Isolation and Characterization of a Gene Involved in Hemagglutination by an Avian Pathogenic
Escherichia coli Strain

DAVID L. PROVENCE* AND ROY CURTISS III

Department of Biology, Washington University, St. Louis, Missouri 63130

Received 11 October 1993/Returned for modification 14 December 1993/Accepted 25 January 1994

In this article, we report the isolation and characterization of a gene that may be important in the adherence of avian pathogenic Escherichia coli to the avian respiratory tract. The E. coli strain HB101, which is unable to agglutinate chicken erythrocytes, was transduced with cosmids libraries from the avian pathogenic E. coli strain χ7122. Enrichment of transductants that could agglutinate chicken erythrocytes yielded 19 colonies. These isolates contained cosmids that encompassed four nonoverlapping regions of the E. coli chromosome. Only one group of cosmids, represented by pYA3104, would cause E. coli CC118 to agglutinate chicken erythrocytes. A 10-kb fragment of this cosmid was subcloned in pACYC184. Transposon mutagenesis of this fragment with Tn5seq1 indicated that a contiguous 4.4-kb region of cloned DNA was required for hemagglutination. In vitro transcription/translation assays indicated that this 4.4-kb region of DNA encoded a protein of approximately 140 kDa. The nucleotide sequence of this region was determined and found to encode one open reading frame of 4,134 nucleotides that would encode a protein of 1,377 amino acids with a deduced molecular weight of 148,226. This gene confers on E. coli K-12 a temperature-sensitive hemagglutination phenotype that is best expressed when cells are grown at 26°C, and we have designated this gene tsh and the deduced gene product Tsh. Insertional mutagenesis of the chromosomal tsh gene in χ7122 had no effect on hemagglutination titers. The deduced protein was found to contain significant homology to the Haemophilus influenzae and Neisseria gonorrhoeae immunoglobulin A1 proteases. These data indicate that (i) a single gene isolated from the avian pathogenic E. coli strain χ7122 will confer on E. coli K-12 a hemagglutination-positive phenotype, (ii) χ7122 contains at least two distinct mechanisms to allow hemagglutination to occur, and (iii) the hemagglutinin Tsh has homology with a class of proteins previously not known to exist in E. coli.

Avian pathogenic Escherichia coli infection of the avian respiratory tract causes respiratory tract lesions and septicemia. This disease is referred to as air-sacculitis, pneumonitis, septicemic colibacillosis, or colisepticemia. These infections are often secondary to primary infections of infectious bronchitis virus, Newcastle disease virus, and Mycoplasma spp. (24). Avian E. coli strains with one of three O antigens, O1, O2, or O78, cause the majority of all E. coli-induced colisepticemias (39).

Infection of poultry by E. coli is thought to occur via inhalation of feces-contaminated dust. The precise location of E. coli deposition within the respiratory tract which leads to disease is unknown. Previous investigators have shown that avian pathogenic E. coli isolates can bind to avian tracheal epithelial cells and that type 1 and type 1-like fimbriae are responsible for this activity (16–18, 27, 38, 52–55, 60). The agglutination of erythrocytes has long been used as an in vitro assay of the expression of many different adhesins by E. coli (for a review, see reference 41). Several investigators have found that avian pathogenic E. coli do not agglutinate chicken erythrocytes in the presence of mannose (2, 20, 58). We have used a more sensitive assay of hemagglutination, the microhemagglutination assay (29), to test avian pathogenic E. coli for the expression of adhesins which mediate mannose-resistant hemagglutination. Using this assay, we recently described the mannose-resistant hemagglutination phenotype of an avian pathogenic E. coli isolate (47).

Hemagglutinins and other adhesins expressed by E. coli are produced in vitro when isolates are grown at 37°C but are often not produced during growth at lower temperatures. In contrast, the hemagglutinin we described was best expressed at lower temperatures (47). Hemagglutination activity was highest when cells were grown at 26°C and lower in cells grown at 37°C, and cells grown at 42°C lacked activity. This temperature-dependent hemagglutination phenotype was termed Tsh, for temperature-sensitive hemagglutinin. The in vitro conditions that cause the Tsh phenotype to be expressed are identical to those that cause certain E. coli strains to express curli, a fibronectin-binding structure (47). We report herein the cloning and characterization of a gene, tsh, isolated from the avian pathogenic E. coli strain χ7122. This gene will confer on E. coli K-12 a hemagglutination-positive phenotype.

**MATERIALS AND METHODS**

**Bacterial strains, plasmids, phage, and growth conditions.** The bacterial strains, plasmids, and phage used are described in Table 1. Phage were propagated in strain LE392. L broth and L agar (31) were used for routine growth of bacteria unless noted otherwise. Bacterial cells were grown on colonization factor antigen (CFA) agar (21) at 26°C for 48 h. The medium used to assay for productive TphthaA insertions was CFA agar containing 120 mM EPPS [N-(2-hydroxyethyl)piperazine-N'- (3-propanesulfonic acid)] (pH 8.2), 40 μg of 5-chloro-4-bromo-3-indolyl-p-toluidine salt (XP) per ml, and kanamycin. Antibi-
Bacterial strains and plasmids of tination. Briefly, bacteria on Biological Lipamide PROVENCE (Sigma) kanamycin, βg/ml; A (gpt-proA)62 gln V44 recA1
Amp′
ΔphoA20 recA1 argCamb
Δ(gpt-proA)62 gln V44 recA13

Plasmids
pYA3102 Cosmid clone of χ7122 in pSCOS1 This study
pYA3103 Cosmid clone of χ7122 in pSCOS1; contains the crl gene This study
pYA3104 Cosmid clone of χ7122 in pSCOS1 This study
pYA3105 Cosmid clone of χ7122 in pSCOS1 This study
pYA3107 10-kb EcoRI subclone of pYA3104 in pACYC184 This study
pYA3108 7-kb ClaI subclone of pYA3107 in pBluescript II SK This study
pYA3270 2.3-kb internal fragment of tsh subcloned in pGP704 This study
pACYC184 Tet′ Cm′ New England BioLabs
pBluescript II SK Amp′ Stratagene
pCRL20 crl in pUC18 40 pGP704 mobR4, Amp′ 34 pSCOS1 Cosmid cloning vector, Amp′ Stratagene
pUC18 Amp′ New England BioLabs
Phage
λ::Tn5seq1 λ::Tn5seq1.lh22c1857.Pam80 37 λ::TphoA λ::TphoA.lh8221.Pam3.rec::TphoA 26

Hemagglutination assays. All hemagglutination assays were done in the presence of 0.5% methyl α-D-mannopyranoside (Sigma) to control for mannose-sensitive hemagglutination by type 1 fimbriae. Chicken erythrocytes were obtained from Lampire Biological Laboratories (Pipersville, Pa.). Hemagglutination activity was assayed with the microhemagglutination test (29) with 96-well round-bottom microtiter plates (Corning Glass Works) as described elsewhere (47). Briefly, bacteria grown on CFA agar were harvested from the plates and serially diluted in 0.85% NaCl containing methyl α-D-mannopyranoside. A suspension of chicken erythrocytes was added to each well and mixed. Wells containing a small pellet of erythrocytes at the bottom of the well after 1 h of incubation on ice were considered hemagglutination negative, and those wells containing an even sheet of erythrocytes across the well were considered hemagglutination positive. Hemagglutination titer is expressed as the reciprocal of the greatest dilution of bacterial cells that resulted in a positive hemagglutination.

DNA and genetic manipulations. All DNA-modifying, ligase, and restriction endonuclease enzymes were used as suggested by the manufacturer (Promega). DNA was isolated from the avian pathogenic strain χ7122 by the procedure of Hull et al. (28). The DNA was partially digested with Sau3A and size fractionated over a sucrose gradient prepared as described elsewhere (5). DNA fragments of the correct size were pooled and ligated to pSCOS1 that had been digested with BamHI and phosphatased with calf intestinal phosphatase. The resulting ligation mixture was packaged in vitro as directed by the manufacturer (Stratagene).

Transposon mutagenesis of pYA3107, a plasmid conferring a hemagglutination-positive phenotype on E. coli strains CC118 and HB101, was done in strain CC118 as described elsewhere (48). After transposon mutagenesis, cells were incubated overnight in L broth containing kanamycin. Plasmids were purified from the overnight cultures. Plasmids isolated from cells mutagenized with λ::Tn5seq1 were used to transform HB101, and plasmids isolated from cells mutagenized with λ::TphoA were used to transform CC118. Individual kanamycin- and ampicillin-resistant transformants were tested for hemagglutination activity. CC118 transformants containing TphoA-mutagenized plasmids were tested for productive fusions between the transposon-encoded PhoA and plasmid-encoded proteins by plating on CFA agar containing XP and incubating at 26°C for 48 h. CFA agar buffered at pH 8.2 and containing XP had no effect on hemagglutination activity.

The plasmid pYA3270 was constructed by digestion of pYA3107.tsh-84 with BamHI and BstEII. pYA3107.tsh-84 contains the transposon Tn5seq1, and this transposon has a BamHI site close to each end (37). Thus, double digestion of pYA3107.tsh-84 with BamHI and BstEII yielded a 2.3-kb internal fragment of tsh. This fragment was treated with the Klenow fragment of DNA polymerase I and ligated into the EcoRV site of pGP704 to yield pYA3270. This plasmid was mobilized into χ7122 by SM10/pipir as described elsewhere (56). Transconjugants were selected by plating the mating mixture on L agar containing ampicillin.

Enrichment for hemagglutination-positive recombinant clones. The in vitro-packaged cosmid library was used to transduce HB101 to ampicillin resistance and plated on CFA agar containing ampicillin. These cells were then used to enrich for those that could agglutinate chicken erythrocytes by a modification of a previously described enrichment procedure (14). Approximately 750 bacterial colonies were harvested from the plates and resuspended in 2 ml of 0.85% NaCl. A 2-ml volume of a 3% erythrocyte suspension was added to this, mixed, and incubated on ice for 1 h. The erythrocytes and associated bacteria were separated from unbound bacteria by briefly pelleting at 200 × g for 5 min. The supernatant was removed, and the pellet was gently resuspended in 4 ml of 0.85% NaCl and incubated on ice for 1 h. This procedure was repeated two more times. After the third centrifugation, the erythrocyte-bacterium pellet was streaked for isolated colonies on L agar containing ampicillin. A total of 37 colonies were assayed for agglutination activity after growth on CFA agar. Nineteen colonies were able to agglutinate chicken erythrocytes.

Colony blot and Southern blot hybridization. Cosmids containing the crl gene were identified by colony blot hybridization by using a 300-bp crl-specific DNA probe (47). Insertional inactivation of tsh in χ7141 was determined by Southern blot hybridization by using a 3.0-bp tsh-specific DNA probe. This internal fragment of tsh was isolated by digestion of pYA3107 with AfII and NdeI and gel purified. DNA probes were labeled with [γ-32P]ATP (NEN Research Products) by using a random-primer DNA-labeling kit (Boehringer Mannheim Biochemicals). Chromosomal DNAs were isolated from χ7122 and χ7141 as described previously (28) and digested with
BamHI and Clal. DNA fragments were resolved on a 0.5% agarose gel. DNA fragments and colonies were transferred to nitrocellulose (Micron Separations, Inc.), and DNA hybridizations were done as described elsewhere (51). Hybridized membranes were exposed to Kodak X-Omat AR film at −70°C with an intensifying screen (Du Pont Cronex Lightning Plus).

In vitro transcription/translation. pYA3107 and three Tn5seq1-mutagenized derivatives, pYA3107-tsh-55, pYA3107-tsh-84, and pYA3107-tsh-98, were each subcloned into pBlueScript II SK. Subcloning was done by digesting pYA3107 and the three mutagenized derivatives with Clal (see Fig. 1). The DNA fragments containing the sequences conferring the hemagglutination-positive phenotype were isolated and ligated to pBlueScript II SK that had been digested with Clal. The orientation of the insert within the vector was determined by restriction endonuclease digestion. The four subclones used for in vitro transcription/translation were in the same orientation. In vitro transcription/translation was done with [35S]methionine (NEN Research Products) and a commercially available kit (Promega). The number of trichloroacetic acid-precipitated counts was determined for each reaction. Approximately 200,000 counts of each reaction were resolved by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis. The molecular weight markers used were [14C]-methyldated proteins (Amersham). The resulting gel was soaked in Amplify (Amersham) as directed by the manufacturer and exposed to Kodak X-Omat AR film at −70°C with an intensifying screen (Du Pont Cronex Lightning Plus).

DNA sequence. The DNA sequence was determined by using pYA3107 and the following Tn5seq1 insertions in pYA3107 as templates: tsh-42, -44, -52, -53, -59, -67, -72, -75, -76, -78, -83, -84, -85, -89, -95, and -98. Tn5seq1 contains T7 and SP6 sequences at the left and right ends of the transposon, respectively, and allows the DNA sequence adjacent to the transposon to be determined (37). Because of the presence of a DNA sequence within pYA3107 that had homology to the T7 promoter, we used a different primer to determine the DNA sequence adjacent to the left end of the transposon. The sequence of this primer was 5′ GCTCGCATCTAAATACGAC 3′. Gaps in the DNA sequence were closed by synthesizing internal sequencing primers as necessary and using pYA3107 as template DNA. The nucleotide sequence of each strand of DNA was determined. The methodology used was standard dideoxynucleotide termination reactions with 7-deaza-dGTP and with [35S]-dATP as the label. The sequencing reactions, data entry, and alignment of the individual DNA sequences into a final consensus sequence were performed by Lark Sequencing Technologies, Inc. (Houston, Tex.).

DNA sequence analysis. DNA sequence analysis was done with the sequence analysis software package by Genetics Computer, Inc. (the GCG package) (15). The algorithm TFASTA was used to detect homology at the amino acid level between the DNA sequence reported herein and the deduced protein sequences in the GenEMBL nucleotide sequence data library (42). The optimal secondary structure of the RNA molecule directly downstream of a putative stop codon was determined by the method of Zucker and Stiegler (61).

Nucleotide sequence accession number. The DNA sequence described here has been submitted to the Genome Sequence Data Base. The accession number is L27423.

RESULTS

Isolation of cosmids clones that confer a hemagglutination-positive phenotype on HB101. We previously described an

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>HB101 HA titer (%)</th>
<th>HB101 Fbn binding (%)</th>
<th>CC118 HA titer (%)</th>
<th>CC118 Fbn binding (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pYA3102</td>
<td>128</td>
<td>55</td>
<td>0</td>
<td>1.3</td>
</tr>
<tr>
<td>pYA3103</td>
<td>16</td>
<td>18.8</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>pYA3104</td>
<td>4,096</td>
<td>7.6</td>
<td>8,192</td>
<td>1.2</td>
</tr>
<tr>
<td>pYA3105</td>
<td>16</td>
<td>32</td>
<td>0</td>
<td>1.2</td>
</tr>
<tr>
<td>pSCOS1</td>
<td>0</td>
<td>7.1</td>
<td>0</td>
<td>1.1</td>
</tr>
<tr>
<td>pCRL20</td>
<td>8</td>
<td>22.2</td>
<td>0</td>
<td>1.3</td>
</tr>
<tr>
<td>pUC18</td>
<td>0</td>
<td>2.2</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>pYA3107</td>
<td>1,024</td>
<td>2.7</td>
<td>2,048</td>
<td>1.2</td>
</tr>
<tr>
<td>pACYC184</td>
<td>0</td>
<td>1.5</td>
<td>0</td>
<td>1.1</td>
</tr>
</tbody>
</table>

- **HA titer**: Bacterial strains were grown on CFA agar at 26°C for 48 h and tested for hemagglutination of chicken erythrocytes and fibronectin binding as described in Materials and Methods.
- **Fbn binding**: Hemagglutination titer is expressed as the reciprocal of the greatest dilution of bacterial cells that resulted in a positive hemagglutination.
- **HA titers**: HA, hemagglutination. HA titer is expressed as the reciprocal of the greatest dilution of bacterial cells that resulted in a positive hemagglutination.
- **Fbn binding**: Fbn, fibronectin. Values are mean percentages of fibronectin bound in triplicate assays.

avian pathogenic *E. coli* strain, designated χ7122, that was capable of agglutinating chicken erythrocytes (47). This hemagglutination phenotype was expressed best when χ7122 was grown on CFA agar at 26°C and has been designated Tsh. We constructed cosmid libraries of χ7122 DNA and enriched for those cosmids that caused HB101 to agglutinate erythrocytes (see Materials and Methods). Previous work with HB101 suggests that HB101 may contain the genes that encode a hemagglutinin but is incapable of expressing these genes (47). This deficiency in expression of the hemagglutination phenotype can be overcome when the *crf* gene is present on a high-copy-number plasmid (47). Consequently, the use of HB101 as host for the isolation of cosmids causing hemagglutination was expected to yield two classes of cosmids. One class would be cosmids that contained genes causing expression of the cryptic genes present in HB101. The other class of cosmids would contain the genes that may encode a hemagglutinin.

Enrichment of hemagglutination-positive cells resulted in the isolation of 19 hemagglutination-proficient isolates. This procedure did not rule out the isolation of siblings, and restriction endonuclease digestion of the 19 cosmid clones isolated indicated that there were four groups of nonoverlapping clones (data not shown). One representative cosmid from each group was chosen, and the four were designated pYA3102, pYA3103, pYA3104, and pYA3105 (Table 1). χ7122 contains DNA sequences that are homologous to the *crf* gene (47); thus, we expected our enrichment procedure to result in the isolation of the χ7122 *crf* gene. Colony blot analysis of the 19 cosmids with a labeled *crf* probe indicated that one cosmid, pYA3103, contained the *crf* gene (data not shown). The isolation of the *crf* gene indicated that this procedure was able to enrich for hemagglutination-proficient cells.

**Phenotypes of the** *E. coli* K-12 strains HB101 and CC118 containing the four cosmids clones. We wanted to distinguish between those cosmids containing regulatory genes causing the expression of the cryptic hemagglutination activity of HB101 and those cosmids containing genes encoding a putative hemagglutinin. We distinguished between these two classes of cosmids by using two assays: a fibronectin-binding assay and a hemagglutination activity assay.

HB101 will bind fibronectin when the *crf* gene is present on
a high-copy-number plasmid (40), and this is also due to the expression of cryptic genes in HB101 (1). The E. coli K-12 strain CC118 will not bind fibronectin when the crl gene is present (Table 2). The ability of the four cosmids to cause HB101 and CC118 to bind fibronectin was determined (Table 2). pYA3103, the cosmid containing crl, caused HB101 to bind fibronectin and did not cause this phenotype to be expressed by CC118, pYA3102 and pYA3105 caused HB101 but not CC118 to bind fibronectin. In contrast, pYA3104 did not confer on HB101 or CC118 the ability to bind fibronectin.

E. coli strain CC118 will not agglutinate erythrocytes when grown on CFA agar at 26°C (Table 2). Also, unlike HB101, CC118 will not agglutinate erythrocytes when it contains the crl gene on a high-copy-number plasmid (Table 2), suggesting that CC118 does not contain the cryptic genes present in HB101. We determined the hemagglutination activity of CC118 when it contained each of the cosmids pYA3102, pYA3103, pYA3104, and pYA3105 (Table 2). Only one cosmid, pYA3104, caused CC118 to agglutinate erythrocytes.

Since pYA3102 and pYA3105 caused the expression of hemagglutination activity and fibronectin binding in HB101 and CC118 in a manner similar to that caused by crl, it is possible that pYA3102 and pYA3105 may contain regulatory genes that act, like crl, to cause the expression of cryptic genes. pYA3104, the only clone that caused both strains to hemagglutinate, did not have the same effect that pYA3102, pYA3105, or the crl gene had on HB101 and CC118. We believe these results are consistent with the hypothesis that pYA3104 contains the structural genes that encode a hemagglutinin. Consequently, this clone was chosen for further analysis.

Subcloning and transposon mutagenesis of pYA3104. EcoRI digestion of pYA3104 resulted in three large DNA fragments (data not shown). When each of these was subcloned into the plasmid pACYC184, only one caused HB101 to agglutinate erythrocytes. This subclone contained a 10-kb DNA fragment and was designated pYA3107 (Fig. 1). This subclone had the same effect on HB101 and CC118 as did pYA3104 (Table 2).

The region of DNA in pYA3107 required for hemagglutination by HB101 was identified by transposon mutagenesis with Tn5seq1 and TnphoA (Fig. 1). All insertions resulted in either no hemagglutination activity or wild-type levels of activity. This analysis indicated that at least 4.4 kb of DNA between the tsh-69 and tsh-97 insertions was necessary for hemagglutination. TnphoA mutagenesis resulted in productive fusions within the region of DNA required for hemagglutination activity, indicating that this region contained at least one gene encoding an exported protein (Fig. 1). The orientation of productive TnphoA insertions was determined by restriction endonuclease digestion. Each productive fusion indicated that the direction of transcription was from the left to the right, as depicted in Fig. 1. Also, productive TnphoA insertions were limited to the 5' end of the 4.4-kb region of DNA important in conferring a hemagglutination-positive phenotype.

Identification of plasmid-encoded polypeptides. pYA3107 and three Tn5seq1 insertions, tsh-55, tsh-84, and tsh-98, were digested with CiaI. The DNA fragment containing the region of DNA important in conferring hemagglutination activity was isolated from each plasmid and subcloned into pBluescript II SK to yield pYA3108, pYA3108.tsh-55, pYA3108.tsh-84, and pYA3108.tsh-98. These plasmids were analyzed for plasmid-encoded polypeptides. Figure 2 shows that the pYA3108 subclone encodes one prominent polypeptide of approximately 140 kDa (lane 2).

The plasmids containing the tsh-84 and tsh-98 insertions each produced three predominant polypeptides (Fig. 2, lanes 4 and 5). Two proteins, migrating at 51.8 kDa and 29 kDa, are the transposon-encoded tnp and aphA gene products, respectively. The 140-kDa polypeptide produced by pYA3108 is not present in the polypeptide profile of pYA3108.tsh-84 or pYA3108.tsh-98 and is replaced by a polypeptide of 29.5 kDa or 118 kDa, respectively. The plasmid containing the tsh-55 insertion produced the 140-kDa peptide, but at a greatly reduced level (Fig. 2, lane 3).

As described above, mutagenesis of pYA3107 with TnphoA resulted in productive fusions between the transposon-encoded phoA gene and the gene present on pYA3107. These TnphoA insertions map to locations within the gene encoding the 140-kDa polypeptide. Mapping the orientation of the six productive TnphoA insertions indicated that the direction of transcription of the gene containing the TnphoA insertions was from left to right.

DNA sequence of the 4.4-kb region required to confer a hemagglutination-positive phenotype. The nucleotide se-
The deduced molecular weight of this protein is 148,226, which is similar to the size of the gene product in the in vitro transcription/translation system (Fig. 2, lane 2). A putative promoter is present. The −35 region (TTCACA, nucleotides 271 to 276) and the −10 region (TCTTTC, nucleotides 293 to 298) are similar to the consensus −35 (TTGACA) and the consensus −10 (TATAAT) regions found in sigma 70 promoters (19, 50). The spacing between the −35 and −10 regions of the putative promoter is 16 bp, the optimal spacing for promoters (50). A putative ribosome binding site (AGGAG) is 8 bp upstream of the first ATG codon of the ORF. The sequence of this putative Shine-Dalgarno site is identical to the consensus, and the spacing between it and the ATG codon is optimal (23). The transposon insertion mutation tsh-55 maps within the putative promoter. This is consistent with the observation that this mutation does not affect the size of the protein in the in vitro transcription/translation system, but only the level of expression (Fig. 3, lane 3). The DNA sequence analysis and transposon mutagenesis strongly suggest that this ORF of 4,134 nucleotides encodes a protein that is important in conferring a temperature-sensitive phenotype on E. coli K-12. Accordingly, this gene was designated tsh, and the gene product was designated Tsh.

TphoA mutagenesis indicates that this protein is secreted (see above). There are several potential signal sequence cleavage sites in the amino-terminal region of Tsh (Fig. 3) that follow the “(−3, −1) rule” (59). The locations of two TphoA insertions were determined by DNA sequence analysis (Fig. 3). The DNA sequence of the tsh-16 and tsh-33 fusion joints indicated that in-frame fusions had occurred with tsh and that the direction of transcription of both insertions was from left to right. These results are consistent with the interpretation that the 4.4-kb region of DNA that confers hemagglutination activity produces one protein of approximately 140 kDa and this protein is exported.

The end of the ORF is followed by an inverted repeat that begins 5 bp downstream of the termination codon and spans nucleotides 4520 to 4556. The stem is 14 bp and the loop is 9 bp. As calculated by the method of Timco et al. (57), this inverted repeat has a ΔG of −32.8 kcal (ca. −137 kJ). This inverted repeat is followed by TCTG immediately after an AT-rich stretch of 4 nucleotides. The TCTG sequence is thought to be a consensus sequence in rho-independent transcription terminators (8). These results indicate that the 4.4-kb region of DNA that confers hemagglutination activity contains one gene and that this gene is preceded by and followed by the appropriate DNA sequence motifs to signal the initiation and termination of transcription and translation of the gene and gene product.

Inactivation of tsh in χ7122 by insertional mutagenesis. We tested the role of tsh in the expression of hemagglutination activity by χ7122. A 2.3-kb internal fragment of χ7122 was subcloned into the suicide plasmid pGEM4 (Fig. 1). This plasmid requires the pir gene product to replicate (34). Transfer of this plasmid into a host that does not contain the pir gene allows for easy selection of those strains in which the plasmid has integrated into the chromosome. If the plasmid contains an internal fragment of a gene, integration into the chromosome by homologous recombination will result in the insertional mutagenesis of that gene. After mobilization of this plasmid into χ7122, ampicillin-resistant transconjugants were selected. One ampicillin-resistant transconjugant was chosen and designated χ7141.

χ7141 was screened for insertion of the plasmid pYA3270 into the chromosomal tsh by Southern blot analysis (Fig. 4). A 3.0-kb NdeI-AflIII internal fragment of tsh hybridized to a 6.8-kb DNA fragment in χ7122. This probe hybridized to 4.3- and 5.2-kb DNA fragments in χ7141. We expected the insertion of pYA3270 into tsh by a single homologous recombination event to result in a 4.3- and a 5.1-kb DNA fragment that would hybridize to the internal tsh probe. The observed results are consistent with the expected molecular weights of the DNA fragments of χ7141 and indicate that tsh has been mutated in χ7141. This allele has been designated tsh-101.

Both χ7122 and the mutant χ7141 agglutinated erythrocytes to the same titer of 64. We interpret these results to indicate that χ7122 has at least two mechanisms to mediate the hemagglutination of erythrocytes.

Comparison of the deduced amino acid sequence of Tsh with those of other proteins. The deduced amino acid sequence encoded by tsh was found to contain homology to the serine-type immunoglobulin A1 (IgaA1) proteases produced by Haemophilus influenzae and Neisseria gonorrhoeae (Fig. 5). The sequences of four serologically distinct H. influenzae iga genes and one N. gonorrhoeae iga gene are in the GenBank data base. Comparison of the full-length proteins reveals that Tsh has approximately 29% identity and 50% similarity with each of the four deduced H. influenzae protein sequences. The identity and similarity between Tsh and the N. gonorrhoeae IgaA1 proteases are 28% and 50%, respectively. As can be seen in Fig. 5, the region of Tsh homologous to the serine-type IgaA proteases is limited mainly to the approximately 900 amino-terminal residues of Tsh.

The deduced amino acid sequence encoded by tsh was also found to contain homology to six regions of the Bordetella pertussis pertactin outer membrane protein (Fig. 6).
FIG. 3. Nucleotide sequence of tsh. The accession number of this DNA sequence is L27423. The sequence of the noncoding strand is shown with a portion of the deduced protein sequence. The entire deduced protein sequence is shown in Fig. 5. Putative promoter −35 and −10 regions and the putative ribosome binding site (RBS) are underlined and marked. The putative initiation codon is underlined and shown in boldface print. The 9-bp regions that three Tn5seq insertions (tsh-55, tsh-84, and tsh-98) map to are shown in boldface print. The insertion sites of two TnphoA insertions that result in productive fusions are shown by the downward-pointing arrows at nucleotides 883 (tsh-16) and 1778 (tsh-33). The potential signal peptidase cleavage sites are shown by the upward-pointing arrows. Only the first 74 amino acids of the ORF are shown. The transcriptional terminus of the cloned DNA was required for hemagglutination of this polypeptide was determined. The nucleotides capable of forming a hairpin structure downstream of the stop codon at nucleotides 4512 to 4514 is shown in boldface. The nucleotides are underlined. The enrichment of cosmids in the E. coli strain CC118, a strain that does not express hemagglutination activity when the crl gene is present on a plasmid. Of the four cosmids, only pYA3104 would cause CC118 to express hemagglutination activity. The two cosmids pYA3102 and pYA3105 had the same effect on HB101 and CC118 as the crl-containing cosmids pYA3103 inasmuch as they caused only HB101 and not CC118 to express hemagglutination activity. The similarity between the three cosmids pYA3102, pYA3103, and pYA3105, the three cosmids that had no effect on CC118, is underscored by the observation that these three cosmids also caused HB101 to bind fibronectin. Fibronectin binding is another phenotype that the crl gene will cause HB101 to express (40). These data suggest that pYA3102 and pYA3105 contain sequences that act in HB101 to cause the expression of a hemagglutination-positive phenotype. Work to delineate the role of pYA3102 and pYA3105 in the expression of hemagglutination activity in HB101 is under way. The region of cloned DNA of pYA3107 required for hemagglutination was determined by transposon mutagenesis to be approximately 4.4 kb in size. The coding capacity of this region

**DISCUSSION**

We report herein the isolation and characterization of a gene from the avian pathogenic E. coli strain χ7122 that will cause the E. coli K-12 strains HB101 and CC118 to hemagglutinate chicken erythrocytes. This gene, tsh, was isolated on a cosmid clone and subcloned as a 10-kb EcoRI fragment. Transposon mutagenesis and in vitro transcription/translation of this 10-kb EcoRI DNA fragment indicated that a 4.4-kb region of the cloned DNA was required for hemagglutination and that this 4.4-kb region encoded one polypeptide of approximately 140 kDa. Transposon mutagenesis with TnphoA also suggested that this polypeptide was exported. The DNA sequence of this gene was determined. The DNA sequence was found to contain a single ORF of 4,134 nucleotides that would encode a protein with the deduced molecular weight of 148,226.

Previous work has shown that HB101 will express a cryptic hemagglutination activity when the crl gene is present on a plasmid, and we have shown that the avian pathogenic E. coli strain χ7122 contains DNA sequences homologous to crl (47). Thus, the use of HB101 as the recipient strain for the cosmid libraries provided for an internal control to determine the utility of the enrichment procedure for the isolation of hemagglutination-positive bacteria. The enrichment procedure used to isolate hemagglutination-positive bacteria resulted in the identification of four groups of nonoverlapping clones. The crl gene was isolated on a cosmid as expected. However, the use of HB101 as the recipient allowed the isolation of two types of cosmids, those containing regulatory genes that will cause the expression of the cryptic hemagglutination activity of HB101 and those containing the structural genes encoding proteins required for hemagglutination activity. Identification of cosmids containing structural genes important in hemagglutination activity was accomplished by screening the cosmids in the E. coli strain CC118, a strain that does not express hemagglutination activity when the crl gene is present on a plasmid. Of the four cosmids, only pYA3104 would cause CC118 to express hemagglutination activity. The two cosmids pYA3102 and pYA3105 had the same effect on HB101 and CC118 as the crl-containing cosmids pYA3103 inasmuch as they caused only HB101 and not CC118 to express hemagglutination activity. The similarity between the three cosmids pYA3102, pYA3103, and pYA3105, the three cosmids that had no effect on CC118, is underscored by the observation that these three cosmids also caused HB101 to bind fibronectin. Fibronectin binding is another phenotype that the crl gene will cause HB101 to express (40). These data suggest that pYA3102 and pYA3105 contain sequences that act in HB101 to cause the expression of a hemagglutination-positive phenotype. Work to delineate the role of pYA3102 and pYA3105 in the expression of hemagglutination activity in HB101 is under way.

The region of cloned DNA of pYA3107 required for hemagglutination was determined by transposon mutagenesis to be approximately 4.4 kb in size. The coding capacity of this region

FIG. 4. Southern blot analysis of insertional inactivation of tsh in χ7122. (A) χ7122 (lane 1) and χ7141 (lane 2) whole-cell genomic DNAs were isolated, digested with BamHI and ClaI, electrophoretically resolved, and transferred to membranes as described in Materials and Methods. The membranes were hybridized to a 32P-labeled internal fragment of tsh. At left are DNA size standards (in kilobases). (B) Schematic of the single homologous recombination event between pYA3270 and χ7122 and the resulting structure of the χ7141 chromosome. The DNA fragments resulting from BamHI and ClaI digestion of χ7122 and χ7141 are labeled A through C and correspond to the bands in panel A as follows: A, band in lane 1; B, upper band in lane 2; C, lower band in lane 2.
FIG. 5. Comparison of the deduced amino acid sequence of Tsh with those of the IgAl protease proteins of *H. influenzae* HK368 (45) and *N. gonorrhoeae* MS11 (44). The amino acids of the deduced Tsh sequence are numbered from the amino-terminal end of the unprocessed protein. The three single arrows indicate the proposed autoproteolytic sites in the IgA protease from *H. influenzae* HK368 (45). The two double arrows indicate the autoproteolytic sites in the *N. gonorrhoeae* IgA protease (44). The triangle indicates an autoproteolytic site in both the *H. influenzae* HK368 and the *N. gonorrhoeae* IgA proteases. Dashes indicate gaps introduced into the amino acid sequence to optimize homology. Amino acids identical between all the protein sequences are shown in boldface and are depicted as the appropriate residue on the line marked Consensus.
of DNA is approximately 161 kDa; thus, it is possible that this region could encode a protein of 140 kDa. This suggests that most of the faint bands migrating below the 140-kDa band in Fig. 2, lane 2, are not encoded by the subcloned DNA and may be the result of premature transcription or translation termination products of the gene encoding the 140-kDa protein. Alternatively, the proteins migrating below 140 kDa may be the result of proteolysis of the 140-kDa protein. In vitro transcription/translation assays of the transposon insertion mutations tsh-84 and tsh-98 indicate that the 140-kDa polypeptide is replaced by peptides of 29.5 kDa and 118 kDa, respectively. This observation is consistent with the interpretation that the tsh-84 and tsh-98 insertions are in a single gene and result in the truncation of the 140-kDa polypeptide. Furthermore, the DNA sequence analysis indicates that transposon insertion mutations tsh-84 and tsh-98 are both located within the 4,134-nucleotide ORF (Fig. 3). If the first ATG codon shown in Fig. 3 is the initiation codon, the predicted molecular weight of the truncated proteins from insertions tsh-84 and tsh-98 are 29.5 kDa and 118 kDa, respectively. The predicted sizes of these truncated proteins are consistent with the sizes observed in the in vitro transcription/translation analysis.

While the transposon insertion mutation tsh-55 will no longer confer a hemagglutination-positive phenotype, the full-length protein is still produced in vitro (Fig. 2, lane 3). The location of the transposon insertion mutation tsh-55 was determined and was found to be in the putative promoter of the
gene tsh (Fig. 3). This is consistent with the observation that the tsh-55 mutation results in decreased levels of the full-length protein Tsh (Fig. 2, lane 3). We believe that in E. coli K-12 strains containing this mutation the greatly decreased levels of Tsh are not enough to confer a hemagglutination-positive phenotype.

Insertional mutagenesis of the gene tsh had no effect on the ability of χ7122 to express hemagglutination activity. We believe that this indicates that χ7122 has more than one mechanism to mediate the hemagglutination of erythrocytes. This interpretation is supported by the results of the enrichment procedure we used to isolate cosmid clones conferring hemagglutination activity on HB101. Four distinct groups of cosmids, represented by pYA3102, pYA3103, pYA3104, and pYA3105, that conferred on HB101 the ability to agglutinate erythrocytes were isolated. This suggests that there are multiple mechanisms by which χ7122 will agglutinate erythrocytes. Assessment of the role of tsh in hemagglutination activity by χ7122 will require that other hemagglutination mechanisms in χ7122 be characterized and mutagenized.

The deduced amino acid sequence of Tsh was found to be homologous to the serine-type IgA1 proteases of H. influenzae and N. gonorrhoeae. The serine-type IgA1 proteases of H. influenzae and N. gonorrhoeae are composed of three functional domains (44, 45): the amino-terminal signal peptide, the protease, and the carboxy-terminal β domain. The IgA1 proteases of H. influenzae and N. gonorrhoeae are secreted into the extracellular space by a novel mechanism (25, 44). After movement of the protein into the periplasmic space, the carboxy-terminal β domain of the prorotease is inserted into the outer membrane. The amino-terminal end of the protein is then exported through the outer membrane by the β domain. When the amino-terminal end has been exported, the protein is autoproteolytically cleaved at three closely spaced sites. This allows the mature protease to be released from the cell and presumably to act on the substrate during the infective process. The long-term fate of the β domain left behind in the outer membrane is unknown.

The deduced amino acid sequences of the H. influenzae and N. gonorrhoeae serine-type IgA1 proteases have been compared (45). The levels of identity and similarity between the two proteins are 49 and 65%, respectively. This homology exists throughout the domain containing the mature protease, and there is also extensive homology in the carboxy-terminal end of the amino acid sequence that encodes the β domain (45). When Tsh is aligned to the IgA1 proteases (Fig. 5), it is apparent that essentially all of the homology between Tsh and the other proteases is in the region of the IgA1 proteases that encodes the mature protease. The levels of identity and similarity between the first 900 amino acids of Tsh and the serine-type IgA1 proteases are 33 and 54%, respectively. We believe that this homology is significant for two reasons. The homology extends over a large region of amino acids that corresponds to a functional domain in the IgA1 proteases. Also, there are four areas of complete identity spanning six to nine consecutive amino acids between Tsh and the five other IgA1 proteases (Fig. 5, amino acids 64 to 72, 163 to 168, 257 to 264, and 441 to 446).

The functional significance of the homology between the deduced Tsh and the serine-type IgA1 proteases is unclear. Since the homology between Tsh and the serine-type IgA1 proteases is localized to the region of the IgA1 proteases that encodes the mature protease, the amino acid sequence comparison suggests that tsh encodes an IgA1 protease. Chickens do contain an Ig heavy chain that appears to be similar to human and mouse IgA (30, 33). However, this chicken Ig does not appear to contain a hinge region (33). This may be significant, because the hinge region of human IgA1 is cleaved by the H. influenzae and N. gonorrhoeae IgA1 proteases (for a review, see reference 43). We are currently evaluating the ability of Tsh to cleave human IgA1 and chicken IgA.

One observation indirectly suggests that the protein encoded by tsh may have proteolytic activity in addition to the observed hemagglutination activity. The H. influenzae and N. gonorrhoeae IgA1 proteases are serine proteases (3). Serine proteases of the chymotrypsin-trypsin family contain a 7-amino-acid sequence that is highly conserved and includes the active site serine (10). The IgA1 proteases contain this amino acid motif (3), and in vitro mutagenesis of an H. influenzae IgA1 protease that changed the active site serine to a threonine completely abolished protease activity (46). Tsh contains this 7-amino-acid sequence (Fig. 5, amino acids 257 to 263) and also contains the serine residue shown to be important in the enzymatic activity of the H. influenzae IgA1 protease.

The significance of the limited homology between Tsh and pertactin is unknown. The pertactin protein of B. pertussis is a member of a family of outer membrane proteins also present in Bordetella bronchiseptica and Bordetella parapertussis (35). The B. pertussis pertactin has been shown to mediate attachment to some eukaryotic cells (9). Since Tsh contains regions of homology to pertactin and the serine-type IgA1 proteases, two features shared by both pertactin and the serine-type IgA1 proteases may be noteworthy. Both are outer membrane proteins and both undergo posttranslational processing (11, 12, 44, 45). We have shown that Tsh is an outer membrane protein and that it is proteolytically cleaved during maturation (49).

tsh was isolated and characterized as a gene that allowed two laboratory E. coli strains to agglutinate chicken erythrocytes. As discussed above, the deduced amino acid sequence of the protein Tsh indicates that this protein has significant homology with the serine-type IgA1 proteases and therefore may have proteolytic activity. We believe that there are two formal possibilities for the role of the tsh gene product in hemagglutination. The gene tsh may encode a bifunctional protein that...
acts as both hemagglutinin and protease. Other bifunctional microbial adhesins have been isolated and characterized. The Hap protein of Vibrion cholerae is a zinc- and calcium-dependent metalloprotease that will cause chicken erythrocytes to agglutinate (6, 22). The Pseudomonas aeruginosa exoenzyme S protein will ADP ribosylate membrane-associated eukaryotic proteins (13) and will also bind to glycosphingolipids (4). Alternatively, the tsh gene product may play an ancillary role in hemagglutination. In this case, Tsh may proteolytically process a cryptic protein on the surface of E. coli which would then mediate hemagglutination or may modify a host erythrocyte surface protein to which the microbe could then bind. Distinguishing between these possible roles of Tsh in hemagglutination will require a greater understanding of the properties of Tsh. To this end we are currently investigating the events that occur with Tsh as it matures during export and the proteolytic properties of Tsh.

ACKNOWLEDGMENTS

Helpful discussions with P. K. Brown, A. L. Honeymoon, and R. S. Munson, Jr., and the critical review of the manuscript by A. L. Honeymoon and R. S. Munson, Jr., are gratefully acknowledged.

This work was supported by grant AI28487 from the National Institutes of Health and an unrestricted grant from Bristol-Meyers Squibb.

REFERENCES


Downloaded from http://iai.asm.org/ on August 28, 2017 by guest