Regulation of *rns*, a Positive Regulatory Factor for Pili of Enterotoxigenic *Escherichia coli*

BARBARA FROELICH, LINDA HUSMANN, JUDY CARON,[†] AND JUNE R. SCOTT*

Department of Microbiology and Immunology, Emory University, Atlanta, Georgia 30322

Received 16 February 1994/Accepted 28 June 1994

Attachment of enterotoxigenic *Escherichia coli* to the human gut is considered an important early step in infection that leads to diarrhea. This attachment is mediated by pili, which belong to a limited number of serologically distinguishable types. Many of these pili require the product of *rns*, or a closely related gene, for their expression. We have located the major promoter for *rns* and found that although its sequence diverges significantly from a sigma-70 promoter consensus sequence, it is very strong. Transcription of *rns* is negatively regulated both at a region upstream of this promoter and at a region internal to the *rns* open reading frame. In addition, *rns* positively regulates its own transcription, probably by counteracting these two negative effects.

Diarrheal diseases caused by enteric bacteria are a serious human health problem worldwide. This is especially true among infants and young children, in which these diseases lead to considerable excess mortality. Enterotoxigenic *Escherichia coli* (ETEC) is one of the major pathogens in this group (19) and is therefore an important focus of study. An early step in the establishment of disease by this organism involves colonization of the gut of the human host. This is generally dependent on pili, which are long proteinaceous appendages extending from the bacterial surface. Among ETEC strains associated with human disease, a limited number of serologically distinguishable types of pili have been identified. These include CS1 and CS2 (which, together with the filamentous appendage CS3, were originally called Cfa/II [10, 20, 40]); Cfa/IV, which, like Cfa/II, consists of three different types of appendages (CS4 and CS5, which are pili, and CS6, whose morphology has not yet been elucidated [25, 42, 45]); and Cfa/I (11).

These pili are produced only at temperatures greater than about 25°C (40), and they depend for their production on a *trans*-acting gene encoding a protein in the AraC family of positive regulators. The prototype of this regulator is Rns, first identified as being needed for expression of CS1 and CS2 (1). A homologous gene, called *cfaR* or *cfaD*, is required for expression of CFA/I pilin (3, 36). Rns and CfaR can substitute for each other functionally (3), and they are also able to cause expression of CS4 pili in clinical isolates that contain the pilin gene but are defective in pilin production (43, 44). DNA from clinical isolates that produced CFA/I, CS1, CS2, or CS3 as well as DNA from those that produced CS4 and CS6 hybridized with an *rns* DNA probe (2). In addition, an open reading frame named *csvR*, which was isolated from a CS5-producing strain but has not been associated with the production of any pili, shows strong homology with the translated *rns* sequence and partially replaces *cfaR* or *rns* (9).

In addition to the extremely homologous Rns-related proteins that are needed for expression of ETEC pili, other members of the AraC family of proteins positively regulate virulence factors of enteric pathogens (38). These include VirF of *Yersinia enterocolitica* (7) and VirF of *Shigella* spp. (34), as well as TcpN of *Vibrio cholerae* (29) and FapR of *E. coli* (18).

Most of the virulence determinants regulated by these AraC-related proteins are not expressed at low temperatures, but in most cases the step affected by temperature has not been identified.

Because of the importance of Rns in regulating virulence determinants, we have investigated some of the factors affecting its expression. In this study, we identified two potential promoters, one of which is internal to the large open reading frame of *rns*. This promoter, P2, may be used for production of a smaller Rns protein produced from the same reading frame. Using *lac* fusions, we studied transcription from these promoters in the presence and absence of Rns, and we compared transcription at 20°C with that at 37°C and that from cells grown in the presence of aeration with that from cells grown in the absence of aeration. We found two DNA regions at which *rns* transcription is independently regulated. At both, Rns causes positive regulation and an unknown factor causes negative regulation of transcription from the *rns* promoter(s). One regulatory region is at least 72 bases upstream of the start site for *rns* transcription, and the other is internal to the *rns* gene.

MATERIALS AND METHODS

Media. Bacteria were grown in Luria-Bertani broth (37). Antibiotics used were ampicillin (50 µg/ml), chloramphenicol (40 µg/ml), and spectinomycin (100 µg/ml).

Bacterial strains and plasmids. *E. coli* K-12 strains JM83 (23) and MC4100 (5) were used for cloning. LMC10 is a *lac* deletion restriction-negative derivative of the O6:K15:H16 ETEC-derived strain C921b-2, which was cured of the plasmid encoding Rns and CS3 (1, 30).

The high-copy-number vectors pUC18(Ap) (26) and pUC19 derivative pNEB193(Ap) (New England Biolabs), the low-copy-number vector pHSG576(Cm) (41), and the single-copy vectors pEU720 and pEU730 (13) were used for cloning (see Table 1).

Plasmid pEU730 is derived from the promoter cloning vector pEU720 (13). A fragment containing the polylinker of pNEB193 (New England Biolabs) was obtained by PCR by using primers with nonhomologous 5' ends containing an *Xho*I site (primer 1, AAGACCTCGAGTGTAAAACGACGGCC AG; primer 2, AAGACCTCGAGTGTGGAATTGTGAGC GG). The polylinker-containing fragment was digested with

* Corresponding author. Phone: (404) 727-0402. Fax: (404) 727-3659. Electronic mail address: scott@microbio.emory.edu.

[†] Present address: Project Management, Oncology, Sandoz Pharmaceuticals, Inc., East Hanover, NJ 07936.

TABLE 1. Cloning vectors and plasmids expressing Rns

Plasmid	Relevant trait(s)	Reference
Cloning vectors		
pUC18	High copy number	26
pNEB193	High copy number	26
pHSG576	Low copy number	41
pEU720	Single copy, promoterless <i>lacZ</i>	13
pEU730	Single copy, promoterless <i>lacZ</i>	This study
Clones expressing Rns		
pEU2030	Rns ⁺ , high copy number	1
pEU2040	Rns ⁺ , low copy number	30
pEU2034	Rns1, high copy number	1
pEU2035	Rns ⁺ , low copy number	This study
pEU2063	Rns2, high copy number	This study

*Xho*I and ligated to pEU720 digested with *Xho*I to produce pEU730 (see Table 1).

Plasmid pEU2030 contains *ms* cloned into pUC18 (1), pEU2040 contains *ms* cloned into pHSG576 (30), and pEU2034 contains the *ms*1 mutation (1) (see Fig. 1). The *ms*1 mutation is a 4-base insertion at the unique *Bgl*II site within the *ms* open reading frame. It was generated from pEU2030 by *Bgl*II digestion followed by filling in of the 3' recessed ends with the Klenow fragment of DNA polymerase and religation of the resulting DNA (1). To make pEU2035, which transcribes *ms* from the *lacZ* promoter, a 1-kb *Rsa*I fragment from pEU2030 (see Fig. 1) was ligated to pHSG576 (41) digested with *Hinc*II. To make pEU2063, which has the *lac* promoter in front of an *ms* mutant allele that is missing the N-terminal portion of *ms*, *ms*2, a 660-bp *Bsm*I fragment was isolated from pEU2030 (see Fig. 1). The *Bsm*I ends were filled in with T4 DNA polymerase, and the fragment was ligated to pUC18 digested with *Sma*I. The orientations of all of the cloned fragments were determined by restriction digestion of the purified plasmids; these plasmids are listed in Table 1.

To make pEU724, the 242-bp *Dra*I fragment of pEU2030, which contains the N terminus of Rns (Fig. 1), was isolated from an intermediate plasmid (pEU2064, a pUC18 derivative with this fragment in the *Sma*I site) by digestion with *Eco*RI and *Hind*III. The ends were filled in and ligated to filled-in *Xho*I ends of pEU720. Plasmid pEU726 was constructed similarly by using a 250-bp *Bam*HI-*Dra*I fragment from pEU2030 (the *Bam*HI site is in the polylinker), which consists of DNA directly upstream from the start site of the Rns open reading frame (Fig. 1). To make pEU728, a similar strategy was used, except that a partial *Dra*I digestion and a complete *Bam*HI digestion of pEU2030 produced the 492-bp fragment cloned. This fragment consists of 250 bases upstream from the start site of Rns and the first 242 bp of Rns (Fig. 1).

To make further transcriptional *lac* fusion plasmids, a different strategy was used. The desired fragment was first cloned into the *Hinc*II site of pNEB193. Primers homologous to polylinker DNA but synthesized with a nonhomologous 5' end containing an *Asp* 718 site were used to isolate the fragment for cloning into the *Asp* 718 site of pEU730. An 815-bp *Hae*III-*Dra*I fragment containing DNA ending directly upstream from the start site of Rns (Fig. 1) was used to make pEU736. A 1,067-bp *Hae*III-*Nsi*I fragment, which contains 815 bp upstream from the start site of Rns and the first 253 bp of Rns (Fig. 1), was used to make pEU738 (the *Nsi*I end was filled in with T4 DNA polymerase).

The orientations of the inserts in all derivatives of pEU720 and pEU730 were determined by PCR, and the inserts were

FIG. 1. Map of the *ms* region showing the extents of cloned fragments in plasmids used. The positions of the P1 and P2 promoters are shown and labelled with their names. The DNA probe used for RNA dot blots is shown above the line. The positions and orientations of repeated sequences are indicated by arrowheads below the line. The extent of the Rns open reading frame and the direction of its transcription and translation are indicated by the arrow labelled *ms*. The extent of the DNA included in each plasmid constructed is indicated by the bar next to the plasmid name. Restriction sites in parentheses are not unique in the piece of DNA shown above. Abbreviations: Bg, *Bgl*II; Bs, *Bsm*I; D, *Dra*I; H, *Hae*III; Hh, *Hha*I; N, *Nsi*I; R, *Rsa*I. *, pEU2034 contains *ms*1, which has an insertion of 4 bases at the *Bgl*II site.

sequenced to be sure that point mutations had not been isolated.

DNA sequencing. Both strands of DNA were sequenced completely by a modification of the Sanger method (35), as described by Zagursky et al. (46).

β -Galactosidase assays. Strains to be tested were grown to mid-log phase in Luria-Bertani broth at various temperatures from overnight cultures grown under the same conditions but in the presence of antibiotics to select for the plasmid. Spectinomycin was added to strains containing derivatives of pEU720 or pEU730, and chloramphenicol was added to strains containing pEU2040. For aerobic growth, the cultures were grown in flasks and shaken vigorously. For growth at low oxygen tension, cells were grown without shaking in closed screw-cap flasks under a layer of mineral oil. The cultures were put on ice to stop growth and assayed as described previously (13, 27).

In vitro transcription and translation. In vitro transcription and translation were performed as described by Caron et al. (1).

RNA. RNA was isolated from log-phase cultures as described by MacDonald and Mosig (22). The RNA was separated from contaminating DNA and protein by CsCl centrifugation, and the resulting RNA was shown to be insensitive to DNase I.

Primer extensions. Primers were ³²P labelled at their 5' ends with polynucleotide kinase (Bethesda Research Laboratories).

FIG. 2. Comparison of the *ms* and *cfaR* DNA sequences. Numbering of the *ms* sequence begins at a *Hind*III site of pEU2005 (1). The *cfaR* sequence (36) and the *ms* sequence starting with base 2320 (1) have been described previously. The comparison was done by using the Bestfit program in the Genetics Computer Group package. The predicted promoters are marked by brackets labelled -35 and -10. The 5' transcription termini are indicated by vertical arrows above the sequence labelled S1 (start site of transcription from P1) and S2 (start site of transcription from P2). Potential ribosome binding sites are boxed under the designation S-D. The start sites of the open reading frames of Rns and Rns2 are indicated by arrows above the DNA sequence. Repeats are indicated by arrows above the sequence labelled A, B, and C. The primers OC18 and OC10 are indicated by arrows labelled with their names.

The labelled primers were annealed to 1 to 90 μ g of RNA in 0.1 M KCl–0.013 M Tris-HCl (pH 8.3) by incubation at 100°C for 2 min, 55°C for 2 min, and 4°C for 15 min. The transcripts were extended with reverse transcriptase (Superscript; BRL) for 1 h at 44°C in 1 $\times$ reverse transcription buffer (50 mM Tris-HCl [pH 7.6], 40 mM KCl, 12 mM magnesium acetate, 2 mM dithiothreitol, 0.2 mM [each] dinucleoside triphosphates, 0.39 U of Rnasin (Promega) per μ l).

RNA dot blots. Total RNA (20 μ g) was denatured in 6 M glyoxal–0.01 M MOPS (morpholinepropanesulfonic acid) for 1 h at 65°C and chilled on ice, and a miniblott apparatus (Bio-Rad) was used to spot duplicate samples (10 μ g per dot) onto nitrocellulose (Schleicher and Schuell) pre-equilibrated in 20 $\times$ SSC (1 $\times$ SSC is 0.15 M NaCl plus 0.015 M sodium citrate). The filters were then washed in 20 $\times$ SSC, air dried, and baked in a vacuum oven at 80°C for 2 h. The filters were prehybridized at 42°C for 6 h in a solution containing 50% formamide, 5 $\times$ SSC, 6 $\times$ Denhardt's solution (23), and 100 μ g of baker's yeast tRNA per ml. Hybridization was performed in the same solution. The probe was labelled by incorporation of [α - 32 P]dATP by random oligonucleotide priming (12) and denatured by being heated at 100°C for 5 min. Following hybridization, the filters were washed in 0.01 $\times$ SSC–1% sodium dodecyl sulfate (SDS) at 42°C.

Nucleotide sequence accession numbers. The *ms* DNA sequence from base 2065 to 2319 is in the GenBank database under accession number L33908. The *ms* sequence starting with base 2320 is under accession number J04166.

RESULTS

Location of promoters for *ms*. To identify the 5' end of the RNA encoding Rns, extension with primer OC18 (Fig. 2) was performed on RNA isolated from LMC10 containing either pEU2030 or pEU728 (which contains the cloned *ms* promoter) and pEU2035 (which expresses Rns from a vector promoter but lacks the *ms* promoter; Fig. 1). In both cases, two strong start sites were observed (Fig. 3 and data not shown). The strongest band corresponds to the G at bp 2395 (labelled S1 in Fig. 2), which represents the furthest upstream 5' end of the primary transcript.

Upstream from the G marking the strong start site, there is a potential -10 sequence that matches the consensus sequence for *E. coli* sigma-70 utilization (15, 32) in five of six bases, but the -35 region matches the consensus sequence in only three of six bases. These two elements are separated by 16 bases, and we will call this presumptive promoter P1. In *cfaR*, the sequence of this region is identical.

FIG. 3. Identification of the 5' end of the *ms* transcript. Primer extension reactions were performed with 1 μ g of total cellular RNA from LMC10/pEU728/pEU2035 and the primer OC18 (Fig. 2). Lanes G, A, T, and C correspond to dideoxy sequencing reactions carried out on pEU2030 DNA by using the same primer. The positions of the 5' termini are designated by arrows (the primary transcriptional start site is indicated in Fig. 2 by the designation S1).

The additional strong primer extension product (Fig. 3) at base 2418 is unlikely to be transcribed by the sigma-70 RNA polymerase because there is no sequence in the -10 and -35 regions resembling the consensus sequence. It may be transcribed by an RNA polymerase containing a different sigma factor, or it may result from RNA processing. There is a potential ribosome binding site (39) between this start site and the *ms* open reading frame which is conserved in the sequence of *cfaR* (S-D in Fig. 2). This is probably the site used to initiate translation.

Possible production of Rns2, a smaller protein encoded by *ms*. Within the Rns open reading frame, there is a second sequence with similarity to a ribosome binding site preceded by a potential promoter sequence (P2; Fig. 2). The -10 and -35 regions of this promoter (P2) both match the consensus at five of six bases. However, instead of the optimal 17 bases between the two promoter elements, P2 has 19 bases. The ribosome binding site is followed by an ATG at the proper distance. This may serve as the start site of a smaller Rns protein (Rns2), which would be produced from the same reading frame as the large protein (Rns). These features are all conserved in the sequence of *cfaR*, except for the ribosome binding site, which differs at 1 base. Extension of primer OC10 (Fig. 2 and 4) with RNA isolated from LMC10/pEU2030 revealed a 5' end of

FIG. 4. Identification of the 5' end of the *ms2* transcript. Primer extension reactions were performed with 90 μ g of total cellular RNA from LMC10/pEU2030 and the primer OC10 (Fig. 2). Lanes G, A, T, and C correspond to dideoxy sequencing reactions carried out on pEU2030 DNA by using the same primer. The position of the 5' terminus is designated by the arrow (it is indicated in Fig. 2 by the designation S2).

FIG. 5. In vitro translation products of plasmids carrying the *ms* gene. Autoradiogram of [35 S]methionine-labelled products synthesized in an *E. coli* K-12 extract (Amersham) and separated by electrophoresis through a 15% polyacrylamide gel containing 0.1% SDS. Protein molecular weight markers (sizes given in thousands) are in lane 1. Lane 2 contains the products of a control reaction with no DNA. Lanes 3 through 5 contain the products made from the following plasmid templates: lane 3, the vector pUC18; lane 4, pEU2030 (Rns $^{+}$); and lane 5, pEU2063 (Rns2). The positions of the Rns protein and the smaller Rns2 protein are indicated by arrows.

RNA at the A indicated in Fig. 2 (S2) by a vertical arrow (base 2754), consistent with the use of P2 as a promoter.

When extracts of *E. coli* K-12 were programmed with plasmid pEU2030 for in vitro transcription and translation, two protein bands that were not produced from the vector (pUC18) alone were visible (Fig. 5, lane 4). One migrated with an apparent molecular weight of 26,000, which corresponds to the size of Rns. The other migrated at about 21,000, which is the correct size for the predicted Rns2 protein. The N-terminal portion of the Rns open reading frame was removed by digesting pEU2030 with *Bsm*I, and the resulting DNA fragment was cloned into pUC18 to produce pEU2063 (Fig. 1; also see Materials and Methods). Although pEU2063 cannot produce the full-length Rns protein, it should encode the entire Rns2 protein. In vitro transcription and translation from this plasmid resulted in a band with an apparent molecular weight of 21,000 (Fig. 5, lane 5), i.e., of the same size as the smaller protein made from pEU2030. We conclude that Rns2 is produced from an *E. coli* K-12 extract in vitro. It therefore seems possible that Rns is produced in two forms: the intact large protein and the smaller protein, Rns2, truncated at the amino end.

Use of *lac* fusions in a one-copy vector to measure transcription from P1 and P2. To study regulation, it is desirable to maintain the gene dosage at a level comparable to that in the wild-type form of the system being investigated. For this reason, we designed and constructed the one-copy vectors pEU720 and pEU730, based on the R100 replicon, to place *lac* under the control of a cloned promoter (13; also this work). These plasmids contain transcriptional terminators surrounding the region containing the multiple cloning site and *lacZ*, and, in addition, they have an RNase III cleavage site between the base at which the insert is placed and the ribosome binding site for β -galactosidase (Fig. 6). The RNase III site serves to

FIG. 6. Relevant region of vectors for transcriptional *lac* fusions. Transcriptional terminators are shown as open boxes labeled term, the closed box labeled lacZ represents the promoterless gene for β -galactosidase, the RNase III site is indicated by the arrow, and the site in pEU720 which is recognized by *XhoI* is shown. In pEU730, a polylinker from pNEB193 was inserted into the *XhoI* site.

ensure that the enzyme is always expressed from the same piece of RNA, and it therefore precludes effects of secondary structure of the transcript and of potential RNA processing sites on the assay (21). All plasmids were transformed by electroporation into the ETEC-derived *lac* deletion strain LMC10, and all assays were performed with this strain.

On the basis of its sequence, P1 was expected to be a weak promoter in the absence of positive regulation (24). The *lac* fusion plasmid pEU726 contains the P1 promoter region but not the P2 region (Fig. 1). Surprisingly, the region cloned in pEU726 serves as a very strong promoter, producing almost 7,000 U of β -galactosidase (Table 2) in LMC10 grown to log phase in Luria-Bertani broth at 37°C.

The existence of a functional promoter in the 242-bp *DraI* fragment, which is downstream of P1 but contains P2 (Fig. 2), was confirmed by analysis of the *lac* fusion plasmid pEU724 (Fig. 1). This region also showed significant promoter activity, although it is only about 1/10 as strong as that of P1 (Table 2).

Autoregulation by Rns. To determine whether Rns promotes its own transcription, RNA dot blots were used. The hybridization probe was a 390-base PCR product from within the *ms* gene (Fig. 1; also see Materials and Methods). RNA prepared from the ETEC-derived strain LMC10 containing pEU2040 (which includes the entire *ms* gene cloned into pHSG576 [30]) served as the positive control to which all values were compared. RNA from LMC10 containing no *ms*-encoding plasmid showed no significant hybridization (0.06 times the amount from the positive control). To test whether the Rns protein is required for *ms* transcription, the *msI* mutant, which is truncated at the *BglII* site (1), was used. When LMC10 carrying pEU2034, which contains this mutation, was used as the RNA source, there was only 1/10 as much hybridization to the *ms* probe as there was when the complete *ms* gene was present in the same vector (LMC10/

pEU2040). This suggests that the amount of *ms*-specific mRNA is positively regulated by the presence of an intact Rns protein.

The *lac* fusion plasmids were used to determine whether Rns regulation of the amount of mRNA occurred at the level of initiation of transcription. Plasmid pEU2040 was used to provide Rns for pEU726 and pEU724 in LMC10 (Table 2). The strain with plasmid pEU726 showed a less than 10% increase in promoter activity in the presence of Rns, which is much less than the 10-fold difference seen between strains containing pEU2040 and pEU2034 in the dot blots. The strain with pEU724 showed no increase in β -galactosidase levels in the presence of Rns. Therefore, it appears that the region at which Rns regulation is exerted was interrupted in the construction of pEU724 and pEU726.

This theory was tested by using pEU728, which includes all the DNA in pEU724 and in pEU726 (Fig. 1). This plasmid expressed only about one-seventh as much β -galactosidase as did pEU726 (Table 2), suggesting that the region between P1 and P2 serves to negatively regulate *ms* transcription. The addition of Rns to the strain containing pEU728 induced transcription to the higher level characteristic of pEU726, indicating that the DNA around the *DraI* site within the *ms* open reading frame is needed for positive regulation by Rns.

Regulation at a region upstream of P1. The 5' end of pEU726 includes only 38 bases upstream of the P1 promoter, and transcriptional regulators often bind further upstream than this (6). Therefore, to investigate the possibility of regulation of *ms* P1, plasmid pEU736 was constructed to include 640 bases upstream of the region cloned in pEU726 as well as the region in pEU726 (Fig. 1). In the absence of Rns, transcription from pEU736 was about 20-fold less efficient than transcription from pEU726 (Table 2). This indicates that the region upstream of pEU726 serves to negatively regulate transcription from P1 of *ms*.

In the presence of pEU2040, which produces Rns, transcription from pEU736 increased more than 10-fold, while that from pEU726 appeared to be constitutive (Table 2). Thus, the region upstream of P1 seems to be needed for positive autoregulation by Rns.

It therefore appears that there are two regions of DNA used independently for regulation of *ms* transcription. Both play a negative role in the absence of Rns and a positive Rns-dependent role. In the strain carrying pEU738, which contains the upstream region, the P1 region, and the P2 region, the low constitutive level of β -galactosidase and its induction by Rns support this conclusion.

TABLE 2. β -Galactosidase production by LMC10 carrying translational fusion plasmids:

Plasmid carried by LMC10	Amt of β -galactosidase (Miller units) expressed by LMC10 grown under the following conditions ^a :					
	37°C		37°C, low oxygen tension		20°C	
	-Rns ^b	+Rns ^c	-Rns	+Rns	-Rns	+Rns
pEU726	6,737 $\pm$ 792	7,476 $\pm$ 333	10,556 $\pm$ 687	10,146 $\pm$ 1,626	7,693 $\pm$ 1,374	8,199 $\pm$ 2,703
pEU724	738 $\pm$ 80	757 $\pm$ 197	1,362 $\pm$ 147	1,435 $\pm$ 314	1,819 $\pm$ 103	1,431 $\pm$ 296
pEU736	355 $\pm$ 40	4,870 $\pm$ 868	540 $\pm$ 58	3,498 $\pm$ 348	144 $\pm$ 8	3,240 $\pm$ 150
pEU728	1,058 $\pm$ 152	7,009 $\pm$ 219	1,140 $\pm$ 80	6,724 $\pm$ 446	912 $\pm$ 40	5,570 $\pm$ 484
pEU738	602 $\pm$ 47	6,858 $\pm$ 1,324	610 $\pm$ 51	4,454 $\pm$ 852	240 $\pm$ 27	6,954 $\pm$ 2,424

^a Assays were performed on mid-log-phase cultures grown under the conditions indicated, and results are means $\pm$ standard errors of the means. In the absence of plasmid, LMC10 produces 2.5 $\pm$ 2 U of β -galactosidase. LMC10/pEU720, the vector for pEU724, pEU726, and pEU728, produces 243 $\pm$ 49 U, and LMC10/pEU730, the vector for pEU732, pEU736, and pEU738, produces 97 $\pm$ 7 U at 37°C with aeration. Units are as described by Miller (27).

^b -Rns, pEU2040 was absent.

^c +Rns, pEU2040 was present.

Low oxygen tension and low temperature. The activity of β -galactosidase from the *lac* fusions was also measured when the bacteria were grown at low oxygen tension and at 20°C (Table 2). Very few differences were detected, and no differences in response to Rns were noticed.

DISCUSSION

P1, the major promoter for *rns*. By primer extension, we have identified the 5' end of the major form of *rns* mRNA. When the *lac* gene is fused to this DNA region, the level of β -galactosidase produced is higher than those made from many very strong *E. coli* promoters (Table 2). This is surprising because the strength of promoters can generally be correlated with the similarity of their sequences to the consensus sequence (24), and that of *rns* P1 differs significantly from the consensus sequence for this promoter. However, the strength of *rns* P1 may be due to the AT richness of the region upstream of its -35 region, since as much as a 30-fold increase in transcription has been observed to be caused by such a sequence (33). Alternatively, a sigma factor other than sigma-70 may be involved in *rns* transcription.

Negative regulation of *rns* at two sites. Transcription from P1 is reduced almost 20-fold when a region beginning 72 bases upstream of the transcription start site is included in the *lac* fusion construct (Table 2; compare pEU736 and pEU726). This suggests that a negative regulatory factor interacts with this upstream region to repress transcription. When *rns* DNA downstream of the 3' end of the DNA cloned in pEU726 is added to the *lac* fusion to produce pEU728 (Fig. 1), transcription is reduced about sevenfold (Table 2). This suggests the existence of another region of DNA downstream of P1 at which negative regulation occurs. Although the downstream negative regulation could be caused by premature termination, as has been found for *purB* (16), the upstream region precedes the start of transcription.

The negative regulator(s) of one or both regions may be a specific repressor, or it may be H-NS or some other protein whose DNA binding site is less specific. H-NS has been demonstrated to negatively regulate expression of several pilin genes, such as the *pap* pilin operon (14), the *sfa* pilin determinant (28), and the *cfa/I* gene cluster (17), which is closely related to CS1.

Most operators for promoters recognized by the sigma-70 form of RNA polymerase in *E. coli* are less than 70 bases upstream of the transcription start site (6), unlike the negative regulatory region upstream of *rns* P1. Those more distantly located are thought to be brought into contact with the RNA polymerase by DNA bending. The AT richness and distribution of the DNA sequence upstream of P1 and of that downstream of it make it seem possible that this DNA is bent (8, 31). The regulatory protein bound at these regions may thus be brought closer to P1 to directly regulate initiation.

Positive regulation of *rns* at two regions. Positive autoregulation of *rns* transcription by Rns is manifested by an at least 10-fold increase in transcription of the *lac* fusions containing the upstream region when Rns is supplied in *trans* (Table 2; compare pEU736 and pEU738). In the case of pEU728, which contains the region downstream of P1, addition of Rns in *trans* increases transcription to the level expressed constitutively from plasmid pEU726, which lacks that region. This indicates that positive autoregulation by Rns occurs independently at two regions.

A plasmid containing both the upstream and the downstream regulatory regions (pEU738; Fig. 1) behaves much like the plasmid with only the downstream region (pEU728),

although the constitutive level of transcription is somewhat lower.

It is possible that Rns serves only to counteract the negative effects mentioned above. In this case, the positive effect of the addition of Rns should not lead to a β -galactosidase level higher than that which is found when the negative regulatory region is absent. This is the case for the upstream region (compare pEU736 with Rns to the constitutive expression of pEU726) and for the downstream region (compare pEU728 or pEU738 with Rns to pEU726). The positive autoregulatory effect may be due to direct binding of Rns, or it may be an indirect effect involving a regulatory cascade.

Potential binding sites for regulatory factors. Regulatory proteins recognize and bind to specific DNA sequences. Therefore, if a single protein is responsible for upstream and downstream positive or negative regulation, a common sequence to which this protein might bind should be present in both DNA regions. Because CfaR and Rns are at least partially interchangeable, important binding sites might be expected to be conserved in both.

Two different repeated sequences common to both the upstream and downstream regulatory regions are present. The first (repeat A; Fig. 1 and 2) is an identical direct repeat of 9 bases located 302 bases upstream (bases 2094 to 2102) and 216 bases downstream (bases 2611 to 2619) of the start of transcription from promoter P1. This 9-base repeat is conserved in *cfaR* (Fig. 2). The second (repeat B; Fig. 2 and 5), which is closer to the P1 promoter (bases 2242 to 2253 and 2463 to 2474; Fig. 2), is an inverted repeat conserved at 11 of 12 bases. In *cfaR*, the downstream copy of repeat B is not conserved.

Within the downstream regulatory region, there is an 11-base inverted repeat (repeat C; Fig. 1 and 2) that overlaps the -10 region of the P1 promoter (bases 2378 to 2388; Fig. 2). This sequence is also found 2 bases downstream from the start site of the P2 transcript (bases 2756 to 2766; Fig. 2). In *cfaR*, repeat C is conserved in location and sequence, except for 1 base in the first copy. No similar sequence is found in the upstream regulatory region, which raises the possibility that one factor needed for downstream regulation is not used upstream.

Lack of effect of low oxygen tension and low temperature. Because CS1 pili are not produced at temperatures below about 25°C, and since ETEC organisms inhabit an environment with low oxygen tension, it seemed possible that *rns* would not be transcribed as well at low temperatures or in the absence of vigorous aeration. Investigation of all the *lac* fusions described (Table 2), with and without Rns supplied in *trans*, indicated little effect on *rns* transcription of growth under these different conditions.

P2 and Rns2. Primer extension and *lac* fusion analysis were used to identify P2, a promoter within the *rns* reading frame (Fig. 4 and Table 2). On the basis of the amount of β -galactosidase activity generated from the *lac* fusion construct, P2 appears to be much less active than P1. This lower level of activity is probably caused by the nonoptimal spacing between the -10 and -35 regions (19 bases instead of 17 bases), since these regions are very similar to the consensus sequence recognized by sigma-70.

Transcription and translation in vitro produced Rns2, a protein encoded in frame by the carboxy terminal part of *rns* which might be transcribed from P2. If Rns2 is produced in vivo, it may play several possible roles. We imagine that Rns2 is able to bind DNA because the helix-turn-helix regions predicted for Rns lie in the C-terminal part of the protein that is retained in Rns2 (1). It is unlikely that this shorter form of the Rns protein activates CS1 production, because the plasmid

pEU2063, which makes Rns2 but not Rns (Fig. 1), does not induce the expression of CS1 (4). However, Rns2 may compete with Rns for binding to DNA and prevent Rns from activating production of CS1. Alternatively, Rns2 may recognize sites different from those recognized by Rns and/or use coregulators different from those used by Rns, thus providing the opportunity to respond to different sets of environmental conditions.

Why is regulation of *ms* so complex? The presence and interaction of the control elements suggested by this work present the opportunity for very tight control of the amount of Rns produced. However, because of the unusual codon usage of *ms* (which has a 28% G+C content) (1), little Rns protein is likely to be produced in spite of the complexity of this regulation and in spite of the unusual strength of the *ms* P1 promoter. The *ms* gene appears likely to be a recent arrival in *E. coli*, since its G+C content and codon usage are so foreign, and its proximity to insertion elements suggests the possibility that it was introduced on a transposon (38). Perhaps in its native organism Rns regulated some protein(s) whose overproduction would be unhealthy, so that a complex system evolved to control Rns expression.

Possible importance for virulence. It has been suggested that pili are advantageous to bacterial pathogens at some stages of their processes of infection and disadvantageous at others, so that regulation of pilus synthesis in response to changes in the microenvironment might be a significant advantage. The two separate regions for negative regulation as well as two regions involved in positive autoregulation of Rns, which is required for CS1 pilus expression, should provide enough control points to allow production of pili only where they are useful to the bacterium.

ACKNOWLEDGMENTS

We thank Christy Kelly for some of the plasmid constructions used in this study.

This work was supported by grant AI24870 from the National Institutes of Health.

REFERENCES

- Caron, J., L. Coffield, and J. Scott. 1989. A plasmid-encoded regulatory gene, *ms*, required for expression of the CS1 and CS2 adhesins of enterotoxigenic *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **86**:963–967.
- Caron, J., D. R. Maneval, J. B. Kaper, and J. R. Scott. 1990. Association of *ms* homologs with colonization factor antigens in clinical *Escherichia coli* isolates. *Infect. Immun.* **58**:3442–3444.
- Caron, J., and J. R. Scott. 1990. A *ms*-like regulatory gene for colonization factor antigen I (CFA/I) that controls expression of CFA/I pilin. *Infect. Immun.* **58**:874–878.
- Caron, J. E. 1990. Characterization of *ms*, a positive regulator of enterotoxigenic *Escherichia coli* pilin genes. Ph.D. thesis. Emory University, Atlanta, Ga.
- Casadaban, M. J. 1976. Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. *Mol. Microbiol.* **104**:541–555.
- Collado-Vides, J., B. Magasanik, and J. D. Gralla. 1991. Control site location and transcriptional regulation in *Escherichia coli*. *Microbiol. Rev.* **55**:371–394.
- Cornelis, G., C. Sluiter, C. Lambert de Rouvroit, and T. Michiels. 1989. Homology between VirF, the transcriptional activator of the *Yersinia* virulence regulon, and AraC, the *Escherichia coli* arabinose operon regulator. *J. Bacteriol.* **171**:254–262.
- Crothers, D. M., T. E. Haran, and J. G. Nadeau. 1990. Intrinsically bent DNA. *J. Biol. Chem.* **265**:7093–7096.
- De Haan, L. A. M., G. A. Willshaw, B. A. M. Van Der Zeijst, and W. Gaastra. 1991. The nucleotide sequence of a regulatory gene present on a plasmid in an enterotoxigenic *Escherichia coli* strain of serotype O167:H5. *FEMS Microbiol. Lett.* **83**:341–346.
- Evans, D. G., and D. J. Evans, Jr. 1978. New surface-associated heat-labile colonization factor antigen (CFA/II) produced by enterotoxigenic *Escherichia coli* of serogroups O6 and O8. *Infect. Immun.* **21**:638–647.
- Evans, D. G., R. P. Silver, D. J. Evans, Jr., D. G. Chase, and S. L. Gorbach. 1975. Plasmid-controlled colonization factor associated with virulence in *Escherichia coli* enterotoxigenic for humans. *Infect. Immun.* **12**:656–667.
- Feinberg, A., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**:6–13.
- Froehlich, B. J., and J. R. Scott. 1991. A single promoter-cloning vector for use in *Escherichia coli*. *Gene* **108**:99–101.
- Goransson, M., B. Sonden, P. Nilsson, B. Dagberg, K. Forsman, K. Emanuelsson, and B. E. Uhlin. 1990. Transcriptional silencing and thermoregulation of gene expression in *Escherichia coli*. *Nature (London)* **344**:682–685.
- Hawley, D. K., and W. McClure. 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res.* **11**:2237–2255.
- He, B., and H. Zalkin. 1992. Repression of *Escherichia coli purB* is by a transcriptional roadblock mechanism. *J. Bacteriol.* **174**:7121–7127.
- Jordi, B. J., B. Dagberg, L. A. De Hann, A. M. Hamers, B. A. Van Der Zeijst, W. Gaastra, and B. E. Uhlin. 1992. The positive regulator CfaD overcomes the repression mediated by histone-like proteins H-NS (H1) in the CFA/I fimbrial operon of *Escherichia coli*. *EMBO J.* **11**:2627–2632.
- Klaasen, P., and F. K. De Graaf. 1990. Characterization of FapR, a positive regulator of expression of the 987P operon in enterotoxigenic *Escherichia coli*. *Mol. Microbiol.* **4**:1779–1783.
- Levine, M. 1987. *Escherichia coli* that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. *J. Infect. Dis.* **155**:377–389.
- Levine, M. M., P. Ristaino, G. Marley, C. Smyth, S. Knutton, E. Boedeker, R. Black, C. Young, M. L. Clements, C. Cheney, and R. Patnaik. 1984. Coli surface antigens 1 and 3 of colonization factor antigen II-positive enterotoxigenic *Escherichia coli*: morphology, purification, and immune responses in humans. *Infect. Immun.* **44**:409–420.
- Linn, T., and R. St. Pierre. 1990. Improved vector system for constructing transcriptional fusions that ensures independent translation of *lacZ*. *J. Bacteriol.* **172**:1077–1084.
- MacDonald, P., and G. Mosig. 1983. Cloning and physical mapping of an early region of the bacteriophage T4 genome. *Genetics* **106**:17–27.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- McClure, W. R. 1985. Mechanism and control of transcription initiation in prokaryotes. *Annu. Rev. Biochem.* **54**:171–204.
- McConnell, M. M., L. V. Thomas, G. A. Willshaw, H. R. Smith, and B. Rowe. 1988. Genetic control and properties of coli surface antigens of colonization factor antigen IV (PCF8775) of enterotoxigenic *Escherichia coli*. *Infect. Immun.* **56**:1974–1980.
- Messing, J., and J. Vieira. 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* **19**:269–276.
- Miller, J. H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Morschhauser, J., B. Uhlin, and J. Hacker. 1993. Transcriptional analysis and regulation of the *sfa* determinant coding for S fimbriae of pathogenic *Escherichia coli* strains. *Mol. Gen. Genet.* **238**:97–105.
- Ogierman, M. A., and P. A. Manning. 1992. Homology of TcpN, a putative regulatory protein of *Vibrio cholerae*, to the AraC family of transcriptional activators. *Gene* **116**:93–97.
- Perez-Casal, J., J. Swartley, and J. R. Scott. 1990. Gene encoding the major subunit of CS1 pili of human enterotoxigenic *Escherichia coli*. *Infect. Immun.* **58**:3594–3600.
- Plaskon, R. R., and R. M. Wartell. 1987. Sequence distributions associated with DNA curvature are found upstream of strong *E. coli* promoters. *Nucleic Acids Res.* **15**:785–796.
- Pribnow, D. 1979. Genetic control signals in DNA, p. 219–227. *In*

- R. F. Goldberger (ed.), Biological regulation and development, vol. 1. Plenum Press, New York.
33. Ross, W., K. K. Gosink, J. Salomon, K. Igarashi, C. Zou, A. Ishihama, K. Severinov, and R. L. Gourse. 1993. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **262**:1407-1413.
 34. Sakai, T., C. Sasakawa, S. Makino, and M. Yoshikawa. 1986. DNA sequence and product analysis of the *virF* locus responsible for Congo red binding and cell invasion in *Shigella flexneri* 2a. *Infect. Immun.* **54**:395-402.
 35. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
 36. Savelkoul, P., G. Willshaw, M. McConnell, H. Smith, A. Hamers, B. Van Der Zeijst, and W. Gastra. 1990. Expression of CFA/I fimbriae is positively regulated. *Microb. Pathog.* **8**:91-99.
 37. Scott, J. R. 1974. A turbid plaque-forming mutant of phage P1 that cannot lysogenize *Escherichia coli*. *Virology* **62**:344-349.
 38. Scott, J. R., and B. J. Froehlich. 1993. CS1 pili of enterotoxigenic *E. coli*, p. 17-30. In C. I. Kado and J. H. Crosa (ed.), *Molecular mechanisms of bacterial virulence*. Kluwer Academic Publishers, Hingham, Mass.
 39. Shine, J., and L. Dalgarno. 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to non-sense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **71**:1342-1346.
 40. Smyth, C. J. 1982. Two mannose-resistant haemagglutinins on enterotoxigenic *E. coli* of serotype O6:K15:H16 or H- isolated from travellers' and infantile diarrhoea. *J. Gen. Microbiol.* **128**:2081-2096.
 41. Takeshita, S., M. Sato, M. Toba, W. Masahashi, and T. Hashimoto-Gotoh. 1987. High-copy-number and low-copy-number plasmid vectors for *lacZ* alpha-complementation and chloramphenicol- or kanamycin-resistance selection. *Gene* **61**:63-74.
 42. Thomas, L. V., M. M. McConnell, B. Rowe, and A. M. Field. 1985. The possession of three novel coli surface antigens by enterotoxigenic *E. coli* strains positive for the putative colonization factor PCF8775. *J. Gen. Microbiol.* **131**:2319-2326.
 43. Willshaw, G., M. McConnell, W. Gastra, A. Thomas, M. Hibberd, and B. Row. 1990. Structural and regulatory genes for coli surface antigen 4 (CS4) are encoded by separate plasmids in enterotoxigenic *Escherichia coli* strains of serotype O25.H42. *FEMS Microbiol. Lett.* **68**:255-260.
 44. Willshaw, G. A., H. R. Smith, M. M. McConnell, and B. Rowe. 1991. Cloning of regulator genes controlling fimbrial production by enterotoxigenic *Escherichia coli*. *FEMS Microbiol. Lett.* **82**:125-130.
 45. Wolf, M. K., G. P. Andrews, B. D. Tall, M. M. McConnell, M. M. Levine, and E. C. Boedeker. 1989. Characterization of CS4 and CS6 antigenic components of PCF8775, a putative colonization factor complex from enterotoxigenic *Escherichia coli* E8775. *Infect. Immun.* **57**:164-173.
 46. Zagursky, R. J., K. Baumeister, N. Lomax, and M. L. Berman. 1985. Rapid and easy sequencing of large linear double-stranded DNA and supercoiled plasmid DNA. *Gene Anal. Tech.* **2**:89-94.