JOURNAL OF BACTERIOLOGY, Feb. 1996, p. 662–667
0021-9193/96/$04.00+0

Vol. 178, No. 3

Salmonella enteritidis agfBAC Operon Encoding Thin, Aggregative Fimbriae

S. KAREN COLLINSON,[1] SHARON C. CLOUTHIER,[1] JAMES L. DORAN,[1,2]
PAMELA A. BANSER,[1] AND WILLIAM W. KAY[1,2]*

Department of Biochemistry and Microbiology and the Canadian Bacterial Diseases Network, University of Victoria, Victoria, British Columbia V8W 3P6,[1] and Microtek International Ltd., Saanichton, British Columbia V8X 3X1,[2] Canada

Received 16 October 1995/Accepted 20 October 1995

Salmonella enteritidis **produces thin, aggregative fimbriae, named SEF17, which are composed of polymerized AgfA fimbrin proteins. DNA sequence analysis of a 2-kb region of *S. enteritidis* DNA revealed three contiguous genes, *agfBAC*. The 453-bp *agfA* gene encodes the AgfA fimbrin, which was predicted to be 74% identical and 86% similar in primary sequence to the *Escherichia coli* curli structural protein, CsgA. pHAG, a pUC18 derivative containing a 3.0-kb *Hin*dIII fragment encoding *agfBAC*, directed the in vitro expression of the major AgfA fimbrin, with an M_r of 17,000, and a minor AgfB protein, with an M_r of 16,000, encoded by the 453-bp *agfB* gene. AgfA was not expressed from pDAG, a pUC18 derivative containing a 3.1-kb *Dra*I DNA fragment encoding *agfA* but not *agfB*. Primer extension analysis identified two adjacent transcription start sites located immediately upstream of *agfB* in positions analogous to those of the *E. coli* curlin *csgBA* operon. No transcription start sites were located immediately upstream of *agfA* or *agfC*. Northern (RNA) blot analysis confirmed that transcription of *agfA* was initiated from the *agfB* promoter region. Secondary-structure analysis of the putative mRNA transcript for *agfBAC* predicted the formation of a stem-loop structure ($\Delta G°$, −22 kcal/mol [−91 kJ/mol]) in the intercistronic region between *agfA* and *agfC*, which may be involved in stabilization of the *agfBA* portion of the *agfBAC* transcript. *agfBAC* and flanking regions had a high degree of sequence similarity with those counterparts of the *E. coli* curlin *csgBA* region for which sequence data are available. These data are demonstrative of the high degree of similarity between *S. enteritidis* SEF17 fimbriae and *E. coli* curli with respect to fimbrin amino acid sequence and genetic organization and, therefore, are indicative of a common and relatively recent ancestry.**

Fimbriae are filamentous surface structures composed of a repeated major subunit protein (fimbrin) and sometimes several minor subunit-like proteins (24). Fimbrins are polymerized by hydrophobic and hydrophilic interactions to form either thick (7- to 8-nm), rigid structures, thin (2- to 4-nm), flexible filaments, or composites of both (13, 16, 19). The characterization of fimbrial biosynthesis has shown that the regulation and assembly of some fimbriae are accomplished by several accessory proteins (17). The genes encoding structural and accessory proteins are typically arranged in 7- to 9-kb clusters organized as a series of operons and individual genes (4, 7, 18, 23).

Salmonella enteritidis is an important, food-borne, enteric pathogen which produces several fimbrial types (5). Characterization of the various fimbriae of *S. enteritidis* 27655-3b led to the discovery of thin, aggregative fimbriae (SEF17) that mediate fibronectin binding (10, 11). SEF17 fimbriae are composed mainly of a fimbrin with an M_r of 17,000, AgfA. These fimbriae are highly stable structures, requiring treatment with 90% formic acid for depolymerization (11). Diarrheagenic *Escherichia coli* strains produce thin, aggregative fimbriae that are biochemically and serologically related to those of SEF17 (12). However, the degree of DNA sequence dissimilarity between the respective fimbrin genes is sufficient that *agfA*-based nucleotide probes hybridize only to *Salmonella* DNA, thereby providing a valuable, genus-specific diagnostic for *Salmonella* spp. (14).

E. coli HB101 also produces an AgfA-related fimbrin (3). The corresponding thin, aggregative fimbriae have been named curli. *csgA*, the gene encoding the major curlin subunit protein, has recently been cloned and sequenced (22). Transcriptional analysis revealed that *csgA* is transcribed as a dicistronic operon from the promoter of the upstream gene, *csgB* (2). The *csgBA* promoter requires an AT-rich upstream activating sequence, which is recognized by both σ^s and σ^{70} sigma factors and is repressed by H-NS, which prevents the formation of transcription initiation complexes with σ^{70} under conditions under which σ^s is not expressed, including temperatures above 26°C, high osmolarity, and non-stationary-phase growth (2).

The structural similarities of curli and SEF17 fimbriae and similarities in the N-terminal fimbrin sequences suggest that these related fimbriae are assembled by analogous mechanisms. In an effort to further understand the biosynthesis of SEF17 fimbriae, the DNA sequence of the *agfBAC* operon was determined and the expression of these genes was examined.

(Some parts of this paper were presented at the First International Rushmore Conference on Mechanisms in the Pathogenesis of Enteric Diseases, 27 to 30 September 1995, and published as an extended abstract [9a]).

* Corresponding author. Mailing address: Department of Biochemistry and Microbiology, Petch Building, University of Victoria, P.O. Box 3055, Victoria, British Columbia V8W 3P6, Canada. Phone: (604) 721-7078. Fax: (604) 721-8882. Electronic mail address: wkay@sol.uvic.ca.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *S. enteritidis* 27655-3b was routinely grown on solid T medium at 37°C (11). To assay for SEF17 production, Congo red (100 μg/ml) was incorporated into T medium (10, 12). *E. coli* recombinant

clones were grown on Luria-Bertani medium containing ampicillin (100 μg/ml) (26).

Plasmid construction. *S. enteritidis* chromosomal DNA, purified by CsCl gradient centrifugation (26), was digested with *Dra*I or *Hin*dIII. AgfA-encoding fragments were identified by hybridization to a ^{32}P-labeled (26), 374-bp, AgfA-encoding DNA fragment amplified from *S. enteritidis* Tn*phoA* 2-7f by PCR (10). Hybridizing fragments were subcloned into pUC18 (29) and introduced into *E. coli* DH5α by transformation (26). Plasmids were purified by standard alkaline lysis procedures (26) or on Qiagen columns according to the manufacturer (Qiagen Inc., Chatsworth, Calif.).

DNA sequence analyses. DNA sequences were determined with Sequenase version 2.0 (United States Biochemical Corp., Cleveland, Ohio) and custom oligonucleotide primers synthesized on a PCR-MATE EP model 391 DNA synthesizer (Applied Biosystems Inc., Foster City, Calif.). GeneWorks version 2.0 (IntelliGenetics Inc., Mountainview, Calif.) was used to order overlapping DNA sequences. DNA Strider 1.1 (20) was used to identify open reading frames within the composite sequence. The programs of the PC/GENE suite (IntelliGenetics Inc.) were used to locate putative promoters (25) and Shine-Dalgarno motifs (15) and to predict RNA and protein secondary structures. The predicted amino acid sequence for each open reading frame was compared with protein sequences listed in the National Center for Biotechnology Information (NCBI) databases with the program BLASTX (1).

SDS-PAGE and Western blot (immunoblot) analysis. AgfA was solubilized from *S. enteritidis* or *E. coli* recombinant clones and detected by Western blot analysis according to the method of Collinson et al. (10, 11). Proteins contained in the sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) sample buffer-glycine extracts of whole cells and the formic acid-digested cell material were separated by SDS-PAGE, electrophoretically transferred to nitrocellulose, and screened with rabbit anti-SEF17 immune serum. Immunoreactive AgfA proteins were detected with goat anti-rabbit immunoglobulin G-alkaline phosphatase conjugates and visualized with 5-bromo-4-chloro-3-indolylphosphate and the enhancer Nitro Blue Tetrazolium.

In vitro transcription-translation analysis. Proteins encoded by *agfBAC* were labeled with [^{35}S]methionine with a coupled, in vitro transcription-translation system (Promega Corp., Madison, Wis.). Plasmid templates (4 μg) were used in reaction mixtures, which were incubated first for 1.5 h at 37°C and then for 10 min at 0°C. A 20-μl aliquot was removed from the transcription-translation reaction mixture, added to 80 μl of acetone, and held on ice for 15 min. The acetone-precipitated proteins were recovered by centrifugation (14,000 × *g*, 5 min, 4°C), dried for 15 min under vacuum, resuspended in 50 μl of SDS-PAGE sample buffer, and heat denatured (100°C, 5 min). [^{35}S]methionine-labeled proteins in a 15-μl aliquot were separated by SDS-PAGE. The 12.5% acrylamide gel was fixed in 7% acetic acid for 1 h, dried onto 3-mm-thick Whatman paper for 2 h at 80°C, and then exposed to Kodak X-Omat AR5 film for autoradiographic detection of labeled proteins.

RNA extraction. Total RNA was prepared from *S. enteritidis* cells grown on T medium overnight at 37°C by a modification of the procedure of McCormick et al. (21) as described by Clouthier et al. (9). Two gene probes for Northern (RNA) blot analysis were prepared by PCR amplification (10) of either a 192-bp *agfB* fragment, by using the primer pair TAF41 (5′GCAAGTCTTCATTTAAT CAG3′) and TAF43 (5′CTGGCATCGTTGGCATTGCC3′), or a 183-bp *agfC* fragment, by using the primer pair TAF38 (5′CGATATTTACACGGTGAT CC3′) and TAF30 (5′GTCCTCTGAAGATATATTATCCA3′). A 394-bp *agfA* probe was prepared from an *Eco*RI fragment of pAGF3 (14) that was purified with the Sephaglas BandPrep kit (Pharmacia Biotech, Uppsala, Sweden). The gene probes were labeled with [α-^{32}P]dATP by random priming (26).

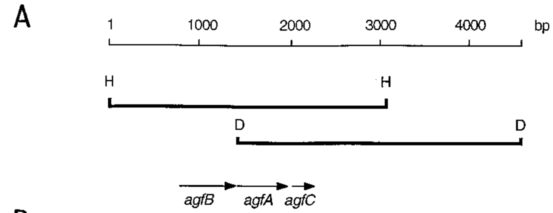
Primer extension analysis. Custom sequencing primers complementary to the coding strands of *agfB* (TAF44, 5′GTCGCGGTTGCAATCCCAG3′), *agfA* (TAF42, 5′GTGGAACGACGCCAGCCA3′), and *agfC* (TAF26, 5′GTCCTCT GAAGATATATTTACACT3′) were end labeled with [γ-^{32}P]ATP and used in primer extension reactions as previously described (9, 27).

Nucleotide sequence accession numbers. The nucleotide sequences reported herein for *agfB*, *agfA*, and *agfC* were submitted to GenBank and given accession no. U43280.

RESULTS

Cloning and sequencing of *agfBAC* region. To identify chromosomal fragments encoding proteins responsible for SEF17 biosynthesis, genomic digests of *S. enteritidis* 27655-3b prepared with *Hin*dIII or *Dra*I were probed with a PCR fragment containing a 333-bp *agfA* sequence. The hybridizing 3.0-kb *Hin*dIII fragment and the hybridizing 3.1-kb *Dra*I fragment were each subcloned into pUC18 to form pHAG and pDAG, respectively (Fig. 1A). Sequencing of approximately 2 kb of DNA surrounding the overlapping region of the *Hin*dIII and *Dra*I fragments resulted in recognition of the *agfA* gene and two flanking open reading frames (Fig. 1B). *agfA* was identified by comparison with a previously sequenced 333-bp region of

A



B

```
CCAGAAGTACTGACAGATGTTGCACTGCTGTGGGTTGAAATAGCCCATTATCCAGAAAGAGAAAAATATTTACGAAAATA   80
CTTTTAACTGTTTTCAATCTAGCCATTACAAATCTTAAAGCAAGTGTTAAAATTGTAACAAGATGTAAAAATATATATTA  160
AAATGTTGTTTTTGGGTTTTTTTGAAGTTTAGATTTGATAGTAAAGTTGTACATTTCGCTGTTATTGCATAGATTTAAAA  240
AATCATACAAATTATAATAATTCATTGATTTTTAATCATTTTAATTATTGTGTGTCATGTTTTGATTTTATTTTTTCTTA  320
AAATTTGAGACGAGGCATTAACCTGGACAGCACAAAGACAAAAAAAACGAAGTGTGTCACGTCTTGTGCGTATTGCCCCC  400
CATAGGAAGCATAAGAACATCCCCATGGCGGCATAACACACACCAACACTTCATTTTTTGGGTTCGCGATACACTATCTT  480
CTATGGCCAAAAATCAATTATAAAAAATCACATGGCTATCGTTTTATTAGCACTTTGGTATGAGCTTAAATAACAAAATA  560
         -35                          -10       **
CCACGCGTGGGTGAGTTATTAAAAATTTTTCCACGGACATACTCTTCATCGTAACGACGCGTTAACAAAAAACGCATGTC  640
                                                 SD
GCTAACAAGGTAATAGATAATTTTCGCTATGTACGACCAGGTCCAGGGTGACAGC ATG AAA AAC AAA TTG TTA   713
                                                   AgfB  M   K   N   K   L   L
                                                            ⇓
TTT ATG ATG TTG ACA ATA CTG GGT GCG CCT GGG ATT GCA ACC GCG ACA AAT TAT GAT CTG   773
 F   M   M   L   T   I   L   G   A   P   G   I   A   T   A   T   N   Y   D   L

GCT CGT TCA GAA TAT AAT TTT GCG GTA AAT GAA TTA AGC AAG TCT TCA TTT AAT CAG GCG   833
 A   R   S   E   Y   N   F   A   V   N   E   L   S   K   S   S   F   N   Q   A

GCC ATT ATT GGT CAA GTC GGC ACG GAT AAT AGT GCC AGA GTA CGC CAG GAA GGA TCA AAA   893
 A   I   I   G   Q   V   G   T   D   N   S   A   R   V   R   Q   E   G   S   K

CTA TTG TCC GTT ATT TCA CAA GAA GGA GGA AAT AAT CGG GCG AAA GTC GAC CAG GCA GGG   953
 L   L   S   V   I   S   Q   E   G   G   N   N   R   A   K   V   D   Q   A   G

AAT TAT AAC TTT GCG TAT ATT GAG CAA ACG GGC AAT GCC AAC GAT GCC AGT ATA TCG CAA  1013
 N   Y   N   F   A   Y   I   E   Q   T   G   N   A   N   D   A   S   I   S   Q

AGC GCT TAC GGT AAT AGT GCA GCT ATT ATC CAG AAA GGT TCT GGA AAT AAG GCC AAT ATT  1073
 S   A   Y   G   N   S   A   A   I   I   Q   K   G   S   G   N   K   A   N   I

ACC CAG TAC GGT ACG CAG AAA ACA GCA GTT GTA GTG CAG AAA CAG TCG CAT ATG GCT ATT  1133
 T   Q   Y   G   T   Q   K   T   A   V   V   V   Q   K   Q   S   H   M   A   I
                                  DraI.           SD
CGC GTC ACC CAA CGC TAATACCGTTACGACTTTTAAATCAATCCGATGGGGGTTTTACC ATG AAA CTT TTA 1204
 R   V   T   Q   R                                        AgfA  M   K   L   L
                                                      ⇓
AAA GTG GCA GCA TTC GCA GCA ATC GTA GTT TCT GGC AGT GCT CTG GCT GGC GTC GTT CCA  1264
 K   V   A   A   F   A   A   I   V   V   S   G   S   A   L   A   G   V   V   P

CAA TGG GGC GGC GGC GGT AAT CAT AAC GGC GGC GGC AAT AGT TCC GGC CCG GAC TCA ACG  1324
 Q   W   G   G   G   G   N   H   N   G   G   G   N   S   S   G   P   D   S   T

TTG AGC ATT TAT CAG TAC GGT TCC GCT AAC GCT GCG CTT GCT CTG CAA AGC GAT GCC CGT  1384
 L   S   I   Y   Q   Y   G   S   A   N   A   A   L   A   L   Q   S   D   A   R

AAA TCT GAA ACG ACC ATT ACC CAG AGC GGT TAT GGT AAC GGC GCC GAT GTA GGC CAG GGT  1444
 K   S   E   T   T   I   T   Q   S   G   Y   G   N   G   A   D   V   G   Q   G

GCG GAT AAT AGT ACT ATT GAA CTG ACT CAG AAT GGT TTC AGA AAT AAT GCC ACC ATC GAC  1504
 A   D   N   S   T   I   E   L   T   Q   N   G   F   R   N   N   A   T   I   D

CAG TGG AAC GCT AAA AAC TCC GAT ATT ACT GTC GGC CAA TAC GGC GGT AAT AAC GCC GCG  1564
 Q   W   N   A   K   N   S   D   I   T   V   G   Q   Y   G   G   N   N   A   A

CTG GTT AAT CAG ACC GCA TCT GAT TCC AGC GTA ATG GTG CGT CAG GTT GGT TTT GGC AAC  1624
 L   V   N   Q   T   A   S   D   S   S   V   M   V   R   Q   V   G   F   G   N

AAC GCC ACG GCT AAC CAG TAT TAATTTAGCGTCTGCGCTAATAAAAAAACAGGGCATAAGCCCTGTTTTTTTT 1697
 N   A   T   A   N   Q   Y
 SD                                                                          ⇓
CGGGAGGAAATT ATG CAT ACT TTA TTG CTC CTT GCC GCA CTT TCA AAT CAG ATT ACG TTT ACC 1760
        AgfC  M   H   T   L   L   L   L   A   A   L   S   N   Q   I   T   F   T

ACG ACT CAG CAA GGC GAT ATT TAC ACG GTG ATC CCT CAG GTC ACA TTA AAC GAA CCC TGC  1820
 T   T   Q   Q   G   D   I   Y   T   V   I   P   Q   V   T   L   N   E   P   C

GTC TGT CAG GTG CAA ATT CTC TCT GTG CGC GAC GGC GTC GGG GGA CAA AGC CAT ACA CAG  1880
 V   C   Q   V   Q   I   L   S   V   R   D   G   V   G   G   Q   S   H   T   Q

CAA AAA CAA ACG CTA TCT TTA CCT GCT AAT CAA CCG ATT GAG TTG TCT CGT CTT AGT GTA  1940
 Q   K   Q   T   L   S   L   P   A   N   Q   P   I   E   L   S   R   L   S   V

AAT ATA TCT TCA GAG GAC TCG GTT AAA ATT ATT GTT ACT GTT TCG GAC GGA CAA TCA CTG  2000
 N   I   S   S   E   D   S   V   K   I   I   V   T   V   S   D   G   Q   S   L

CAT TTA TCA CAA CAA TGG CCG CCT TCT GCA CAG TAGTTTTTGATGGTGGCGGAAATGGATTGGCTGA    2067
 H   L   S   Q   Q   W   P   P   S   A   Q
```

FIG. 1. Cloning and sequence determination of the region of the *S. enteritidis* chromosome surrounding *agfA*. (A) Schematic indicating the locations and orientations of *agfBAC* on two recombinant pUC18 plasmid derivatives, pHAG and pDAG, containing overlapping 3.0-kb *Hin*dIII (H) and 3.1-kb *Dra*I (D) fragments, respectively. (B) DNA sequences and predicted amino acid sequences of *agfBAC*. The Shine-Dalgarno motifs (15) (SD) and −35 and −10 promoter motifs (25) are overscored. Arrows mark the predicted signal sequence cleavage sites in AgfB, AgfA, and AgfC. The AgfA sequence previously determined by N-terminal amino acid sequence analysis is underscored. Transcriptional start sites (Fig. 4) are noted by asterisks.

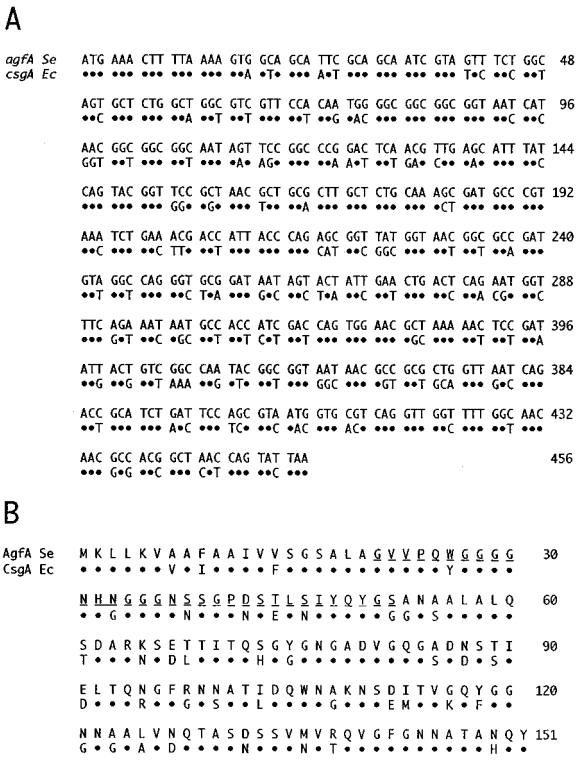


FIG. 2. Sequence alignments of the *S. enteritidis* (*Se*) *agfA* and *E. coli* (*Ec*) *csgA* (22) genes (A) and the predicted structural proteins, AgfA and CsgA (B). Sequence residue identities are denoted by dots.

agfA (14). The 453-bp *agfA* gene encoded a 15,305-Da AgfA precursor protein (Fig. 1B). Comparison of the predicted AgfA sequence with the N-terminal amino acid sequence of the mature AgfA protein (11, 14) indicated that the precursor form of AgfA consisted of 151 residues and included a 20-amino-acid signal sequence. The mature AgfA protein had a predicted molecular mass of 13,330 Da. Comparison of the predicted amino acid sequence of AgfA with protein sequences listed in the NCBI databases showed that AgfA was highly related to the *E. coli* curli subunit fimbrin, CsgA, thereby reflecting 72% DNA sequence identity (Fig. 2A). The AgfA primary sequence was 74% identical and 86% similar to that of CgsA (Fig. 2B). The AgfA and CsgA secretory sequences were 86% conserved (Fig. 2B). No other characterized fimbrial protein had notable sequence similarity to AgfA.

The two open reading frames flanking *agfA* were designated *agfB* and *agfC* (Fig. 1). *agfB*, *agfA*, and *agfC* were found to have the same polarity, and each was preceded by a Shine-Dalgarno motif for translation initiation (Fig. 1B). The *agfB* and *agfC* open reading frames were 453 and 324 bp, respectively, and encoded unique proteins without counterparts in the NCBI databases. *agfB* encoded a 151-amino-acid, 16,146-Da protein that included a 21-amino-acid, N-terminal sequence typical of a prokaryotic signal sequence, the cleavage of which would result in a mature AgfB protein of 130 amino acids and 13,932 Da (Fig. 1B). *agfC* encoded a protein with a putative 17-residue prokaryotic signal sequence and, if processed, would result in a 9,823-Da mature protein of 91 amino acids (Fig. 1B).

In vitro expression of *agfBAC*. To show that *agfBAC* encoded proteins of the predicted sizes, in vitro transcription-translation was performed with pHAG as the DNA template. A major [^{35}S]methionine-radiolabeled protein with an M_r of 17,000 to 18,000 was considered to be the precursor form of AgfA, since this protein was absent from in vitro transcription-translation directed by pUC18 (Fig. 3A) and typically migrates at an M_r of 17,000 in SDS-PAGE (11). Whole-cell lysates of *E. coli* DH5α harboring pHAG were analyzed by Western blotting with polyclonal antisera to SEF17. Clones carrying pHAG were found to direct the production of AgfA (Fig. 3B).

AgfB migrated as a faint [^{35}S]methionine-labeled protein with an estimated M_r of 16,000 on SDS-PAGE following in vitro transcription-translation directed by pHAG (Fig. 3A). pDAG, which encodes *agfA* but not *agfB*, did not direct expression of AgfA in the in vitro transcription-translation system (Fig. 3B). This result indicated either that the transcriptional start site for *agfA* resided upstream of the *Dra*I site, within the *agfB* coding sequence, or that *agfA* was transcribed as part of an operon along with the upstream gene, *agfB*. No [^{35}S]methionine-labeled band which corresponded to the predicted, unprocessed AgfC of an estimated 11,690 Da could be recognized by SDS-PAGE analysis because of the comigration of several *E. coli* proteins with this M_r (Fig. 3A).

Transcriptional analysis of *agfBAC*. To determine whether *agfB*, *agfA*, and *agfC* were organized as an operon, transcription start sites were mapped. Primer extension analysis revealed two major transcription start sites located 85 and 86 nucleotides (nt) upstream of the AUG translational start codon for the *agfB* transcript (Fig. 1B; Fig. 4). No transcription start sites were found immediately upstream of *agfA* or *agfC*, indicating that transcription initiated upstream of *agfB* and extended through *agfA* and *agfC*.

To determine the lengths of the transcripts initiated upstream of *agfB*, total RNA, extracted from *S. enteritidis*, was analyzed by Northern hybridization with gene probes for *agfB*, *agfA*, or *agfC*. The *agfB*-specific probe hybridized to two transcripts of approximately 1,100 and 700 nt, whereas the *agfA*-specific probe hybridized to three transcripts of approximately 1,100, 700, and 500 nt (Fig. 5A). No transcripts were detected with the *agfC*-specific probe. These results confirmed that *agfBA* were cotranscribed and indicated that the dicistronic transcript was processed to the smaller *agfA* transcripts. A 30-nt stem-loop structure (ΔG° of −22 kcal/mol [−91 kJ/mol]) was predicted to form in the intercistronic region between *agfA* and

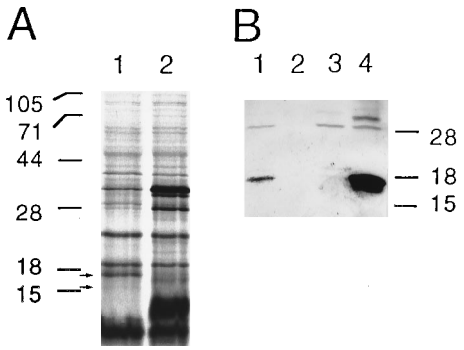


FIG. 3. In vitro transcription-translation analysis of recombinant plasmids and Western blot analysis of *E. coli* recombinant clones. (A) [^{35}S]methionine-labeled proteins expressed in vitro from pHAG (lane 1) and pUC18 (lane 2) and separated by SDS-PAGE. Arrows indicate the two pHAG encoded proteins. (B) Western blot analysis of whole-cell extracts of *E. coli* DH5α harboring pHAG (lane 1), *E. coli* DH5α harboring pDAG (lane 2), *E. coli* DH5α (lane 3), and *S. enteritidis* 27655-3b (lane 4). Polyclonal antiserum to SEF17 was used to detect AgfA. The relative mobilities of the prestained standard proteins are noted (in kilodaltons) to the left and right of panels A and B, respectively.

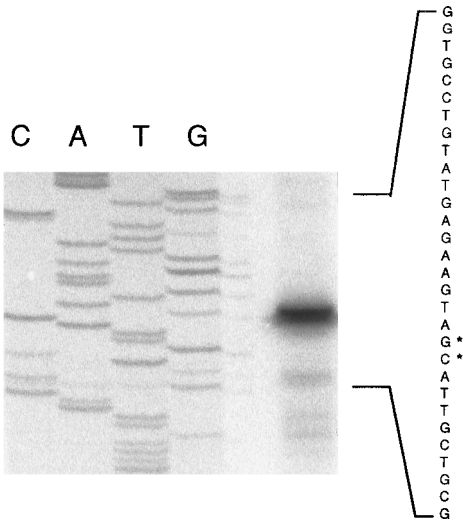


FIG. 4. Primer extension analysis to determine the transcription initiation site(s) for *agfBAC*. The primer TAF 44 (5′GTCGCGGTTGCAATCCCAG3′), located 46 bp downstream of the *agfB* start codon, was used to direct sequencing reactions of *agfB* encoded by pHAG and to direct the primer extension reaction on total *S. enteritidis* RNA prepared as indicated in Materials and Methods. Asterisks mark the transcriptional start sites, which are located 85 and 86 nt upstream of the *agfB* start codon (Fig. 1).

agfC (Fig. 5B), which might stabilize the *agfBA* and *agfA* transcripts.

Comparison of the *S. enteritidis agfBAC* and *E. coli csgBA* regions. The DNA sequences of *agfBAC* and flanking regions were compared with the partial DNA sequences determined for the *E. coli csgBA* region (2, 22). The approximately 700-bp region immediately upstream of *agfB* had 63% similarity with sequences immediately upstream of *csgB* (Fig. 6A). The level of similarity was 76% over the 5′ 275-bp regions, 39% over the central regions, and 82% in the 246-bp regions adjacent to the translational start codons of *agfB* and *csgB* (Fig. 6A). Notably, *agfBA* and *csgBA* shared the same two transcriptional start sites (Fig. 6A). The predicted N-terminal amino acid sequence of AgfB was identical to the 21-amino-acid N-terminal sequence predicted for CsgB (Fig. 6A). Similarly, 10 of the 12 C-terminal residues predicted for CsgB were identical to those of AgfB (Fig. 6B). The limited sequence data published for *csgB* (2) prevented further comparisons. The *agfBA* intergenic region had 93% similarity to the *csgBA* intergenic region (Fig. 6B). The sequence immediately downstream of *agfA* was 71% identical to the *agfAC* intergenic region (Fig. 6C). Comparison of *agfC* with the partially sequenced region downstream of *agfA* indicated that *E. coli* encoded an amino acid sequence of which 15 of 17 residues matched the N-terminal amino acid signal sequence predicted for AgfC (Fig. 6C).

DISCUSSION

This study has resulted in the characterization of the *S. enteritidis agfBA* gene sequences which are required for the expression of AgfA, the major structural fimbrin component of an unusual, thin, aggregative fimbria, SEF17. *agfA* is transcribed as part of an operon which includes the upstream gene, *agfB*. A third open reading frame, *agfC*, also belongs to this operon since no transcriptional start site between *agfA* and *agfC* was found. Consistent with these experimental observations, no sequences similar to prokaryotic promoter motifs were recognized within the *agfBAC* intragenic regions. The inability to detect an *agfC* transcript was likely indicative of a very short half-life of a portion of a transcript (4, 6). A potentially stable stem-loop structure in the intercistronic region between *agfA* and *agfC* could contribute to the stabilization of the *agfBA* transcript and the individual *agfA* transcripts (6, 8). The need for production of large amounts of the structural fimbrin AgfA proteins of SEF17 fimbriae is a plausible biological rationale for the processing and enhanced half-life of the *agfA* region of the *agfBAC* transcript.

The role of AgfA as the major structural fimbrin subunit protein of SEF17 is unequivocal (10, 11). The function of AgfB is unknown. According to the predicted amino acid sequences, AgfB matches AgfA in size and resembles AgfA in primary amino acid sequence. The differences between the two proteins include the presence of two tryptophan and proline residues in AgfA which are lacking in AgfB and the fact that AgfB has half the glycine residues of AgfA and more than double the number of the basic residues arginine and lysine. Although these differences between the primary structures exist, secondary-structure predictions indicate that these two proteins fold similarly (12a). These results suggest that AgfB is a fimbrin-like protein. *agfB* appears to encode a signal sequence characteristic of an exported protein, but AgfB did not copurify with AgfA in intact SEF17 in amounts detectable by N-terminal amino acid sequence analysis (11). This cannot be taken as proof that AgfB is absent from the fimbrial fiber, as AgfB may be a structural component of SEF17 that is present in the minute amounts characteristic of minor fimbrial tip proteins (17, 24) or as AgfB may be loosely associated with SEF17 filaments and therefore be readily dissociated during the rigorous, unconventional purification procedure used to obtain SEF17 filaments containing AgfA (11).

At present, it is unclear how many genes are required for the biosynthesis of SEF17. The *agfBAC* operon does not encode proteins characteristic of chaperone or usher proteins typically required for fimbrial biosynthesis (17), and analysis of the DNA sequences 0.7 kb upstream or 2.3 kb downstream of

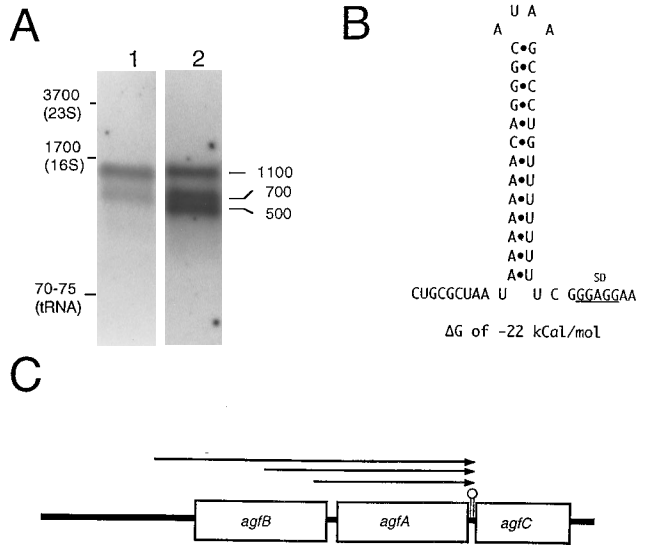


FIG. 5. Transcriptional analysis of *agfBAC*. (A) Total *S. enteritidis* RNA was probed by Northern blot analysis with gene probes to *agfB* (lane 1) and *agfA* (lane 2). The migrations of 23S RNA, 16S RNA, and tRNA species are noted on the left, and the approximate sizes (in nucleotides) of the transcripts are noted on the right. (B) The predicted stem-loop structure within the transcript arising from the *agfAC* intercistronic region is presented. (C) Schematic of the relative positions of the detected transcripts.

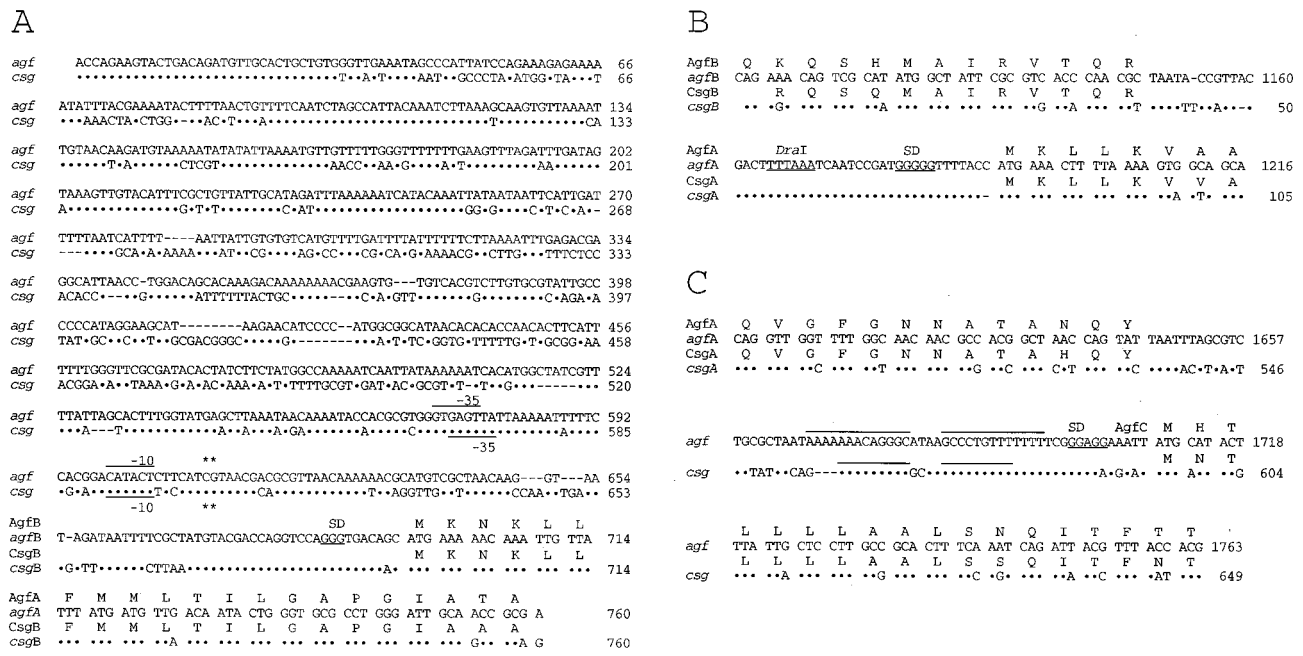


FIG. 6. DNA sequence comparison of the *S. enteritidis* 27655-3b *agfBAC* region with the *E. coli csgBA* region. Spaces (dashes) were introduced into both sequences to maximize alignments. Bases in the *E. coli* sequence matching those in the *S. enteritidis* sequence are denoted by dots. (A) Comparison of the *agfB* upstream region with that upstream of *csgB* (2). Promoter motif −10 and −35 regions (25) and a Shine-Dalgarno (SD) motif (15) for ribosome binding are underlined. The transcriptional start sites are noted with asterisks in both sequences. (B) Comparison of the *agfBA* intergenic region with the sequence upstream of *E. coli csgA* (2, 22). (C) Comparison of the *agfAC* intergenic region with the sequence downstream of *csgA* (22). Palindrome sequences indicative of potential stem-loop structures are overlined in both sequences.

agfBAC similarly did not identify open reading frames encoding other obvious fimbrial biosynthetic proteins (12a), suggesting that assembly and biosynthesis of SEF17 may proceed by a novel mechanism.

The data presented here support a close relationship between *S. enteritidis* SEF17 fimbriae and *E. coli* filaments called curli. The genetic organization of the *agfBAC* operon, which is partially responsible for SEF17 biosynthesis, closely resembles that of the *E. coli agfBA* region, which is necessary for the production of curli. The major structural protein genes, *agfA* and *csgA*, are homologous. Both genes are transcribed as part of an operon from promoters upstream of the respective flanking genes, *agfB* and *csgB* (2), which direct transcription from identical start sites. Although the complete DNA sequences for the genes flanking *csgA* of *E. coli* have not been published, it is clear from the sequence data available (2, 22) that *agfB* is homologous to *csgB*. These findings are not completely unexpected, since SEF17 and curli are morphologically, structurally, biochemically, and serologically related and are likely functionally similar as well (4, 11, 12, 22). However, the degree and extent of the DNA sequence conservation is somewhat surprising. The highly related regions of DNA sequence include intragenic regions, the 5′ sequences of open reading frames considered to encode signal peptides typical of those of exported prokaryotic proteins (28), and large portions of the upstream regions, including that encompassing or corresponding to the *csgA* "upstream activating sequence" (2). These results provide strong evidence that the genes for SEF17 and curlin biosynthesis have a common ancestry and that there was a relatively recent gene transfer to *E. coli*, to *S. enteritidis*, or to both from a common ancestor.

ACKNOWLEDGMENTS

This project was supported by operating grants to W.W.K. provided by the Natural Sciences and Engineering Research Council of Canada and the Canadian Bacterial Diseases Network, and by an operating grant to J.L.D. from the Science Council of British Columbia. S.C.C. was the recipient of a Graduate Research Engineering and Technology Awards scholarship from the Science Council of British Columbia.

REFERENCES

1. **Altschel, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Arnqvist, A., A. Olsén, and S. Normark.** 1994. σ^S-dependent growth-phase induction of the csgBA promoter in *Escherichia coli* can be achieved in vivo by σ^{70} in the absence of the nucleoid-associated protein H-NS. Mol. Microbiol. **13:**1021–1032.
3. **Arnqvist, A., A. Olsén, J. Pfeifer, D. G. Russell, and S. Normark.** 1992. The Crl protein activates cryptic genes for curli formation and fibronectin binding in *Escherichia coli* HB101. Mol. Microbiol. **6:**2443–2452.
4. **Båga, M., M. Göransson, S. Normark, and B. E. Uhlin.** 1988. Processed mRNA with differential stability in the regulation of *E. coli* pilin gene expression. Cell **52:**197–206.
5. **Bäumler, A. J., and F. Heffron.** 1995. Identification and sequence analysis of *lpfABCDE*, a putative fimbrial operon of *Salmonella typhimurium*. J. Bacteriol. **177:**2087–2097.
6. **Belasco, J. G., and C. F. Higgins.** 1988. Mechanisms of mRNA decay in bacteria: a perspective. Gene **72:**15–23.
7. **Bilge, S. S., J. M. Apostol, Jr., M. A. Aldape, and S. L. Moseley.** 1993. mRNA processing independent of RNaseIII and RNaseE in the expression of the F1845 fimbrial adhesin of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **90:** 1455–1459.
8. **Chen, C.-Y. A., J. T. Beatty, S. N. Cohen, and J. G. Belasco.** 1988. An intercistronic stem-loop structure functions as an mRNA decay terminator necessary but insufficient for *puf* mRNA stability. Cell **52:**609–619.
9. **Clouthier, S. C., K.-H. Müller, J. L. Doran, S. K. Collinson, and W. W. Kay.** 1993. Characterization of three fimbrial genes, *sefABC*, of *Salmonella enteritidis*. J. Bacteriol. **175:**2523–2533.

9a. **Collinson, S. K., S. C. Clouthier, J. L. Doran, P. A. Bamser, and W. W. Kay.** Characterization of the *agfBA* fimbrial operon encoding thin aggregative fimbriae of *Salmonella enteridis*. *In* P. S. Paul, D. H. Francis, and D. Benfield

(ed.), Mechanisms in the pathogenesis of enteric diseases. Proceedings of the First International Rushmore Conference on Pathogenic Mechanisms in Enteric Diseases. Rapid City, S.Dak.
10. **Collinson, S. K., P. C. Doig, J. L. Doran, S. C. Clouthier, T. J. Trust, and W. W. Kay.** 1993. Thin aggregative fimbriae mediate binding of *Salmonella enteritidis* to fibronectin. J. Bacteriol. **175:**12–18.
11. **Collinson, S. K., L. Emödy, K.-H. Müller, T. J. Trust, and W. W. Kay.** 1991. Purification and characterization of thin, aggregative fimbriae from *Salmonella enteritidis*. J. Bacteriol. **175:**4773–4781.
12. **Collinson, S. K., L. Emödy, T. J. Trust, and W. W. Kay.** 1992. Thin aggregative fimbriae from diarrheagenic *Escherichia coli*. J. Bacteriol. **174:**4490–4495.
12a. **Collinson, S. K., and W. W. Kay.** Unpublished data.
13. **De Graaf, F. K.** 1988. Fimbrial structures of enterotoxigenic *E. coli*. Antonie van Leeuwenhoek **54:**395–404.
14. **Doran, J. L., S. K. Collinson, J. Burian, G. Sarlós, E. C. D. Todd, C. K. Munro, C. M. Kay, P. A. Banser, P. I. Peterkin, and W. W. Kay.** 1993. DNA-based diagnostic tests for *Salmonella* species targeting *agfA*, the structural gene for thin, aggregative fimbriae. J. Clin. Microbiol. **31:**2263–2273.
15. **Gold, L.** 1988. Posttranscriptional regulatory mechanisms in *Escherichia coli*. Annu. Rev. Biochem. **57:**199–233.
16. **Gong, M., and L. Makowski.** 1992. Helical structure of P pili from *Escherichia coli*. J. Mol. Biol. **228:**735–742.
17. **Hultgren, S. J., S. Normark, and S. N. Abraham.** 1991. Chaperone-assisted assembly and molecular architecture of adhesive pili. Annu. Rev. Microbiol. **45:**383–415.
18. **Jordi, B. J. A. M., I. E. L. op den Camp, L. A. M. de Haan, B. A. M. van der Zeijst, and W. Gaastra.** 1993. Differential decay of RNA of the CFA/I fimbrial operon and control of relative gene expression. J. Bacteriol. **175:**7976–7981.
19. **Kuehn, M. J., J. Heuser, S. Normark, and S. J. Hultgren.** 1992. P pili in uropathogenic *E. coli* are composite fibres with distinct fibrillar adhesive tips. Nature (London) **356:**252–255.
20. **Marck, C.** 1988. "DNA Strider": a "C" program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16:**1829–1836.
21. **McCormick, J. R., J. M. Zengel, and L. Lindahl.** 1991. Intermediates in the degradation of mRNA from the lactose operon of *Escherichia coli*. Nucleic Acids Res. **19:**2767–2776.
22. **Olsén, A., A. Arnqvist, M. Hammar, S. Sukupolvi, and S. Normark.** 1993. The RpoS sigma factor relieves H-NS-mediated transcriptional repression of *csgA*, the subunit gene of fibronectin-binding curli in *Escherichia coli*. Mol. Microbiol. **7:**523–536.
23. **Oudega, B., and F. K. De Graaf.** 1988. Genetic organization and biogenesis of adhesive fimbriae of *Escherichia coli*. Antonie van Leeuwenhoek **54:**285–299.
24. **Paranchych, W., and L. S. Frost.** 1988. The physiology and biochemistry of pili. Adv. Microb. Physiol. **29:**53–114.
25. **Reznikoff, W. S., D. A. Siegele, D. W. Cowing, and C. A. Gross.** 1985. The regulation of transcription initiation in bacteria. Annu. Rev. Genet. **19:**355–387.
26. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Sawers, G., and A. Böck.** 1989. Novel transcriptional control of the pyruvate formate-lyase gene: upstream regulatory sequences and multiple promoters regulate anaerobic expression. J. Bacteriol. **171:**2485–2498.
28. **von Heijne, G.** 1990. The signal peptide. J. Membr. Biol. **115:**195–201.
29. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.