excluded the *ler* gene Shine-Dalgarno sequence (Figure 1). The resulting product was restricted with EcoRI and HindIII and cloned into pRW224. The different fragments in the LEE10 and LEE20 sets of nested deletions, illustrated in Figure 2, were made using primers listed in the supplementary material (Table S1). Each fragment is described by the location of its upstream end, in terms of the number of bases upstream from the functional *ler* ATG codon defined by Yerushalmi *et al.* (2008). Different mutations were introduced into the LEE20-275 and LEE20-203 fragments cloned in pRW224 by using error prone PCR (Leung *et al*., 1989) with flanking primers D10520 and D53463, or by using megaprimer PCR (Perrin and Gilliland, 1990) as in Chismon *et al*. (2010). The different bases at the *LEE1* P1 promoter in the LEE20-275 fragment are numbered 1-118, as shown in Figure 3. We used this numbering system to describe different P1 promoter mutations, in both the LEE20-275 and LEE20-203 fragments. Note that we did not adopt the customary convention of numbering bases with respect to the transcript start point because of uncertainty about its location.

Cloning of grlRA *and mutational analysis*

PCR was used to amplify a HindIII-SalI fragment carrying the *grlRA* operon using the D62895 and D62897 oligos and genomic DNA from the O157:H7 Sakai 813 strain. The resulting product was restricted with HindIII and SalI and cloned into pACYC184 to give pSI01. To construct pSI02, which is a derivative of pSI01 carrying a large in-frame deletion in *grlR*, we used PCR with primers D63209 and D63210 and pSI01 as a template. The resulting product was cut with BamHI and circularised by ligation to give pSI02. To construct pSI03, which is a derivative of pSI01 deleted for *grlA*, we used primers D62895 and D63698 and pSI01 as a template. The resulting product was restricted with HindIII and SalI and cloned into pACYC184 to give pSI03. Different mutations were introduced into pSI02 by

using error prone PCR (Leung *et al*., 1989) with flanking primers D63048 and D63049, or by using megaprimer PCR (Perrin and Gilliland, 1990).

β-Galactosidase assays

To assay promoter activity in the *LEE1* regulatory region we used the different LEE10 and LEE20 fragments (Figure 2) cloned in pRW224. *E. coli* strains carrying different pRW224 recombinants were grown aerobically with shaking at 37°C in LB medium or Dulbecco's modified Eagle's medium (DMEM) containing 35 μg ml^{-1} tetracycline. In experiments to measure the effects of *grlR* and *grlA*, encoded by pACYC184 derivatives, 35 μg ml^{-1} chloramphenicol was also included. β-galactosidase levels were measured by the Miller (1972) method. Recorded activities are shown in Miller units and are the average of at least three independent experiments.

Permanganate footprinting and run-off transcription assays

Purified PstI–BamHI DNA fragments were derived from caesium chloride preparations of plasmid pSR carrying the LEE20-275 fragment. Fragments were labelled at the BamHI end using [γ-^{32}P]-ATP and polynucleotide kinase. Permanganate footprints were performed following the protocol of Savery *et al.* (1996) using holo *E. coli* RNA polymerase, which was purchased from Epicenter (Madison). Each reaction contained approximately 3 nM labelled DNA fragment in 20 mM HEPES, pH 8.0, 5 mM $MgCl_2$, 50 mM potassium glutamate, 1 mM DTT and 0.5 mg ml^{-1} BSA and 50 nM holo RNA polymerase as required. Products of footprinting reactions were analysed on denaturing 6% polyacrylamide sequencing gels, calibrated with Maxam Gilbert "G+A" sequencing reactions.

The *in vitro* transcription experiments were performed as described by Browning *et al*. (2009) using PstI–BamHI DNA fragments purified from caesium chloride preparations of pSR carrying the LEE20-275 fragment. These fragments served as a template for multiple round *in vitro* transcription assays in which 20 ng fragment was incubated in transcription buffer containing 40 mM Tris pH 7.9, 10 mM $MgCl_2$, 1 mM DTT, 100 mM KCl, 100 μg ml^{-1} bovine serum albumin, 200 μM GTP, 200 μM ATP, 200 μM CTP, 10 μM UTP and 5 μCi [α^{32}P]-UTP. Reactions were started by adding holo *E. coli* RNA polymerase purchased from Epicenter (Madison). RNA products were analysed on a denaturing 5.5 % polyacrylamide gel

and visualised using a Fuji phosphor screen and BioRad Molecular Imager FX. Gels were calibrated with Maxam Gilbert "G+A" sequencing reactions.

Helix-turn-helix prediction for GrlA

The DNA-binding helix-turn-helix of GrlA, suggested by Deng *et al.* (2004), was confirmed computationally by the Dodd and Egan (1990) method, http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hth.html, the GYM2.0 method, http://users.cis.fiu.edu/~giri/bioinf/GYM2/prog.html (Narasimhan *et al.*, 2002), and the Jpred structure prediction tool, http://www.compbio.dundee.ac.uk/www-jpred/index.html (Cole *et al*., 2008).

Acknowledgements

This work was supported by a BBSRC project grant and by a Commonwealth PhD studentship to M.S.I. We are grateful to Kerry Hollands for help with the footprinting and transcription experiments.

References

Barba, J., Bustamante, V.H., Flores-Valdez, M.A., Deng, W., Finlay, B.B., and Puente, J.L. (2005) A positive regulatory loop controls expression of the locus of enterocyte effacement-encoded regulators Ler and GrlA. *J Bacteriol* **187**: 7918-7930.

Barnard, A., Wolfe, A., and Busby, S. (2004) Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. *Curr Opin Microbiol* **7:** 102-108.

Bhavsar, A.P., Guttman, J.A., and Finlay, B.B. (2007) Manipulation of host-cell pathways by bacterial pathogens. *Nature* **449:** 827-834.

Bingham, A.H., and Busby, S.J. (1987) Translation of *galE* and coordination of galactose operon expression in *Escherichia coli*: effects of insertions and deletions in the non-translated leader sequence. *Mol Microbiol* **1**: 117-124.

Brown, N.L., Stoyanov, J.V., Kidd, S.P., and Hobman, J.L. (2003) The MerR family of transcriptional regulators. *FEMS Microbiol Rev* **27**: 145-163.

Browning, D.F., and Busby, S.J. (2004) The regulation of bacterial transcription initiation. *Nat Rev Microbiol* **2**: 57-65.

Browning, D., Savery, N., Kolb, A., and Busby, S. (2009) Assays for transcription factor activity. In *Methods in Molecular Biology: DNA-Protein interactions.* 3rd Edition. Moss, T., and Leblanc, B. (eds). Heidelberg: Springer Science, volume 543, pp 369-387.

Buchet, A., Nasser, W., Eichler, K., and Mandrand-Berthelot, M.A. (1999) Positive co-regulation of the *Escherichia coli* carnitine pathway *cai* and *fix* operons by CRP and the CaiF activator. *Mol Microbiol* **34**: 562-575.

Bustamante, V.H., Santana, F.J., Calva, E., and Puente, J.L. (2001) Transcriptional regulation of type III secretion genes in enteropathogenic *Escherichia coli*: Ler antagonizes H-NS-dependent repression. *Mol Microbiol* **39**: 664-678.

Casadaban, M.J. (1976) Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. *J Mol Biol* **5:** 541-555.

Chang, A.C., and Cohen, S.N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol* **134**: 1141–1156.

Chismon, D.L., Browning, D.F., Farrant, G.K., and Busby, S.J. (2010) Unusual organisation, complexity and redundancy at the *Escherichia coli hcp-hcr* operon promoter. *Biochem J* **430**: 61-68.

Cole, C. , Barber, J.D., and Barton, G.J. (2008) The Jpred 3 secondary structure prediction server. *Nucl Acids Res*. **36:** W197-W201.

Darst, S. (2009) A structural atlas of RNAP. In *RNA polymerases as molecular motors*. (Buc, H. and Strick, T., eds) Royal Society of Chemistry Publishing, Cambridge, UK, pp. 1-10.

Deng, W., Puente, J.L., Gruenheid, S., Li, Y., Vallance, B.A., Vazquez, A., *et al*. (2004) Dissecting virulence: systematic and functional analyses of a pathogenicity island. *Proc Natl Acad Sci USA* **101**: 3597-3602.

Dodd, I.B., and Egan, J.B. (1990) Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. *Nucl Acids Res* **18:** 5019-5026.

Elliott, S.J., Wainwright, L.A., McDaniel, T.K., Jarvis, K.G., Deng, Y.K., Lai, L.C., *et al*. (1998) The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Mol Microbiol* **28:** 1-4.

Grainger, D.C., Goldberg, M.D., Lee, D.J., and Busby, S.J. (2008) Selective repression by Fis and H-NS at the *Escherichia coli dps* promoter. *Mol Microbiol* **68:** 1366-1377.

Haack, K.R., Robinson, C.L., Miller, K.J., Fowlkes, J.W., and Mellies, J.L. (2003) Interaction of Ler at the *LEE5* (*tir*) operon of enteropathogenic *Escherichia coli*. *Infect Immun* **71:** 384-392.

Hansen, A.M., and Kaper, J.B. (2009) Hfq affects the expression of the LEE pathogenicity island in enterohaemorrhagic *Escherichia coli*. *Mol Microbiol* **73**: 446-465.

Huang, L.H., and Syu, W.J. (2008) GrlA of enterohemorrhagic *Escherichia coli* O157:H7 activates *LEE1* by binding to the promoter region. *J Microbiol Immunol Infect* **41**: 9-16.

Iyoda, S., Koizumi, N., Satou, H., Lu, Y., Saitoh, T., Ohnishi, M., and Watanabe, H. (2006) The GrlR-GrlA regulatory system coordinately controls the expression of flagellar and LEE-encoded type III protein secretion systems in enterohemorrhagic *Escherichia coli*. *J Bacteriol* **188**: 5682-5692.

Jimenez, R., Cruz-Migoni, S.B., Huerta-Saquero, A., Bustamante, V.H., and Puente, J.L. (2010) Molecular characterization of GrlA, a specific positive regulator of *ler* expression in enteropathogenic *Escherichia coli*. *J Bacteriol* **192**: 4627-4642.

Jobichen, C., Li, M., Yerushalmi, G, Tan, Y.W., Mok, Y-K., Rosenshine, I., *et al*. (2007) Structure of GrlR and the implication of its EDED motif in mediating the regulation of type III secretion system in EHEC. *PLoS Pathog* **3**: e69.

Kendall, M.M., Rasko, D.A., and Sperandio, V. (2010) The LysR-type regulator QseA regulates both charatcterised and putative virulence genes in enterohaemorrhagic *Escherichia coli* O157:H7. *Mol Microbiol* **76**: 1306-1321.

Kolb, A., Kotlarz, D., Kusano, S., and Ishihama, A. (1995) Selectivity of the *Escherichia coli* RNA polymerase E sigma 38 for overlapping promoters and ability to support CRP activation. *Nucleic Acids Res* **23**: 819–826.

Leung, D.W., Chen, E., and Goeddel, D.V. (1989) A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1**: 11–15.

Lodge, J., Fear, J., Busby, S., Gunasekaran, P., and Kamini, N-R. (1992) Broad host range plasmids carrying the *Escherichia coli* lactose and galactose operons. *FEMS Microbiol Lett* **74:** 271-276.

McClure, W.R. (1985) Mechanism and control of transcription initiation in prokaryotes. *Annu Rev Biochem* **54**: 171-204.

McDaniel, T.K., Jarvis, K.G., Donnenberg, M.S., and Kaper, J.B. (1995) A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc Natl Acad Sci USA* **92**: 1664–1668.

Mellies, J.L., Elliott, S.J., Sperandio, V., Donnenberg, M.S., and Kaper, J.B. **(**1999) The Per regulon of enteropathogenic *Escherichia coli*: identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE)-encoded regulator (Ler). *Mol Microbiol* **33:** 296-306.

Mellies, J.L, Barron, A.M.S., and Carmona, A.M. (2007) Enteropathogenic and enterohemorrhagic *Escherichia coli* virulence gene regulation. *Infect Immun* **75**: 4199-4210.

Miller, J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Minchin, S.D., and Busby, S.J. (2009) Analysis of mechanisms of activation and repression at bacterial promoters. *Methods* **47**: 6-12.

Narasimhan, G., Bu, C., Gao, Y., Wang, X., Xu, N., and Mathee, K., (2002) Mining protein sequences for motifs. *J Comput Biol* **9**: 707-720.

Nataro, J.P., and Kaper, J.B. (1998) Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* **11**: 142-201.

Perna, N.T., Mayhew, G.F., Pósfai, G., Elliott, S., Donnenberg, M.S., Kaper, J.B., and Blattner, F.R. (1998) Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* **66**: 3810–3817.

Perrin, S., and Gilliland, G. (1990) Site-specific mutagenesis using asymmetric polymerase chain reaction and a single mutant primer. *Nucleic Acids Res* **18**: 7433-7438.

Porter, M.E., Mitchell, P., Free, A., Smith, D.G.E., and Gally, D.L. (2005) The *LEE1* promoters from both enteropathogenic and enterohemorrhagic *Escherichia coli* can be activated by PerC-like proteins from either organism. *J Bacteriol* **187**: 458-472.

Russell, R.M., Sharp, F.C., Rasko, D.A., and Sperandio, V. (2007) QseA and GrlR/GrlA regulation of the locus of enterocyte effacement genes in enterohemorrhagic *Escherichia coli*. *J Bacteriol* **189**: 5387-5392.

Saitoh, T., Iyoda, S., Yamamoto, S., Lu, Y., Shimuta, K., Ohnishi, M., *et al*. (2008) Transcription of the *ehx* enterohemolysin gene is positively regulated by GrlA, a global regulator encoded within the locus of enterocyte effacement in enterohemorrhagic *Escherichia coli*. *J Bacteriol* **190**: 4822–4830.

Savery, N.J., Belyaeva, T., and Busby, S. (1996) Introduction to protein: DNA interactions, DNase I footprinting, hydroxyl radical footprinting, permanganate footprinting and

supplementary protocols. In *Essential Techniques: Gene Transcription*. (Docherty, K., ed.) Oxford: BIOS Scientific Publishers, pp. 1–5 and 21–33.

Schauer, D.B., and Falkow, S. (1993) Attaching and effacing locus of a *Citrobacter freundii* biotype that causes transmissible murine colonic hyperplasia. *Infect Immun* **61**: 2486–2492.

Sclavi, B. (2009) Opening the DNA at the promoter: the energetic challenge. In *RNA polymerases as molecular motors*. (Buc, H. and Strick, T., eds) Royal Society of Chemistry Publishing, Cambridge, UK, pp. 38-68.

Sharp F.C., and Sperandio, V. (2007) QseA directly activates transcription of *LEE1* in enterohemorrhagic *Escherichia coli*. *Infect Immun* **75**: 2432-2440.

Sperandio, V, Li, C.C., and Kaper, J.B. (2002) Quorum-sensing *Escherichia coli* regulator A: a regulator of the LysR family involved in the regulation of the locus of enterocyte effacement pathogenicity island in enterohemorrhagic *E. coli*. *Infect Immun* **70**: 3085-3093.

Stoebel, D.M., Free, A., and Dorman, C.J. (2008) Anti-silencing: overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. *Microbiology* **154**: 2533-2545.

Tree, J.J., Wolfson, E.B., Wang, D., Roe, A.J., and Gally, D.L. (2009) Controlling injection: regulation of type III secretion in enterohaemorrhagic *Escherichia coli*. *Trends Microbiol* **17**: 361-370**.**

Umanski, T., Rosenshine, I., and Friedberg, D. (2002) Thermoregulated expression of virulence genes in enteropathogenic *Escherichia coli*. *Microbiology* **148**: 2735-2744.

Yerushalmi, G., Nadler, C., Berdichevski, T., and Rosenshine, I. (2008) Mutational analysis of the locus of enterocyte effacement-encoded regulator (Ler) of enteropathogenic *Escherichia coli*. *J Bacteriol* **190**: 7808-7818.

Figure Legends

Fig. 1. Schematic representation of the LEE region.

The upper line shows the organisation of the different LEE transcription units. The lower left part of the figure shows an expanded sketch of the *LEE1* regulatory region and illustrates the LEE10-568 fragment and how it was cloned into the pRW224 *lac* expression vector. The locations of the two *LEE1* regulatory region promoters P1 and P2 are indicated, together with the translation initiation region (TIR that contains the Shine-Dalgarno sequence and the translation initiation codon, see supplementary Figure S1). The lower right part of the figure shows an expanded sketch of the *grlRA* transcription unit and illustrates how it was cloned into vector plasmid pACYC184 to give pSI01, together with two deletion derivatives to give pSI02 and pSI03. E, H, B and S indicate the location of the EcoRI, HindIII, BamHI and SalI sites that were used in cloning (see experimental procedures).

Fig. 2. Nested deletions of the *LEE1* regulatory region.

A. Schematic representation of a set of EcoRI-HindIII DNA fragments covering the *LEE1* regulatory region. The coordinates of the upstream and downstream end of each fragment refer to the number of base pairs upstream from the functional ATG start codon of the *ler* gene reported by Yerushalmi *et al*. (2008). The P1 and P2 promoters are indicated by bent arrows and the shaded black boxes represent the cognate -10 and -35 hexamer elements.

B. Bar chart to illustrate measured β-galactosidase activities in *E. coli* K-12, EHEC Sakai and EHEC EDL933 cells carrying pRW224 with each of the different fragments. Vector refers to empty pRW224 with a short linker between the EcoRI and HindIII sites. Activities were measured after growing the cells in LB medium to an optical density at 650 nm of ~0.5 at 37°C. The values are the average of three independent assays. Standard errors are shown with error bars.

Fig. 3. Mutational analysis of the LEE20-275 fragment.

The figure shows the nucleotide sequence of the upper strand of LEE20-275 fragment. The base sequences are numbered 1-118, starting with the first cloned *LEE1* regulatory region base (that is 275 base pairs upstream from the *ler* translation start codon). The positions of randomly generated single mutations that reduced expression from this fragment are illustrated by showing the substituted base and, in each case, the adjacent number records the number of times that the particular substitution was obtained. The locations of the P1 promoter -35 and -10 hexamer elements, deduced from this study, are shown by grey shading

and the transcript start, at position 107A, determined from data in Figure 4B is indicated. The upstream end of the smaller LEE20-203 fragment is indicated by a bent solid/dotted line.

Fig. 4. *In vitro* permanganate footprinting and transcription assays.

A. The figure shows an autoradiogram that identifies the potassium permanganate sensitive sites in complexes of holo RNA polymerase with a DNA fragment that includes the LEE20-275 sequence (WT) or a derivative carrying the 98C mutation (98C). Lanes 1 and 3 show the result after control incubations without RNA polymerase, whilst lanes 2 and 4 show the analysis of samples with 50 nM RNA polymerase. The gel was calibrated using a Maxam-Gilbert sequence reaction (GA) and relevant positions are indicated. The location of the *LEE1* P1 promoter -10 and -35 elements are indicated by black boxes and the asterisks identify 6 consecutive residues just downstream of the -10 element that display RNA polymerase-dependent reactivity to permanganate.

B. Autoradiogram of an analysis by gel electrophoresis of ^{32}P-labelled RNA transcripts made by RNA polymerase holoenzyme from PstI-BamHI fragments carrying wild type LEE20-275 sequences (lanes 1-3) and the 98C derivative (lanes 4-6). The RNA polymerase concentration was: lanes 1 and 4, no enzyme; lanes 2 and 5, 200 nM; lane 3 and 6, 400 nM. The gel was calibrated with the pSR plasmid-encoded 108 base RNA-I transcript (lane 7) and a Maxam-Gilbert sequence reaction (GA). The proposed LEE1 P1 transcript is indicated by an asterisk.

Fig. 5. Activation of expression from the *LEE1* regulatory region by GrlA.

The bar chart illustrates measured β-galactosidase expression in M182 cells containing pRW224 with the LEE10-568 promoter fragment together with plasmid pACYC184ΔHN (R-A-), pSI01 (R+A+), pSI03 (R+A-) or pSI02 (R-A+). Measurements were made after growing the cells in LB medium at 37ºC to an optical density of ~0.5 at 650 nm. Each bar represents the mean of three independent experiments together with the standard error.

Fig. 6. GrlA-dependent activation at different *LEE1* regulatory region fragments.

The left part of the figure illustrates different fragments covering the *LEE1* regulatory region that were cloned into pRW224, using the same conventions as in Figure 2. The right part of the figure presents β-galactosidase activities measured in M182 cells carrying pRW224 containing each of the fragments. Cells also contained either pSI02 (+GrlA) or empty vector, pACYC184ΔHN (-GrlA). Measurements were made after growth in LB at 37ºC to an optical

density of ~0.5 at 650 nm. Standard deviations were obtained from at least three independent experiments. The fold activation by GrlA for each fragment is shown in brackets.

Fig. 7. Repression by H-NS and activation by GrlA at the *LEE1* regulatory region.

A. Schematic representation of the LEE10-568 and LEE20-203 fragments of *LEE1* operon regulatory region.

B. The bar chart illustrates measured β-galactosidase activities in *E. coli* K-12 strain MG1655 (WT) and a Δ*hns* derivative carrying pRW224 with either the LEE10-568 or LEE20-203 fragments. Cells also contain either empty vector pACYC184ΔHN (-GrlA: open bars) or pSI02 (+GrlA: shaded bars). Measurements were made after growth the cells in LB at 37ºC to an optical density of ~0.5 at 650 nm. Values are the average of at least three independent assays, and standard errors are shown with error bars.

Fig. 8. Mutational analysis of the LEE20-203 fragment.

A. Base sequence of the LEE20-203 fragment, numbered and annotated as in Figure 3. The asterisks and shading indicate bases where substitutions suppress GrlA-dependent.

B. The figure illustrates β-galactosidase activities measured in M182 cells carrying pRW224 containing the starting LEE20-203 fragment (WT) or derivatives with different single base substitutions, indicated on the X-axis. Measurements were made in cells containing either pSI02 (+GrlA: shaded bars) or empty vector, pACYC184ΔHN (-GrlA: open bars).

Fig. 9. Analysis of GrlA-dependent activation using suppression genetics.

A. The figure shows the GrlA amino acid sequence. The helix-turn-helix DNA binding motif predicted by Deng *et al.* (2004), and confirmed by the Dodd & Egan, GYM2.0 and Jpred prediction tools (see experimental procedures), is indicated by gray shading. Substitutions at R53 and I59, identified after random mutagenesis and selection for suppressors of the 92A mutation in the *LEE1* P1 promoter, are indicated. The number adjacent to each substitution is number of times that it was isolated.

B. The figures show the effect of GrlA substitutions on the activity of *LEE1* P1 promoter mutants. The bar charts illustrate β-galactosidase activities in M182 cells carrying pRW224 containing the LEE20-203 fragment with single base substitutions, indicated on the X-axis. Measurements were made in cells containing either pSI02 encoding wild type GrlA (WT, open bars) or GrlA with the RK53, IT59 or IV59 substitutions, as indicated (grey bars).

Fig. 10. Epistasis analysis of GrlA interactions.

The figure shows bar charts that illustrate measured β-galactosidase activities in M182 cells carrying pRW224 containing the LEE20-203 fragment with the starting *LEE1* P1 promoter sequence (panel A), and with the 89G (panel B), 90C (panel C) or 91C (panel D) mutations. Measurements were made as in Figure 5 with cells carrying pSI02 encoding wild type GrlA (WT) or the different alanine substitutions indicated on the X-axis. For each promoter, the data are expressed as a percentage of the activity with wild type, the values are the average of three independent assays, and standard errors are shown with error bars.

Fig. 11. Model for activation of *LEE1* P1 promoter by GrlA.

A. Weak recognition of the promoter by RNA polymerase due to sub-optimal spacer length.

B. Efficient recognition of the promoter by RNA polymerase due to optimised spacer length.

C. GrlA interacts with the spacer sequence and facilitates RNA polymerase activity.

Table 1: Mutational analysis of the *LEE1* P1 promoter.

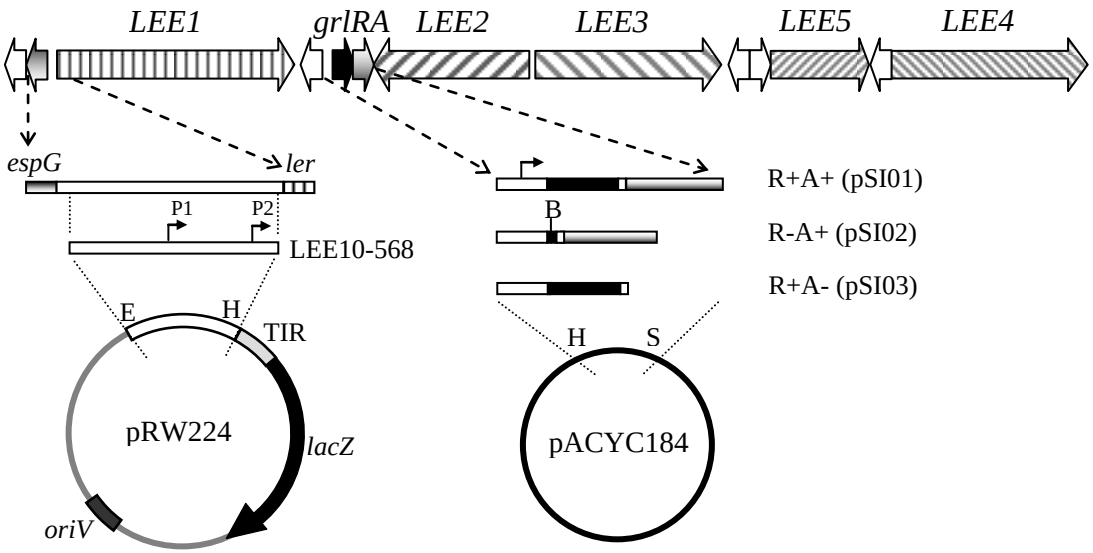
Fragment	Promoter sequence from positions 71-102	β-galactosidase activity (Miller units ± SD)
Starting fragment	71 80 90 100	
LEE20-275	5′-TGTTGACATTTAATGATAATGTATTTTACACA-3′	5477 ± 300
Point mutation upstream of the -35 element		
LEE20-275 71A	5′-**A**GTTGACATTTAATGATAATGTATTTTACACA-3′	1034 ± 123 (18.9)
Point mutations in the -35 element (TTGACA)		
LEE20-275 73A	5′-TG**A**TGACATTTAATGATAATGTATTTTACACA-3′	1103 ± 20 (20.1)
LEE20-275 73C	5′-TG**C**TGACATTTAATGATAATGTATTTTACACA-3′	1128 ± 28 (20.6)
LEE20-275 74C	5′-TGT**C**GACATTTAATGATAATGTATTTTACACA-3′	991 ± 90 (18.1)
LEE20-275 75A	5′-TGTT**A**ACATTTAATGATAATGTATTTTACACA-3′	690 ± 27 (12.6)
LEE20-275 75C*	5′-TGTT**C**ACATTTAATGATAATGTATTTTACACA-3′	962 ± 141 (17.6)
LEE20-275 76G	5′-TGTTG**G**CATTTAATGATAATGTATTTTACACA-3′	1478 ± 120 (27.0)
LEE20-275 77A*	5′-TGTTGA**A**ATTTAATGATAATGTATTTTACACA-3′	1532 ± 40 (28.0)
LEE20-275 77G*	5′-TGTTGA**G**ATTTAATGATAATGTATTTTACACA-3′	1136 ± 80 (20.7)
LEE20-275 77T	5′-TGTTGA**T**ATTTAATGATAATGTATTTTACACA-3′	1185 ± 60 (21.6)
Point mutations in the spacer region		
LEE20-275 93T	5′-TGTTGACATTTAATGATAATGT**T**TTTTACACA-3′	812 ± 86 (14.8)
LEE20-275 InsT (93-94)	5′-TGTTGACATTTAATGATAATGTA**T**TTTTACACA-3′	205 ± 26 (3.7)
LEE20-275 Δ94T*	5′-TGTTGACATTTAATGATAATGTA—TTTACACA-3′	11622 ± 291 (212.2)
Point mutations in the -10 element (TACACA)		
LEE20-275 97C	5′-TGTTGACATTTAATGATAATGTATTT**C**ACACA-3′	318 ± 12 (5.8)
LEE20-275 98C*	5′-TGTTGACATTTAATGATAATGTATTTT**C**CACA-3′	233 ± 4 (4.3)
LEE20-275 98G	5′-TGTTGACATTTAATGATAATGTATTTT**G**CACA-3′	202 ± 15 (3.7)
LEE20-275 98T	5′-TGTTGACATTTAATGATAATGTATTTT**T**CACA-3′	234 ± 9 (4.3)
LEE20-275 99A*	5′-TGTTGACATTTAATGATAATGTATTTTA**A**ACA-3′	10169 ± 298 (185.7)
LEE20-275 99G*	5′-TGTTGACATTTAATGATAATGTATTTTA**G**ACA-3′	2721 ± 16 (50.0)
LEE20-275 100G	5′-TGTTGACATTTAATGATAATGTATTTTAC**G**CA-3′	305 ± 9 (5.6)
LEE20-275 101G*	5′-TGTTGACATTTAATGATAATGTATTTTACA**G**A-3′	2493 ± 132 (45.5)
LEE20-275 102C*	5′-TGTTGACATTTAATGATAATGTATTTTACAC**C**-3′	4375 ± 105 (80.0)
LEE20-275 102G*	5′-TGTTGACATTTAATGATAATGTATTTTACAC**G**-3′	1843 ± 72 (33.6)

The table shows measured β-galactosidase activities in cultures of *E. coli* strain M182 carrying pRW224 containing the LEE20-275 fragment and different mutations. Cultures were grown aerobically at 37°C in LB medium to an optical density of ~0.5 at 650 nm. Activities were measured in triplicate, giving a mean and standard deviation (SD). Activities expressed as a percentage of activity with the starting LEE20-275 fragment are shown in parentheses. The central part of the table shows the fragment base sequence from position 71 to position 102, with the P1 promoter -10 and -35 hexamer elements underlined. Base substitutions and insertions in the different fragments are highlighted in boldface type whilst the Δ94 deletion is shown by a dash. Mutations made by site directed mutagenesis are indicated by asterisks, whilst the other mutations came from the random PCR mutagenesis experiment illustrated in Figure 3.

Table 2: Effects of spacer length on activation by GrlA at the *LEE1* P1 promoter.

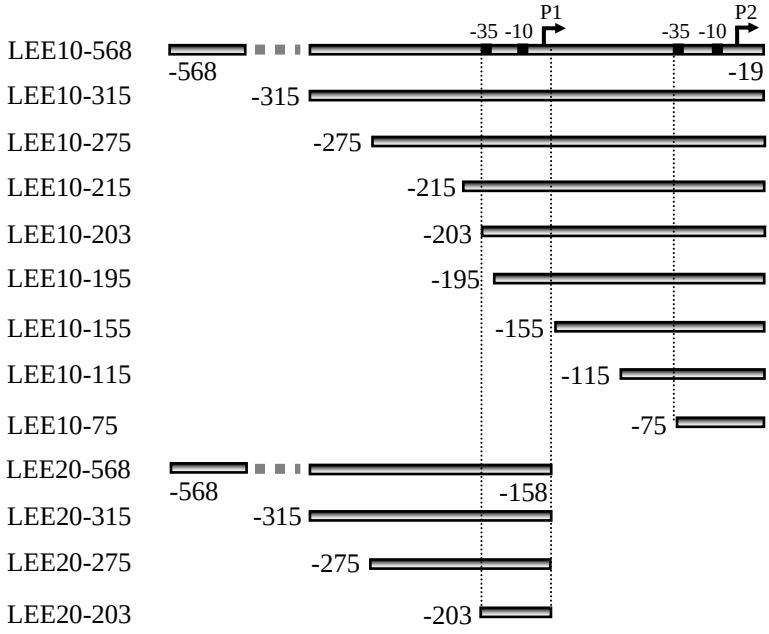
Derivative	Promoter sequence from positions 73-102	Spacer length (bp)	β-galactosidase activity (Miller units ± SD)	
			- GrlA	+GrlA
	73 80 90 100			
LEE20-203 (WT)	5'-TTGACATTTAATGATAATGTATTTTACACA-3'	18	1479 ± 36	4797 ± 140 (3.2)
Δ79T	5'-TTGACA–TTAATGATAATGTATTTTACACA-3'	17	6255 ± 87	7213 ± 149 (1.2)
Δ94T	5'-TTGACATTTAATGATAATGTA–TTTACACA-3'	17	3918 ± 79	4097 ± 55 (1.0)
Δ79/80T	5'-TTGACA––TAATGATAATGTATTTTACACA-3'	16	473 ± 21	491 ± 29 (1.0)
InsT (78-79)	5'-TTGACA**T**TTTAATGATAATGTATTTTACACA-3'	19	1054 ± 123	947 ± 12 (1.0)
InsT (93-94)	5'-TTGACATTTAATGATAATGTA**T**TTTTACACA-3'	19	588 ± 6	621 ± 13 (1.0)

The table shows measured β-galactosidase activities in cultures of *E. coli* strain M182 carrying pRW224 containing the LEE20-203 fragment and different derivatives with insertions or deletions in the P1 promoter spacer. Cells also contained either plasmid pACYC184ΔHN (-GrlA) or pSI02 (+GrlA). Cultures were grown aerobically at 37°C in LB medium to an optical density of ~0.5 at 650 nm. Activities were measured in triplicate, giving a mean and standard deviation (SD). Fold activation by GrlA is indicated in parentheses after the data. The central part of the table shows the fragment base sequence from position 73 to position 102, with the P1 promoter -10 and -35 hexamer elements underlined. Base deletions in the different fragments are indicated by dashes whilst insertions are shown in boldface type.

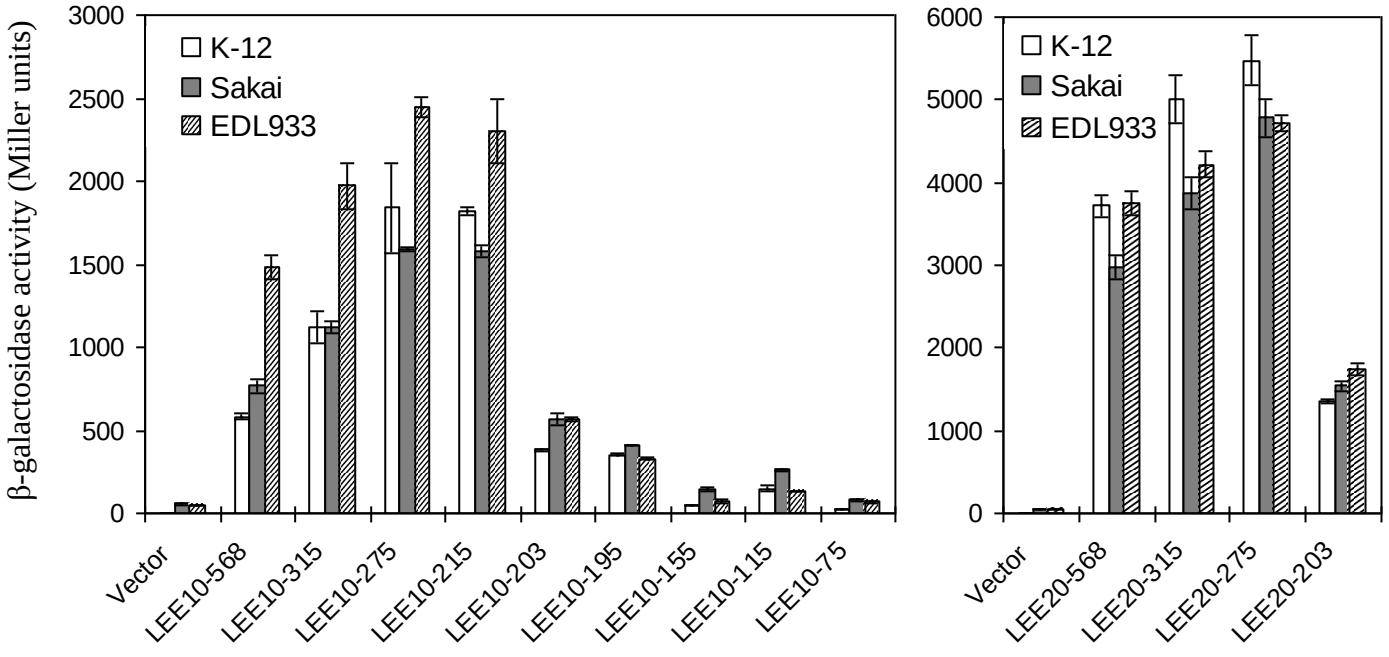


Islam *et al*. Figure 1

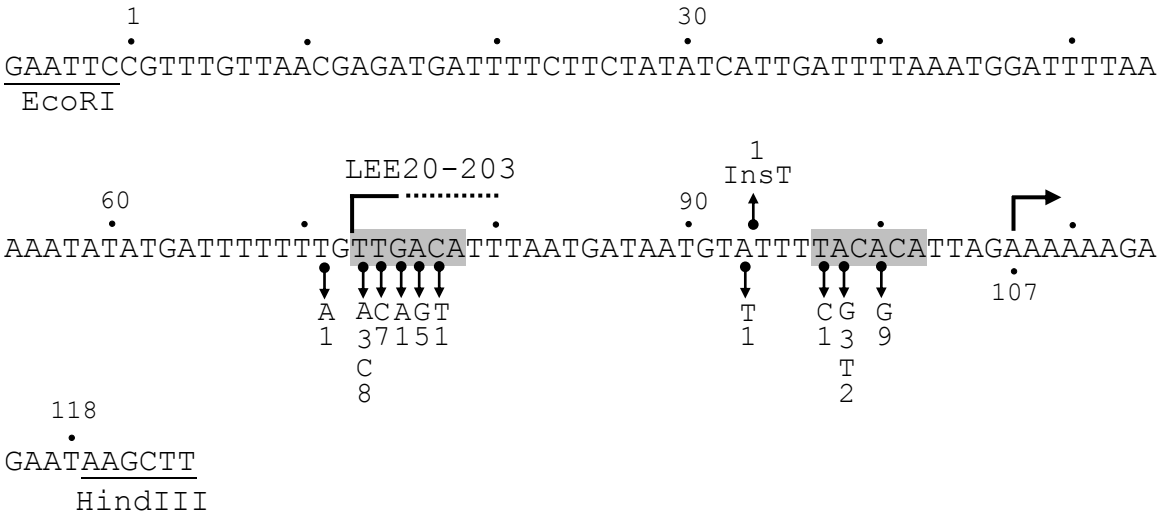
(A)



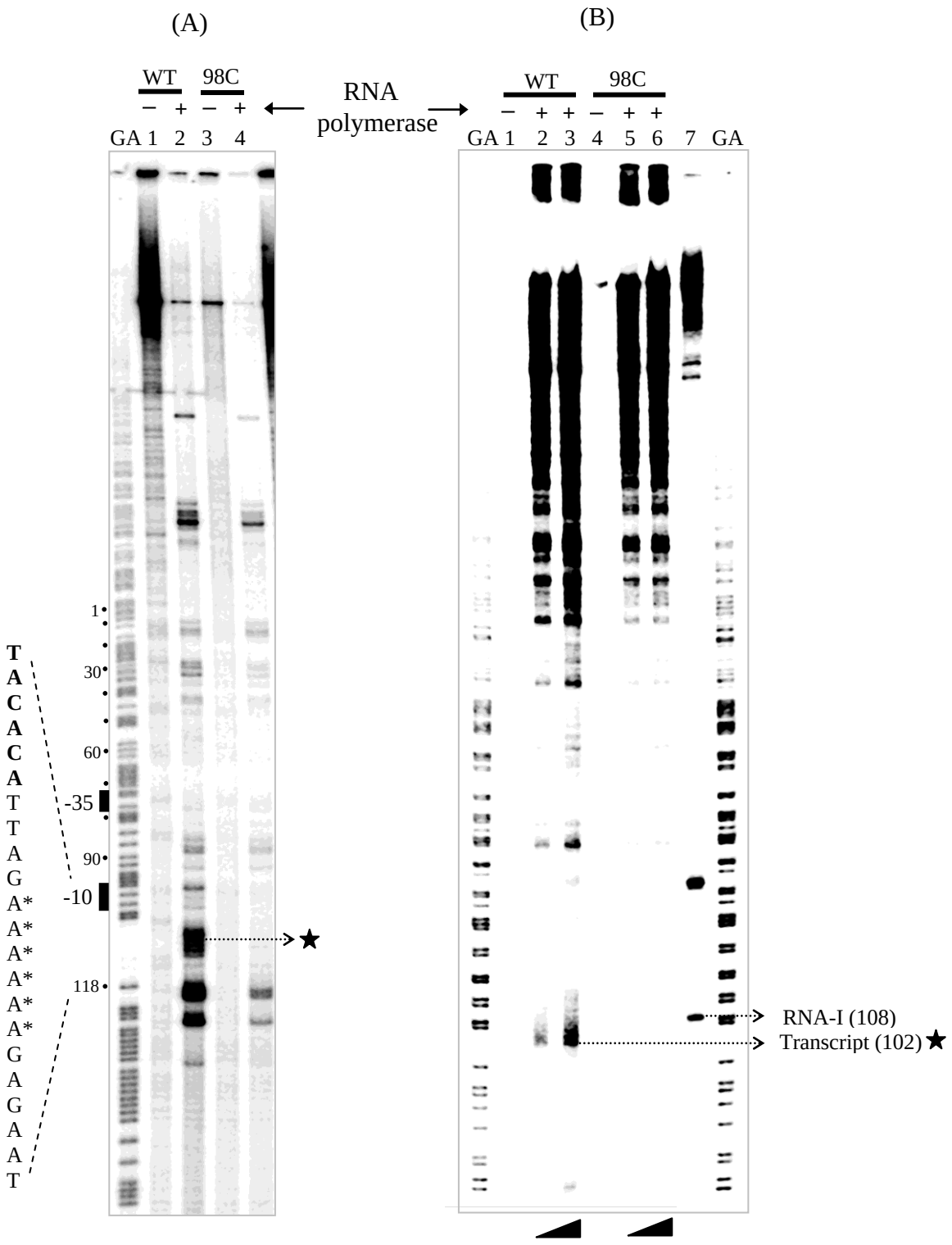
(B)



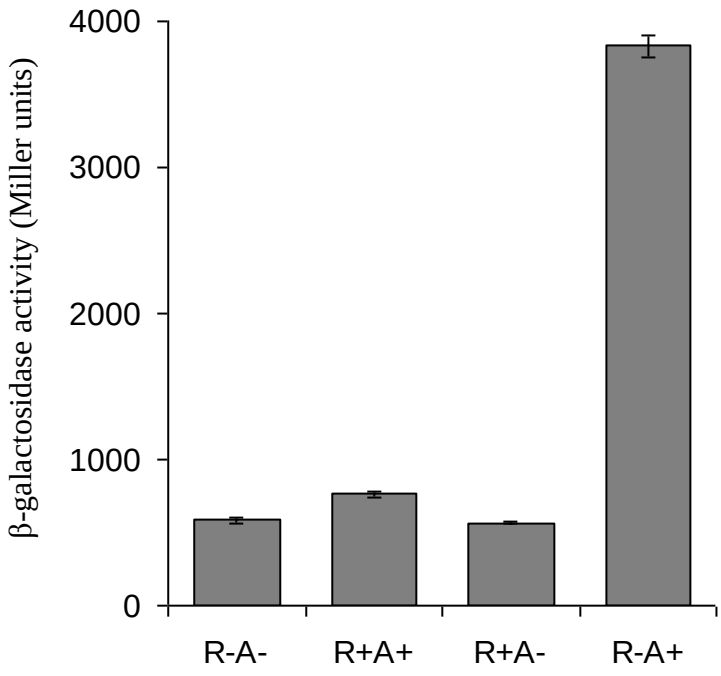
Islam *et al.* Figure 2



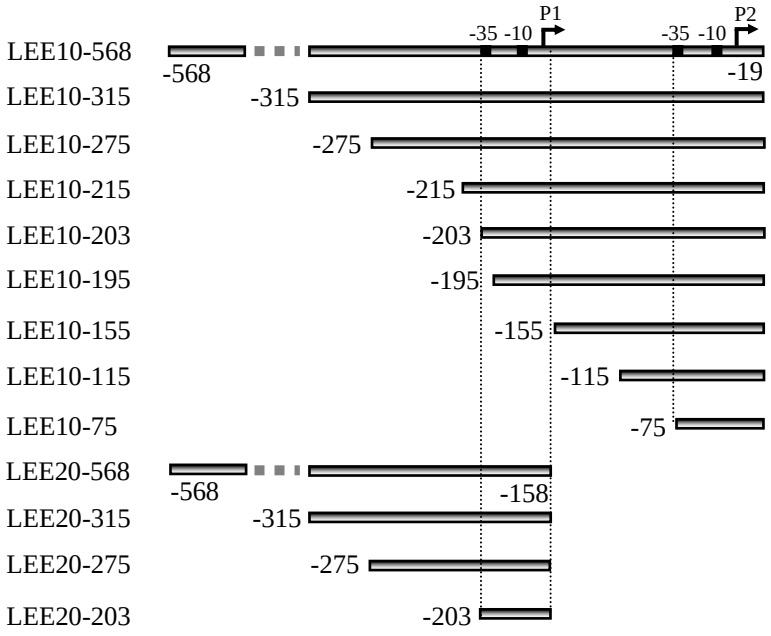
Islam *et al.* Figure 3



Islam *et al.* Figure 4

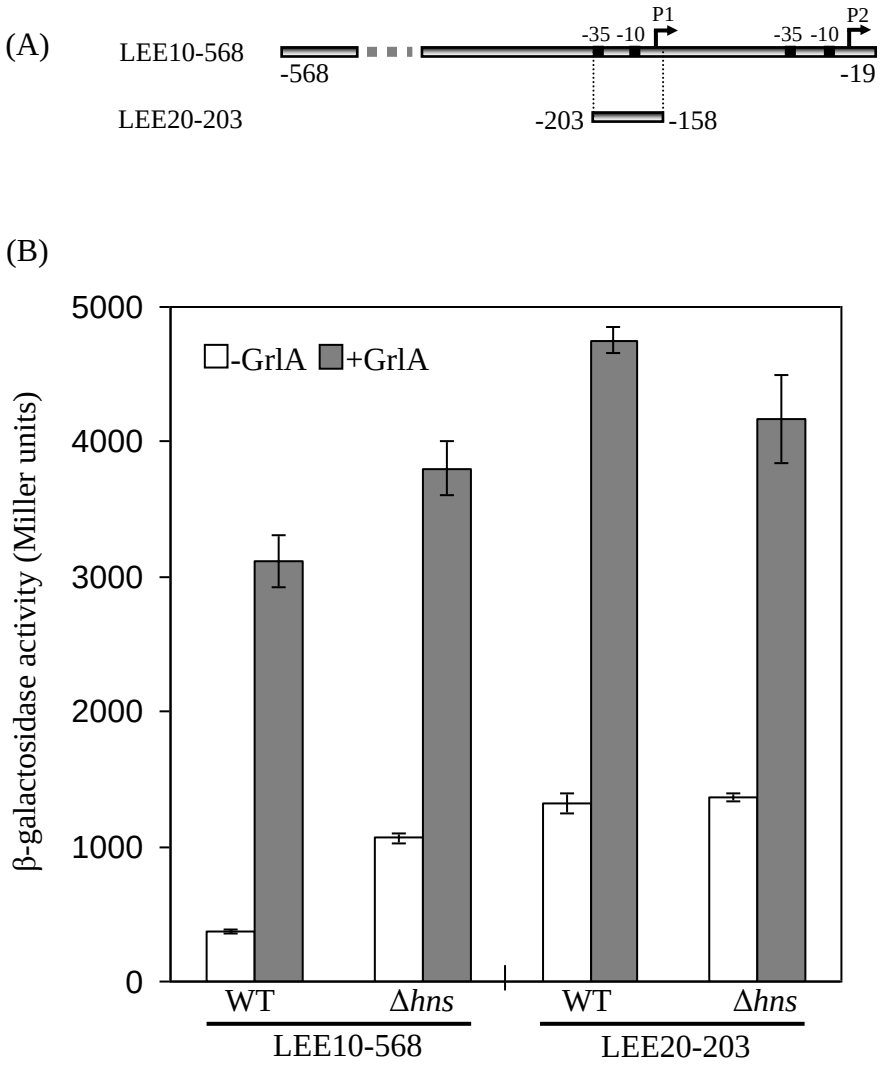


Islam *et al.* Figure 5

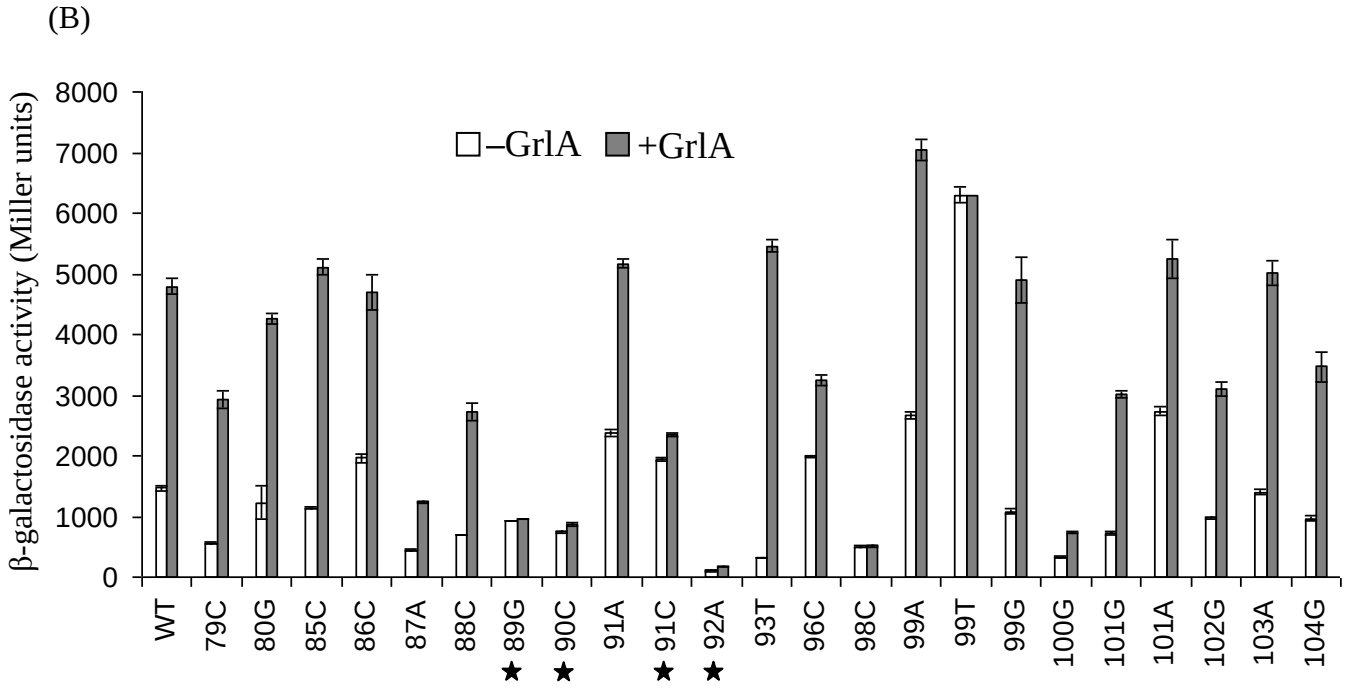


Expression from different promoters β-galactosidase activity (Miller Units ± Standard Deviation)		
-GrlA	+GrlA	Fold activation
584 ± 21	3831 ± 72	6.6
1120±93	5325±81	4.8
1838±273	6178±162	3.4
1821±28	5848±231	3.2
381±5	2278±260	6
354±5	356±3	1
49±1	49±1	1
146±17	153±17	1
26±0	24±1	1
3721±130	7645±67	2.1
5012±297	10772±616	2.1
5477±300	9125±237	1.7
1353±16	4448 ± 169	3.3

Islam *et al.* Figure 6

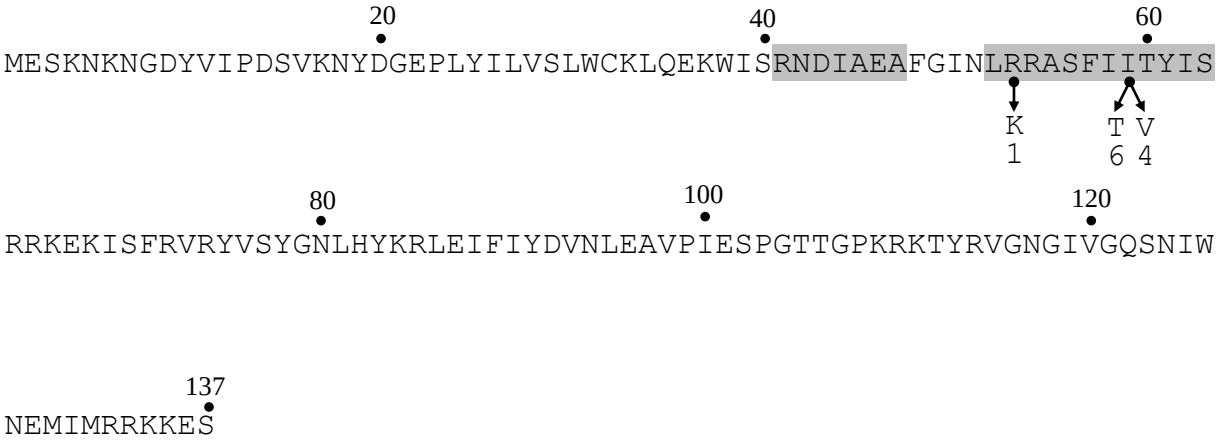


Islam *et al.* Figure 7

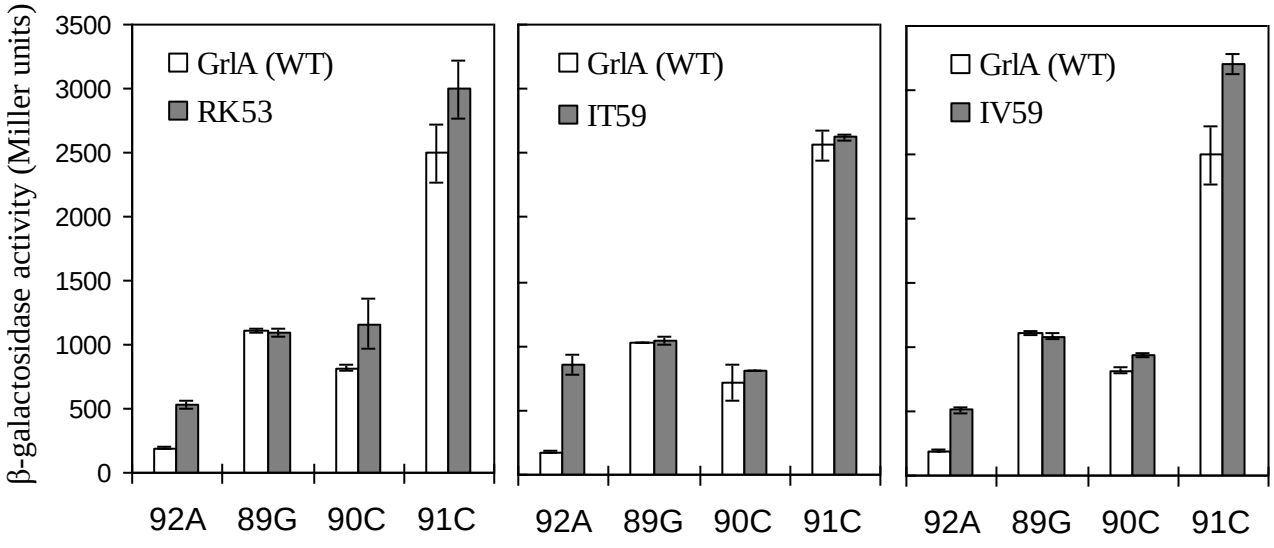


Islam *et al.* Figure 8

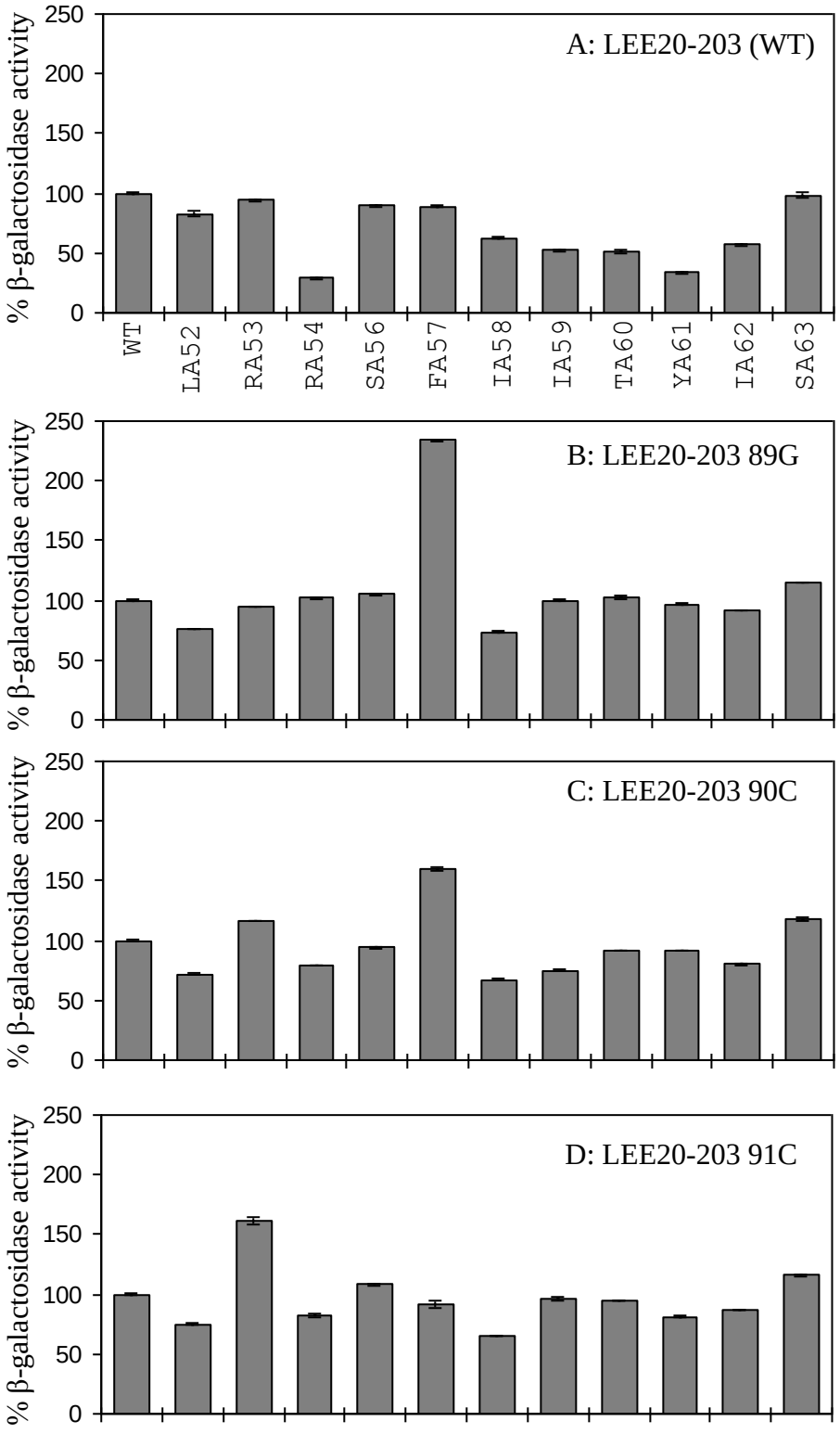
(A)



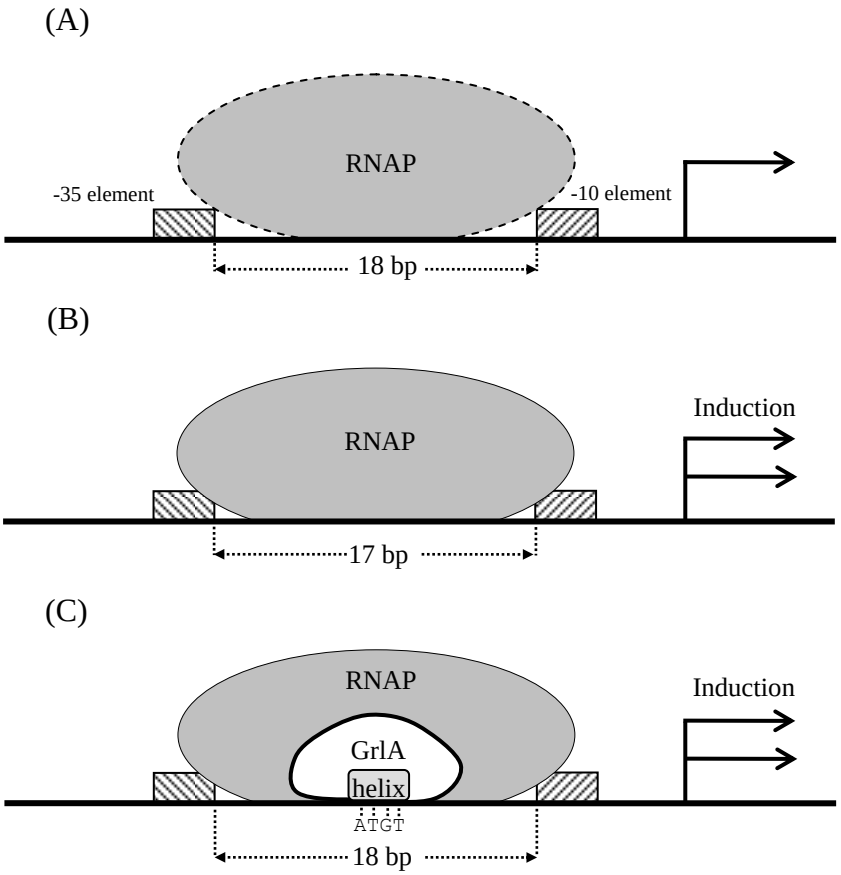
(B)



Islam *et al.* Figure 9



Islam *et al.* Figure 10



Islam *et al.* Figure 11