JOURNAL OF BACTERIOLOGY, Oct. 1988, p. 4706–4713
0021-9193/88/104706-08$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 170, No. 10

Transcriptional Analysis of the Major Surface Array Gene of *Caulobacter crescentus*

JAMES A. FISHER,† JOHN SMIT,‡ AND NINA AGABIAN§*

Naval Bioscience Laboratory, School of Public Health, University of California-Berkeley, Berkeley, California 94720

Received 27 June 1988/Accepted 1 July 1988

The major component of the paracrystalline surface array of *Caulobacter crescentus* CB15 and one of the most abundant cellular proteins is a protein designated 130K. We have determined the DNA sequence of the 5′ portion of the 130K gene, including the N-terminal one-third of the protein coding region, and analyzed the transcription of the gene. The site of transcription initiation was determined by S1 mapping of *Caulobacter* RNA. Although the DNA sequence upstream from the transcription start site showed significant homology to the consensus promoter sequences of *Escherichia coli*, S1 analysis of RNA from *E. coli* carrying the 130K gene on a plasmid indicated that the 130K promoter was not transcribed by *E. coli* RNA polymerase in vivo. Quantitative S1 analysis of RNA isolated from synchronously growing *Caulobacter* cells suggested that this promoter was not under developmental regulation; the amount of 130K transcript varied no more than 1.5-fold during the cell cycle. The length of the 130K mRNA was determined to be 3.3 kilobases by Northern (RNA blot) analysis, indicating that the 130K mRNA is not part of a polycistron. The amino acid sequence predicted from the DNA sequence agreed well with the N-terminal amino acid sequence determined by sequencing of the 130K protein. The 130K protein appears to be synthesized without an N-terminal leader sequence, but the N-terminal 20 amino acids are relatively hydrophobic and may function like a signal sequence during transmembrane translocation.

Paracrystalline surface arrays consisting of regularly arranged protein subunits are commonly found in many bacterial genera (28, 29). The subunits assemble on the surface of the outer membrane or cell wall, and as such might be considered excreted proteins. The arrays probably serve a variety of functions relating to protecting the bacterial cell from external influences such as lytic enzymes or phagocytosis. Yet in most genera, specific roles have not been conclusively demonstrated. Considering the relatively large amount of the cell's total protein synthesis capacity which is devoted to producing these layers and the observation that they are frequently lost in laboratory culture (28, 29), it is often assumed that these layers serve specific and vital roles for the bacteria in their natural environment.

The periodic surface array of *Caulobacter crescentus* is a complex layer with three proteins somehow involved in its structure (32). The M_r-130,000 protein (130K protein), howver, is the major component of the *Caulobacter* surface array and is probably the most abundant protein in the cell, constituting about 5 to 7% of total cellular protein (32). A dimorphic bacterium, *C. crescentus* exists as a flagellated (swarmer) cell and a nonmotile stalked cell at different stages in its life cycle. Assembly of the surface array during morphogenesis apparently occurs by at least two distinct patterns of subunit incorporation: random addition of subunits occurs over the cell body within the preexisting array, while localized array formation occurs at the site of stalk elongation and along the cell division plane (31). The spatially regulated array formation is temporally regulated as well, occurring at precisely defined times during *Caulobacter* development. Underlying the spatial and temporal patterns of surface layer growth is an apparently constant rate of 130K protein synthesis during the cell cycle (1). The 130K protein therefore provides an interesting molecular paradox. How can the expression of a single-copy gene be regulated so that its assembly into the supramolecular surface array occurs both randomly during cell growth and localized during stalk formation and cell division?

We have undertaken a study of the synthesis and assembly of this component of the surface array and here analyze the transcription of the 130K gene and the structure of its promoter. We have determined the DNA sequence of the 5′ portion of the 130K gene, including sequence upstream from the protein coding region, and determined the site of transcription initiation. Only a few *Caulobacter* promoters have been characterized, and no general rules of structure and function have emerged. It has been assumed that the structure of *Caulobacter* promoters is significantly different from that of *Escherichia coli* promoters because the majority of *Caulobacter* genes cloned into *E. coli* are not transcribed by *E. coli* (4, 18, 30, 36). Of the *Caulobacter* promoters which have been sequenced (2, 4, 19, 20, 25), the majority do not have sequences similar to the *E. coli* consensus promoter sequence. In contrast, the sequence of the 130K gene upstream region showed considerable similarity to the *E. coli* consensus promoter sequence. Nevertheless, direct evidence was obtained to show that the 130K promoter is not transcribed in *E. coli*. We further show that transcription of the 130K gene is essentially constant throughout the cell cycle.

MATERIALS AND METHODS

Bacterial strains and growth media. *C. crescentus* CB15A (ATCC 19089), a surface array-producing strain, was used in

* Corresponding author.
† Present address: Applied Biosystems, Inc., Foster City, CA 94404.
‡ Present address: Department of Microbiology, University of British Columbia, Vancouver, British Columbia V6T 1W5, Canada.
§ Present address: Intercampus Program on Molecular Parasitology, University of California-San Francisco, Suite 150, Laurel Heights Campus, Box 1204, San Francisco, CA 94143-1204.

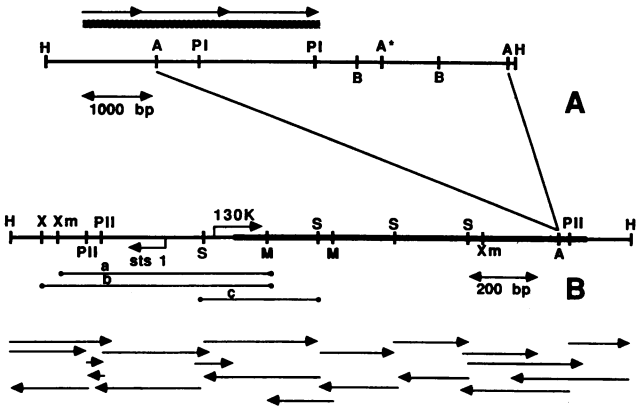

FIG. 1. Restriction map of the 130K gene. (A) Map of the A16 fragment. Cross-hatched area represents the extent of the 130K mRNA, transcribed left to right. Restriction sites are as defined below. A* refers to the presence of one or more *Ava*I sites between the two outermost *Ava*I sites. The precise map locations of these extra *Ava*I sites have not been determined. (B) Map of the E_1 *Ava* fragment and sequencing strategy. The E_1Ava subclone was constructed from the A16 insert by deletion of DNA between the outermost *Ava*I sites as described above. The line connecting parts A and B of the figure shows the extent of the DNA which has been deleted. The bold-face line denotes the 130K open reading frame, read left to right. The rightward arrow denotes the start site of the 130K transcript. The lower leftward arrow denotes the start site of the *sts1* transcript from the opposite strand (see text). Line segments a to c depict fragments used for probes in S1 or Northern analysis (see appropriate sections). The sequencing strategy shows to scale the number of bases read from each subclone. Restriction sites: A, *Ava*I; B, *Bam*HI; H, *Hin*dIII; M, *Mlu*I; PI, *Pvu*I; PII, *Pvu*II; S, *Sal*I; X, *Xho*II; Xm, *Xmn*I.

this work. Cells were grown in PYE medium (23) or in M_3HIGG minimal salts medium (32) at 30°C with shaking.

For analysis of developmental changes in 130K mRNA levels, CB15A cells were synchronized by the Ludox gradient technique (7). Cells were grown in M_3HIGG medium; differentiation was monitored by light microscopy. At specified times after synchronization, portions of cells were removed, pelleted by centrifugation, and frozen at −70°C until used for RNA extraction.

E. coli JM109 (21) was the host for M13 phage used in DNA sequencing, and strain K-12ΔHΔtrp (24) was used as the host for pLC28-E_1Ava. *E. coli* cells were grown in LB medium (14) or 2× YT medium (16). JM109 was grown at 37°C, and K-12ΔHΔ*trp* (which contains a λ lysogen with a temperature-sensitive *c*I857 repressor) was grown at 30°C. Plasmid pLC28 is a λ p_L promoter vector (24) and was used for expression of 130K protein in *E. coli*.

Purification and sequencing of 130K protein. The 130K protein was purified from shed cell surface complexes and used to raise antibodies in rabbits as described previously (32). N-terminal protein sequence analysis was performed on an Applied Biosystems model 477A/120A protein sequencer.

Sequencing of pLC28-E_1Ava. Plasmid pLC28-E_1Ava is derived by deletion of an internal *Ava*I fragment from plasmid pLC28-A16, which contains a 6.7-kilobase (kb) *Hin*dIII-*Hin*dIII fragment of *Caulobacter* DNA (30). The E_1 *Ava* fragment and various restriction fragments thereof were cloned into M13mp18 or mp19 (21). DNA sequencing by the dideoxy method (26) with [^{35}S]dATP was performed by standard protocols (3). The sequencing strategy is shown in Fig. 1B. The contiguity of two *Sal*I fragments was shown by high-resolution restriction mapping. Sequences were analyzed with the programs developed by the University of California–San Francisco Biomathematics Computing Center. Protein homology analysis was by the FASTP program (13), searching the National Biomedical Research Foundation protein sequence data base (November 1986 edition).

Northern (RNA blot) analysis. RNA was isolated from fresh or frozen *Caulobacter* cell pellets by proteinase K digestion, followed by extraction with phenol-chloroform (1:1). DNA was removed by precipitation with 3 M sodium acetate (pH 6), and RNA was collected by subsequent ethanol precipitations. RNA pellets were dried and suspended in distilled water.

RNA was electrophoresed in 1.2% agarose–2.2 M formaldehyde gels as described by Maniatis et al. (14). After electrophoresis, RNA was electroblotted onto a Nytran membrane (Schleicher & Schuell). Hybridization and washing were carried out under high-stringency conditions (14). Probes for Northern analysis were made by labeling single-stranded M13 templates with Klenow extension from the hybridization probe primer (12). From 1×10^7 to 2×10^7 cpm of probe was added to the hybridization mix.

S1 mapping. S1 mapping was carried out essentially as described by Gilman and Chamberlain (9). From 20,000 to 100,000 cpm of probe (see below) was mixed with various amounts of *Caulobacter* RNA and carrier tRNA and precipitated with ethanol. Pellets were suspended in 12 μl of hybridization buffer (80% formamide, 400 mM NaCl, 40 mM PIPES [pH 6.4], 1 mM EDTA) and denatured at 90°C for 10 to 12 min. Tubes were immediately transferred to a 60°C waterbath and incubated submerged for 3 to 4 h, and 300 μl of cold S1 buffer (280 mM NaCl, 30 mM sodium acetate (pH 4.6), 5 mM $ZnSO_4$, 20 μg of calf thymus DNA per ml, 200 U of S1 nuclease per ml) was then added, and the tubes were placed on ice. After cooling, the tubes were incubated for 30 min at 37°C to digest unhybridized RNA and probe. The S1 reaction was quenched by the addition of 75 μl of 2.5 M ammonium acetate–50 mM EDTA and 2.5 volumes of ethanol.

The ethanol-precipitated reaction products were analyzed by electrophoresis on 5% polyacrylamide (29:1, acrylamide-bisacrylamide)–8.3 M urea gels. Precipitated reaction products were suspended in gel loading buffer (80% formamide, 1 mM EDTA, 5 mM NaOH, 0.25% each bromphenol blue and xylene cyanol), heated to 65°C, and loaded on the gel. For high resolution, 60-cm sequencing gels were run at 2,000 V. For lower resolution (e.g., quantitative analysis), 15-cm gels were run at 170 V. After electrophoresis, gels were dried and autoradiographed.

Double-stranded 5′-end-labeled restriction fragments were used as probes in the S1 experiments. The map locations of the probe fragments are shown in Fig. 1B. The 612-base-pair (bp) *Xho*II-*Mlu*I fragment was dephosphorylated with bacterial alkaline phosphatase and end-labeled with [γ-^{32}P]ATP by using T4 polynucleotide kinase (14). The labeled fragment was then digested with *Xmn*I to generate two fragments, each labeled at only one 5′ end. The 573-bp *Xmn*I-*Mlu*I fragment was purified by agarose gel electrophoresis onto DEAE membranes (6). In some experiments, the end-labeled *Xho*II-*Mlu*I fragment probe was not restricted, allowing simultaneous visualization of transcripts proceeding in opposite directions from promoters on the same fragment.

In quantitative S1 experiments, autoradiograms were quantitated by densitometry with a Bio-Rad model 620 densitometer. The area under the peaks was determined with a Beckman 427 integrator.

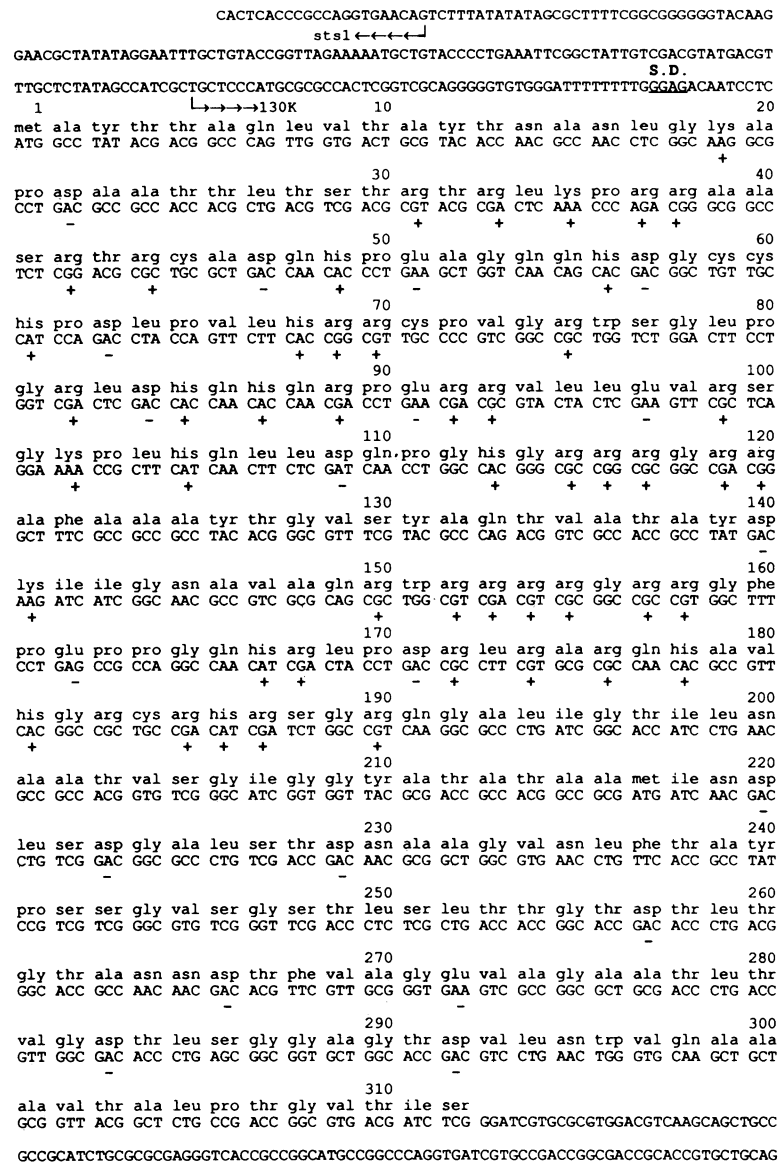

FIG. 2. Sequence of the 130K open reading frame. The DNA sequence of the E_1 *Ava* fragment was determined, and the single long open reading frame was translated into amino acid sequence The open reading frame extends 25 amino acids beyond the last residue shown here, but those residues are beyond the *Ava*I deletion site and are not present in the full-length 130K protein. Shown upstream of the initiator methionine is the putative Shine-Delgarno sequence (S.D.), the 130K transcription start site (rightward arrow), and the *sts1* transcription start site (leftward arrow). + and −, Charges of indicated amino acids.

RESULTS

Partial sequence of the 130K gene and protein. The 130K gene was initially cloned into λ1059 and identified by expression of immunoreactive 130K protein (30). The entire 130K coding sequence was located within a 6.7-kb *Hin*dIII-*Hin*dIII fragment, the restriction map of which is shown in Fig. 1A. To facilitate analysis of the promoter, we constructed a smaller plasmid by deleting sequence between the outermost *Ava*I sites on the 6.7-kb fragment. A detailed map of the 1.8-kb insert of the smaller plasmid, designated pLC28-E_1Ava, is shown in Fig. 1B. *E. coli* K-12ΔHΔtrp containing pLC28-E_1Ava, when grown at 42°C, expressed a protein of about 30 kilodaltons (kDa) which was immunoreactive with anti-130K antibody in a Western blot immunoassay (data not shown). Transcription was presumed to be from the λ p_L promoter on the vector.

The sequence of the 1.8-kb E_1 *Ava* fragment (Fig. 2) revealed a single long open reading frame. The protein encoded in the open reading frame had a predicted molecular weight of 35,684, consisting of 337 amino acids. Figure 2 shows the amino acid sequence up to the *Ava*I restriction site where the deletion creating the E_1 *Ava* fragment was made. Residues 313 to 337 of the 35-kDa protein were downstream from the deletion site and were not present in the full-length 130K protein.

The sequence of the N-terminal 20 amino acids of purified 130K protein was identical to the predicted amino acid sequence of the long open reading frame, with the exception

	1	2	3	4	5	6	7	8	9	10	
130K Protein	ala	tyr	thr	thr	ala	gln	leu	val	thr	ala	
	1	2	3	4	5	6	7	8	9	10	11
130K ORF	met	ala	tyr	thr	thr	ala	gln	leu	val	thr	ala
	11	12	13	14	15	16	17	18	19	20	
130K Protein	tyr	*nc*	asn	ala	asn	leu	*nc*	lys	ala	pro	
	12	13	14	15	16	17	18	19	20	21	
130K ORF	tyr	thr	asn	ala	asn	leu	gly	lys	ala	pro	

nc= no call

FIG. 3. Comparison of N-terminal sequences of the 130K protein and 130K open reading frame (ORF). The 130K protein was isolated and its N-terminal sequence was determined. *nc*, Residues could not be assigned due to low signal-to-noise ratio in the sequencer output. The N-terminal amino acid sequence predicted from the DNA sequence is shown in alignment with the protein sequence.

of the initiator methionine predicted in the DNA sequence and two residues that could not be assigned by the protein sequencer (Fig. 3). We conclude that the open reading frame of the E_1 *Ava* fragment encodes the 130K gene and that the protein expressed by pLC28-E_1Ava is a truncated 130K protein. Since the insert contains nearly 800 bases upstream of the 5′ end of the 130K protein-coding region, we inferred that the insert contained the promoter for the 130K gene as well.

A sequence homology search of the National Biomedical Research Foundation protein sequence data base did not reveal significant similarity between the partial 130K sequence and any other proteins. To our knowledge, no other bacterial surface proteins have been sequenced, so it was not possible to compare the 130K sequence with those proteins. Codon usage in the 130K sequence was unremarkable, except that the third base of degenerate codons tended to be G or C, which is not surprising given the high G+C content of *Caulobacter* DNA.

The determination of the translation initiation site was based on the presence of a likely Shine-Dalgarno sequence (GGAG) between 10 and 13 bases upstream of the presumed initiator methionine (27). Another AUG codon was located inframe 54 bases upstream, but it was only 9 bases downstream from the transcription start site (see below) and was not preceded by a sequence resembling the Shine-Dalgarno sequence. The downstream methionine was therefore assigned as the translation initiation site, although we cannot completely exclude the other possibility.

Assuming that translation of the 130K protein initiates at the predicted position, the amino acid sequence of the primary translation product was the same as that of the mature 130K protein, except for the N-terminal methionine, which is presumably cleaved off after translation. Although there was no putative signal sequence, the N-terminal 18 amino acids were either neutral or hydrophobic. A plot of hydrophobicity (not shown) indicated that the N-terminal 18 amino acids were somewhat less hydrophobic than signal sequences from other membrane and secreted proteins.

Downstream of the first 18 residues of the 130K protein, two distinct domains of amino acid composition were observed. Between residues 19 and 190, the protein was rich in basic amino acids, especially arginine. From residues 191 to 312, the protein consisted mostly of neutral and hydrophobic amino acids and a few acidic amino acids. No basic residues were found in this latter region.

Determination of the site of transcription initiation. The site

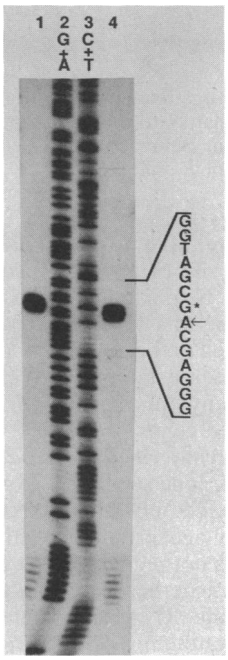

FIG. 4. High-resolution S1 mapping of 130K transcription initiation site. S1 mapping of *Caulobacter* RNA was performed with probe fragment a (shown in Fig. 1B) 5′-end labeled only at the *Mlu*I end. Probe was also subjected to chemical sequencing reactions (14). S1-protected and chemically sequenced probes were electrophoresed on sequencing gels. Lanes 2 and 3 show the G+A and C+T sequencing tracks, respectively. Lanes 1 and 4 show the S1-protected fragments from identical reactions. Alongside lane 4 is the DNA sequence of the probe in the area around the S1-protected bands. The asterisk denotes the G residue in the probe which migrated closest to the major S1-protected fragment. The transcription initiation site was determined after correction for differential migration of S1-protected and chemically sequenced fragments (discussed in text). This correction shows the initiation site to be a T in the sense strand, corresponding to the A base in the probe strand (arrow).

of initiation of the 130K mRNA was determined by S1 mapping. A 5′-end-labeled *Xmn*I-*Mlu*I fragment labeled only at the *Mlu*I end was used as the probe (fragment b in Fig. 1B). Figure 4 shows the results of this experiment. The major S1-protected fragment migrated slightly more slowly than the indicated G residue. S1-protected fragments have been found to migrate approximately 1.5 bp more slowly than chemically sequenced fragments (33). When corrected for this difference in migration, the data indicate that the 130K transcript initiated at a T residue on the sense strand, corresponding to the A 1 base downstream from the indicated G on the probe (antisense) strand. When *Caulobacter* RNA was omitted from the hybridization, no protected fragment of any size was seen, save a small amount of undigested full-length probe (data not shown).

The sequence of bases −50 to +10 relative to the transcription start site is shown in Fig. 5. Examination of this sequence revealed two 6-base segments around bases −10 and −35 which were similar to the *E. coli* consensus −10 and −35 promoter sequences (10). In the 130K sequence, three of the six bases in the −35 region and four of the six bases in the −10 region were the same as in the *E. coli* consensus. Spacing between the two regions was 17 bases, which is the optimal spacing between the −10 and −35 sequences in *E. coli* (10).

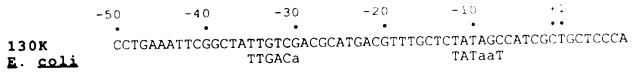

FIG. 5. Sequence upstream of 130K transcription initiation site. Bases −50 to +10 relative to the transcription initiation site of the 130K gene are shown. Below the 130K sequence is the consensus promoter sequence for *E. coli* (10).

At base −133 relative to the 130K start site was the transcription start site of a newly discovered gene, termed *sts1* (J. Fisher and N. Agabian, manuscript in preparation). Transcription of *sts1* was developmentally regulated, and the gene was transcribed in the opposite direction from 130K. Its function has not been established. We have not determined whether there is any functional overlap of regulatory sequences for the two genes.

130K gene is not transcribed in *E. coli*. In the original phage isolates of the 130K gene sequence, expression of the protein in *E. coli* was detected only when the gene was placed under the control of a phage promoter (30), suggesting that the 130K promoter was not recognized by *E. coli* RNA polymerase. We examined whether the 130K gene is transcribed in *E. coli* by S1 analysis. RNA was isolated from *C. crescentus* and from *E. coli* containing pLC28-E_1Ava; the *E. coli* cells were grown at 30°C to repress transcription from the p_L promoter, which could have potentially competed with the 130K promoter for RNA polymerase molecules. S1 mapping was carried out with 2.5 or 10 μg of *E. coli* RNA and 0.25 or 2.5 μg of *Caulobacter* RNA. The 130K mRNA was expected to protect a 147-base fragment from S1 digestion. Whereas the expected fragment was observed in both *Caulobacter* RNA samples (Fig. 6), that fragment was not detected in the *E. coli* samples. In fact, no protected fragment was detected with as much as 25 μg of *E. coli* RNA (data not shown), indicating that little or no 130K mRNA is transcribed in *E. coli* from the authentic *Caulobacter* promoter.

The experiment described above used a double-stranded *Xho*II-*Mlu*I fragment (fragment b) labeled on both 5′ ends, in contrast to the S1 mapping experiment shown in Fig. 4. This was done in order to detect transcripts of both the 130K and

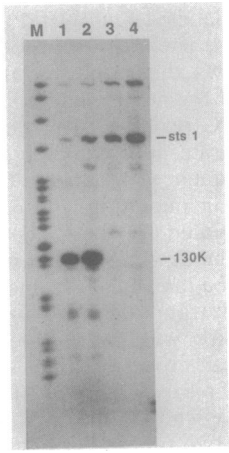

FIG. 6. 130K gene is not transcribed in *E. coli*. RNA was isolated from *E. coli* K-12ΔHΔtrp carrying pLC28-E_1Ava and from *C. crescentus*. S1 mapping was carried out on 0.25 μg (lane 1) or 2.5 μg (lane 2) of *Caulobacter* RNA and on 2.5 μg (lane 3) or 10 μg (lane 4) of *E. coli* RNA. Probe fragment b (Fig. 1B) was labeled on both ends and was therefore able to detect both the 130K and *sts1* transcripts going in opposite directions. The protected fragments corresponding to the 130K and the *sts1* transcripts are shown.

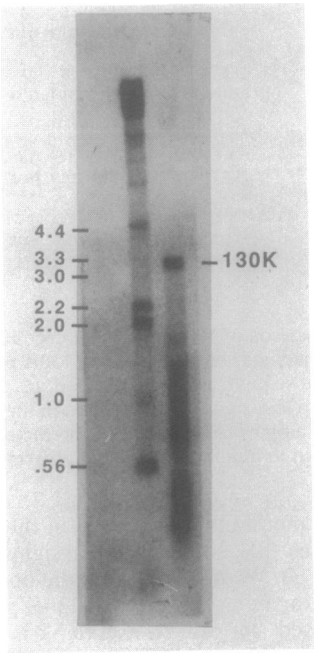

FIG. 7. Northern analysis of 130K mRNA. *Caulobacter* RNA (5 μg) was electrophoresed on 1% agarose–2.2 M formaldehyde gels and electrophoretically transferred to Nytran membranes. Single-stranded M13 containing fragment c (Fig. 1B) was used as the probe. Markers are in adjacent lanes with molecular sizes indicated (in kilobases).

sts1 genes. Because the *sts1* promoter is located 133 bp 5′ of the 130K promoter and initiates transcription in the opposite direction, it provided an internal control for transcription in *E. coli*. *sts1* mRNA should protect a 309-base fragment. That fragment was observed in the *Caulobacter* RNA samples and also in the *E. coli* RNA samples, indicating that *sts1* is transcribed in *E. coli* whereas 130K is not.

If significant transcriptional initiation had occurred at the p_L promoter, we would have expected to see protection of full-length probe. Although a small amount of full-length probe was in fact protected, the amount of full-length fragment was much less than the amount protected by the *sts1* transcript.

Cell cycle regulation of 130K gene expression. The size of the 130K mRNA was determined by Northern blot analysis of *Caulobacter* RNA by using as the probe a single-stranded M13 clone of fragment c (Fig. 1B). The length of the mRNA was 3.3 kb (Fig. 7), which is adequate to code for a protein of 130 kDa. Since there was little excess coding capacity on the mRNA, we infer that it is not polycistronic. The location of the transcript relative to the restriction map is shown in Fig. 1A. The map location was determined by the transcription start site and the length of the mRNA.

To determine whether 130K mRNA was present at a constant level during the cell cycle, we performed quantitative S1 mapping on RNA isolated from synchronously growing *Caulobacter* cells at different times during the cell cycle. The results (Fig. 8) indicate that the level of 130K mRNA did not vary by more than 1.5-fold during the cell cycle, in agreement with the protein synthesis data (1). No protected fragment was seen when *Caulobacter* RNA was omitted from the hybridization (Fig. 8B, lane 2).

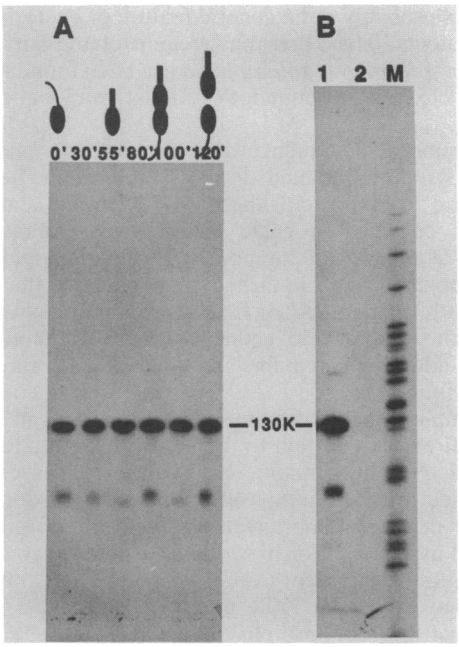

FIG. 8. (A) Quantitative S1 analysis of 130K mRNA levels. Quantitative S1 mapping was performed with 1.7 μg of RNA isolated from synchronized *Caulobacter* cells at the indicated times (minutes) during the cell cycle. The cell types present at different times were swarmer (0 and 30 min), stalk (55 and 80 min), and predivisional (100 and 120 min). Cell division occurred between 120 and 140 min. Probe fragment a was used in these experiments. The band protected by 130K mRNA is shown. (B) Assay performed with (lane 1) or without (lane 2) 5 μg of *Caulobacter* RNA. Both assays contained 20 μg of carrier yeast tRNA. Lane M shows end-labeled pBR322 *Hpa*II digest as molecular weight markers.

DISCUSSION

This report presents an analysis of the transcription of the 130K gene. We have determined the sequence of a segment of *Caulobacter* DNA which contains the N-terminus of the 130K protein-coding region, as well as the site of transcription initiation of 130K mRNA, and sequence upstream of the promoter. Although the upstream sequences are similar to the *E. coli* consensus promoter, the gene is not transcribed by *E. coli* RNA polymerase.

Synthesis of the 130K protein is essentially constant during the *Caulobacter* life cycle (1), and our results show that transcription of the 130K mRNA is essentially constant as well. In contrast, stage-specific *Caulobacter* genes show large variations in levels of their mRNAs. Expression of the flagellar hook protein (4) and the flagellin structural proteins (17, 19) is regulated at the level of transcription. Although synthesis of the 130K protein is not regulated, its assembly into the surface array is temporally and spatially regulated by some as yet unknown mechanism.

Since 130K is the most abundant protein in the *Caulobacter* cell and is encoded by a single-copy gene (30), we expected the 130K promoter to be relatively strong. Since the 130K and *sts1* promoters are only 133 bp apart, we were able to directly compare the levels of these two transcripts by S1 mapping with a double-stranded probe that spanned both transcription start sites. From experiments similar to that shown in Fig. 7, we estimated that the levels of 130K mRNA were at least 30 times higher than the levels of *sts1* mRNA in unsynchronized cells. In order to determine promoter strength from the relative transcript levels, it is necessary to consider transcript stability. The half-life of the 130K mRNA was 10 to 15 min, and the average half-life for *Caulobacter* mRNAs is 2 min (1). Assuming that the half-life for *sts1* mRNA is also 2 min, we estimate that the 130K promoter is indeed significantly stronger than the *sts1* promoter, perhaps by as much as six times. A more exact determination would depend on measuring the stability of the *sts1* transcript.

By determining the transcription start site of the 130K mRNA, we were able to begin analysis of its promoter and upstream regulatory sequences. The striking feature of the upstream region of the 130K promoter is its similarity to the consensus *E. coli* promoter sequence, a feature which was unexpected owing to the inability of *E. coli* RNA polymerase to transcribe the 130K gene. The 130K sequence is the same at several of the bases which are most highly conserved in the *E. coli* consensus sequence, e.g., the TTG in the −35 region and the TAT in the −10 region (10). Spacing between the −10 and −35 hexamers in the 130K promoter was 17 bp, which is optimal for *E. coli* promoters (10).

The experiments presented in Fig. 6 indicate that the 130K gene is not transcribed in vivo by *E. coli* RNA polymerase, in spite of the similarity of its promoter to the *E. coli* consensus promoter. These experiments extend earlier results (30) which showed that 130K protein could be expressed in *E. coli* only if the gene was under the control of an *E. coli* promoter. The inability to detect 130K mRNA in *E. coli* by the S1 analysis could possibly be explained by rapid degradation of the mRNA, particularly if the mRNA were not being translated. If, however, the p_L promoters within the λ1059 phage vector or pLC28 plasmid vector are derepressed, high levels of either 130K protein (30) or the 35-kDa truncated protein (J. Smit, unpublished observations) are produced, indicating that the mRNA is functional in translation. We conclude that rapid degradation of the 130K mRNA is an unlikely explanation for its absence in *E. coli*.

High levels of transcription from the p_L promoter in pLC28 could potentially consume available RNA polymerase molecules, thus preventing transcription of 130K. However, the RNA used in this experiment was extracted from *E. coli* grown at 30°C, at which temperature p_L transcription is repressed. Moreover, transcription from the upstream p_L promoter would have been detected as protection of full-length probe. In fact, little full-length probe was detected, indicating that there is no significant transcription from the p_L promoter.

Another possible explanation for the lack of 130K transcription in *E. coli* is that transcription from the *sts1* promoter blocks transcription from the nearby 130K promoter in *E. coli*, possibly by steric hindrance of RNA polymerase binding or by some effect on DNA structure. A possible rationale for this effect would be that 130K is a more efficient promoter in *C. crescentus* than is *sts1* and could therefore overcome the transcriptional block. In *E. coli*, on the other hand, 130K is a less efficient promoter and thus susceptible to the blockage induced by transcription from *sts1*. Alternatively, auxiliary transcription factors may be required in order for 130K to overcome the block by *sts1* in *C. crescentus*; presumably those factors would be absent in *E. coli*. These possibilities cannot be excluded.

Given the small number of *Caulobacter* promoters characterized heretofore, few generalizations regarding *Caulobacter* promoter structure can be made. Most of the characterized promoters fall into one of two groups. The first group, which includes the 130K promoter, the *sts1* promoter, and the internal 16S rRNA internal promoter (2),

have −10 and −35 regions similar to those of *E. coli* promoters. The *sts1* and rRNA promoters are transcribed in *E. coli*, while 130K is not. The structural differences that lead to differential recognition by *E. coli* polymerase are not known. The two bases at −7 and −8 in the 130K promoter, C and G, are not found in any *E. coli* promoters (10), whereas neither the *sts1* nor rRNA promoter has G or C in those positions. This observation tempts us to speculate that the −7/−8 dinucleotide in the 130K promoter prevents recognition by the *E. coli* polymerase. The possibility that bases outside the consensus region may confer selectivity cannot be excluded. Clearly, the critical differences between the 130K and the *E. coli*-like promoters are subtle.

The *trpFBA* operon promoter, which is not transcribed in *E. coli* (36), has upstream sequences around −30 which are similar to the *E. coli* −35 region consensus (25). The sequence around −6 shows somewhat less similarity to the −10 sequence of *E. coli*. Both the −30 and −6 regions are similar to 5′ regions of genes from other bacteria evolutionarily related to *C. crescentus*.

The second group of *Caulobacter* promoters includes the 25- and 27-kDa flagellin structural genes (19) and two promoters in the flagellar hook gene cluster (4, 20). These promoters resemble the promoters of the *nif* and *ntr* operons of various enteric bacteria (5) rather than the *E. coli* consensus promoter. Not surprisingly, these *Caulobacter* genes are not transcribed by *E. coli* RNA polymerase (4, 18). Since transcription of the *nif* and *ntr* operons involves an RNA polymerase holoenzyme containing minor sigma factors (5), it has been speculated that a minor sigma factor may be involved in the transcription of the *Caulobacter* flagellar gene as well (19). It has recently been reported that a minor sigma factor, sigma-28, controls flagellar expression in *Bacillus subtilis* and that sigma-28 RNA polymerase recognizes the promoter sequences of several flagellar and chemotaxis genes of *E. coli* and *Salmonella typhimurium* (11).

The 29-kDa flagellin promoter and a third hook cluster promoter show only limited similarity to the *nif*/*ntr* consensus promoter (20). Both promoters, however, have sequences similar to another putative regulatory element contained in the *nif*-like flagellar promoters (20).

From the partial DNA sequence of the 130K gene, the amino acid sequence of the N-terminal approximately one-third of the protein was predicted. A major conclusion from the protein sequence is that the 130K protein is not synthesized with an N-terminal leader peptide sequence. The relatively hydrophobic N-terminal 18 amino acids may, however, function as an uncleaved signal sequence, such as is found in several bacterial and eucaryotic proteins (35). Similarly, the flagellin structural genes (8; Milhausen et al., manuscript in preparation) and the flagellar hook protein (22) also do not appear to have cleaved N-terminal presequences. While the N-terminal 20 to 30 amino acids of these proteins are generally hydrophobic, the sequences are not exceedingly hydrophobic, and none have the long unbroken stretches of hydrophobic residues characteristic of cleaved signal sequences. However, neither 130K nor the flagellins contain acidic residues in the first 20 amino acids, which is consistent with known signal sequences (34). It is possible, therefore, that the N-terminal amino acids play a role in *Caulobacter* analogous to a signal sequence for transport of proteins to the surface of the cell, but other internal amino acid segments may also be involved in targeting this protein to the surface.

While the lack of a cleaved signal sequence may be common among *Caulobacter* membrane and secreted proteins, it is probably not a general feature of bacterial surface layer proteins. The N-terminal region of the surface layer protein of *Aeromonas salmonicida* has been found to contain a typical cleaved presequence (T. Trust, personal communication).

The amino acid sequence data also provide evidence for multiple structural domains in the 130K protein. Beyond the N-terminal 20 amino acids there is a 170-amino-acid segment which is enriched in basic amino acids, followed by a segment of at least 120 amino acids which is largely neutral and nonpolar. What role these domains play in the structure of the 130K protein and surface layer is unknown. Completion of the amino acid sequence of 130K should reveal whether additional domains are present and may suggest possible functions.

It is important to determine what sequences in the 130K protein direct this protein to the cell surface in a temporally and spatially defined manner and cause it to assemble into the surface array. It is possible that different domains are involved in transmembrane translocation, spatial locaization, and assembly. Construction of a heterologous expression-secretion system for *Caulobacter* spp. may therefore be possible by using the 130K promoter and signal sequence portions of the 130K protein.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant GM34556 from the National Institutes of Health and by the Office of Naval Research.

We gratefully acknowledge Pau Yuan and Sylvia Yuen of the Protein Sequencing Group at Applied Biosystems for performing the amino acid sequencing of the 130K protein.

ADDENDUM IN PROOF

After submission of this paper we learned of several instances of surface array proteins from other bacteria, in addition to *Aeromonas salmonicida*, in which an apparently typical signal leader peptide is cleaved to form a mature protein. These include the surface layer proteins of *Bacillus brevis* (H. Yamagata, T. Adachi, A. Tsuboi, M. Takao, T. Sasaki, N. Tsukagoshi, and S. Udaka, J. Bacteriol. **169:**1239–1245, 1987), *Deinococcus radiodurans* (J. Peters, M. Peters, F. Lottspeich, W. Schafer, and W. Baumeister, J. Bacteriol. **169:**5216–5223, 1987), and *Halobacterium halobium* (J. Lechner and M. Sumper, J. Biol. Chem. **262:** 9724–9729, 1987).

LITERATURE CITED

1. **Agabian, N., M. Evinger, and G. Parker.** 1979. Generation of asymmetry during development: segregation of type-specific proteins in *Caulobacter*. J. Cell Biol. **81:**123–136.
2. **Amemiya, K., V. Bellofatto, L. Shapiro, and J. Feingold.** 1986. Transcription initiation *in vitro* and *in vivo* at a highly conserved promoter within a 16S ribosomal RNA gene. J. Mol. Biol. **187:** 1–14.
3. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80:**3963–3965.
4. **Chen, L.-S., D. Mullin, and A. Newton.** 1986. Identification, nucleotide sequence, and control of developmentally regulated promoters in the hook operon region of *Caulobacter crescentus*. Proc. Natl. Acad. Sci. USA **83:**2860–2864.
5. **Doi, R. H., and L.-F. Wang.** 1986. Multiple procaryotic ribonucleic acid polymerase sigma factors. Microbiol. Rev. **50:**227–243.
6. **Dretzen, G., M. Bellard, P. Sassone-Corsi, and P. Chambon.** 1981. A reliable method for the recovery of DNA fragments

from agarose and acrylamide gels. Anal. Biochem. **112**:295–298.
7. **Evinger, M., and N. Agabian.** 1977. The envelope-associated chromosome from *Caulobacter* stalked and swarmer cells. J. Bacteriol. **132**:294–301.
8. **Gill, P. R., and N. Agabian.** 1983. The nucleotide sequence of the M_r 28,500 flagellin gene of *Caulobacter crescentus*. J. Biol. Chem. **258**:7395–7401.
9. **Gilman, M. Z., and M. J. Chamberlain.** 1983. Developmental and genetic regulation of *Bacillus subtilis* genes transcribed by σ^{28} RNA polymerase. Cell **35**:285–293.
10. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
11. **Helman, J. D., and M. J. Chamberlain.** 1987. DNA sequence analysis suggests that expression of flagellar and chemotaxis genes in *Escherichia coli* and *Salmonella typhimurium* is controlled by an alternative sigma factor. Proc. Natl. Acad. Sci. USA **84**:6422–6424.
12. **Hu, N.-T., and J. Messing.** 1982. The making of strand-specific M13 probes. Gene **17**:271–277.
13. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
14. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1983. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
16. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
17. **Milhausen, M., and N. Agabian.** 1983. *Caulobacter* flagellin mRNA segregates asymmetrically at cell division. Nature (London) **302**:630–632.
18. **Milhausen, M., P. R. Gill, G. Parker, and N. Agabian.** 1982. Cloning of developmentally regulated flagellin genes from *Caulobacter crescentus* via immunoprecipitation of polyribosomes. Proc. Natl. Acad. Sci. USA **79**:6847–6851.
19. **Minnich, S. A., and A. Newton.** 1987. Promoter mapping and cell cycle regulation of flagellin gene transcription in *Caulobacter crescentus*. Proc. Natl. Acad. Sci. USA **84**:1142–1146.
20. **Mullin, D., S. Minnich, L.-S. Chen, and A. Newton.** 1987. A set of positively regulated flagellar gene promoters in *Caulobacter crescentus* with sequence homology to the *nif* gene promoters of *Klebsiella pneumoniae*. J. Mol. Biol. **195**:939–943.
21. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene **26**:101–106.
22. **Ohta, N., L.-S. Chen, E. Swanson, and A. Newton.** 1985. Transcriptional regulation of a periodically controlled flagellin gene operon in *Caulobacter crescentus*. J. Mol. Biol. **186**:107–115.
23. **Poindexter, J. S.** 1964. Biological properties and classification of the *Caulobacter* group. Bacteriol. Rev. **28**:231–295.
24. **Remaut, E., P. Stanssens, and W. Fiers.** 1981. Plasmid vectors for high-efficiency expression controlled by P_L promoter of coliphage lambda. Gene **15**:81–93.
25. **Ross, C., and M. E. Winkler.** 1988. Structure of the *Caulobacter crescentus trpFBA* operon. J. Bacteriol. **170**:757–768.
26. **Sanger, F., S. Nicklen, and A. R. Coulsen.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
27. **Shine, J., and L. Dalgarno.** 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) **254**:34–38.
28. **Sleytr, U. B., and P. Messner.** 1983. Crystalline surface layers on bacteria. Annu. Rev. Microbiol. **37**:311–339.
29. **Smit, J.** 1986. Protein surface layers of bacteria, p. 343–376. *In* M. Inouye (ed.), Outer membranes as model systems. John Wiley & Sons, New York.
30. **Smit, J., and N. Agabian.** 1984. Cloning of the major protein of the *Caulobacter crescentus* periodic surface layer: detection and characterization of the cloned peptide by protein expression assays. J. Bacteriol. **160**:1137–1154.
31. **Smit, J., and N. Agabian.** 1982. Cell surface patterning and morphogenesis. Biogenesis of a periodic surface array during *Caulobacter* development. J. Cell Biol. **95**:41–49.
32. **Smit, J., D. A. Grano, R. M. Glaeser, and N. Agabian.** 1981. Periodic surface array in *Caulobacter crescentus*: fine structure and chemical analysis. J. Bacteriol. **146**:1135–1150.
33. **Sollner-Webb, B., and R. H. Reeder.** 1979. The nucleotide sequence of the initiation and termination sites for ribosomal RNA transcription in *X. laevis*. Cell **18**:485–499.
34. **Watson, M. E. E.** 1984. Compilation of published signal sequences. Nucleic Acids Res. **12**:5145–5164.
35. **Wickner, W. T., and H. F. Lodish.** 1985. Multiple mechanisms of protein insertion into and across membranes. Science **230**: 400–407.
36. **Winkler, M. E., P. V. Schoenlein, C. M. Ross, J. T. Barrett, and B. Ely.** 1984. Genetic and physical analysis of *Caulobacter crescentus trp* genes. J. Bacteriol. **160**:279–287.