A multifactor regulatory circuit involving H-NS, VirF and an antisense RNA modulates transcription of the virulence gene icsA of Shigella flexneri

Chi Nhan Tran1, Mara Giangrossi1, Gianni Prosseda2, Anna Brandi1, Maria Letizia Di Martino2, Bianca Colonna2 and Maurizio Falconi1,*

1Laboratory of Molecular Genetics, School of Bioscience and Biotechnology., University of Camerino, 62032 Camerino (MC) and 2Istituto Pasteur-Fondazione Cenci Bolognetti, Department of Biology and Biotechnology ‘C. Darwin’, Sapienza, University of Rome, 00185 Roma, Italy

ABSTRACT
The icsA gene of Shigella encodes a structural protein involved in colonization of the intestinal mucosa by bacteria. This gene is expressed upon invasion of the host and is controlled by a complex regulatory circuit involving the nucleoid protein H-NS, the AraC-like transcriptional activator VirF, and a 450 nt antisense RNA (RnaG) acting as transcriptional attenuator. We investigated on the interplay of these factors at the molecular level. DNase I footprints reveal that both H-NS and VirF bind to a region including the icsA and RnaG promoters. H-NS is shown to repress icsA transcription at 30°C but not at 37°C, suggesting a significant involvement of this protein in the temperature-regulated expression of icsA. We also demonstrate that VirF directly stimulates icsA transcription and is able to alleviate H-NS repression in vitro. According to these results, icsA expression is derepressed in hns- background and overexpressed when VirF is provided in trans. Moreover, we find that RnaG-mediated transcription attenuation depends on 80 nt at its 5’-end, a stretch carrying the antisense region. Bases engaged in the initial contact leading to sense–antisense pairing have been identified using synthetic RNA and DNA oligonucleotides designed to rebuild and mutagenize the two stem–loop motifs of the antisense region.

INTRODUCTION
Shigella is a highly adapted human pathogen that causes the bacillary dysentery. This enteric syndrome, mainly found in the developing world, is extremely contagious and, although self-limiting, may be life threatening and often fatal in children (1). Shigellosis is the result of the coordinated action of several genes encoded by both the large virulence plasmid (pINV) and the chromosome. Upon host invasion, the primary event consequent to the upshift of bacteria to the host temperature (37°C) is the synthesis of two plasmid-encoded regulators: VirF and VirB. First, VirF, an AraC-like activator, triggers a regulatory cascade involving the activation of the icsA and virB genes. Then, VirB activates several operons-encoding factors required to invade and colonize the intestinal mucosa (2,3).

At this stage, an important virulence factor is represented by the icsA protein (also known as VirG), which sponsors the intra- and intercellular spread of bacteria (4). IcsA is an outer membrane protein, which asymmetrically accumulates at one pole of the bacterial cell, catalyzing the direct elongation of an actin tail which propels Shigella through the cytoplasm and facilitates the intercellular dissemination toward neighboring cells (5). Due to its prime role in the invasion process, it is reasonable to expect the regulation of the icsA gene to be a sophisticated multilayered process. VirF, the master activator of the plasmid-encoded virulence cascade, has been shown to induce the expression of icsA in vivo (2,6,7), although little is known on its molecular mechanism. Another player in the regulation of icsA is the nucleoid associated protein H-NS, implicated in chromosome organization and also acting as a global regulator (8–10). H-NS has been shown to repress icsA at low temperature (11) and to specifically interact with the icsA promoter (12). This protein also binds and represses the transcription of virF and virB as a function of temperature, playing a direct role in silencing the Shigella virulence regulon outside the host (13–16).

*To whom correspondence should be addressed. Tel: +39 737 403265; Fax +39 737 40390; Email: maurizio.falconi@unicam.it

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
We have recently discovered a small non-coding antisense RNA of ∼450 nt, termed RnaG, controlling icsA expression and transcribed in cis from the icsA sequence (17). Small RNAs are emerging as a powerful tool in gene regulation and particularly in modulating adaptive stress—responses that are important for bacteria to survive within the host (18–21). Different RNA–RNA-mediated mechanisms, entailing changes in processing and degradation of the target messenger and alterations of the efficiency of transcription and translation have been described (22–24). RnaG regulates icsA transcription by two independent mechanisms; transcriptional interference and transcriptional attenuation. As for the first mechanism, we have recently shown that silencing of the RnaG promoter results in a higher activity of the icsA promoter (17). On the other hand, we also have shown that RnaG represents the first evidence of a RNA-mediated transcriptional attenuation of a bacterial virulence gene. In fact, binding of RnaG to the nascent icsA transcript induces the formation of a stem–loop motif at the 5’-end of the mRNA, mimicking an intrinsic terminator and resulting in the synthesis of a truncated icsA mRNA (17).

Here, we report on the identification of the H-NS and VirF-binding sites on the icsA and RnaG promoter regions and on the in vivo and in vitro relevance of their interactions on the expression of icsA. VirF directly stimulates icsA transcription while H-NS contributes to the thermoregulation of icsA, severely reducing its transcription at 30°C. We also found that the RnaG-mediated premature termination of the icsA transcript depends on about 80 out of 120 nt of the antisense region of RnaG. In particular, by the use of synthetic RNA and DNA molecules and a site-specific mutagenesis approach, we have been able to identify the RnaG regions initially contacting with the icsA mRNA to form the so-called kissing complex and needed for the transcriptional attenuation.

MATERIALS AND METHODS

Bacterial strains and general procedures

Escherichia coli K12 strain HMG9 is a HMG11 derivative defective in the hns gene. Strains ULS1127 and ULS129 are P90C monolysogens for a λ transducing phage carrying a P_{icsA-lacZ} transcriptional fusion with a functional or silenced P_{RnaG} convergent promoter, respectively (17). Proteins were expressed in E. coli strain BL21(DE3) containing the pLysS plasmid (Promega).

Plasmid pMYSH6601 is a pBR322-derivative containing the virA-icsA region of S. flexneri 2a pINV (31). A 478-bp icsA fragment (from positions −262 to +216) was amplified by PCR using the primer pair G2Z4/ACCC9 on the pMYSH6601 template and cloned into the BamHI site of pKK232-8, thus obtaining plasmid pKG450.

Plasmids pGT1127 is a pGEMT-easy derivative containing a 866 bp from the virA-icsA region (−448 to +419) of the S. flexneri 2a. Plasmids pGT1129 and pGT1083 contain the same fragment lacking the P_{RnaG} or P_{icsA} promoter, respectively.

Plasmid pGT80T contains the first 80 bp of the sequence-encoding RnaG followed by a transcription termination signal. It has been obtained by cloning into pGEMT-easy a fragment obtained by amplifying pGT1127 DNA with the oligo pair G+120H/G40T (or ACC9/G40T).

pMYSH6504 and pDIA510 plasmids are both pBR322 derivatives containing, respectively, the entire virF gene from S. flexneri 2a pINV (25) and the entire hns gene from E. coli (26).

β-galactosidase assays were performed with sodium dodecyl sulfate/chloroform-permeabilized cells grown in LB medium supplemented with ampicillin. β-galactosidase activity was determined as described by Miller (1992) (48) and the results were expressed as averages of three independent experiments.

Extraction of plasmid DNA, restriction digestion, electrophoresis, PCR and purification of DNA fragments were carried out as described previously (16). RNA extraction was carried out as previously described by La Teana et al. (27) except that hot phenol and chloroform:isoamyl alcohol (24:1) extractions were performed in place of purification in CsCl.

Radioactivity associated with DNA or RNA was detected and quantified by Molecular Imager (Bio-Rad, mod. FX) and oligos used in this study are listed in Supplementary Data and Table S1.

VirF purification

To obtain overexpression of VirF, we constructed pULS22F, a plasmid containing the entire virF gene of S. flexneri 2a. To this end, an ampiclon obtained upon PCR amplification of the template plasmid pMYSH6504 DNA (25) with oligo pair FsinL and SW04 (modified to contain NdeI or XhoI sites, respectively) was introduced into the pET22b vector (Novagen) linearized by NdeI/XhoI digestion. pULS22F was then transformed into BL21(DE3) pLysS strain selecting for Apr. Expression of the C-terminally His-tagged VirF protein was induced with 1 mM IPTG in exponentially growing 500 ml of culture. Bacterial cells were then sonicated and the protein was purified under native conditions using immobilized-metal affinity chromatography (IMAC) with a Ni-nitrilotriacetic acid-agarose column according to the manufacturer protocol (Qiagen). The protein was finally eluted at 1 ml fractions with 8-ml elution buffer (50 mM Na_2HPO_4, 300 mM NaCl, 10% Glycerol, 0.1% Tween-20, 10 mM β-mercaptoethanol) containing from 50 to 500 mM imidazole. Fractions were analyzed by SDS–PAGE and those containing the VirF protein were pooled and dialyzed against 11 of 20 mM Tris–HCl, pH 8, 30 mM Na_2HPO_4, 50 mM NaCl, 5% glycerol. The His-tagged wild type VirF protein was found to be fully functional in vivo (data not shown).

RnaG80 synthesis and purification

Plasmid pTZRnaG80 was obtained by cloning 80 nt of the RnaG gene (from positions +40 to +120 with respect to the icsA transcription start) into the HindIII/EcoRI restriction sites of pTZ19R, under the control of the T7
promoter. This DNA fragment was obtained by PCR using pGT1127 as DNA template and the primer G+120H and G+40E. To synthesize RnaG80 transcript, pTZRnaG80 was linearized with EcoRI and used as template in an in vitro transcription reaction with T7 RNA polymerase as described by Brandi et al. (28).

**DNase I footprinting**

Supercoiled plasmids pKG673 (17) and pKG450 (~150 ng/sample) were preincubated 5 min at 25°C with the indicated concentrations of H-NS or VirF in 30 μl of binding buffer (40 mM HEPES–HCl, pH 8.0, 100 mM KCl, 10 mM magnesium acetate and 0.5 mM dithiothreitol). The DNA–protein complex was incubated with 1 U of DNase I for 40 s. After stopping the reaction, the DNA was precipitated and extended as described by Giangrossi et al. (29).

**In vitro transcription**

**In vitro** transcription from supercoiled plasmids (~200 ng/sample) as DNA templates was carried out at 30 and at 37°C for 20 min. Each reaction mixture (40 μl) contained 40 mM Tris–HCl, pH 7.5, 150 mM KCl, 10 mM MgCl2, 10 mM dithiothreitol, 0.01% Triton X-100, 0.5 mM each of NTPs, 2 U of ribonuclease inhibitor and 0.2 U of E. coli RNA polymerase (USB). The reaction was stopped on ice and RNA was precipitated with ethanol in the presence of 1 μg tRNA as carrier. The transcription product was detected by primer extension using a [γ-32P]-labeled oligonucleotide essentially as described by Giangrossi et al. (17).

**Chemical RNA probing**

Chemical modification of RNA was performed as described by Giangrossi et al. (17) using the single-strand-specific reagents DMS (A and C specific) and CMCT (U and G specific). Control samples (−) were treated identically with the exception that no modifying reagents were added. The modified RNA was subjected to primer extension using the oligo G+40E.

**RESULTS**

**H-NS and VirF modulate transcription of icsA and RnaG by a direct interaction at promoter regions**

Although H-NS and VirF are known to be involved in the regulation of virulence genes in *S. flexneri* (2,30), very little is known on their role in the modulation of icsA expression, especially at the molecular level. While band shift assays indicate that H-NS recognizes the icsA promoter (12), no evidence supports a binding of VirF to the icsA promoter region.

We investigated on the interaction of H-NS and VirF with the regulatory region of icsA by DNase I footprinting. H-NS protects three sites, numbered I, II and III and centered at positions +25, +115 and +330, respectively (Figure 1A and B), while VirF recognizes four regions, numbered I, II, III and IV and centered at positions −3, +62, +160 and +245, respectively (Figure 1C and D). DNase I protected regions are reported on the schematic map shown in Figure 1E. H-NS-binding sites have an average length of ~50 bp and, within the limit of accuracy of this type of analysis, were bound with similar affinities by H-NS (~250–500 nM). As opposed to other H-NS-binding sites have been mapped on upstream regulatory regions (13–15), in the case of the icsA promoter H-NS binds downstream the transcriptional initiation site. Interestingly, H-NS site II includes the −10 conserved promoter element of RnaG whose occlusion might be relevant for transcriptional inhibition of P_{RnaG} by H-NS (see below). VirF sites, spanning from ~40 to 60 bp in length, are detectable at low protein concentrations (~80–130 nM). VirF protection at site I covers the TATA box of P_{RnaG} and partially overlaps the left end of H-NS site I, while the right end of the latter site is superimposed on the VirF site II. Moreover, H-NS bound to site II and VirF bound to site III may contact the RnaG promoter (Figure 1E).

The position of H-NS and VirF-binding sites within the icsA promoter region suggests that both genes icsA and RnaG might be controlled by these regulators and provides a physical basis for a possible functional competition between the two proteins. Initially, we analyzed the effect of H-NS on the transcription of icsA and RnaG at 30 and 37°C in an in vitro system (Figure 2), using plasmids pGT1127 (wt), pGT1129 (inactive P_{RnaG}) and pGT1083 (inactive P_{icsA}) as DNA templates. The latter two constructs carry mutations in the TATA box of RnaG and icsA, respectively, therefore the transcription of these genes is abolished. This allowed us to study the transcription of these genes discriminating the effects caused by H-NS from those due to the transcriptional interference (TI) between the convergent promoters of icsA and RnaG (17). At 30°C, H-NS is able to inhibit transcription of both icsA and RnaG. Such repression is more pronounced on P_{icsA} than on P_{RnaG}. At ~200 nM H-NS, the icsA mRNA is no longer detectable and transcription attains ~5% of the basal level obtained in the absence of protein (Figure 2B and D). Double amounts of H-NS (300–400 nM) are required to produce a comparable effect on RnaG transcription (Figure 2A and D). At 37°C, H-NS almost completely loses its capacity to act as transcriptional silencer, strongly suggesting that this protein plays a key role in the temperature-dependent regulation of icsA and RnaG genes. According to TI, the inactivation of icsA promoter stimulates the transcriptional activity of the convergent RnaG promoter. Under this condition of hyper-expression, RnaG seems less susceptible to H-NS repression even at 30°C. In fact, at 234 and 312 nM H-NS transcription is, respectively, 85 and 40% of the level without protein (Figure 2C and D).
VirF concentrations the activity of the icsA promoter is strongly stimulated attaining an ~10-fold peak at 17 nM VirF. Transcription returns to basal levels when an excess of VirF is used (34 and 51 nM), possibly due to VirF occupying all its target sites and to the consequent formation of a transcriptionally inactive DNA–protein complex. As opposed to icsA, transcription of RnaG displays a progressive decrease with increasing VirF concentration: the activity of the RnaG promoter attains 58 and 22% in the presence of 34 and 51 nM VirF, respectively (Figure 3B and C). Taken together, these results indicate that VirF is able, depending on concentration, to stimulate icsA and repress PRnaG.

Since both H-NS and VirF bind the icsA-RnaG locus (Figure 1), and antagonistically influence their transcription (Figures 2 and 3), an active competition between these regulators can be surmised. Therefore, this hypothesis has been tested in vitro by monitoring icsA promoter activity in the presence of both H-NS and VirF. The results indicate that VirF is able to significantly counteract the H-NS-dependent inhibition, thus almost completely relieving the transcriptional repression of icsA at the lowest H-NS concentration tested (Supplementary Figure S2).

**H-NS and VirF modulate the in vivo expression of icsA and RnaG**

The in vitro functional assays (Figures 2 and 3) suggest that H-NS is able to inhibit transcription of both icsA and RnaG, as a function of temperature, while VirF behaves as an activator of icsA. In addition, the basal activities of Prcsa and PrnaG (they are convergent and 120 nt apart) are not completely independent, since these promoters respond to transcriptional interference (TI) regulation (17). Hence, we investigated on the possible interplay among H-NS, VirF, temperature and TI in vivo. This was done by monitoring both the level of icsA mRNA and the expression of β-galactosidase from Prcsa-lacZ transcriptional fusions carrying or lacking a functional PrnaG promoter (Figure 4A and B).
The icsA promoter displays a natural thermoregulation and its transcription is higher at 37°C (20 A.U.) than at 30°C (11 A.U.) also in cells lacking an active H-NS. Moreover, a marked regulatory action can be attributed to this protein since mutations of the hns gene cause a further de-repression of icsA independently of temperature.

Indeed, when both icsA and RnaG are transcribed (i.e. using plasmid pGT1127) the steady-state level of the icsA mRNA in hns defective strains increases about 3-fold as compared to the wt (from 4 to 11 A.U. at 30°C and from 7 to 20 A.U. at 37°C). In agreement with the transcriptional pattern of the icsA mRNA, the activity of the lacZ reporter gene in the E. coli strain ULS1127, carrying a single chromosomal copy of the PicsA-lacZ fusion, shows an ~3-fold reduction in cells overproducing H-NS from the pDIA510 plasmid. As expected, H-NS-mediated repression is more pronounced at 30°C than at 37°C (Figure 4B). Similar conclusions about the role of H-NS in the temperature-mediated regulation of icsA can
be drawn using the pGT1129 construct, which carries a modified icsA regulatory region preventing RnaG synthesis. Under these conditions, the basal activity of P\textsubscript{icsA} is increased due to the absence of TI (compare hns\textsuperscript{+} cells transformed with pGT1129 or pGT1127 in Figure 4A). Consistently, TI also accounts for the higher expression of icsA in ULS1129 (an E. coli strain carrying a P\textsubscript{icsA}-lacZ fusion with an inactivated RnaG promoter) as compared to ULS1127. This is particularly evident in strains overproducing neither H-NS nor VirF (Figure 4B). The target of H-NS inhibition is mainly icsA rather than RnaG, in agreement with the results shown in Figure 2. Indeed, in an hns defective background, as an elevated synthesis of icsA mRNA is achieved (11–20 A.U.) RnaG becomes undetectable (Figure 4A). It is reasonable to assume that this is the outcome of TI, which leads to decreased accessibility of P\textsubscript{RnaG} to RNA polymerase when P\textsubscript{icsA} is overexpressed. Finally, as shown in Figure 4B, providing VirF in \textit{trans} by introducing plasmid pMYSH6504, results in a stimulation of icsA expression. This activation mainly takes place at 37°C and in strain ULS1127, i.e. in the presence of a functional RnaG promoter.

Establishment of a kissing complex between sense and antisense RNAs mediates the transcriptional attenuation of icsA by RnaG

To gain deeper insight into the transcriptional attenuation mechanism governing the RnaG-mediated repression of icsA, we have studied how RnaG interacts with its target sequence on the icsA mRNA. RnaG is a 450 nt sRNA. We have demonstrated that its antisense region (RnaG120, i.e. nucleotides 1–120) still preserves the ability to repress \textit{icsA} transcription in \textit{vivo}, suggesting that this stretch constitutes a single functional domain. In particular, RNA probing showed that RnaG120 is highly structured and is characterized by three stem–loop motifs, GH1, GH2 and GH3 (17) (Supplementary Figure S3).

To identify the RnaG regions promoting the initial pairing with the icsA transcript, we first shortened RnaG120 by deleting 40 nt at its 3’-end, giving RnaG80. This smaller variant was then subjected to chemical probing using DMS and CMCT to verify whether the extended deletion had caused an extensive modification of the native structure. As seen in Figure 5A and B, the original structure of RnaG120 is essentially conserved in RnaG80; the GH1 and GH2 motifs are still present and structurally unchanged while the GH3 hairpin is lost.

This result prompted us to examine the ability of RnaG80 to repress \textit{icsA} transcription in \textit{vivo} and in \textit{vivo}. First, we compared the entire RnaG (expressed from pGT1083) to the shorter variant consisting of 80 nt (expressed from pGT80T) for their ability to inhibit the expression of a chromosomal single-copy P\textsubscript{icsA}-lacZ fusion (ULS119). As shown in Figure 6A both the full-length RnaG and RnaG80, induce a comparable decrease of \textit{icsA} expression. RnaG80 was added to the transcription mixture either at beginning (samples B) or at the end (samples E) of the reaction; this strategy allowed us to probe for differences in accessibility of PRnaG to RNA polymerase either at start (samples C) or at the end (samples F) of transcription.

To identify the RnaG regions promoting the initial pairing with the icsA transcript, we first shortened RnaG120 by deleting 40 nt at its 3’-end, giving RnaG80. This smaller variant was then subjected to chemical probing using DMS and CMCT to verify whether the extended deletion had caused an extensive modification of the native structure. As seen in Figure 5A and B, the original structure of RnaG120 is essentially conserved in RnaG80; the GH1 and GH2 motifs are still present and structurally unchanged while the GH3 hairpin is lost.

This result prompted us to examine the ability of RnaG80 to repress \textit{icsA} transcription in \textit{vivo} and in \textit{vivo}. First, we compared the entire RnaG (expressed from pGT1083) to the shorter variant consisting of 80 nt (expressed from pGT80T) for their ability to inhibit the expression of a chromosomal single-copy P\textsubscript{icsA}-lacZ fusion (ULS119). As shown in Figure 6A both the full-length RnaG and RnaG80, induce a comparable decrease of \textit{icsA} expression. RnaG80 was added to the transcription mixture either at beginning (samples B) or at the end (samples E) of the reaction; this strategy allowed us to
Figure 5. Secondary structure at 5'-end of RnaG. (A) Chemical probing of RnaG80 (nucleotides 1–80). Purified RnaG80 (2 pmol) was treated with increasing amounts of the single-strand specific reagents dimethyl sulfate (DMS) (0%, lanes 1 and 2; 0.22%, lane 3; 0.46%, lanes 4; 0.78%, lane 5) and 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate (CMCT) (0 mg/ml, lanes 6 and 7; 1.5 mg/ml, lane 8; 3 mg/ml, lane 9; 6 mg/ml, lane 10) as described in ‘Materials and Methods’ section. Modified nucleotides were detected by primer extension using the oligo G+40H and accessible sites were evaluated comparing samples incubated in the absence (−) and in the presence of either DMS or CMCT. Lanes C, T, A and G correspond to DNA sequencing ladders made with the same primer. (B) Schematic representation of the secondary structure of RnaG80. Numbering is according to 5'-end of the antisense RNA corresponding to position +120 on icsA sequence. Nucleotides reactive to DMS or CMCT in Panel A are circled. Chemical probing data were superimposed on a computer prediction generated by the MFOLD program (47).

Figure 6. RnaG80 is able to repress icsA expression. (A) The expression of a P_{wca}-lacZ fusion carried by E. coli strain ULS1129 was monitored in cells transformed with pGT80T, pGT1083 or pGT1129 as control. The data (±SD) represent the average of three independent experiments. (B) In vitro transcription was carried out at 37°C using 200 ng of supercoiled plasmid pGT1129 as template. Increasing amounts of RnaG80 were added to the transcription mixture either at the beginning of the reaction (prior to RNA polymerase addition, samples B) or at the end (immediately before RNA precipitation, samples E). The in vitro transcribed icsA mRNA was primer-extended with the oligo G+110.
discriminate the effect of the antisense RNA on transcription initiation from that on cDNA elongation. The assay showed that in the presence of low amounts of RnaG80 (up to 20 fmol), the icsA mRNA level is lower in samples B than in samples E. This result indicates that the shortened RnaG specifically causes termination of the icsA transcript. Higher amounts of RnaG80 (40 fmol) are required to interfere also with the detection of icsA mRNA by primer extension. By plotting icsA transcription, expressed as ‘sample B/sample E’ ratio, as a function of the amount of RnaG80 (Figure 7F), it can be seen that the icsA promoter activity attains 85, 64 and 34% at 5, 10 and 20 fmol of RnaG80, respectively. Since RnaG80 conserves its ability to properly function in vivo and in vitro, despite the loss of the GH3 hairpin, we can conclude that this motif is not required to interact with the icsA mRNA and that the first 80 nt confer RnaG the complete capability to terminate icsA transcription.

We then investigated which regions within RnaG80 play a key role in establishing the initial contact with the icsA

Figure 7. Effects on icsA transcription of synthetic RNA and DNA oligonucleotides reproducing the structural motifs of RnaG. Secondary structure of DNA oligos used to construct the GH1 (left) and GH2 (right) stem–loops of RnaG (A). Mutations carried by oligos GM18, GM30 and GM60 are reported below the native sequence. In vitro transcription of icsA gene was performed, as described in the legend of Figure 6, using as templates the pGT1129 (B, C and D) and the pGT1129M (E) which synthesizes the mutated icsA81/4 mRNA (17). The increasing amounts of RNA oligo pair GR4–48/GR49–80 (B) and DNA oligo pairs G1–50/G49–80 (C and E) and G14–48/G49–80 (D) are indicated. After quantization, the radioactivity associated with icsA mRNA is expressed as ratio sample B/sample E and values are plotted as percentage assuming 100% the transcription of DNA template without oligos (F). Data from Figure 6B and from panels B, C, D and E are indicated with closed squares, open squares, closed triangles, closed circles and open circles, respectively.
mRNA. For this purpose, we designed a set of RNA and DNA oligonucleotides partially or entirely reproducing the GH1 (oligos GR4–48, G1–50, G14–48 and G1–35) and GH2 (oligos GR49–80, G49–80) motifs which characterize the secondary structure of RnaG80 (Figures 5B and 7A). The ability of these oligos to block the synthesis of the icsA messenger was monitored by means of the previously illustrated in vitro transcription assay. While none of the oligos was individually able to affect icsA transcription (not shown), we found that the use of the RNA oligo pair GR4–48/GR49–80, which fully reconstructs RnaG80, inhibits transcription even more severely (60 and 41% at 5 and 10 fmol of oligos, respectively) than the purified RnaG80 (Figure 7B and F). Likewise, alternatively combining the DNA oligos G1–50 and G14–48 with G49–80 caused a reduction of icsA transcription, comparable to that observed with the RNA oligos (Figure 7C, D and F), suggesting that the secondary structure, rather than the nature of the molecule, is essential in determining the RnaG-mediated transcriptional repression. Recently, we showed that mutations on icsA gene (from positions 81 to 84) abolish the transcriptional attenuation by destabilizing the intrinsic terminator triggered on the nascent mRNA by the sense–antisense interaction (17). As expected, the oligo pair G1–50/G49–80 is almost totally inactive in repressing transcription from a plasmid carrying the icsA81/4 mutation (Figure 7E and F), which impedes the formation of the RnaG-mediated icsA mRNA terminator.

Altogether these results indicate that the minimum size necessary to maintain the RnaG attenuation activity on icsA is ~65 nt spanning from positions 14 to 80 and that both motifs, GH1 and GH2, are required. To further learn about which portions within this region primarily contribute to recognize the icsA mRNA, we used DNA oligos carrying base-exchanges localized at the level of unpaired nucleotides (apical loops and bulge) of the GH1 (oligos GM18 and GM30) and GH2 (oligo GM60) hairpins (Figure 7A). In vitro transcription assays were set up using each mutated GH1 or GH2 oligo in combination, respectively, with the wt GH2 (G49–80) or GH1 (G1–50) oligos, so as to cover the entire RnaG80 sequence. While the GM18 oligo conserves a residual repressive activity, GM30 and GM60 mutations completely abate the ability of these synthetic DNA molecules to terminate icsA transcription, suggesting that unpaired bases at the apical loops GH1 (nucleotides 30–33) and GH2 (nucleotides 60–64) provide the primary nucleation points, known as kissing complex, for sense–antisense RNAs pairing (Supplementary Figure S4 and Figure 8).

DISCUSSION

The actin-based motility of Shigella, the causative agent of bacillary dysentery, depends on IcsA, an outer membrane protein that accumulates at one pole of the bacterial cell (4). IcsA is capable of directly recruiting and activating host cell factors, thus catalyzing the assembly of an actin tail which provides the propulsive force for intracellular movement and intercellular dissemination of bacteria through the human intestinal epithelium (5).

As is the case for other Shigella virulence genes, also the icsA gene is located on the large virulence plasmid (pINV) typically harbored by this pathogen (31). Over the years, evidence has built up indicating that the regulation of icsA depends on the nucleoid protein H-NS, which acts as a repressor, and on the AraC-like protein VirF, the major activator of the Shigella invasivity regulon (7,11,12). Nevertheless, the molecular details of this regulatory mechanism have remained elusive. We have recently shown that an additional player is involved: RnaG, a non-coding sRNA which acts as an antisense regulator on the icsA transcript (17). This study is an attempt to clarify the interplay among these factors in order to evaluate their contribution to the expression of icsA, not only as...
H-NS modulates the expression of ~5% of the *E. coli* genome. Approximately one-third of these genes are linked to the adaptation of bacteria to stress conditions and to the regulation of pathogenicity processes (9,32–34). Actually, it is emerging that H-NS participates to the regulation of many genes whose expression is sensitive to temperature shifts and specifically to the virulence genes of *Shigella*. For example it has been found that H-NS represses transcription of *virF*, the primary regulator of the invasivity functions, at 30°C but not at 37°C (14,16). In the present study, we have shown that H-NS can also silence the virulence gene *icsA* in a temperature-dependent manner (Figures 2 and 4). This effect, exerted at the transcriptional level, is likely achieved through a temperature-mediated occupancy by H-NS of its binding sites on the *icsA* promoter region (Figure 1). This hypothesis is supported by the unsuccessful attempts to obtain clear protections by H-NS in footprints carried out at 37°C (Supplementary Figure S5).

Unlike other H-NS-regulated genes, in which binding sites have been mostly identified on the upstream regulatory regions (8,35), the three H-NS recognition sites on the *icsA* promoter are located downstream the *icsA* transcription start. Interestingly, two sites reside in the DNA tract between *icsA* and the RnaG promoters (Figure 1E). This arrangement likely favors the repressive action of H-NS on both *icsA* and RnaG genes. Indeed, the transcription of both genes is reduced by the presence of H-NS (Figure 2). The inhibition is more pronounced on *icsA*, suggesting that it may represent a preferential target as compared to RnaG. Nonetheless, the ability of H-NS to repress also a sRNA promoter further stresses the significance of this nucleoid protein as global regulator. It is not surprising that H-NS and sRNA molecules cooperate, since both are known to regulate bacterial gene expression (36). Moreover, DsrA can also counteract the virulence gene *icsA* as a function of protein amount, might cause the curvature

flanked by multiple target sites might provide a nucleation point for protein oligomerization, resulting in the occlusion of the entire *virA-icsA* region to RNA polymerase. This hypothesis is supported by the fact that both *icsA* and *virA* are derepressed in hns- cells [Figure 4 and (15)]. The intrinsically curved DNA, localized between *virA* and *icsA* genes, might respond to temperature shifts, thereby controlling the binding of H-NS as demonstrated in the regulation of *virF* (14,16). Alternatively, it has been proposed that a thermal increase causes modifications in the quaternary structure of H-NS, leading to the decrease of high-order oligomers versus the appearance of discrete dimers. Such conformational change, in turn, would reduce the ability of H-NS to bind DNA (41). However, this issue has not been settled conclusively since a large number of H-NS-repressed genes remain silenced at 37°C, suggesting that a simple effect of temperature on the state of aggregation of H-NS cannot account for the temperature-dependent derepression of this subset of genes. Moreover, opposed to the temperature-mediated aggregation model, H-NS oligomerization has been reported to be constant in the range 28–42°C (33).

Our data reveal that, besides H-NS, also VirF is able to specifically interact with the *icsA*-RnaG region, by recognizing four binding sites (Figure 1) and it plays a direct role in activating *icsA* transcription (Figure 3). *In vivo* data indicate that VirF is able to stimulate *icsA* expression at 37°C and that this also occurs at 30°C, albeit to a much lesser extent (Figure 4). In most cases, VirF does not bind to only one site but rather simultaneously interacts with two sites. By different footprinting techniques, the 13-bp conserved sequence TTTaGYcTat (nucleotides with a frequency >60% in the consensus sequence are in uppercase and Y indicates pyrimidines) was found in VirF protected regions and it has been proposed that VirF tightly binds to this sequence when present as an inverted repeat and weakly when the sequence is present only once (49). Although the mechanism adopted by VirF to activate transcription is still questionable, VirF bound at its site I, which includes the −10 conserved element, presumably contacts the RNA polymerase thus facilitating its access to the *icsA* promoter. This condition can occur assuming that first VirF preferentially interacts with a high affinity site (i.e. site III). Subsequently, the occurrence of DNA looping, which is commonly involved in the AraC-mediated regulation (50), may favor the contact of VirF with the promoter proximal site I. This hypothesis is validated by some facts: (i) this two step binding model has been described in the regulation of *yop* genes of *Yersinia* by VirF (49); (ii) protection at VirF site III is quite pronounced even with the lowest concentrations of VirF used (42–84 nM VirF, Figure 1C and D), indicating that this site might be the first to be occupied by the protein; (iii) three regions, showing a high homology with the 13-bp consensus motif and inversely orientated, are present within VirF site III while only one or two have been found within sites I, II and IV (Supplementary Figure S1). This binding pathway can also explain repression of transcription by VirF. In fact, progressive filling-in of all available sites, as a function of protein amount, might cause the formation of a VirF–DNA complex covering a wide
region which incorporates both the icsA and the RnaG promoters. This, in turn, would block transcription by trapping the RNA polymerase. Besides acting at the transcriptional level by directly stimulating $P_{icsA}$, VirF increases $IcsA$ expression by two additional strategies: (i) reduction of the intracellular level of RnaG, as observed in vivo by analyzing $P_{RnaG-lacZ}$ fusions (data not shown); (ii) lowering of the frequency of transcription initiation at $P_{RnaG}$ thus affecting the extent of transcriptional interference (TI) between RnaG and icsA. These mechanisms are direct consequences of the repressive action exerted by VirF on $P_{RnaG}$. We have previously shown that TI occurs between the convergent RnaG and icsA promoters (17). Hence, VirF would limit the transit events and consequently the occupancy of $P_{icsA}$ by elongation complexes moving from the convergent $P_{RnaG}$, thus enabling icsA mRNA synthesis.

VirF binding sites I and II are partially overlapping the H-NS box I while the VirF site III is adjacent to the H-NS box II (Figure 1E). Such an organization of target sites on the icsA promoter is consistent with a possible competition between the icsA-activating VirF and the repressing H-NS. Since virF expression rapidly increases when temperature is raised $>32^\circ$C (14) and H-NS is able to recognize the icsA-RnaG region only at lower temperature, it could be argued that VirF and H-NS do not interfere with each other because their opposite action on icsA promoter is, somehow, spaced by a gap of temperature. On the contrary, transcriptional competition assays between VirF and H-NS reveal that, in the presence of H-NS, VirF is capable to restore, albeit not completely, the basal activity of $P_{icsA}$ (Supplementary Figure S2), thus proposing a more complex scenario. As the temperature raises, the interaction of VirF with sites I, II and III may disrupt the H-NS-stabilized complex by forming a putative H-NS-icsA-VirF intermediate, thus promoting a switch from a repressed to an activated state. In agreement with this observation, an antagonistic role between H-NS and VirF has been found in the regulation of virB (13).

Regulatory antagonism between H-NS and members of the AraC-like protein family is emerging as a common strategy that bacterial pathogens adopt to control the virulence genes [reviewed in (42)], as exemplified, for instance, by (i) the V. cholerae TCP system, essential for the biogenesis of the toxin-co-regulated pilus (ToxT/H-NS) (43); (ii) the E. coli gad genes, involved in the adaptive response to acid stress (GadX/H-NS) (29); (iii) the ler gene of enteropathogenic E. coli, required to initiate the regulatory cascade activating the LEE locus (PerA/H-NS) (44). While the AraC-like protein represents the primary regulator, the involvement of H-NS might also reflect the need to fine-tune gene expression adapting it to the environmental changes that bacterial cells face upon host invasion (32,45).

The interplay between VirF and H-NS in the regulation of icsA is made more complex by the presence of RnaG, one of the largest (450 nt) non-coding RNAs in bacteria. We have shown that RnaG is a cis-encoded antisense RNA whose 120 nt at the 5′-end (RnaG120) are complementary to the 5′-UTR icsA transcript and behave as a functional domain inducing premature termination of the icsA mRNA by transcriptional attenuation (17). In general, the initial sense–antisense RNA contact, also known as kissing complex, takes place between two complementary loops (multistep pathway) or between a loop and a single-stranded region (one-step pathway). In both cases, such interactions, each one involving 4–7 exposed nucleotides, lead to a helix progression until a fully paired RNA is formed (22,46). We have recently described the secondary structure of RnaG120 and of the 5′-end (~150 nt) of the icsA mRNA (17); while the first is characterized by three hairpin structures, GH1, GH2 and GH3, the latter only shows two structural motifs, AH1 and AH2 (Figure 8 and Supplementary Figure S3). In the present study, we demonstrate that only GH1 and GH2 participate in establishing the kissing complex between RnaG and icsA mRNA, thus forming a functional domain responsible for icsA regulation by RnaG. This conclusion is supported by several lines of evidence. In vitro and in vivo experiments (Figure 6) indicate that out of 120 nt at the 5′-end of RnaG, the initial 80 nt (RnaG80) represent the minimum stretch able to repress icsA transcription. This finding is in agreement with the observation that a limited number of base pairs and not full-duplex formation between sense and antisense RNAs is usually required for control as shown in the case of R1, ColE1 and pLP501 (22). Despite the lack of GH3 in RnaG80, the secondary structures of GH1 and GH2 (Figure 5) are conserved in RnaG120 (Supplementary Figure S3), as well as at the 5′-end of the entire RnaG (not shown), stressing the relevance of the spatial organization of RnaG for its function. The need for both motifs, GH1 and GH2, is also evidenced by the finding that synthetic molecules carrying either GH1 or GH2 are unable to terminate icsA transcription and that this ability is restored as the native structure of RnaG80 is reconstituted by aligning the two motifs head-to-tail using suitable oligos (Figure 7). Termination of icsA mRNA, however, does not depend on the nature of the oligo molecule used (DNA or RNA).

The analysis of icsA mRNA and 5′-end of RnaG structures reveals the following potential points of interaction: (i) the apical loop of GH2 can pair with the basal bulge of AH1; (ii) the apical loop and the internal bulge of GH1 can anneal with the structurally similar motifs of AH2 (Figure 8). The relevance of contact points emerges from experiments using mutated RnaGs. Changing the exposed bases on the GH1 and GH2 loops to hinder their pairing with the complementary icsA mRNA regions, affects the repression activity of RnaG (Supplementary Figure S4). Complete inactivation occurs with the GM30 and GM60 mutations, indicating that these stretches of bases are essential for the formation of the kissing complex. Pairing might follow a two-step binding pathway, starting at the critical sequence of either the GH1 or the GH2 loop and using the remaining loop–loop interaction as nucleation site for RNA duplex propagation. Instead, GM18 mutation only partially alters RnaG functioning, suggesting that the internal bulge of GH1 may have an auxiliary role in establishing the early contact between the GH1 and AH2 structural motifs. Irrespective of the binding kinetics and loop–loop initiating system, the functional
characterization of RnaG80 and its variants confirms the model based on the formation of an intrinsic terminator on the nascent icxA mRNA, triggered by the sense–antisense pairing (17). Indeed, both native and synthetic molecules carrying the 5′-end of RnaG fails to halt transcription of the icxA81/4 mRNA in which a four bases mutation in the stem of the terminator impedes the formation of an intramolecular RNA duplex (Figure 7 and data not shown).

Besides the relevance of IcsA as a virulence determinant, the fact that its expression is thermoregulated and finely controlled by VirF, H-NS and the antisense RnaG makes this gene a very interesting candidate to comprehend how complicated regulatory circuits are functioning and how they have evolved. The aim of the present study has been to address some crucial points of this regulation, particularly those related to the antagonism between the two regulators VirF and H-NS and to their interplay with RnaG. Furthermore, icxA offers a chance to clarify several molecular aspects of sense–antisense pairing providing insights into the rules driving the formation of the kissing complex and, more generally, into the mechanisms of riboregulation.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
The authors thank G. Micheli and R. Spurio for critical reading of the manuscript.

FUNDING
Fondi Ateno Ricerca (FAR) and Progetti di ricerca di Rilevante Interesse Nazionale (PRIN) to M.F. and B.C. Funding for open access charge: FAR to M.F.

Conflict of interest statement. None declared.

REFERENCES


