Cloning, characterization and sequencing of two haemagglutinin genes from *Eikenella corrodens*

VENKATARAMA K. RAO, JOAN A. WHITLOCK and ANN PROGULSKE-FOX*

Department of Oral Biology and Periodontal Disease Research Center, University of Florida, Gainesville, Florida 32610, USA

(Received 10 July 1992; accepted 11 August 1992)

*Eikenella corrodens* is emerging as an important human pathogen, in both extra-oral and periodontal infections. From a clone bank of *Eikenella corrodens* chromosomal DNA produced in *Escherichia coli* JM109, twenty-two clones expressed *Eikenella* antigens and of these, two expressed functional haemagglutinins. By virtue of different restriction maps and a lack of homology by Southern hybridization, the two cloned fragments encoding the two haemagglutinins have been shown to be distinct. Maxicell analysis revealed that clone 1, carrying plasmid pVKR201, produces three *Eikenella* proteins, one of 31.5 kDa and two of approximately 14 kDa each. Expression of each of the proteins appears to be under the control of an *Eikenella* promoter(s). Clone 2, carrying plasmid pVKR301, produces two proteins, one of 93 kDa and the second of 17 kDa. Expression of both of these proteins in *E. coli* requires the *lac* promoter in the vector. By preparing a series of subclones and testing each by maxicell analysis and for haemagglutination activity, a functional map of the insert of clone 1 was deduced and the 31.5 kDa polypeptide identified as the haemagglutinin. Using similar methods, the 17 kDa protein was found to be the haemagglutinin of clone 2. The nucleotide sequences of both haemagglutinin genes were determined and are presented. Computer analysis revealed no homology between the two haemagglutinins, and no homology to any previously sequenced proteins. These are the first genes of this genus to be cloned and sequenced.

Introduction

*Eikenella corrodens* is a Gram-negative, microaerophilic rod which was first described by Henriksen (1948) and later studied in more detail by Eiken (1958). Characterized by the unusual ability to pit or 'corrode' the agar surface, this bacterium was regarded mainly as a microbiological curiosity until relatively recently. Partly because of improved culture techniques, *Eik. corrodens*, which colonizes the upper respiratory tract and oral cavity of humans (Henriksen, 1969; Tanner et al., 1979), has, more recently, been recognized as an important opportunistic pathogen (Brooks et al., 1974; DeMello & Leonard, 1979; Stoloff & Gillies, 1986). It has been isolated from a variety of infectious processes, including abscesses of the thyroid (Vichyanond et al., 1983; Cheng et al., 1988a), brain (Drake & Fischer, 1986; Cheng et al., 1988b), liver (Hofstad & Horn, 1989; Massey, 1989), osteomyelitis (Jones & Romig, 1979; Farrington et al., 1983; Stoloff & Gillies, 1986), chorioamnionitis (Sporken et al., 1985), septic arthritis (Flesher & Bottone, 1989) and endocarditis (Sobel et al., 1981; Landis & Korver, 1983; Decker et al., 1986). In addition, *Eikenella* is regarded as an important cause of infection following human bite wounds (Schmidt & Herkman, 1983; Stoloff & Gillies, 1986). Although frequently isolated as part of a polymicrobial process, *Eik. corrodens* has been isolated as the sole infecting agent in cases of meningitis, endocarditis and osteomyelitis (Jones & Romig, 1979). *Eik. corrodens* is also associated with certain forms of periodontal disease and is considered an important contributor to tissue destruction and alveolar bone loss in this complex infectious process (Listgarten et al., 1978; Tanner et al., 1979; Dzink et al., 1985; Mattison et al., 1987). Despite increasing evidence as to the importance of *Eikenella* as a human pathogen, relatively little information is available on the mechanisms by which it causes disease. Among the potential factors critical in the
virulence of *Eik. corrodens* are its endotoxin, which is a potent stimulator of bone resorption (Progulske et al., 1984; Mattison et al., 1987), and slime, which has a strong immunosuppressive effect (Behling et al., 1979). For mucosal pathogens such as *Eik. corrodens*, the first step in the pathogenic process is the attachment or adherence of the bacterium to the mucosal surface (Gibbons & Houte, 1975). For many pathogens, including the uropathogenic *Escherichia coli* (Labigne-Roussel et al., 1985), *Bordetella pertussis* (Weiss & Hewlett, 1986), *Vibrio cholerae* (Jones & Freter, 1976), and *Salmonella typhimurium* (Jones & Richardson, 1981), the ability to adhere to host tissues has been correlated with the ability to agglutinate erythrocytes. These studies have suggested that the haemagglutinin(s) of these organisms function as adhesins during the infectious process. Ebisu & Okada (1983) showed that *Eik. corrodens* can agglutinate neuraminidase-treated erythrocytes. Neuraminidase, which is present in human saliva, is believed to potentiate bacterial adherence to human cells by exposing underlying carbohydrate receptors, known as cryptotopes, on host cells (Gibbons, 1989). In addition, *Eik. corrodens* has been shown to adhere in a similar fashion to neuraminidase-treated human buccal epithelial cells (Yamazaki et al., 1981) and guinea-pig macrophages (Miki et al., 1986). More recently, Yamazaki et al. (1988) have reported the partial purification of a lectin-like substance believed to be the haemagglutinin and adhesin of *Eik. corrodens*. This substance was characterized as a high-molecular-mass complex, which upon reduction with β-mercaptoethanol could be separated into several protein bands, suggesting either that the haemagglutinin was only partially pure, or that it was a multimeric complex linked by disulphide bonds.

With the advent of molecular biology techniques, it has become possible to identify and characterize in more detail putative virulence factors from a variety of pathogenic micro-organisms. We here report the cloning, characterization, sequencing and expression in *E. coli* of two distinct genes from the chromosome of *Eik. corrodens* ATCC 23834 which are capable of conferring upon *E. coli* JM109 the ability to agglutinate neuraminidase-treated erythrocytes.

**Methods**

**Bacterial strains, plasmids, media and growth conditions.** *Eikenella corrodens* ATCC 23834 (the type strain) was grown and maintained on blood agar plates (40 g trypticase soy agar l⁻¹ and 5%, v/v, defibrinated sheep blood). Growth was at 37 °C in an atmosphere containing 10% CO₂, and cells were transferred to fresh plates every 2-3 d. Liquid cultures were prepared using BY broth as described previously (Progulske & Holt, 1987). Purity was assessed by Gram-staining, darkfield microscopy, and culture on blood agar plates.

**Construction of genomic libraries.** Purified chromosomal DNA from *Eik. corrodens* was partially digested with *HindIII* so as to obtain fragments ranging in size from 2 to 10 kb as determined by agarose gel electrophoresis. These fragments were then ligated to dephosphorylated *HindIII*-cut pUC9 using T4 DNA ligase according to standard methods (Sambrook et al., 1989). Recombinant plasmids were used to transform *E. coli* JM109 made competent by the method of A. Das (personal communication) as described previously (Progulske-Fox et al., 1989). White colonies were picked and plated onto LB agar containing ampicillin and X-Gal to confirm the Lac⁻ phenotype. All transformants could be separated into several protein bands, suggesting either that the haemagglutinin was only partially pure, or that it was a multimeric complex linked by disulphide bonds.

**DNA preparation.** Chromosomal DNA was isolated from *Eik. corrodens* ATCC 23834 by a modification of the method of A. Das (personal communication) as previously described (Progulske-Fox et al., 1989).

**Table 1. Bacterial plasmids used in this study**

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Characteristics†</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUC series</td>
<td>ApR lacZ⁺</td>
</tr>
<tr>
<td>pVKR201</td>
<td>ApR HA⁺; pUC9 Ω(HindIII 3.6 kb insert from <em>Eik. corrodens</em> 23834)</td>
</tr>
<tr>
<td>pVKR202</td>
<td>ApR HA⁺; pUC19 Ω(HindIII 3.6 kb, orientation opposite to that of pVKR201)</td>
</tr>
<tr>
<td>pVKR203</td>
<td>ApR HA⁺; pUC18 Ω(EcoRI–HindIII 1.9 kb)</td>
</tr>
<tr>
<td>pVKR204</td>
<td>ApR HA⁺; pUC19 Ω(HindIII–EcoRI 1.7 kb)</td>
</tr>
<tr>
<td>pVKR206</td>
<td>ApR; pUC19 Ω(HindIII–KpnI 1.2 kb)</td>
</tr>
<tr>
<td>pVKR207</td>
<td>ApR HA⁺; pUC19 Ω(KpnI 0.7 kb)</td>
</tr>
<tr>
<td>pVKR210</td>
<td>ApR HA⁺; pUC18 Ω(SstI–HindII 1.2 kb)</td>
</tr>
<tr>
<td>pVKR301</td>
<td>ApR HA⁺; pUC9 Ω(HindIII 3.7 kb insert from <em>Eik. corrodens</em> 23834)</td>
</tr>
<tr>
<td>pVKR302</td>
<td>ApR; pUC19 Ω(HindIII 3.7 kb, orientation opposite to that of pVKR301)</td>
</tr>
<tr>
<td>pVKR303</td>
<td>ApR; pUC19 Ω(HindIII–HindII 1.6 kb)</td>
</tr>
<tr>
<td>pVKR304</td>
<td>ApR HA⁺; pUC18 Ω(HindIII–HindII 1.4 kb)</td>
</tr>
<tr>
<td>pVKR305</td>
<td>ApR; pUC19 Ω(HindIII–EcoRI 0.7 kb)</td>
</tr>
<tr>
<td>pVKR306</td>
<td>ApR HA⁺; pUC18 Ω(EcoRI–HindII 2.4 kb)</td>
</tr>
<tr>
<td>pVKR308</td>
<td>ApR; pUC19 Ω(EcoRI 0.6 kb)</td>
</tr>
</tbody>
</table>

† Characteristics described by Yanisch-Perron et al. (1985). All the other plasmids were produced during this work.

* Characteristics described by Yanisch-Perron et al. (1985). All the other plasmids were produced during this work.
containing recombinant plasmids were stored at -70 °C in LB broth containing ampicillin (50 μg ml⁻¹) and glycerol (20%, v/v).

Preparation of antisera. Broth-grown cells of *Eik. corrodens* ATCC 23834 were prepared as described previously (Progulske-Fox et al., 1989) and used to inoculate adult New Zealand White rabbits. Each rabbit was given a booster dose 50–60 d later. Antisera were collected prior to the first immunization and at various intervals beginning one week after the booster dose. Sera were stored at -20 °C. Polyclonal rabbit anti-*Eik. corrodens* sera were absorbed four times with *E. coli* JM109(pUC9) and the titre determined as described previously (Progulske-Fox et al., 1989).

Colonies immunoblotting assay. Ampicillin-resistant transformants containing recombinant plasmids were spotted onto LB plates containing ampicillin. After 24 h of growth, the colonies were blotted onto nitrocellulose filters (0.45 μm). *Eik. corrodens* and *E. coli* JM109(pUC9) were also spotted onto each filter as positive and negative controls respectively. Duplicate nitrocellulose filter imprts of each set of colonies were prepared and the cells on one of each pair were lysed by a 15 min exposure to chloroform vapour. Antigen-positive clones were detected as previously described (Progulske-Fox et al., 1989), but utilizing rabbit polyclonal *Eik. corrodens* antisera as the primary antibody. Clones which reacted positively in this assay were picked, regrown and restested by the same procedure.

Haemagglutination assay. Haemagglutination assays were carried out generally as described by Ebisu & Okada (1983), using V-bottomed microtitre plates (Dynatech Laboratories). *E. coli* transformants expressing *Eik. corrodens* antigens as well as an appropriate *E. coli* control were grown overnight in LB broth containing ampicillin (50 μg ml⁻¹). Cells were pelleted by centrifugation, washed twice with PBS, pH 7.2, and resuspended in one-tenth the original volume of PBS, pH 7.2. The final cellular concentration of each suspension, measured as OD₅₅₀, was made equivalent by the addition of PBS. Broth-grown *Eik. corrodens* cells were similarly washed with PBS, pH 7.2, and the concentration of the suspension adjusted to an OD₅₅₀ of approximately 0.5. Fresh human erythrocytes from a single donor were washed several times with PBS, pH 7.2, once with PBS, pH 5.8, and resuspended to a concentration of 20% (w/v) in PBS, pH 5.8. Neuraminidase type V from *Clostridium perfringens* (Sigma) was added to a concentration of 1 U ml⁻¹ and the mixture was allowed to incubate for 1 h at 37 °C with gentle agitation. The erythrocytes were then washed several times with PBS, pH 7.2, and resuspended in the same buffer at a concentration of 1% (w/v). The bacterial cell suspensions were diluted in a twofold series in PBS, pH 7.2, across the microtitre plate, leaving 0.05 ml in each well. An equal volume (0.05 ml) of washed, neuraminidase-treated erythrocytes was added to each well and gently mixed with the bacterial cells. The plates were stored for 12–14 h at 4 °C and then examined for evidence of haemagglutination.

Restriction mapping and subcloning. CsCl-gradient-purified plasmid DNA was used for all mapping experiments so as to ensure consistent and efficient digestion by restriction enzymes. Enzyme reactions were carried out under conditions recommended by the manufacturer so as to allow complete digestion and the products were analysed by agarose gel electrophoresis using standard methods (Sambrook et al., 1989). For subcloning, vector DNA was digested with the appropriate restriction enzyme(s) and purified by 0.7% agarose gel electrophoresis followed by isolation of the desired fragment from the gel using standard techniques (Silhavy et al., 1984). Fragments to be cloned were likewise purified by electrophoresis prior to ligation. Fragments with asymmetric terminal restriction-enzyme-generated ends were subcloned in either pUC18 or pUC19 so as to maintain the original orientation of the fragment with respect to the plasmid-based promoter.

Maxicell analysis. Maxicell analysis of clones and subclones was performed using the method of Sancar et al. (1979) as described by Silhavy et al. (1984). Following SDS-PAGE (12.5% separating gel) the proteins in the gel were fixed by treatment for 30 min in 2-propanol/water/acidic acid (25:65:10, by vol.). The gel was then soaked for 30 min in the fluorographic reagent Amplify (Amersham) prior to drying using a slab gel dryer (Bio-Rad). The dried gel was then subjected to autoradiography using Kodak XAR-5 film at -70 °C for 12–14 h.

Southern hybridization. DNA samples to be analysed were digested and subjected to electrophoresis in a 1.0% (w/v) agarose gel. The DNA was transferred to nylon membranes (Sigma) by the capillary alkaline transfer method (Reed & Mann, 1985). 32P-labelled probe was prepared from gel-electrophoresis-purified fragments using the Multiprime DNA Labelling Kit (Amersham) and [α-32P]dCTP (Amersham) according to the directions of the manufacturer. Hybridization was carried out according to the method of Hardy et al. (1983) at 65 °C in the presence of 10% (w/v) dextran sulphate (Sigma) and 1% (w/v) SDS for 18 h using the POR800 Hybridization Chamber (Hoeffer Scientific Instruments). The membrane was subjected to a wash of low stringency (2 × SSPE, 1% SDS) for 60 min at 65 °C followed by a high-stringency wash (0.1 × SSPE and 1% SDS) for 15 min at 65 °C. (1 × SSPE is 150 mM-NaCl, 10 mM-sodium phosphate, 1 mM-EDTA, pH 7.4.) The membrane was subjected to autoradiography for 6–12 h at -70 °C using a Lightning Plus intensifying screen (DuPont) and Kodak XAR-5 film.

DNA sequencing. Fragments to be sequenced were cloned in either pUC19 or pUC18 and the recombinant plasmids purified by centrifugation twice in a gradient of CsCl. Double-stranded sequencing was carried out by the University of Florida DNA Core Sequencing Facility using a modification of the dideoxy chain-termination method of Sanger et al. (1977), incorporating fluorescently tagged dideoxy nucleotides as described by Prober et al. (1987). Fluorescent sequencing fragments generated by the Sequenase enzyme were resolved by PAGE and the gels then scanned by the Genesis 2000 detector system (Dupont Biotechnology Systems). Where necessary, oligonucleotides 17 bases in length were synthesized by the University of Florida DNA Core Synthesis Facility and used as a primer in sequencing reactions. Sequence data were compiled and analysed using programs in the Genetics Computer Group Sequence Analysis Software Package version 6.2 (Devereux et al., 1984).

Results

Construction of a genomic library and identification of antigen-expressing clones

Approximately 1400 *E. coli* transformants containing recombinant plasmids were picked and screened for expression of *Eik. corrodens* antigens using a colony immunoblotting method. In this manner, 22 clones were found to cross-react reproducibly with polyclonal antisera raised in rabbits against *Eik. corrodens*. *E. coli* containing the vector pUC9 exhibited no reaction with the antisera in this assay. In all but three cases, whole cells of the transformants gave a reaction equivalent to that seen with chloroform-lysed cells, suggesting that in all except these clones, the cloned antigen is expressed on the surface of *E. coli*. Preliminary restriction analysis of recombinant plasmids isolated from each of these clones revealed that the majority harbour single-fragment inserts ranging in size from 1.9 to 6.2 kb. Several clones
V. K. Rao, J. A. Whitlock and A. Progulske-Fox

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>HA</th>
<th>Proteins encoded (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pVKR201</td>
<td>4+</td>
<td>31.5, 14</td>
</tr>
<tr>
<td>pVKR202</td>
<td>4+</td>
<td>31.5, 14</td>
</tr>
<tr>
<td>pVKR203</td>
<td>4+</td>
<td>14</td>
</tr>
<tr>
<td>pVKR204</td>
<td>3+</td>
<td>31.5</td>
</tr>
<tr>
<td>pVKR206</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pVKR207</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pVKR210</td>
<td>3+</td>
<td>31.5</td>
</tr>
</tbody>
</table>

Fig. 1. Localization of the haemagglutinin gene of pVKR201. The restriction map of the insert of pVKR201 and the relative positions of various subclones are shown. The large arrow indicates the position of the open reading frame. Small arrows denote the position and direction of transcription from the plasmid lac promoter as well as the putative Eikenella promoter found on this fragment. Sizes of plasmid-encoded proteins (kDa) are as determined by SDS-PAGE analysis of [35S]methionine-labelled maxicells. HA, haemagglutination activity.

contained multiple HindIII fragments of Eik. corrodens DNA.

Identification of two haemagglutinating clones

Because bacterial adhesins (haemagglutinins) are frequently major surface antigens, we tested each of the antigen-expressing clones in a standard haemagglutination assay using neuraminidase-treated human erythrocytes. In this manner, two clones were found which could agglutinate erythrocytes, while *E. coli* JM109 harbouring pUC9 possessed no such activity. To confirm this activity, the plasmid DNA was isolated from each clone and used to transform fresh cultures of *E. coli* JM109. In each case 5/5 of the new transformants tested possessed haemagglutination activity equivalent to that of the original clone, confirming that the activity is conferred on *E. coli* by the cloned *Eik. corrodens* DNA. Clone 1 harboured plasmid pVKR201, which contains a single 3.6 kb HindIII fragment. Clone 2 harboured plasmid pVKR301, which contains a single 3.2 kb HindIII fragment. By digestion with a variety of restriction enzymes, both alone and in combination, restriction maps of pVKR201 and pVKR301 were deduced; these are shown in the top portions of Figs 1 and 2 respectively. Comparison of the two restriction maps failed to reveal any obvious similarities between the two cloned DNA fragments. To investigate the regulation of expression of these genes in *E. coli*, we ligated the inserts of pVKR201 and pVKR301 into pUC19 in the reverse orientation, thereby producing pVKR202 and pVKR302 respectively. Transformants harbouring pVKR202 exhibited haemagglutination activity equivalent to the original clone. In contrast, pVKR302 was unable to confer haemagglutination activity on *E. coli* JM109 (Figs 1 and 2).

Southern hybridization

Southern hybridization analysis was used in order to assess subtle relationships, if any, between the two cloned fragments. The insert from pVKR201 was purified by agarose gel electrophoresis, labelled with 32P, and used as a probe in a standard Southern hybridization experiment. As shown in Fig. 3, under the high-stringency conditions used, there was no detectable homology to any DNA fragment from *E. coli* JM109 chromosomal DNA or to the insert from plasmid
Cloning of two E. corrodens haemagglutinins

Fig. 2. Localization of the haemagglutinin gene of pVKR301. The restriction map of the insert of pVKR301 and the relative positions of various subclones are shown. The large arrow indicates the position of the open reading frame. Small arrows denote the position and direction of transcription from the plasmid lac promoter. Other details as for Fig. 1.

Fig. 3. Southern hybridization analysis of the two haemagglutinin clones from Eik. corrodens. The 3.6 kb insert of pVKR201 was used as the probe. Both the agarose gel (a) and the corresponding autoradiograph (b) are shown. Lanes: 1 and 6, λ DNA cut with HindIII and EcoRI; 2, Eik. corrodens ATCC 23834 chromosomal DNA cut with HindIII; 3, E. coli JM109 chromosomal DNA cut with HindIII; 4, pVKR201 isolated from E. coli clone 1 digested with HindIII; 5, pVKR301 isolated from E. coli clone 2 digested with HindIII.

pVKR301. There was a 3.6 kb hybridizing band in the lane containing HindIII-digested chromosomal DNA from E. corrodens (Fig. 3, lane 2), thereby confirming the origin of the cloned fragment. These data indicated that there was no sequence similarity between the two cloned haemagglutinin genes.

Maxicell analysis

In order to identify the proteins encoded by these fragments, we used the maxicell technique to selectively label plasmid-encoded proteins. As shown in Fig. 4, lane 3, pVKR201 encodes three proteins which are expressed in E. coli. Proteins A and B migrate very close together and each has a molecular mass of approximately 14 kDa as estimated by SDS-PAGE. Protein C is significantly larger, with an estimated molecular mass of 31-5 kDa. Interestingly, this protein is visible as a very intense band in Coomassie-blue-stained SDS-PAGE gels of lysates of clone 1, while proteins A and B are not clearly visible in such gels (data not shown). In contrast, pVKR301 encodes two proteins not seen in E. coli containing pUC9 alone (Fig. 4, lane 4). Protein 1 has a molecular mass of 93 kDa as estimated by SDS-PAGE, while protein 2 has a molecular mass of 17 kDa. Neither of these proteins
could be visualized on Coomassie-blue stained SDS-PAGE gels of lysates of clone 2 (data not shown). Maxicell analysis of *E. coli* harbouring pVKR202 revealed the production of identical proteins to those seen with pVKR201, suggesting that all three proteins are expressed in *E. coli* under the control of an *Eikenella* promoter(s) (Fig. 1). As was suggested by the haemagglutination data, maxicells of *E. coli* harbouring pVKR302 produced no proteins other than those seen with pUC19 alone as visualized by maxicell analysis. These data suggested that the 31.5 kDa protein is the cloned haemagglutinin and that DNA contained in the *KpnI–EcoRI* fragment is required for activity of the haemagglutinin. This fragment, however, is not sufficient for activity since pVKR207, which contains a 0.7 kb *KpnI–KpnI* fragment that includes the *KpnI–EcoRI* fragment, does not confer haemagglutination activity on *E. coli*. Since the orientation of the fragments in the original clone was preserved in the subclones, the data further suggest that transcription occurs in a direction opposite to that from the plasmid lac promoter. If transcription were to occur in phase with the lac promoter, deletion of the *KpnI–EcoRI* fragment would represent deletion of the 3' end of the gene. One would therefore expect to obtain a fusion (which may or may not be functional) with the β-galactosidase gene just downstream of the polylinker of pUC19. No such protein is apparently encoded by pVKR206. Along the same lines, since pVKR207 should contain the 5' end and upstream regulatory elements of the gene encoding the 31.5 kDa protein, one would expect to see production of a truncated protein, the size of which would depend upon the amount of the open reading frame contained on the *KpnI–EcoRI* fragment. Although such a band is not visible on Coomassie-blue stained gels, maxicells of transformants harbouring pVKR207 do produce a faint band of low molecular mass (< 10 kDa) not seen with pUC19 alone. pVKR203 contains a 1.9 kb *EcoRI–HindIII* insert representing the right half of the original insert as depicted in Fig. 1. Maxicells of transformants harbouring this plasmid produced a 14 kDa protein. A detailed characterization of this gene as well as that encoding the third cloned protein on this fragment is reported in the following paper (Rao & Progulske-Fox, 1993).

**Localization of the haemagglutinin gene of clone 1**

In order to identify and localize the haemagglutinin gene carried on the insert of pVKR201, the haemagglutination pattern of a series of subclones (Table 1) was determined. Fragments were cloned into appropriate sites in either pUC18 or pUC19 so as to maintain the orientation, with respect to the plasmid promoter, in which the fragment is found in the original clone. Each subclone was tested in the haemagglutination assay and analysed by the maxicell technique. These data, and the relative positions of each subcloned fragment in the original clone, are shown schematically in Fig. 1. pVKR204, containing a 1.7 kb *HindIII–EcoRI* fragment from the lefthand portion of the insert, as depicted in Fig. 1, conferred haemagglutination activity on *E. coli* JM109, and also directed the production of a 31.5 kDa protein as visualized both by Coomassie-blue-stained SDS-PAGE gels of lysates and by maxicells harbouring this plasmid. In contrast, transformants containing plasmid pVKR206, obtained by deletion of the terminal *KpnI–EcoRI* fragment (0.5 kb) from pVKR204, were unable to agglutinate erythrocytes. Moreover, transformants harbouring pVKR206 produced no proteins other than those seen with pUC19 alone as visualized by maxicell analysis. These data suggested that the 31.5 kDa protein is the cloned haemagglutinin and that DNA contained in the *KpnI–EcoRI* fragment is required for activity of the haemagglutinin. This fragment, however, is not sufficient for activity since pVKR207, which contains a 0.7 kb *KpnI–KpnI* fragment that includes the *KpnI–EcoRI* fragment, does not confer haemagglutination activity on *E. coli*. Since the orientation of the fragments in the original clone was preserved in the subclones, the data further suggest that transcription occurs in a direction opposite to that from the plasmid lac promoter. If transcription were to occur in phase with the lac promoter, deletion of the *KpnI–EcoRI* fragment would represent deletion of the 3' end of the gene. One would therefore expect to obtain a fusion (which may or may not be functional) with the β-galactosidase gene just downstream of the polylinker of pUC19. No such protein is apparently encoded by pVKR206. Along the same lines, since pVKR207 should contain the 5' end and upstream regulatory elements of the gene encoding the 31.5 kDa protein, one would expect to see production of a truncated protein, the size of which would depend upon the amount of the open reading frame contained on the *KpnI–EcoRI* fragment. Although such a band is not visible on Coomassie-blue-stained gels, maxicells of transformants harbouring pVKR207 do produce a faint band of low molecular mass (< 10 kDa) not seen with pUC19 alone. pVKR203 contains a 1.9 kb *EcoRI–HindIII* insert representing the right half of the original insert as depicted in Fig. 1. Maxicells of transformants harbouring this plasmid produced a 14 kDa protein. A detailed characterization of this gene as well as that encoding the third cloned protein on this fragment is reported in the following paper (Rao & Progulske-Fox, 1993).

**Localization of the haemagglutinin gene of clone 2**

As with clone 1, a series of subclones of clone 2 was constructed (see Table 1) and tested for haemagglutination activity and studied by maxicell analysis in order
Cloning of two E. corrodens haemagglutinins

Fig. 5. Nucleotide sequence of the 1675 bp insert of pVKR204 containing the gene for a 31.5 kDa haemagglutinin from Eik. corrodens. The deduced amino acid sequence is shown below the nucleotide sequence. Potential -35 and -10 promoter sequences as well as a potential ribosome-binding site (S-D) are indicated above the sequence. Lines above the sequence denote inverted repeats which may function in transcription termination.

to localize the haemagglutinin gene on the cloned fragment. Once again, fragments were subcloned in either pUC19 or pUC18 so as to maintain their original orientation. Orientation was a particularly important factor in the case of clone 2 because of our earlier results which suggested that expression was dependent upon promoter sequences in the vector. These data, along with the relative positions of each subcloned fragment on the original insert are shown schematically in Fig. 2. Transformants containing pVKR306, which harbours a 2.4 kb EcoRI-HindIII fragment from the right side of the insert as depicted in Fig. 2, were capable of haemagglutination. In addition, maxicells prepared from these transformants were able to agglutinate erythrocytes and produced the 17 kDa protein. Deletion of a terminal 1.0 kb EcoRI-EcoRV fragment from the insert of pVKR306 produced plasmid pVKR304. Transformants containing this plasmid were also able to agglutinate erythrocytes and produced the 17 kDa protein, as determined with maxicells. Plasmids pVKR303, 305 and 308, containing inserts from various portions of the left side of the insert, were all unable to confer haemagglutination activity on E. coli and directed the production of no proteins other than those seen with pUC19 alone. These data suggested that the 17 kDa protein is the cloned haemagglutinin and that the gene encoding this protein is located within the 1.4 kb EcoRV-HindIII fragment. Since expression of both proteins appears to be under the control of the lac promoter, the two genes must be transcribed from left to right as drawn in Fig. 2. The gene encoding the 93 kDa protein must be located to the left of the haemagglutinin gene since there is insufficient DNA within the 1-4 kb EcoRV–HindIII fragment. Since expression of both proteins appears to be under the control of the lac promoter, the two genes must be transcribed from left to right as drawn in Fig. 2. The gene encoding the 93 kDa protein must be located to the left of the cloned haemagglutinin gene since there is insufficient DNA within the 1-4 kb.
EcoRV–HindIII fragment to encode both proteins. Production of a C-terminal fusion protein is likewise unlikely since the 93 kDa protein is not seen with either pVKR306 or pVKR304. A detailed characterization of the 93 kDa protein will be reported elsewhere.

**Nucleotide sequence of the haemagglutinin gene from clone 1**

The complete nucleotide sequence of the 1.7 kb insert of pVKR204 found to contain the haemagglutinin gene (Table 1) was determined and is shown in Fig. 5. The G+C content of this fragment was 50.4 mol%, which agrees well with that previously reported (54–56 mol%) for the chromosome of *Eik. corrodens* (Jackson & Goodman, 1984). We identified a major open reading frame (ORF) on this fragment beginning at position 334 and extending 903 nucleotides to position 1237. Transcription occurs in a direction opposite to that of the lac promoter, as suggested by the experimental data and as diagrammed in Fig. 1. The protein encoded by this ORF is composed of 300 amino acid residues and has a deduced molecular mass of 33.5 kDa. This value is consistent with that (31.5 kDa) observed by SDS-PAGE and maxicell analysis (Fig. 4). The proposed translation start site is preceded by a potential ribosome-binding site, AGGG, at positions −10 to −13 (Gold et al., 1981). From the subcloning data presented above, it was clear that the *Eikenella* promoter which regulates expression of the haemagglutinin genes is included in the cloned fragment and that this promoter is functional in *E. coli*. Upon examination of regions of the sequence upstream of the ORF, we found that from position 267 to 272, there is a hexanucleotide sequence, TATATT, which differs by only a single nucleotide from the consensus sequence, TATAAT, for the Pribnow or TATA box portion of prokaryotic promoters (Rosenberg & Court, 1979). In addition, from position 229 to 234, there is a hexanucleotide sequence, TTGATA, which differs by a single nucleotide from the consensus sequence, TTGACA, for the −35 region of prokaryotic promoters (Rosenberg & Court, 1979). While it is likely that these two regions comprise the promoter of this gene, precise and definitive localization of the upstream regulatory elements will await the results of primer extension and/or DNA footprinting studies. Beginning 17 nucleotides downstream of the TAA termination codon, at position 1253, and extending to position 1290, is a potential stem–loop structure (lines drawn above sequence in Fig. 5), which may function in termination of transcription (Rosenberg & Court, 1979). Given these data, we constructed an additional subclone, pVKR210 (Fig. 1) in which a 1.2 kb SspI–SstI fragment, extending from position 228 to 1418, was ligated to the *HincII–SstI* sites of pUC18. This subclone contains the entire open reading frame of the haemagglutinin with little flanking DNA other than that implicated in regulation of expression, i.e. putative promoter and termination sequences. In addition, the fragment was ligated in an orientation opposite that of the lac promoter. As expected, transformants harbouring pVKR210 were able to agglutinate erythrocytes. Moreover, large amounts of the 31.5 kDa protein were clearly visible in SDS-PAGE gels of lysates of these transformants. These data provide additional evidence as to the proposed promoter sequences and direction of transcription of the haemagglutinin.

Secreted proteins, including those which ultimately remain associated with the bacterial outer membrane, generally possess a leader sequence consisting of 15–20 hydrophobic amino acids (Inouye & Halegoua, 1980). Interestingly, the leader resulting from initiation at position 334 as proposed here is remarkably hydrophilic, with 13/20 of the amino acids being polar residues. This potential leader sequence would therefore not be similar to any known signal sequences. Confirmation of the proposed start site and hence the composition of the leader sequence will, of course, require N-terminal sequencing of the protein. Beginning at residue 200 is a hydrophobic stretch where 17/21 of the amino acids are nonpolar. The average hydrophyt of this region as determined by the method of Kyte & Doolittle (1982) is +1.6, making it consistent with properties of the membrane-spanning region of a protein. Such a stretch could therefore be important in the anchoring of this protein in the outer membrane of the cell. A computer search of the EMBL and GenBank databases failed to reveal any significant relationship between the haemagglutinin of clone 1 and any previously sequenced proteins.

**Nucleotide sequence of the haemagglutinin gene from clone 2**

The complete nucleotide sequence of the insert of pVKR304, which was found to contain the haemagglutinin gene, was determined and is shown in Fig. 6. The G+C content of this fragment was 50.6 mol%, which is remarkably close to that found for the insert of pVKR204 containing the haemagglutinin gene from clone 1. We identified a major ORF on this fragment, beginning at position 489 and extending 639 nucleotides to position 1128. The protein encoded by this ORF is 212 amino acid residues in length with a deduced molecular mass of 23712 Da and isoelectric point of 8.25. The deduced molecular mass compares fairly well with that (17 kDa) obtained by maxicell analysis (Fig. 4). Prior to the start site for translation is a potential Shine–Dalgarno
sequence, AGCGGCA, from position -12 to -6 (Gold et al., 1981). There is a second potential start site beginning at position 510, which would result in a protein six amino acids shorter. However, this start codon is not preceded by a potential ribosome-binding site. Therefore, translation is more likely to begin at position 489. Experimental evidence appears to indicate that the haemagglutinin of clone 2 is expressed under the control of the plasmid lac promoter. We identified no sequences with similarity to the consensus sequence for the E. coli o7O promoter. Downstream of the TAA termination codon are two sequences which can potentially form stem-loop structures. The first occurs at position 1143 and extends to position 1177. The second begins at position 1206 and extends 41 nucleotides to position 1246. One or both of these structures may function in rho-independent termination of transcription. As with the haemagglutinin of clone 1, examination of the N-terminal sequence of the haemagglutinin as derived from the nucleotide sequence failed to reveal any characteristics similar to known signal peptides. A computer search of the EMBL and GenBank databases (releases through March 1991) failed to reveal any significant relationship between the haemagglutinin of clone 2 and any previously sequenced proteins. Moreover, computer-assisted comparison of the amino acid sequences of the haemagglutinins from clones 1 and 2 failed to reveal any significant regions or domains of homology between the two proteins.

Discussion

It has been shown previously that Eik. corrodens can agglutinate human erythrocytes, and it has been hypothesized that as with other mucosal pathogens, haemagglutination is strongly correlated with adherence
(Ebisu & Okada, 1983). In order to facilitate the study of the haemagglutinin(s) and other putative virulence factors of this fastidious micro-organism, we constructed a library of the chromosomal DNA of *Eik. corrodens* in *E. coli* JM109 using the plasmid vector pUC9.

Two clones were identified which could agglutinate erythrocytes. On the basis of different restriction maps, differences in the number and sizes of proteins encoded, and a lack of homology by Southern hybridization, we concluded that two distinct DNA fragments had been cloned and that *Eikenella* therefore possesses at least two proteins capable of agglutinating erythrocytes. This result was somewhat surprising since previous work (Yamazaki *et al.*, 1988) had indicated that this organism possessed only one such protein. Yamazaki *et al.* (1988) reported the partial purification of a high-molecular-mass haemagglutinating factor from *Eik. corrodens*. Reduction of this protein preparation, which migrated as a single band in SDS-PAGE, with β-mercaptoethanol resulted in the production of several protein bands. No details as to the sizes or relative proportions of these smaller proteins were given. It is nevertheless tempting to speculate that the two proteins which we have cloned are subunits of a much larger complex on the *Eikenella* surface, with two of the proteins observed by Yamazaki *et al.* (1988) being identical to those described here. Alternatively, these two proteins could occupy completely distinct sites on the bacterial outer membrane and might therefore carry out distinct functions (i.e. have different receptor molecules) in the process of adherence.

In order to characterize the two haemagglutinin genes further, we determined the nucleotide sequences of the two cloned fragments. As suggested by experimental evidence, we could detect no relationship between the two cloned elements at either the nucleotide or amino acid levels. Moreover, these two haemagglutinating proteins both appear to possess novel sequences, as there was no significant similarity between either gene and any previously sequenced proteins in the EMBL and GenBank databases, including haemagglutinins identified in other bacteria.

Experimental evidence suggested that the haemagglutinin gene from clone 1 is expressed under the control of an *Eikenella* promoter. Sequences immediately upstream of the haemagglutinin gene from clone 1 show striking homology to the consensus sequences for bacterial promoters (Rosenberg & Court, 1979). Moreover, the translation start site is preceded by a fairly good ribosome-binding site. Taken together, these characteristics would suggest that this protein should be efficiently expressed in *E. coli*. Indeed, this protein is clearly visible as a broad, intense band in Coomassie-blue-stained SDS-PAGE gels of lysates of transformants harbouring this gene. It is interesting that a heterologous promoter can direct the production of much higher levels of a protein than are often seen with foreign proteins expressed with an *E. coli* promoter. For example, the haemagglutinin of clone 2, which requires the lac promoter for expression, is not visible on Coomassie-blue-stained SDS-PAGE gels. Codon usage is also believed to play a role in the level of protein expression (Grosjean & Fiers, 1982). A number of codons are used infrequently in genes of highly expressed proteins in *E. coli* (Grosjean & Fiers, 1982), including AUA, CGG, AGA, AGG, CUA, CGA, GGA and GGG. In the gene encoding the haemagglutinin from clone 1, with the exception of the arginine codon, CGG, each of these codons was the most infrequently used of those available for the particular amino acid. These findings are consistent with a high level of expression in *E. coli*. In contrast, of the six possible codons encoding arginine, CGG is the one preferentially used (6/14 Arg codons). The use of this codon, which is believed to be an especially minor species in *E. coli*, may be important in the modulation of expression of this protein. However, it is also possible that *Eikenella* produces significant amounts of the tRNA which recognizes CGG and the use of this codon in this gene may therefore reflect a difference in the distribution of tRNA species between *Eikenella* and *E. coli*.

The haemagglutinin gene from clone 2 requires the lac promoter for expression in *E. coli*. Examination of sequences upstream of this gene failed to reveal regions with clear homology to bacterial promoter consensus sequences. It is likely therefore that the haemagglutinin gene is read as part of a polycistrionic message with one or more upstream genes. Because the upstream gene (which encodes a 93 kDa protein) also appears to require the lac promoter for expression, it is possible either that the *Eikenella* promoter was not included in the cloned fragment or that the promoter cannot be used by the transcriptional apparatus of *E. coli*. The second possibility would indicate that *Eikenella* possesses more than one class of promoter, only some of which will function in *E. coli*. Such a scenario is not without precedent. Promoters of the oral pathogen *Porphyromonas (Bacteroides) gingivalis* appear to be equally heterogeneous with regard to function in *E. coli* (Progulske-Fox *et al.*, 1989). Examination of codon usage in the haemagglutinin gene from clone 2 revealed several important differences from that observed with the haemagglutinin gene from clone 1. There was no clear bias against the use of the eight rare codons of *E. coli* (Grosjean & Fiers, 1982). Three of these, CGG, AGA and AGG, are used frequently in this gene. Interestingly, however, in the case of the four threonine codons, in which A and C occupy the first two positions, there is a clear preference for G or C in the third position (17/17
cases). Similarly, of the four alanine codons, there is a strong bias against the use of GCA (0/19 cases). The significance of this pattern is unclear, but may once again reflect differences between the tRNA content of Eikenella and that of E. coli.

In both cases, the G+C content of the Eikenella insert was consistent with that determined previously for the chromosome of Eik. corrodens (Jackson & Goodman, 1984). This suggests that these two genes are either purely of Eikenella origin or were acquired by this species long ago in its evolutionary past. The fact that no homology could be found between these proteins and any others previously sequenced further suggests an Eikenella origin for both of them. However, this lack of homology may also be a reflection of our limited knowledge, at the molecular level, of other Gram-negative anaerobes and facultative aerobes to which Eikenella may be closely related.

It is especially significant that the two proteins described here are functionally expressed (as haemagglutinins) in E. coli, to which Eikenella is not closely related. Surprisingly, upon examination of the amino acid sequences of the two proteins as derived from the nucleotide sequence, we could not detect a pattern consistent with a conventional signal or leader sequence. These proteins must nevertheless be successfully transported to the outer surface of the outer membrane of the cell since they are apparently able to interact effectively with the erythrocyte membrane. The subtle signals encoded in the N-terminus of each protein must therefore be recognized by the E. coli machinery involved in the transport and secretion of proteins.

The two genes described in this report represent the first to be cloned and sequenced from the chromosome of Eik. corrodens. In addition to showing that genes from this bacterium can be efficiently expressed in E. coli, these studies will assist in the complete characterization of these unique proteins at the molecular level. Furthermore, by facilitating the construction of selective mutations in each protein, the roles of each in the pathogenic process can be assessed.

This work was supported in part by Public Health Service grant DE-07117 from the National Institute for Dental Research and by NIDR BSRG grant RR05728. V.R. is the recipient of NIH predoctoral training grant DE-07200.

The University of Florida DNA Sequencing Core Laboratory was supported by Biomedical Research grant 110246212 from the University of Florida Division of Sponsored Research and the Interdisciplinary Center for Biotechnology Research.

References


Jones, G. W. & Richardson, L. A. (1981). The attachment to, and


