JOURNAL OF BACTERIOLOGY, July 1996, p. 4273–4280 Vol. 178, No. 14
0021-9193/96/$04.00+0
Copyright © 1996, American Society for Microbiology

Identification of the Promoter Region for the Colanic Acid Polysaccharide Biosynthetic Genes in *Escherichia coli* K-12

VALERIE STOUT*

Department of Microbiology and Molecular and Cellular Biology Graduate Degree Program, Arizona State University, Tempe, Arizona 85287-2701

Received 19 June 1995/Accepted 1 May 1996

The colanic acid polysaccharide capsule biosynthetic genes (*cps* genes) are primarily clustered at one site located at about 45 min on the *Escherichia coli* chromosome. The network of proteins involved in regulating the expression of these genes includes the two positive regulators RcsA and RcsB. This work describes the site of action of these two activator proteins and the promoter of the *cps* genes. It is likely that the *cps* genes are arranged in a single long operon that is at least 13.5 kb. The promoter region was identified with fusions to *lacZ* that resulted in regulated expression by the Rcs network of regulatory proteins, and the start site of transcription was identified by primer extension. The operator region was cloned from Kohara phage to multicopy plasmids and identified by titrating RcsA or RcsB. Sequence analysis of the promoter and operator region revealed homology to the JUMPstart element found in the untranslated region of many exopolysaccharide biosynthetic operons. In addition, the deduced amino acid sequence of the amino terminus of the first open reading frame of the *cps* operon was found to be homologous to proteins encoded by the exopolysaccharide biosynthetic operons of *Klebsiella pneumoniae* and *Erwinia amylovora*.

Escherichia coli is able to synthesize many different extracellular polysaccharide capsules that can be classified into serotype-specific and nonspecific capsules. The specific capsules are called K antigens, and over 70 biochemically distinct polysaccharide structures have been identified (20). The nonspecific capsule is called colanic acid (or M antigen) and is produced by many enteric bacteria (11). Because colanic acid appears to be loosely attached to cells, in the past it has been designated as a slime rather than a capsule. This may be a distinction that is difficult to make, as there appears to be a continuum in the degree of attachment of exopolysaccharides, rather than just two categories. Although colanic acid is normally synthesized at low levels in most strains, some stimulation can be achieved at low growth temperatures in the presence of carbohydrate carbon sources (35). In addition, preliminary evidence indicates that desiccation may stimulate colanic acid capsule synthesis (15). Because of the desiccation and low temperature stimulation of capsule synthesis, it is more likely that colanic acid capsule functions in the natural environment, rather than the mammalian host.

In addition to the environmental signals, several regulators that also affect the level of capsule expression have been identified. This system of regulators, called the Rcs system (13), has homologs that have been identified in other bacteria, such as *Klebsiella pneumoniae* (1, 27), *Erwinia* spp. (3, 6, 7, 29), and *Salmonella typhi* (19). Because these bacteria are normally found in diverse environmental backgrounds, they probably do not respond to the same environmental signals; however, they do use the same molecular pathway(s) to stimulate expression of their respective capsule biosynthetic genes.

Capsules produced by *E. coli* have been divided into two groups based on structure, map location of the biosynthetic genes, conditions of synthesis, and other factors (20). Group II capsule synthesis genes map at 64 min on the chromosome near *ser* and encode capsules such as K1 and K5. The group I capsules were divided into two subgroups based on their composition: group IB capsules contain amino sugars and group IA capsules do not (21). The biosynthetic genes for the group IA capsules map near *his* and *rfb* at 45 min on the chromosome and include K30 and colanic acid polysaccharides. Keenleyside et al. (22) suggested that the genes encoding biosynthesis of the colanic acid and K30 capsules may, in fact, be allelic and therefore regulated in similar ways.

The biosynthetic pathway for colanic acid begins with the synthesis of the component sugars as nucleotide sugar intermediates, and this is followed by assembly, transport, and modification of the polymer (25). The genes thought to be involved in the latter steps of assembly, transport, and modification are called the *cps* genes. These genes were identified by isolating nonmucoid derivatives of mucoid strains of *E. coli* after insertional mutagenesis (14). The level of capsule synthesis is limited by the level of transcription of the *cps* genes: when *lacZ* fusions were made to some of the *cps* genes, the level of transcription correlated with the amount of capsule produced (38).

The *cps* genes fall into six complementation groups at two chromosomal loci. At least some of these *cps* genes may encode biosynthetic enzymes that were assayed by Markovitz (25). One of the complementation groups (*cpsF*) maps by itself at 90 min, and *cpsF* mutants only reduce capsule synthesis, suggesting that it is not an essential gene. The other five complementation groups, designated *cpsA*, *cpsB*, *cpsC*, *cpsD*, and *cpsE*, map together at 45 min and are adjacent to the *rfb* genes that are involved in lipopolysaccharide biosynthesis. All six of these *cps* complementation groups were identified as nonmucoid derivatives (of a mucoid *lon* mutant) that were isolated by transposon mutagenesis using either Tn*10* or Mu*lac* insertions. Therefore, it is possible that the complementation groups may represent five operons rather than five genes. However, because all of the *cps*::*lac* fusions created by Mu*lac* insertions showed activation in a *lon* mutant strain (38) and therefore are

* Phone: (602) 965-4617. Fax: (602) 965-0098. Electronic mail address: atvgs@asuvm.inre.asu.edu.

regulated similarly, it seems more likely that *cpsABCDE* represent a single operon.

The simplest model for the data is that the single operon represented by *cpsABCDE* is regulated by the Rcs regulatory network of proteins. Specifically, the two positive regulators, RcsA and RcsB, may act as a heterodimer to activate transcription at the single *cps* promoter upstream of the single *cps* operon. This work begins to define the promoter region of the *cps* genes.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. *E. coli* K-12 strains, plasmids, and their sources are listed in Table 1. All strains are derivatives of MC4100 (31) unless otherwise noted. Generally, cells were grown in Luria-Bertani broth at 37°C with shaking or on Luria-Bertani agar or MacConkey lactose agar plates. When necessary, media were supplemented with the following antibiotics at the indicated concentrations: ampicillin, 125 μg/ml; kanamycin, 30 μg/ml; and tetracycline, 15 μg/ml.

DNA manipulations. Lambda phage DNA was purified from the Kohara phage by the procedure described by Silhavy et al. (31). The phage DNA was digested with *Eco*RI, and all resulting DNA fragments were ligated with either pUC18 or pUC19 that also was digested with *Eco*RI. Restriction endonucleases and other enzymes were purchased from either Bethesda Research Laboratories, Promega, or New England Biolabs and used according to the manufacturers' instructions. Ligation mixes were transformed into JM101 so that color screening for the cells that harbored plasmids with inserts could be performed on lactose MacConkey agar plates. DNA purification using the alkaline-sodium dodecyl sulfate method (31) was performed, and DNA was analyzed on 1.0% agarose gels. The resulting plasmids are shown in Fig. 1.

Subcloning of the 3.0-kb *cps* promoter region in pST103 was performed with *Hin*cII, *Sal*I, *Nru*I, and *Eco*RI. The plasmid subclones and brief descriptions of how they were constructed are listed in Table 1.

Sequencing of DNA was performed by the dideoxy-chain termination method (30) by using the Sequenase (version 2.0; U.S. Biochemical Corp.) sequencing kit. The sequence was analyzed by using the Genetics Computer Group package (9), and database searches were performed with Genetics Computer Group and BLAST software. The accession number in GenBank for the sequence shown in Fig. 3 is U52666.

Primer extension. RNA was isolated from an exponentially growing strain (SG20250) overexpressing RcsA from a multicopy plasmid (pATC400) using a hot-phenol extraction procedure (method II) of Hinton (17). Primer extension analysis was performed by the extension of the ^{32}P-, 5′-end-labeled oligonucleotide 5′CACGCTACCGCCCCTGGC (cps8) using the avian myeloblastosis virus reverse transcriptase primer extension system (Promega) following the manufacturer's protocol. The DNA sequencing was performed with the same end-labeled oligonucleotide (cps8).

Genetic manipulations. A titration assay was performed to determine which of the seven pST plasmids and the subclones from pST103 contained regions where limiting positive regulators bound. The plasmids were transformed into a *lon cpsB*::*lacZ* strain (SG20780), plated on lactose MacConkey agar media, and incubated at 37°C for 12 to 15 h. Plasmids that were able to titrate RcsA or RcsB resulted in pink colonies, and those unable to titrate these products were red.

In order to obtain the newly created fusions in single copies in cells so that more consistent regulation results could be obtained, two of the *cps*::*lacZ* plasmid fusions (pDG3 and pVS137) were transferred to the lambda derivative λRS45 (32) and plaques were screened in an *rcsC137* mutant (VS20306) for recombination. This strain was chosen because previous results have shown that the *cps* genes are transcribed well in this *rcsC137* mutant background (4), and the fusions resulted in high levels of activity in an *rcsC137* strain when they were on multicopy plasmids. Therefore, the fusions were expected to be transcribed in this host when found on lambda phage grown lytically (or lysogenically). Lac$^+$ plaques were purified and used for further study. The λRS45 derivatives λcps1 and λcps2 integrate into the chromosome at the attachment site for lambda so that lysogens have the wild-type *cps* genes in addition to the fusion.

RESULTS

Identification of *cps* region on Kohara phage. Trisler and Gottesman (38) previously mapped the *cpsABCDE* genes to 45 min on the *E. coli* chromosome. To identify a more precise location for these genes and to identify a source of DNA for cloning, rescues of *cps* mutations by using phage in the Kohara set (23) were conducted. Kohara phages 350 to 363, which cover the 45-min region of the chromosome, were used to rescue the nonmucoid phenotype in strains that had insertions in several locations in the *cps* gene cluster. These *cps* mutant cells were also *lon* mutants, so that when rescue occurred, the cells would become mucoid. Figure 1 shows that phage 353 rescued a *cpsB* mutation, phage 354 rescued *cpsD* and *cpsE* mutations, and phage 355 rescued a *cpsE* mutation. No rescue of *cpsC* was seen with any phage, and no attempt to rescue a *cpsA* mutation was made.

Cloning of *cps* region to plasmids. The three Kohara phages that rescued the *cps* mutations were used as sources of DNA for cloning the promoter-operator region of the *cps* genes. *Eco*RI fragments from the phage were cloned into pUC18 and pUC19 vectors (Fig. 1). The seven resulting plasmids, pST101 to pST107 were tested for complementation of the same *cps* mutations that were rescued by the phage. No complementation of the mutations was observed. Two explanations for this are (i) that the genes span the *Eco*RI sites or (ii) that the *cps* mutations may be polar and therefore genes downstream of the mutation are not expressed. The shorter segments of DNA seen on the plasmids may be too short to complement the inactive set of genes.

Identification of operator region. To identify plasmid clones that carry the promoter-operator region of the *cps* genes, the clones were analyzed for titration of a limiting positive regulator (RcsA or RcsB) for transcription of the chromosomal *cps* genes in a *lon* mutant strain. The rationale for this assay is that if RcsA or RcsB binding sites are present on a multicopy plasmid, these sites will bind all of the protein in the cell so that there will not be any available to bind and activate the single copy of the *cpsB*::*lacZ* fusion on the chromosome. This assay demonstrated that of the seven plasmids tested (pST101 to pST107), only the pST103 plasmid titrated the product of *cpsB*::*lacZ*.

Since pST103 apparently titrated RcsA or RcsB, subclones of this plasmid were constructed and tested in the titration assay (Fig. 2). pST103 was divided into two (slightly overlapping) fragments, and one of these fragments, a 2.2-kb *Eco*RI-*Nru*I fragment (pVS124), titrated. Similarly, pVS124 was divided into two fragments, and one of these, a 1.2-kb *Sal*I-*Sal*I fragment (pVS126), titrated. Lastly, the DNA fragment cloned into pVS126 was divided into two smaller fragments, and one of these, a 0.5-kb *Hin*cII-*Hin*cII fragment (pVS129), titrated the chromosomally encoded *cpsB*::*lacZ*.

Plasmid fusions. To determine whether a promoter region was located on the DNA fragments that were able to titrate RcsA or RcsB activity, these DNA fragments were inserted in front of a promoterless *lacZ* gene to create transcriptional fusions between the putative *cps* gene promoter and *lacZ* on a plasmid. The *cps* fragments were cloned into a fusion vector, either pRS415 or pRS528 (32), that has a cloning site in front of a full-length promoterless *lacZ* gene. The only difference between pRS415 and pRS528 is the orientation of the multiple-cloning site. In addition to testing for the presence of a promoter, this allowed the testing of the orientation of the *cps* fragments to confirm the direction of transcription of the *cps* genes. Table 2 shows the fusion plasmids that were constructed. The 3.0-kb *Eco*RI-*Eco*RI fragment was cloned in one direction only, and the 2.2-kb *Eco*RI-*Nru*I fragment and the 0.5-kb *Hin*cII-*Hin*cII fragment were cloned in both orientations. These plasmids were transformed into four *lacZ* mutant strains. These strains included one that was wild type for the regulators of capsule synthesis (SG1132), an *rcsC137* mutant (VS2038), a *lon* mutant (JK1261A) and a *lon rcsA* double mutant (VS2032). These mutations have previously been tested with the chromosomal *cpsB*::*lacZ* fusion and been shown to result in high-level *cps* expression in the *rcsC137* and *lon* mutants and low-level *cps* expression in the wild-type and *lon rcsA* double mutant.

TABLE 1. Strains, phages, and plasmids used in this study

Strain, plasmid, or phage	Relevant genotype or description	Source or reference
Strains		
VS2032[a]	*Δlon-510 rcsA51*::*Δkan*	JK1261A + P1(VS20182)
VS2038[a]	*rcsC137 ompC*::Tn*5*	SG1132 + P1(SG20803)
VS20095	*rcsC137 ompC*::Tn*5*	SG20250 + P1(SG20803)
VS20183	*rcsB62*::*Δkan cps-3*(*E*)::Tn*10 Δlon-100*	36
VS20306	*rcsC137 ompC*::Tn*5 cpsE*::Tn*10*	VS20095 + P1(SG20043)
VS20595	*cpsP1*::*lacZ*	SG20250 + λcps1
VS20596	*cpsP2*::*lacZ*	SG20250 + λcps2
VS20598	*cpsP1*::*lacZ lon-146*::ΔTn*10*	VS20595 + P1(SG20322)
VS20599	*cpsP2*::*lacZ lon-146*::ΔTn*10*	VS20596 + P1(SG20322)
VS20601	*cpsP1*::*lacZ rcsC137 rcsB11*::ΔTn*10 ompC*::Tn*5*	VS20595 + P1(SG12025)
VS20602	*cpsP2*::*lacZ rcsC137 rcsB11*::ΔTn*10 ompC*::Tn*5*	VS20596 + P1(SG12025)
VS20604	*cpsP1*::*lacZ rcsC137 ompC*::Tn*5*	VS20595 + P1(SG20803)
VS20605	*cpsP2*::*lacZ rcsC137 ompC*::Tn*5*	VS20596 + P1(SG20803)
VS20607	*cpsP1*::*lacZ rcsB62*::*Δkan*	VS20595 + P1(VS20183)
VS20608	*cpsP2*::*lacZ rcsB62*::*Δkan*	VS20596 + P1(VS20183)
SG1132[a]		S. Gottesman
SG12025[b]	*rcsC137 rcsB11*::ΔTn*10 ompC*::Tn*5*	4
SG20043	*cps-3*(*E*)::Tn*10 Δlon-100*	38
SG20045	*cps-5*(*D*)::Tn*10 Δlon-100*	38
SG20250	wild type	14
SG20322	*cps-11*::*lac*-Mu d1 *lon-146*::ΔTn*10*	26
SG21016	*cpsB10*::*lac*-Mu d1 *Δlon-510*	4
SG20780	*cpsB10*::*lac*-Mu-immλ *Δlon-510*	4
SG20803	*rcsC137 ompC*::Tn*5*	4
JK1261A[a]	*Δlon-510*	S. Gottesman
Phages		
λRS45	imm^{21}; Δlac_{SC}	32
λcps1	*cpsP1*::*lacZ*	λRS45 + pDG3
λcps2	*cpsP2*::*lacZ*	λRS45 + pVS129
Plasmids		
pST101	8.8-kb *Eco*RI fragment from Kohara 353 + pUC19 (*Eco*RI)	
pST102	9.5-kb *Eco*RI fragment from Kohara 353 + pUC18 (*Eco*RI)	
pST103	3.0-kb *Eco*RI fragment from Kohara 354 + pUC19 (*Eco*RI)	
pST104	4.0-kb *Eco*RI fragment from Kohara 354 + pUC19 (*Eco*RI)	
pST105	8.0-kb *Eco*RI fragment from Kohara 354 + pUC18 (*Eco*RI)	
pST106	2.6-kb *Eco*RI fragment from Kohara 355 + pUC19 (*Eco*RI)	
pST107	4.1-kb *Eco*RI fragment from Kohara 355 + pUC19 (*Eco*RI)	
pVS113	2.2-kb *Eco*RI-*Nru*I fragment from pST103 + pRS528 (*Sma*I-*Eco*RI)	
pVS121	0.8-kb *Eco*RI-*Sal*I fragment from pDG3 + pUC18 (*Eco*RI-*Sal*I)	
pVS122	0.9-kb *Eco*RI-*Sal*I fragment from pDG3 + pUC18 (*Eco*RI-*Sal*I)	
pVS124	2.2-kb *Bam*HI-*Eco*RI fragment from pVS113 + pUC18 (*Eco*RI-*Bam*HI)	
pVS125	1.2-kb *Sal*I fragment from pDG3 + pUC18 (*Sal*I)	
pVS126	1.2-kb *Sal*I fragment from pDG3 + pUC18 (*Sal*I) (opposite orientation of pVS125)	
pVS128	2.2-kb *Eco*RI-*Bam*HI fragment from pVS113 + pRS415 (*Eco*RI-*Bam*HI)	
pVS129	0.5-kb *Hinc*II fragment from pVS125 + pUC18 (*Hinc*II)	
pVS134	0.5-kb *Hinc*II fragment from pVS129 + pUC18 (*Sma*I)	
pVS136	0.5-kb *Bam*HI-*Eco*RI fragment from pVS134 + pRS528 (*Bam*HI-*Eco*RI)	
pVS137	0.5-kb *Bam*HI-*Eco*RI fragment from pVS134 + pRS415 (*Bam*HI-*Eco*RI)	
pRS415	*lacZ* operon fusion vector, *Bam*HI-*Eco*RI-*lacZ* orientation	32
pRS528	*lacZ* operon fusion vector, *Eco*RI-*Bam*HI-*lacZ* orientation	32
pATC400	*rcsA*	37
pDG3	3.0-kb *Eco*RI fragment from pST103 + pRS528 (*Eco*RI)	

[a] HB101-derived strain that is also a *Δlac lue pro* mutant.
[b] C600-derived strain.

The fragments that were cloned in one orientation relative to the promoterless *lacZ* gene (pDG3, pVS128, and pVS137) resulted in β-galactosidase expression that was regulated in a manner similar to that of the chromosomal *cpsB*::*lacZ* fusion when assayed by plating plasmid-bearing strains on lactose MacConkey agar (Table 2). The fusion generated in pDG3, with the 3.0-kb fragment directing *lacZ* transcription, was named *cpsP1*::*lacZ*, and the fusion generated in pVS137, with the 0.5-kb fragment directing *lacZ* transcription, was named *cpsP2*::*lacZ*. The fragments that were cloned in the opposite orientation (pVS113 and pVS136) resulted in constitutive expression of β-galactosidase on lactose MacConkey agar, regardless of the host strain, suggesting that a constitutive promoter exists in the opposite orientation to the *cps* genes. As expected, the vectors with no insertions were white on MacConkey lactose plates, indicating that they produced little or

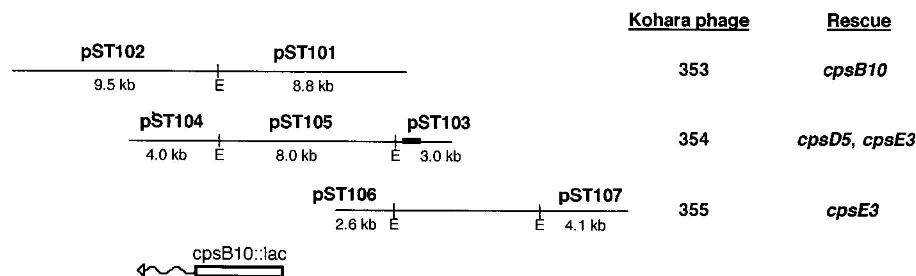

FIG. 1. Plasmid construction and rescue of *cps* mutations. Seven plasmids were constructed by inserting the *Eco*RI DNA fragments from the three Kohara phages into pUC18 or pUC19. The black box on pST103 indicates the 0.5-kb *Hin*cII fragment in pVS129 (Fig. 2). The depiction of *cpsB10*::*lacZ* shows the junction between the *cpsB* gene (box) and the *lacZ* gene (wavy line) and the orientation of transcription of the *cpsB* gene (right to left). Three Kohara phage clones from the 45-min region of the chromosome rescued the three *cps* mutations shown. E, *Eco*RI.

no β-galactosidase (data not shown). Quantitation of the levels of β-galactosidase in these cells was inconsistent, most likely because the fusions were present on multicopy plasmids, which often interferes with regulatory mechanisms.

Transfer of fusions to phage and expression. In order to test the fusions found in single copies in cells, the fusions found on the plasmids pDG3 (*cpsP1*::*lacZ*) and pVS137 (*cpsP2*::*lacZ*) were transferred to λRS45 (Materials and Methods). The phage derived from pDG3 was named λcps1, and the phage derived from pVS137 was named λcps2.

Lysogens were constructed with the two fusion phages λcps1 and λcps2. To ensure that the same number of prophage were inserted into the different strains tested, a strain that was wild type for the *cps* regulators (SG20250) was lysogenized and mutants were made by transduction of the regulatory mutations into this lysogenized wild-type strain. The lysogens were assayed for β-galactosidase-specific activity, expressed from the *cps*::*lacZ* fusion prophage. Table 3 shows the results of the β-galactosidase assays. In a wild-type background, very low levels of β-galactosidase were synthesized. The *lon* mutation increased expression 25- to 30-fold, and the *rcsC137* mutation increased expression approximately 100-fold. The increased expression in the *rcsC137* host was suppressed to wild-type levels by a mutation in *rcsB*. And, as expected, the *rcsB* mutant expresses very low amounts of β-galactosidase. Similar results were obtained for the λcps1 and λcps2 lysogens.

Sequence analysis. The sequence of the entire 0.5-kb *Hin*cII fragment that is present in pVS134 was determined (Materials and Methods) (Fig. 3). This region is approximately 63% AT rich. Although no obvious σ-70 promoter regions were found, two good −10 regions were detected at bp 47 to 51 and bp 162 to 167 (six and five of six matches, respectively, to the consensus TATAAT) (16), as would be expected for an AT-rich region. In addition to these features, there are two inverted repeat structures at bp 124 to 139 and bp 169 to 182. The conserved JUMPstart sequence that is found in the noncoding region of several polysaccharide gene clusters (18) is also found in the *cps* 0.5-kb *Hin*cII fragment at bp 243 to 281.

An open reading frame (ORF) starts at bp 460 and extends through the end of the sequence shown. An additional 314 bp of sequence downstream from that in the 0.5-kb *Hin*cII fragment was determined in order to confirm that this ORF extends beyond the 30 codons within the 0.5-kb *Hin*cII fragment. In fact, the reading frame continues through the DNA sequence shown and beyond (data not shown). The deduced amino acid sequence of the amino terminus of this ORF (Cps ORF) encoded by the first gene in the *cps* operon is shown in Fig. 3. A search of the database revealed high degrees of homology to two proteins involved in exopolysaccharide biosynthesis in *Erwinia amylovora* (AmsH) (5) and *Klebsiella pneumoniae* (ORF4) (2) and limited homologies to three other proteins involved in exopolysaccharide biosynthesis in *Pseudomonas solanacearum* (EpsA), *Haemophilus influenzae* (BexD) (24), and *Neisseria meningitidis* (CtrA) (8) (Fig. 4). Although the functions of all five of these proteins are unknown, they have been implicated in the export of capsular polysaccharides and CtrA has been shown to be an outer membrane protein. When aligned, all of these proteins contain a cysteine residue at approximately amino acid 21. It has been suggested that this residue may function as a lipid attachment site for AmsH and BexD (5). The entire exopolysaccharide capsule gene complexes from *Erwinia*, *Klebsiella*, *Pseudomonas*, *Neisseria*, and *Haemophilus* species have been sequenced. In *Erwinia* and *Klebsiella* species, the homologous proteins (AmsH and ORF4, respectively) are the second ORFs in the operon, whereas the *Pseudomonas*, *Neisseria*, and *Haemophilus* homologous proteins, EpsA, CtrA, and BexD, are the first ORFs in their operons, as shown in this manuscript for *E. coli.*

Transcription start site. The start site of transcription was determined by primer extension using RNA extracted from cells overexpressing RcsA from a plasmid. These cells were used because overexpressed RcsA has been shown to increase transcription of the *cps* genes by approximately 400-fold (37). Figures 3 and 5 show that the 5′ end of the *cps* transcript is a G residue that is 340 bp upstream of the first ORF (Cps ORF) and upstream of the JUMPstart sequence. This start site shows that the two best −10 regions are not used and that the actual

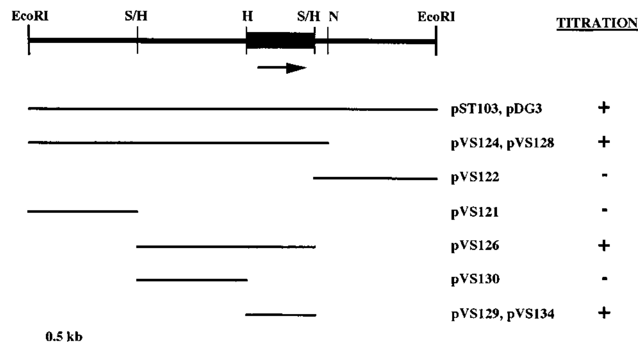

FIG. 2. Subcloning of the *cps* promoter-operator region. The results in the titration column indicate which of the fragments, when cloned into a multicopy plasmid, can (+) and cannot (−) reduce the expression of the single copy *cpsB10*::*lacZ* gene fusion. The solid box represents the 0.5-kb fragment in which the promoter and operator are located. The arrow indicates the direction of transcription of the *cps* genes. S, *Sal*I; H, *Hin*cII; N, *Nru*I.

TABLE 2. Construction and expression of fusion plasmids

Plasmid	Parent plasmid	Insert size (kb)	Orientation[a]	Color of expression on lactose MacConkey medium in mutant host strain[b]		
				Wild type	*rcsC137*	*lon rcsA*
pDG3	pST103	3.0	→	White	Pink	White
pVS128	pVS124	2.2	→	White	Pink	White
pVS113	pVS124	2.2	←	Red	Red	Red
pVS137	pVS129	0.5	→	White	Red	White
pVS136	pVS129	0.5	←	Red	Red	Red

[a] Arrows indicate orientation of *cps* fragment relative to *lacZ*. The arrow pointing to the right is in the same orientation as the arrow in Fig. 2.
[b] Wild type, SG1132; *rcsC137*, VS2038; *lon rcsA*, VS2032. All colonies expressed in the mutant host strain *lon* (JK1261A) were red.

−35 and −10 regions of this operon must be TTGCCT and AACACT, which only have four and three of six matches, respectively, to the consensus promoter sites (16).

DISCUSSION

Regulation of the colanic acid polysaccharide synthesis (*cps*) genes has been well studied (4, 10, 13, 14, 33, 34, 37, 39), although the *cps* gene cluster at 45 min on the *E. coli* chromosome has not been as thoroughly characterized (38). Five complementation groups, *cpsA*, *cpsB*, *cpsC*, *cpsD*, and *cpsE*, have been mapped to this site (38). It was not clear from Trisler and Gottesman's study (38) if these complementation groups represented five genes or five operons. The data presented in this work demonstrate that there is a regulated promoter region approximately 13.5 kb upstream of one of the *cps* genes, *cpsB*, that has been used to define the Rcs regulatory network. No other operator regions were detected by the titration assay, suggesting either that no other regulated promoters are present in this region or that they were not detected by this assay because the restriction enzyme disrupted them or that lethal fusions were created. It is very possible that in addition to the regulated promoter detected here, there are other promoters downstream that may act constitutively or be regulated by proteins other than the Rcs system.

The source of the DNA used for cloning the *cps* gene region was the Kohara phage library. Despite the library being constructed in lytic phage, it was possible to rescue several *cps* mutations with the library. It often took more than 16 h for the rescues to occur, suggesting that they were not the result of complementation. These rescues were likely to have been the result of recombination or homogenotization rather than simple complementation. The results of the rescues indicated that *cpsD* and *cpsE* lie upstream of *cpsB*, but it was not possible to assign an order to *cpsD* and *cpsE* relative to each other. It was surprising that *cpsD* was only rescued by 354, since there is no DNA that is unique to this phage and not carried by either 353 or 355. Two possible explanations for these results are that the insertion mutations in the *cps* genes are polar and therefore may eliminate transcription of the downstream genes and that the recombinations do not allow complete rescue. Because no rescue of *cpsC* or *cpsA* was accomplished, it is not possible to assign any order to these genes relative to the *cpsB*, *cpsD*, and *cpsE* genes.

The assignment of the *cpsB* gene to the site indicated in Fig. 1 was done by comparing restriction maps of the Kohara library and the *cpsB10*::*lacZ* clone. This fusion clone has not exhibited activity in any strain, including the *rcsC137* mutant that normally shows constitutive expression of the capsule (35). This suggests that the clone may be missing a promoter. If the other *cps* genes are upstream of the *cpsB* and there is a single promoter for the *cps* genes, as suggested by the study described in this work, then one would expect that the *cpsB*::*lacZ* fusion clone would be missing a promoter.

There are clearly sequences in both directions of the cloned fragments that can act as promoters as indicated by the plasmid fusion data (Table 2). However, one of these promoters is expressed constitutively and the other one is regulated by the Rcs system of regulators. Nothing is known about the putative promoter or genes in the opposite orientation. When either the 0.5-kb *Hin*cII fragment (*cpsP2*::*lacZ*) or the 3.0-kb *Eco*RI fragment (*cpsP1*::*lacZ*) was found in single copies, the levels of transcription directed by the fusion were approximately equal in all the strains tested, indicating that the promoter for the *cps* gene cluster was found within the 0.5-kb fragment.

Qualitatively, these fusions were regulated in a manner similar to that of the previously defined chromosomal *cpsB10*::*lacZ* fusion. However, the increases were not as great with the *cpsP2*::*lacZ* fusion or the *cpsP1*::*lacZ* fusion (100-fold in an *rcsC137* mutant and 30-fold in a *lon* mutant) as has been seen with the *cpsB10*::*lacZ* fusion (1,000-fold in an *rcsC137* mutant and approximately 100-fold in a *lon* mutant [14]). Therefore, there may be additional promoters or additional positive regulatory elements upstream or downstream from the 3.0-kb fragment that aid in activation of capsule expression in *rcsC137* and *lon* mutant strains. None of the results in this work exclude the possibility of other promoters downstream of the 3.0-kb fragment that could affect only some of the *cps* genes in this long operon.

Initial identification of the operator region by titration of a limiting positive regulator of the *cps* genes indicated that the binding site of a limiting positive regulator was on the 0.5-kb *Hin*cII fragment. The limiting regulator in this assay is likely to be RcsA. Clearly RcsA is limiting in a cell with a wild-type *lon* gene that encodes the Lon protease that degrades RcsA (35). Even in a *lon* mutant strain that lacks active protease, so that the half-life of RcsA is lengthened from 3 to 30 min (36), RcsA appears to be more limiting than RcsB. In addition, when two copies of the *rcsA* gene are present in the cell, capsule expression is increased significantly, whereas when two copies of *rcsB* are present in a cell, capsule expression is essentially unchanged. These results make it likely that the binding site found within the 0.5-kb *Hin*cII fragment is an RcsA binding site.

TABLE 3. Regulation of cloned promoter pieces

Strain	Phage[a]	Relevant genotype	β-Gal sp act (Miller units)[b]
VS20595	λcps1	Wild type	3
VS20596	λcps2	Wild type	4
VS20598	λcps1	*lon*	89
VS20599	λcps2	*lon*	102
VS20604	λcps1	*rcsC137*	305
VS20605	λcps2	*rcsC137*	350
VS20607	λcps1	*rcsB*	3
VS20608	λcps2	*rcsB*	4
VS20601	λcps1	*rcsC137 rcsB*	3
VS20602	λcps2	*rcsC137 rcsB*	2

[a] λcps1 is the result of a cross between λRS45 and pDG3. It has the 3.0-kb piece promoting transcription of *lacZ* (*cpsP1*::*lacZ*). λcps2 is the result of a cross between λRS45 and the equivalent of pVS129. It has the 0.5-kb piece promoting transcription of *lacZ* (*cpsP1*::*lacZ*).
[b] Specific activity of β-galactosidase expressed in Miller units (28).

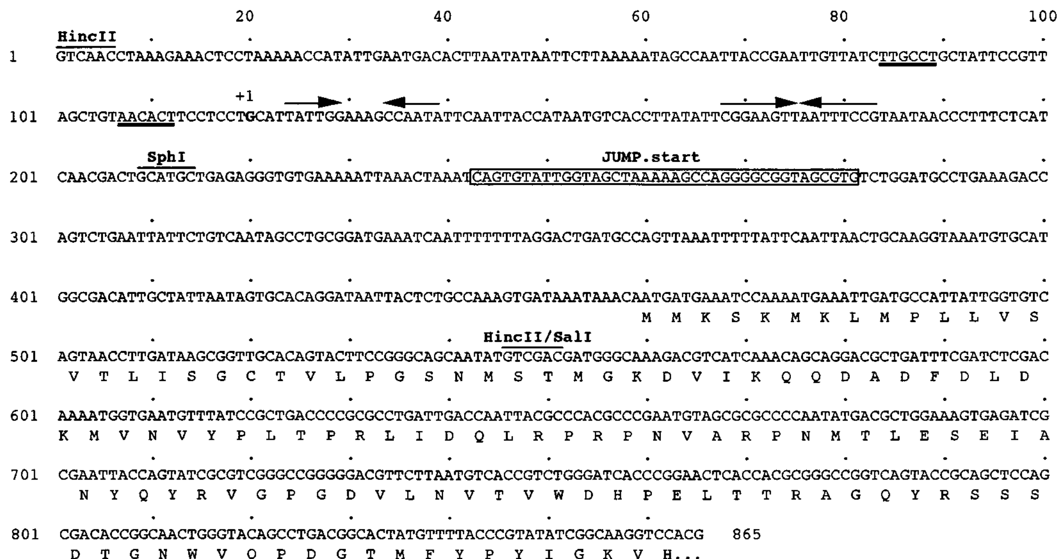

FIG. 3. Sequence of the *cps* promoter-operator region. Underlined sequences indicate −10 and −35 sites of the promoter. The G residue in boldface type with the +1 above it at bp 120 is the start site of transcription. The arrows indicate inverted repeats. The JUMPstart sequence is boxed. The deduced amino acid sequence of the beginning of the first ORF is shown in one-letter abbreviations below the DNA sequence.

The precise binding site of RcsA and RcsB has not been biochemically determined yet. However, sequence analysis of the 0.5-kb fragment showed that there are two inverted repeats with stem sizes of 6 and 7 bases. One of these may serve as the binding site of RcsA and/or RcsB, since the binding sites of many DNA-binding proteins are characterized by inverted repeats that vary in length. This seems unlikely, however, since both of these inverted repeats are downstream of the transcriptional start site. In addition, if these are sites where RcsA and RcsB bind, then they would be expected to be found in other organisms that have genes regulated by these two proteins, such as *K. pneumoniae* (1, 27) and *Erwinia* spp. (3, 6, 7, 29). However, when upstream regions of the *Klebsiella cps* and *Erwinia ams* operons were searched, no identical inverted repeats were found. The closest sequences that could be found are several half sites that were identical or very similar to the inverted repeat centered at bp 131. These may provide enough of a domain for the RcsA and/or RcsB proteins to bind. Rather than being binding sites for RcsA and/or RcsB, the inverted repeats are more likely to be sites where negative regulators bind and repress *cps* transcription. For example, several global regulators, such as cyclic AMP receptor protein and H-NS affect transcription of the *cps* genes (12).

Also found within the 0.5-kb *Hin*cII fragment is the conserved domain (39 bp) of a sequence that recently has been found just upstream of many bacterial polysaccharide gene

FIG. 4. Amino acid sequence similarities. Alignment of the amino terminus of the first Cps ORF with the amino termini of AmsH from *E. amylovora* (5), ORF4 from *K. pneumoniae* (2), EpsA from *P. solanacearum* (accession number U17898), BexD from *H. influenzae* (24), and CtrA from *N. meningitidis* (8). Amino acid residues are designated with the single-letter code. Identical amino acid residues are indicated with a solid vertical line (|), and conserved amino acid residues are indicated with two dots (:). The classification scheme for the conserved amino acids is as follows: nonpolar, ILMVAFW; polar, uncharged, STY; aromatic, FWY; amide, NQ; basic, HKR; acidic, DE. The asterisk shows the conserved cysteine residue that has been proposed to be the site of lipid attachment.

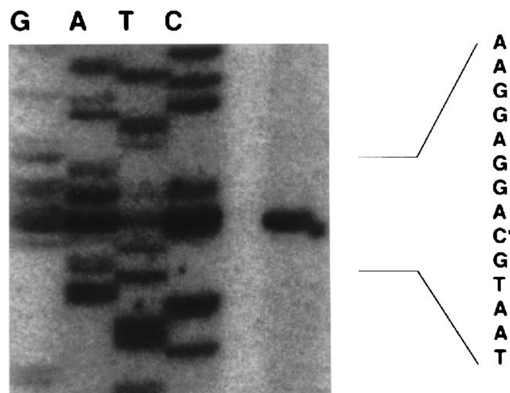

FIG. 5. Start site of transcription. The results of primer extension analysis of the 5′ end of the *cps* transcript from a strain overexpressing RcsA from a plasmid are shown. An end-labeled primer (cps8) complementary to the coding strand was annealed to total RNA and extended with reverse transcriptase. The products were electrophoresed in parallel with a sequence ladder (lanes G, A, T, and C) generated with the same primer.

clusters (18). This JUMPstart sequence (named JUMP for just upstream of many bacterial polysaccharide gene clusters) is often found between the promoter and translational start sites of O antigen and capsular polysaccharides in *E. coli* (group II capsules) and *Shigella*, *Salmonella*, *Yersinia*, *Klebsiella*, *Erwinia*, and *Vibrio* species. All of these species contain JUMPstart sequences, suggesting that there may be a conserved regulatory element(s) among these exopolysaccharide biosynthetic operons. (There is no JUMPstart sequence located upstream of the other three ORFs from *Pseudomonas*, *Neisseria*, and *Haemophilus* species, shown in Fig. 4.) Identification of the role of this conserved element in capsule regulation will be an important objective to pursue in the future. The JUMPstart sequence is found in the unusually long (340-nuceotide) noncoding region of the transcript. There are no ORFs in this region, although there could be a transcript in the opposite direction (as suggested by the unregulated fusions in this orientation). Examination of the possible function(s) of this long leader sequence in the future will be interesting.

In addition to the homologies detected among the upstream, untranslated regions of the *E. coli cps* operon and *Klebsiella* and *Erwinia* operons, there are homologies in the ORFs encoding the Cps ORF, ORF4, and AmsH proteins (respectively). The amino terminus of the first ORF of *E. coli* was very similar to the amino termini of the second ORFs from *Klebsiella* and *Erwinia* exopolysaccharide biosynthetic operons. It is possible that divergent evolution has occurred among these three enterobacterial species, so that a deletion occurred in the *E. coli* colanic acid biosynthetic operon or one additional gene was inserted into the other two operons. The first *Erwinia* ORF encodes an enzyme (AmsG) involved in the transfer of a galactose residue from UDP to a phosphorylated lipid carrier in the inner membrane. No function for the first ORF (ORF3) of the *Klebsiella* operon has been identified, and it does not demonstrate homology to the *Erwinia* AmsG. Thus, the second explanation of insertions occurring in *Klebsiella* and *Erwinia* is more likely. The homology seen between the *E. coli* Cps ORF and the other exopolysaccharide biosynthetic gene products may suggest that capsule biosynthesis may proceed along some similar pathways in many bacterial species. Although the precise function of any of these proteins is not known, BexD has been shown to be an integral outer membrane protein. Further sequencing of the *cps* operon should yield valuable information that may aid in evaluation of the divergence of capsule synthesis genes.

ACKNOWLEDGMENTS

I thank Susan Gottesman for strains and valuable discussions; Sumati Murli, Jennifer Huang, and Stacy Nakamura for technical assistance; and Geeta Gupte for reviewing this manuscript. I am indebted to Darren Sledjeski for his help with the primer extension assay.

This work was supported by grant MCB 9221085 from the National Science Foundation.

REFERENCES

1. **Allen, P., C. A. Hart, and J. R. Saunders.** 1987. Isolation from *Klebsiella* and characterization of two *rcs* genes that activate colanic acid capsular biosynthesis in *Escherichia coli*. J. Gen. Microbiol. **133:**331–340.
2. **Arakawa, Y., R. Wacharotayankun, T. Nagatsuka, H. Ito, N. Kato, and M. Ohta.** 1995. Genomic organization of the *Klebsiella pneumoniae cps* region responsible for serotype K2 capsular polysaccharide synthesis in the virulent strain Chedid. J. Bacteriol. **177:**1788–1796.
3. **Bernhard, F., K. Poetter, K. Geider, and D. L. Coplin.** 1990. The *rcsA* gene from *Erwinia amylovora*: identification, nucleotide sequence, and regulation of exopolysaccharide biosynthesis. Mol. Plant-Microbe Interact. **3:**429–437.
4. **Brill, J., C. Quinlan-Walshe, and S. Gottesman.** 1988. Fine-structure mapping and identification of two regulators of capsule synthesis in *Escherichia coli* K-12. J. Bacteriol. **170:**2599–2611.
5. **Bugert, P., and K. Geider.** 1995. Molecular analysis of the *ams* operon required for exopolysaccharide synthesis of *Erwinia amylovora*. Mol. Microbiol. **15:**917–933.
6. **Chatterjee, A., W. A. Chun, and A. K. Chatterjee.** 1990. Isolation and characterization of an *rcsA*-like gene of *Erwinia amylovora* that activates extracellular polysaccharide production in *Erwinia* species, *Escherichia coli* and *Salmonella typhimurium*. Mol. Plant-Microbe Interact. **3:**144–148.
7. **Coleman, M., R. Pearce, E. Hitchin, J. Busfield, J. W. Mansfield, and I. S. Roberts.** 1990. Molecular cloning, expression and nucleotide sequence of the *rcsA* gene of *Erwinia amylovora*, encoding a positive regulator of capsule expression: evidence for a family of related capsule activator proteins. J. Gen. Microbiol. **136:**1799–1806.
8. **Frosch, M., U. Edwards, K. Bousset, B. Krausse, and C. Weisgerber.** 1991. Evidence for a common molecular origin of the capsule gene loci in Gram-negative bacteria expressing group II capsular polysaccharides. Mol. Microbiol. **5:**1251–1263.
9. **Genetics Computer Group.** 1991. Program manual for the Wisconsin package, version 7, April 1991. 575 Science Dr., Madison, Wis.
10. **Gervais, F. G., P. Phoenix, and G. R. Drapeau.** 1992. The *rcsB* gene, a positive regulator of colanic acid biosynthesis in *Escherichia coli*, is also an activator of *ftsZ* expression. J. Bacteriol. **174:**3936–3971.
11. **Geobel, W. F.** 1963. Colanic acid. Proc. Natl. Acad. Sci. USA **49:**464–471.
12. **Gottesman, S.** Personal communication.
13. **Gottesman, S., and V. Stout.** 1991. Regulation of capsular polysaccharide synthesis in *Escherichia coli* K-12. Mol. Microbiol. **5:**1599–1606.
14. **Gottesman, S., P. Trisler, and A. Torres-Cabassa.** 1985. Regulation of capsular polysaccharide synthesis in *Escherichia coli* K-12: characterization of three regulatory genes. J. Bacteriol. **162:**1111–1119.
15. **Gutnick, D.** Personal communication.
16. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11:**2237–2255.
17. **Hinton, D.** 1989. Transcript analyses of the *uvsX*-40-41 region of bacteriophage T4. J. Biol. Chem. **264:**14432–14439.
18. **Hobbs, M., and P. Reeves.** 1994. The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters. Mol. Microbiol. **12:**855–856.
19. **Houng, H.-S. H., K. F. Noon, J. T. Ou, and L. S. Baron.** 1992. Expression of Vi antigen in *Escherichia coli* K-12: characterization of ViaB from *Citrobacter freundii* and identity of ViaA with RcsB. J. Bacteriol. **174:**5910–5915.
20. **Jann, K., and B. Jann.** 1987. Polysaccharide antigens of *Escherichia coli*. Rev. Infect. Dis. **9:**S517–S526.
21. **Jayaratne, P., W. J. Keenleyside, P. R. MacLachlan, C. Dodson, and C. Whitfield.** 1993. Characterization of *rcsB* and *rcsC* from *Escherichia coli* O9:K30:H12 and examination of the role of the *rcs* regulatory system in expression of group I capsular polysaccharides. J. Bacteriol. **175:**5384–5394.
22. **Keenleyside, W. J., P. Jayaratne, P. R. MacLachlan, and C. Whitfield.** 1992. The *rcsA* gene of *Escherichia coli* O9:K30:H12 is involved in the expression of the serotype-specific group I K (capsular) antigen. J. Bacteriol. **174:**8–16.
23. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50:**495–505.

24. **Kroll, J. S., B. Looynds, L. N. Brophy, and E. R. Moxon.** 1990. The *bex* locus in encapsulated *Haemophilus influenzae*: a chromosomal region involved in capsular polysaccharide export. Mol. Microbiol. **4:**1853–1862.
25. **Markovitz, A.** 1977. Genetics and regulation of bacterial capsular polysaccharide biosynthesis and radiation sensitivity, p. 415–462. *In* I. Sutherland (ed.), Surface carbohydrates of the prokaryotic cell. Academic Press, Inc. (London), Ltd., London.
26. **Maurizi, M., P. Trisler, and S. Gottesman.** 1985. Insertional mutagenesis of the *lon* gene in *Escherichia coli*:*lon* is dispensable. J. Bacteriol. **164:**1124–1135.
27. **McCallum, K. L., and C. Whitfield.** 1991. The *rcsA* gene of *Klebsiella pneumoniae* O1:K20 is involved in expression of the serotype-specific K (capsular) antigen. Infect. Immun. **59:**494–502.
28. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Poetter, K., and D. L. Coplin.** 1991. Structural and functional analysis of the *rcsA* gene from *Erwinia stewartii*. Mol. Gen. Genet. **229:**155–160.
30. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
31. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **Simons, R. W., W. F. Houman, and N. Kleckner.** 1987. Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. Gene **53:**85–96.
33. **Stout, V.** 1994. Regulation of capsule synthesis includes interactions of the RcsC/RcsB regulatory pair. Res. Microbiol. **145:**43–46.
34. **Stout, V., and S. Gottesman.** 1990. RcsB and RcsC: a two-component regulator of capsule synthesis in *Escherichia coli*. J. Bacteriol. **172:**659–669.
35. **Stout, V., and S. Gottesman.** Unpublished data.
36. **Stout, V., A. Torres-Cabassa, M. Maurizi, D. Gutnick, and S. Gottesman.** 1991. RcsA, an unstable positive regulator of capsular polysaccharide synthesis. J. Bacteriol. **173:**1738–1747.
37. **Torres-Cabassa, A. S., and S. Gottesman.** 1987. Capsule synthesis in *Escherichia coli* K-12 is regulated by proteolysis. J. Bacteriol. **169:**981–989.
38. **Trisler, P., and S. Gottesman.** 1984. *lon* transcriptional regulation of genes necessary for capsular polysaccharide synthesis in *Escherichia coli* K-12. J. Bacteriol. **160:**184–191.
39. **Zinkewich-Peotti, K., and J. M. Fraser.** 1988. New locus for exopolysaccharide overproduction in *Escherichia coli* K-12. J. Bacteriol. **170:**1405–1407.