Antigen 43, the major phase-variable protein of the *Escherichia coli* outer membrane, can exist as a family of proteins encoded by multiple alleles

Andrew J. Roche, Joanna P. McFadden and Peter Owen

Abbreviations: Agn43, antigen 43; Dam, deoxyadenosine methylase; EPEC, enteropathogenic *E. coli*; DIG, digoxigenin.

The GenBank accession numbers for the sequences reported in this paper are AF233271–AF233273.

INTRODUCTION

A number of years ago, workers in this laboratory identified and partially characterized a prominent protein termed antigen 43 (Ag43) in the outer membrane of *Escherichia coli* ML308-225 (Owen, 1986; Owen et al., 1987, 1996; Caffrey & Owen, 1989). This strain is a derivative of the human intestinal isolate ML3 and has been extensively used in the study of transport phenomena and membrane architecture (Kaback, 1971, 1976; Owen & Kaback, 1978; Owen, 1983). Ag43 is species-specific, is arguably the most abundant phase-variable surface antigen in *E. coli*, and exists in copy numbers exceeding 50,000 per cell. It exists in situ as a hetero-oligomeric complex composed of two proteinaceous subunits, α43 and β43, present in 1:1 stoichiometry. β43 is a heat-modifiable integral outer-membrane protein. In contrast, the α43 subunit has pronounced surface expression, is bound to the surface via non-covalent interaction with β43 and can be selectively detached from the outer membrane by brief heating to 60 °C (Owen, 1986; Owen et al., 1987, 1996; Caffrey & Owen, 1989). Ag43 has been shown to play a role in autoaggregation and colony morphology of certain *E. coli* strains and may be implicated in cell adhesion and biofilm formation (Diderichsen, 1980; Owen et al., 1996; Henderson et al., 1997a; Hasman et al., 1999, 2000; Danese et al., 2000).

The gene (agn43 or flu) encoding Ag43 has been sequenced from both *E. coli* K-12 and ML308-225 strains (Blattner et al., 1997; Henderson & Owen, 1999) and is located on the K12 chromosome between min 44.6 and min 44.8. The primary translation product of *agn43* is processed via an N-terminal signal peptidase and by internal cleavage to generate the mature α43 [predicted

---

`agn43` encodes a major phase-variable outer-membrane protein, antigen 43 (Ag43), involved in autoaggregation of *Escherichia coli* cells. The gene is present in single copy on the chromosome of *E. coli* K-12. In contrast, Southern hybridization and gene inactivation studies demonstrate that control producer strain *E. coli* ML308-225 possesses duplicate copies of *agn43* (*agn43A* and *agn43B*). Construction and analyses of single and double knockout mutants clearly show that both alleles are capable of expressing antigen in a phase-variable manner, with observed differences in the ON/OFF switch frequencies appearing to favour expression of Ag43B under conditions of normal laboratory growth. Comparative analysis of *agn43A* and *agn43B* gene sequences revealed 98% identity at the nucleotide and predicted protein levels, with differences in the protein sequence of the surface-expressed α43 subunit altering the surface probability of one of the predicted epitopes. Analysis of a panel of enteropathogenic *E. coli* strains by Southern hybridization using *agn43*-specific gene probes provided strong evidence for the presence of varying numbers of *agn43* alleles within clinical isolates. Taken together, the results indicate the presence of a family of distinct Ag43 proteins encoded by multiple chromosomal alleles.

Keywords: antigen 43, *Escherichia coli*, phase variation, multiple alleles
The chromosome of E. coli was digested in both wild-type (E. coli) ML308-225 and agn43 mutants, which were isolated according to standard procedure (Henderson et al., 1998) with α43 representing the passenger domain involved in autoaggregation and β43 the translocation domain organized as an 18-stranded β-barrel pore (Henderson & Owen, 1999). Reversible phase switching of agr43 is regulated by a novel mechanism involving competition between deoxyadenosine methylase (Dam) and OxyR (a LysR-type transcriptional activator/repressor) for three unmethylated 5′-GATC-3′ sites in the regulatory region of the gene. Thus, in contrast to the phase-ON/OFF state observed for parental strains, dam mutants are locked-OFF for Ag43 expression whereas oxyR mutants are locked-ON (Henderson et al., 1997b; Henderson & Owen, 1999; Haagmans & van der Woude, 2000).

Although agr43 is known to be present in single copy on the chromosome of E. coli K-12 (Blattner et al., 1997), the situation in wild-type E. coli strains is less clear. In this context, previous analysis of a panel of enteropathogenic E. coli (EPEC) strains by immuno-
fluorescence microscopy and Western immunoblotting revealed the presence of phase-variable outer-membrane proteins that cross-reacted with anti-α43 antiserum (Owen et al., 1996). Significantly, strains were identified that expressed multiple α43-like subunits in the M<sub>r</sub> 54000–60000 and M<sub>r</sub> 94000 ranges. Cross-reactive subunits in the M<sub>r</sub> 54000–60000 range displayed properties reminiscent of native α43 and could be released from outer membranes when heated at 60 °C. In contrast, those of M<sub>r</sub> 94000 displayed features anticipated of a hypothetical uncleaved (covalently bonded) α43β43 monomer. Based on these data, it has been suggested that Ag43 may be a member of a family of related phase-variable outer-membrane proteins (Owen et al., 1996).

In this study, we demonstrated that E. coli ML308-225 possesses duplicate copies of agr43, both of which are subject to reversible phase variation. Sequencing and Western blotting revealed the genes to be 98% identical at both nucleotide and amino acid levels and to express size-variable subunits. Furthermore, we showed by Southern blot analysis of EPEC strains known to express variable outer-membrane proteins (Owen et al., 1996)., 1999) with α43 and β43 the translocation domain organized as an 18-stranded β-barrel pore (Henderson & Owen, 1999).

**METHODS**

**Bacterial strains, plasmids and growth conditions.** The bacterial strains and plasmids used in this study are listed in Table 1. All strains were routinely grown aerobically at 37 °C on Luria–Bertani agar or Luria–Bertani broth supplemented with ampicillin (100 µg ml<sup>−1</sup>), kanamycin (50 µg ml<sup>−1</sup>), nalidixic acid (15 µg ml<sup>−1</sup>) and tetracycline (10 µg ml<sup>−1</sup>) where appropriate (Lech & Brent, 1987).

**DNA manipulations and genetic techniques.** Total genomic DNA (Wilson, 1987) and plasmid DNA (Felicello & Chinali, 1993) were isolated according to standard protocol. PCR-based DNA amplifications (Kramer & Coen, 1987) and general DNA analyses and manipulations (Sambrook et al., 1989) were conducted by established procedures. Southern blotting was performed as described by Sambrook et al. (1989) and involved Vacu-blot (Vacugene; LKB) transfer of electrophoretically separated restricted DNA fragments (prepared by overnight incubation of 10 µg DNA with 20 U of appropriate restriction enzyme) onto a nylon membrane. High-stringency probing was carried out at 68 °C overnight using digoxigenin (DIG)-labelled DNA probes (Roche) at a concentration of 5 ng ml<sup>−1</sup>. Washing steps were conducted at both room temperature and at 68 °C according to the manufacturer’s recommendations. Conjugations involved spreading 100 µl aliquots of overnight cultures of donor and recipient strains onto non-selective agar and incubating resultant plates at 37 °C for 6 h. Bacteria were then harvested and resuspended in sterile PBS, and serial dilutions were plated onto selective agar prior to overnight incubation at 37 °C. Serum agglutination, routinely used to confirm the origin of putative ML308-225 transconjugants, involved emulsifying selected colonies with a loopful of O13 antiserum. E. coli strains ML308-225 (O13 : O68 : H<sup>+</sup>) and NCTC 9044 (O44) were used as positive and negative controls, respectively.

**Genetic nomenclature.** agr43 alleles were labelled alphabetically in order of decreasing size of their respective EcoRV/SphI restriction fragments. Ag43 proteins were assigned the same letter as the encoding allele.

**Construction of agr43 fragments and DNA probes.** Selected fragments of the agr43 coding region (F1–F5; see Fig. 1) were used in the construction of knockout mutants and of probes for Southern hybridization experiments. The following forward and reverse oligonucleotide primers (Sigma-Genosys) were selected to hybridize to selected agr43 sequences and containing, where appropriate, engineered 5′-Sac1, 5′-EcoRI, 3′-Kpn1 and 3′-SalI restriction enzyme cleavage sites (underlined) were used to amplify agr43 fragments F1 and F3–F5: F1 (551 bp), 5′-CCTGAGCTCTGCTGGCTGCTAGCATGTTTGTGC-3′ (forward) and 5′-GGGGTACCAGTCGCTGATTTTCCGG-3′ (reverse); F3 (400 bp), 5′-GATAGGTTATCATGTTGTTG-3′ (forward) and 5′-CCCGGAACCTCATCCTTGTC-3′ (reverse); F4 (569 bp), 5′-CCGAACTCCAGGGAACGTCGTTGC-3′ (reverse); F5 (1065 bp), 5′-ACGGTAAATGGCCGACTGTT-3′ (forward) and 5′-CCGGCAACCTCTGTTCTCATC-3′ (reverse).

Total genomic DNA from E. coli ML308-225 and E. coli BD1302 was used as templates for construction of F5, and F1, F3 and F4, respectively. F2 (see Fig. 1) was generated by HinCII digestion of F5 followed by gel purification of the desired 252 bp product. agr43 and plasmid probes for Southern hybridization experiments were generated by DIG-labelling fragments F1–F5 and linearized vector pJP5608, respectively, according to the manufacturer’s instructions (Roche).

**Construction of plasmids and knockout mutants.** Amplified fragment F1 (see Fig. 1) was digested overnight with KpnI and SacI and subsequently ligated with KpnI/SacI-digested prip suicide vector pJP5603 or pJP5608 to give plasmid constructs designated pAJR003 (3.7 kb) or pAJR005 (7.2 kb), respectively. In a similar manner, EcoRI/SalI-digested fragment F4 was ligated with appropriately digested prip suicide vector pGP704 to generate pAJR004 (4.2 kb).

Single and double knockout mutants in agr43 were constructed in both wild-type (oxyR<sup>+</sup>) and oxyR backgrounds by
Table 1. E. coli strains and plasmids used

<table>
<thead>
<tr>
<th>Strain/plasmid</th>
<th>Relevant characteristic*</th>
<th>Reference/source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AJR1</td>
<td>ML308-225 agn43B::pAJR003 (Km&lt;sup&gt;+&lt;/sup&gt;)</td>
<td>This study</td>
</tr>
<tr>
<td>AJR2</td>
<td>ML308-225 agn43A::pAJR004 (Ap&lt;sup&gt;+&lt;/sup&gt;)</td>
<td>This study</td>
</tr>
<tr>
<td>AJR3</td>
<td>AJR1 agn43A::pAJR004 (Km&lt;sup&gt;+&lt;/sup&gt; Ap&lt;sup&gt;+&lt;/sup&gt;)</td>
<td>This study</td>
</tr>
<tr>
<td>AJR4</td>
<td>ML308-225 oxyR agn43B::pAJR005 (Km&lt;sup&gt;+&lt;/sup&gt; Tc&lt;sup&gt;+&lt;/sup&gt;)</td>
<td>This study</td>
</tr>
<tr>
<td>AJR5</td>
<td>ML308-225 oxyR agn43A::pAJR004 (Km&lt;sup&gt;+&lt;/sup&gt; Ap&lt;sup&gt;+&lt;/sup&gt;)</td>
<td>This study</td>
</tr>
<tr>
<td>AJR6</td>
<td>AJR5 agn43B::pAJR005 (Km&lt;sup&gt;+&lt;/sup&gt; Tc&lt;sup&gt;+&lt;/sup&gt; Ap&lt;sup&gt;+&lt;/sup&gt;)</td>
<td>This study</td>
</tr>
<tr>
<td>BD1302</td>
<td>metB Δ(ppc–argECBH) oxyR relA spoT (Δ&lt;sup&gt;+&lt;/sup&gt;) Hfr P4X</td>
<td>Diderichsen (1980)</td>
</tr>
<tr>
<td>ML308-225</td>
<td>lacZ mobRP4 oriR6K</td>
<td>Owen et al. (1987)</td>
</tr>
<tr>
<td>ML308-225 oxyR</td>
<td>ML308–225 ΔoxyR::km (Δ&lt;sub&gt;225&lt;/sub&gt; oxyS–galK) (Km&lt;sup&gt;+&lt;/sup&gt;)</td>
<td>Henderson &amp; Owen (1999)</td>
</tr>
<tr>
<td>S17.1 i pir</td>
<td>thi thi leu tonA lacY supE recA::RP4-2 Tc:: Mu i pir R6K Km&lt;sup&gt;+&lt;/sup&gt;</td>
<td>S. Yashuda†</td>
</tr>
<tr>
<td>NCTC 8007</td>
<td>O111, anti-α&lt;sup&gt;+&lt;/sup&gt; CPs of M&lt;sub&gt;s&lt;/sub&gt; 94000 and 54500</td>
<td>Owen et al. (1996)</td>
</tr>
<tr>
<td>NCTC 8603</td>
<td>O55, no anti-α&lt;sup&gt;+&lt;/sup&gt; CPs</td>
<td>Owen et al. (1996)</td>
</tr>
<tr>
<td>NCTC 8621</td>
<td>O86, no anti-α&lt;sup&gt;+&lt;/sup&gt; CPs</td>
<td>Owen et al. (1996)</td>
</tr>
<tr>
<td>NCTC 8622</td>
<td>O126, anti-α&lt;sup&gt;+&lt;/sup&gt; CP of M&lt;sub&gt;s&lt;/sub&gt; 54500</td>
<td>Owen et al. (1996)</td>
</tr>
<tr>
<td>NCTC 9026</td>
<td>O26, anti-α&lt;sup&gt;+&lt;/sup&gt; CPs of M&lt;sub&gt;s&lt;/sub&gt; 94000, 60000 &amp; 56000</td>
<td>Owen et al. (1996)</td>
</tr>
<tr>
<td>NCTC 9040</td>
<td>O44, anti-α&lt;sup&gt;+&lt;/sup&gt; CPs of M&lt;sub&gt;s&lt;/sub&gt; 94000 and 60000</td>
<td>Owen et al. (1996)</td>
</tr>
<tr>
<td>NCTC 9114</td>
<td>O114, anti-α&lt;sup&gt;+&lt;/sup&gt; CP of M&lt;sub&gt;s&lt;/sub&gt; 58000</td>
<td>Owen et al. (1996)</td>
</tr>
<tr>
<td>NCTC 9708</td>
<td>O128, anti-α&lt;sup&gt;+&lt;/sup&gt; CP of M&lt;sub&gt;s&lt;/sub&gt; 54500</td>
<td>Owen et al. (1996)</td>
</tr>
<tr>
<td>NCTC 10089</td>
<td>O142, anti-α&lt;sup&gt;+&lt;/sup&gt; CP of M&lt;sub&gt;s&lt;/sub&gt; 94000</td>
<td>Owen et al. (1996)</td>
</tr>
<tr>
<td>NCTC 10863</td>
<td>O18ac, anti-α&lt;sup&gt;+&lt;/sup&gt; CPs of M&lt;sub&gt;s&lt;/sub&gt; 57500 and 56000</td>
<td>Owen et al. (1996)</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAJR003</td>
<td>pJP5603 containing 551 bp agn43 fragment F1</td>
<td>This study</td>
</tr>
<tr>
<td>pAJR004</td>
<td>pGP5004 containing 569 bp agn43 fragment F4</td>
<td>This study</td>
</tr>
<tr>
<td>pAJR005</td>
<td>pJP5608 containing 551 bp agn43 fragment F1</td>
<td>This study</td>
</tr>
<tr>
<td>pGP704</td>
<td>mobRP4 oriR6K (Ap&lt;sup&gt;+&lt;/sup&gt;)</td>
<td>S. Yashuda†</td>
</tr>
<tr>
<td>pJP5603</td>
<td>lacZ mobRP4 oriR6K (Km&lt;sup&gt;+&lt;/sup&gt;)</td>
<td>S. Yashuda†</td>
</tr>
<tr>
<td>pJP5608</td>
<td>lacZ mobRP4 oriR6K (Tc&lt;sup&gt;+&lt;/sup&gt;)</td>
<td>S. Yashuda†</td>
</tr>
</tbody>
</table>

* Abbreviations: Ap<sup>+</sup>, ampicillin-resistant; CP, cross-reactive protein; Km<sup>+</sup>, kanamycin-resistant; Tc<sup>+</sup>, tetracycline-resistant.
† The Cloning Vector Collection, Dept Microbial Genetics, National Institute Genetics, Yatal, 111 Mishima, Shizuoka-ken 411, Japan (cvector@lab.nig.ac.jp).

Fig. 1. Scaled diagram of agn43 showing the location of fragments F1–F5 in relation to the regions encoding the signal sequence (ss), and the α<sup>+</sup> and β<sup>+</sup> subunits.

Fig. 2. Southern blot analysis of EcoRV/SpeI-digested total genomic DNA from E. coli strains ML308-225 (lanes 1 and 10), AJR1 (agn43B; lanes 2 and 11), AJR2 (agn43A; lanes 3 and 12), AJR3 (agn43AB; lanes 4 and 13), ML308-225 oxyR (lanes 5 and 14), AJR4 (oxyR agn43B; lanes 6 and 15), AJR5 (oxyR agn43A; lanes 7 and 16) and AJR6 (oxyR agn43AB; lanes 8 and 17). Lanes 1–8 and 10–17 were probed with DIG-labelled F1 (see Fig. 1) and DIG-labelled pJP5608, respectively. Lane 9, DIG-labelled DNA ladder no. VII (8.6, 7.4, 6.1, 4.9 and 3.6 kb; Roche).

mating appropriate E. coli ML308-225 derivatives with E. coli S17.1 i pir containing one of the three suicide vectors pAJR003–pAJR005 (see Table 1). Vector integration via single crossover events resulted in formation of interrupted alleles. The relevant genotypes of these mutants were confirmed by Southern blot analysis of randomly chosen transconjugants, using EcoRV/SpeI-restricted total genomic DNA with F1 and plasmid probes (see Fig. 2). Thus, disruption of agn43A in both wild-type and oxyR backgrounds using pAJR004 (4.2 kb; Ap<sup>+</sup>) resulted in disappearance of the larger (7.5 kb) of
the two F1-hybridizing bands and the appearance of a band of the anticipated size (12 kb) hybridizing with both F1 and plasmid probes (see Fig. 2, lanes 1, 3, 5, 7, 10, 12, 14 and 16). Similarly, disruption of *agn43B* in a wild-type background using pAJR003 (3.7 kb; Km<sup>r</sup>) resulted in selective increase in size of the smaller EcoRV/SpeI fragment to generate an approximately 8 kb doublet, the upper component of which (disrupted *agn43B*) hybridized with both F1 and plasmid probes. The lower component of this doublet corresponds to the original *agn43A* restriction fragment (Fig. 2, lanes 1, 2, 10 and 11). Inactivation of *agn43B* in an oxyR (<Km<sup>r</sup>) background required the use of an alternative suicide vector (pAJR005; 7.2 kb; Te<sup>r</sup>) carrying a different antibiotic selection marker. This vector possesses an EcoRV site which results in mutated *agn43B* allelic giving rise to two restriction fragments (approx. 3.6 and 7 kb) capable of hybridizing with both F1 and plasmid-specific probes. The upper band is similar in size to the native *agn43A* restriction fragment (Fig. 2, lanes 5, 6, 14 and 15). Southern blot analysis of double (*agn43AB*) mutants constructed in wild-type and oxyR backgrounds using pAJR003/4 and pAJR004/5, respectively, again yielded the predicted pattern of hybridizing bands (see Fig. 2, lanes 1, 4, 5, 8, 10, 13, 14 and 17). The slight increase in anticipated size of the mutated *agn43A* fragment (Fig. 2, lanes 4 and 13) is most likely the result of a duplication event involving the integrated vector.

**Strategy for sequencing the *agn43* genes.** The *agn43*-hybridizing fragments of approximately 7.5 kb (*agn43A*) and 4.5 kb (*agn43B*) obtained following digestion of ML308-225 total genomic DNA with EcoRV and SpeI were individually gel purified and used as templates in PCR amplification reactions to generate product for sequencing. Cross-contamination of the *agn43A*- and *agn43B*-containing fractions was ruled out in Southern blots and experiments which revealed (a) gel-purified products in the size range 6–7 kb obtained following digestion of genomic DNA with EcoRV and SpeI [i.e. material between the 7.5 kb (*agn43A*) and 4.5 kb (*agn43B*) fragments] failed to hybridize with the *agn43*-specific F5 probe or to generate amplified *agn43*-specific PCR products; and (b) that the *agn43A* - and *agn43B*-containing fractions each generated a single hybridizing band of the anticipated size (7.5 kb and 4.5 kb, respectively).

The upstream regions of each gene could not be amplified using the above strategy, presumably due to the presence of an EcoRV or SpeI cleavage site at the 5’ end of the gene. Accordingly, a different strategy was devised to amplify the region encoding the signal sequence of both genes. This involved use of purified total genomic DNA from single knockouts AJR1 (ML308-225 *agn43A*::pAJR003) and AJR2 (ML308-225 *agn43A*::pAJR004) as templates in PCR reactions utilizing forward and reverse primers complementary to sequences upstream of the signal codons and downstream of the mutated *agn43* allele, respectively, and extension periods selectively favouring amplification of the (smaller) wild-type allele. The clear size difference between products derived from mutated and wild-type alleles additionally ensured the unambiguous identification of *agn43A*-specific PCR fragments (produced from AJR1 template DNA) and *agn43B*-specific PCR fragments (produced from AJR2 template DNA) following their analysis by agarose gel electrophoresis.

DNA sequencing, performed with custom-made primers, was carried out on PCR-generated gel-purified DNA fragments according to the manufacturer’s instructions using the Thermosequenase kit (Amersham). Sequencing reactions were resolved using an ABI 373A sequencer and resultant sequences were analysed by Seqed (Applied Biosystems). Analysis of completed sequences was carried out using the GCG programs (University of Wisconsin Genetics Computer Group).

**Protein analysis.** SDS-PAGE was performed using 12.5% (w/v) polyacrylamide separating gels and a 4.5% polyacrylamide stacking gel (Laemmli, 1970). Samples were routinely heated for 3 min at 100 °C in Laemmli sample buffer (Laemmli, 1970) prior to electrophoresis. Proteins were detected by staining with Coomassie brilliant blue. Molecular masses were determined from the relative mobilities of 15 standard molecular mass marker proteins (BenchMark protein ladder; Gibco-BRL). Immunofluorescence microscopy, colony and Western immunoblotting, and determination of switch frequencies were carried out using anti-χ<sup>2</sup> antiserum in accordance with established procedures (Eisenstein, 1981; Caffrey et al., 1988; Henderson et al., 1997a). The Antigenicity Index program devised by Jameson & Wolf (1988) was used to identify potential epitopes. This method is based on an algorithm that integrates the predicted influence of hydropathy and surface probability with flexibility factors.

**RESULTS**

**E. coli** ML308-225 possesses two *agn43*-hybridizing fragments

*E. coli* K-12 possesses a single copy of *agn43* (Blattner et al., 1997), which shows 98% sequence identity with *agn43* from our control producer strain *E. coli* ML308-225 (Henderson & Owen, 1999). Surprisingly, however, when restricted with certain (combinations of) enzymes (viz. EcoRV, EcoRV plus SpeI, and EcoRV plus HindIII), the DNA from ML308-225 and the K-12 derivative BD1302 showed significant differences in Southern hybridization experiments conducted with *agn43*-specific probe F5. The most notable feature was the appearance of two bands in ML308-225 digests compared with only a single band for the corresponding BD1302 digests (Fig. 3). From a knowledge of gene

![Fig. 3. Southern blot analysis of total genomic DNA from E. coli ML308-225 (odd-numbered lanes) and E. coli K-12 derivative BD1302 (even-numbered lanes) following digestion with EcoRV (lanes 1 and 2), EcoRV plus SpeI (lanes 3 and 4), EcoRV plus HindIII (lanes 5 and 6) and SpeI plus EcoRI (lanes 7 and 8). The blot was probed with F5. Identical profiles were obtained with probes F2 and F3 (see Fig. 1). No strongly hybridizing bands characteristic of plasmids were observed for undigested DNA samples (not shown). The positions of molecular size standards are shown to the right.](image-url)
sequence and assuming a single copy of the gene, one hybridizing band would have been anticipated in each instance. The fact that identical hybridization patterns were also obtained using probes F2 and F3 (see Fig. 1) eliminates intragenic cleavage as a possible reason for this phenomenon. These results strongly suggest that there is an additional \textit{agn43} homologue on the ML308-225 chromosome. The two putative alleles in ML308-225 were named \textit{agn43A} and \textit{agn43B} in order of decreasing size of their respective \textit{EcoRV}/\textit{SpeI} restriction fragments.

**Construction and analysis of \textit{agn43} mutants**

To confirm the presence of two functional \textit{agn43} alleles in \textit{E. coli} ML308-225, \textit{agn43A} and \textit{agn43B} were disrupted by insertional inactivation using \textit{agn43}-specific \textit{p}ir suicide vectors pAJR003–pAJR0005. Single (\textit{agn43A} and \textit{agn43B}) and double (\textit{agn43AB}) knockout mutants were constructed in both wild-type (phase-variable) and \textit{oxyR} (locked-ON) backgrounds (see Methods and Table 1), and their genotypes confirmed by Southern blot analysis of \textit{EcoRV}/\textit{SpeI}-restricted total genomic DNA using F1 and plasmid probes (see Methods and Fig. 2).

Analysis of the various knockout mutants by Western immunoblotting (Fig. 4), immunofluorescence microscopy (Fig. 5) and colony immunoblotting (not shown) indicated that disruption of \textit{agn43A} or \textit{agn43B} alone was not sufficient to suppress expression of immunoreactive antigen. Thus, phase-ON populations of single mutants AJR1 (\textit{agn43B}) and AJR2 (\textit{agn43A}), as well as cells of AJR4 (\textit{agn43B} \textit{oxyR}) and AJR5 (\textit{agn43A} \textit{oxyR}) clearly expressed a polypeptide (apparent M, 60 000) capable of reacting with anti-\textit{x}43 antisera, whereas phase-OFF populations of the single mutants, together with double mutants AJR3 (\textit{agn43AB}) and AJR6 (\textit{agn43AB} \textit{oxyR}) did not (Fig. 4). These results were confirmed and extended in immunofluorescence and colony-immuno-

 blotting experiments which clearly revealed that AJR1 (\textit{agn43B}) and AJR2 (\textit{agn43A}) expressed \textit{Ag43} in a phase-variable manner similar to that observed for control strains, whereas expression of the antigen in AJR4 (\textit{agn43B} \textit{oxyR}) and AJR5 (\textit{agn43A} \textit{oxyR}) was locked-ON in a manner similar to that evidenced by the parental (\textit{oxyR}) derivative, with all cells in the population expressing antigen (see Fig. 5). These data clearly demonstrate that ML308-225 possesses duplicate copies of \textit{agn43} viz. \textit{agn43A} and \textit{agn43B}, both of which are capable of expressing \textit{Ag43} homologues in a phase-variable manner. They also indicate that the \textit{agn43A}::pAJR004 constructs (see Table 1) containing insertions within the \beta^{343A}-encoding region are incapable of producing stable \textit{x}43A, as anticipated from consideration of the roles and properties attributed to the translocation (\beta) domain of autotransporters (Henderson et al., 1998; Maurer et al., 1999).

Interestingly, both phase-ON and phase-OFF populations of AJR1 (\textit{agn43B}) typically possessed higher proportions of cells in the OFF state than similar populations from AJR2 (\textit{agn43A}; see Fig. 5). Further analysis of this phenomenon revealed that, when grown aerobically in L-broth at 37°C, \textit{agn43A} switched from ON → OFF with a frequency (1.4 × 10^{-5}) approximately twofold faster than \textit{agn43B} (5.9 × 10^{-3}; \textit{P} = 0.038) and underwent OFF → ON switching with a frequency (9.7 × 10^{-3}), almost eight times slower than that of \textit{agn43B} (7.7 × 10^{-4}; \textit{P} = 0.003).

**Sequence analysis of \textit{agn43A}_{ML308-225} and \textit{agn43B}_{ML308-225}**

A sequence for \textit{agn43A}_{ML308-225} based on analysis of cloned PCR products amplifed from total genomic DNA, has previously been published by workers in this laboratory (Henderson & Owen, 1999). However, in view of the demonstration of two \textit{agn43} homologues in this strain and the obvious potential for cross-amplification, it became necessary to refine this work and independently sequence each gene using more stringent procedures. The strategy involved is detailed in Methods. Comparative analysis of the resulting \textit{agn43A} and \textit{agn43B} sequences (GenBank accession nos AF233271 and AF233272, respectively) revealed 98\% identity at both nucleotide and predicted protein levels. Both sequences showed minor differences to that previously published for \textit{agn43} (U24429; Henderson & Owen, 1999). The 3120 bp coding region of each allele (\textit{agn43A} and \textit{agn43B}) differed at 40 nucleotide positions, nine occurring in the \textit{x}43A-encoding region, four in the \beta^{13A}-encoding region and the remainder in the region encoding the signal sequence. This resulted in similarly sized (1039 residue) primary translation products of predicted \textit{M}s 106 979 (\textit{Ag43A}) and 106 940 (\textit{Ag43B}) containing, respectively, 4, 3 and 12 amino acid differences between the two \textit{x}43A domains, the two \beta^{13A} domains and the two signal sequences. The predicted \textit{M}s for the \textit{x}43A, \textit{x}43B, \beta^{13A} and \beta^{13B} subunits were 49 904, 49 795, 51 554 and 51 633, respectively. Of interest are differences observed in the two surface-expressed \textit{x}43A...
subunits. Notably, similar changes from alanine residues (α13B) to threonine residues (α13A) at two neighbouring positions (226 and 230) of the primary translation product significantly increase the surface probabilities [from 0.7 (α13B) to 1.5 (α13A)] of one of the predicted (Jameson & Wolf, 1988) cell-surface epitopes. The small sequence variations may also account for the minor differences in the electrophoretic mobilities of the two α subunits detected following SDS-PAGE (see Fig. 4). The differences predicted in the amino-terminal sections of the two precursor proteins do not appear to radically alter the properties of the principal domains/motifs characteristic of the extended signal sequence observed for autotransporters (Henderson et al., 1998).

Both genes contained previously identified motifs, viz., those for an aspartyl protease active site, O-glycosylation attachment, a P-loop and RGD motifs (Henderson & Owen, 1999). Additional motifs identified as a result of this study include (a) a leucine zipper motif (L200–L211) implicated in protein dimerization (Phizicky & Fields, 1995) and located at the C terminus of each α subunit and (b) the presence, within 53 residues of the P-loop motif of both β14 subunits, of a sequence (R692DSDES) characteristic of a tyrosine phosphorylation site. These observations suggest possible mechanisms for subunit–subunit associations/auto-aggregation and autotransport, respectively. Numerous (17) potential N-glycosylation sites were dispersed throughout both α13 and β14 subunits.

**Southern blot analysis of EPEC strains**

Previous studies from this laboratory have documented the presence within certain EPEC strains of multiple anti-α13 cross-reactive proteins (Owen et al., 1996). In view of the current demonstration of duplicate *agn43* genes in ML308-225, this panel of EPEC strains was analysed by Southern blotting using *agn43*-specific gene probes (F1, F3 and F5). Resultant blots revealed several strains to possess multiple (up to four) hybridizing *EcoRV/SpeI* fragments. Some strains [e.g. NCTC 10089 (Fig. 6), NCTC 8603 and NCTC 9114 (data not shown)] possessed only one hybridizing fragment and NCTC 8621 none at all (see Fig. 6). Identical hybridization patterns were obtained for all three probes except in the
case of NCTC 8007, where the smallest hybridizing band had an approximate size of 6 kb and 3 kb when probed with F3/5 and F1, respectively, indicating the presence of an EcoRV or SpeI site between regions F1 and F5 of the agr43D allele of this strain. These data strongly suggest the presence of varying numbers of agr43 alleles within clinical EPEC isolates. The number of predicted gene copies was always equal or greater than the number of anti-α3 cross-reactive proteins detected by Western immunoblotting (see Fig. 6 and Table 1). Taken together, these results suggest the presence of a family of Ag43 proteins encoded by multiple chromosomal alleles.

Preliminary evidence suggests considerable structural diversity within the surface-exposed cross-reactive α3 subunits of the various strains. Thus, sequence analysis of the (single) α3-encoding region from NCTC 9114 (GenBank accession no. AF233273) revealed (a) 95% identity with the corresponding region of agr43B_{ML308-225}, with 21 of the 25 amino acid differences being located within the first N-terminal third of the subunit, and (b) the absence of the RGD motif implicated in integrin binding.

**DISCUSSION**

It seems clear from analysis of mutants and of the gene sequence that *E. coli* ML308-225, unlike *E. coli* K-12, possesses duplicate copies of agr43. Gene duplication is believed to have a major role in evolution, one gene copy maintaining its original function in response to selective constraints, thereby freeing the other to generate possibly advantageous mutations and new functions. These processes are believed to facilitate formation of antigenically variant families of proteins and of proteins with novel functions, thus providing organisms with the ability to evade the host immune response, to facilitate transmission to the next host, and to adapt to different microenvironments. Failure of a duplicated gene to evolve a beneficial function results in it being lost from the genome (Force et al., 1999). A number of such protein families have been documented (Dehio et al., 1998; Barbour, 1993; Schmitt et al., 1991). Examples include the Opa proteins of *Neisseria meningitidis* and *Neisseria gonorrhoeae*, where gene copy numbers of 4 and 11 have been observed, respectively (Dehio et al., 1998), and the two Shiga-like toxin II proteins of enterohaemorrhagic *E. coli* (EHEC) strain O157:H- (Schmitt et al., 1991). In both families the function of multiple alleles is to express antigenically/functionally variant forms of the protein. We believe this to be the case for agr43. The mechanism of gene duplication in the current instance remains unclear. However, it should be noted that analysis of the K12 genome in the vicinity of agr43 (GenBank accession no. AE000291) reveals the presence of two IS elements and an elevated G+C content, features often associated with mobile genetic elements (Hacker et al., 1997). These and other issues related to agr43 duplication may become clearer following completion of the genome sequencing project for the uropathogenic *E. coli* isolate CFT073 (Blattner et al., 2000).

The differences between the two agr43 alleles of *E. coli* ML308-225 are subtle rather than striking. Both genes are clearly capable of expressing product. Both are subject to phase variation, and both appear to be regulated by a similar mechanism in which OxyR competes with Dam for unmethylated 5’-GATC-3’ sites in the regulatory region of agr43 (Henderson et al., 1997a; Henderson & Owen, 1999; Haagmans & van der Woude, 2000). Certainly, the locked-ON phenotype observed for both oxyR agr43A and oxyR agr43B derivatives (Fig. 5) and the total absence of any Ag43-expressing cells in *dam agr43A*B- derivatives (Henderson & Owen, 1999) are fully compatible with this proposition. In the absence of allele/product-specific probes it is difficult to determine whether agr43A and agr43B are capable of simultaneous expression in any one cell. However, comparison of fluorescent intensities in parental and single knockout mutants suggests that this may be possible (A. J. Roche, M. Meehan & P. Owen, unpublished data). The observed differences in OFF → ON and ON → OFF switch frequencies for the two alleles may reflect a mechanism that allows for selection of populations expressing either Ag43A or Ag43B *in vivo*, and strongly suggests that Ag43B is the more dominant of the two antigens during growth under standard laboratory conditions.

Whether or not the Ag43A and Ag43B of *E. coli* ML308-225 are antigenically/functionally variant has not been definitively established. Autoaggregation and colony morphology changes induced by Agr43 (Diderichsen, 1980; Henderson et al., 1997a) occur only in certain genetic backgrounds (e.g. *E. coli* K-12) and are not pronounced properties of *E. coli* ML308-225 derivatives, a phenomenon probably related to the presence of different types of surface structures, for example fimbriae and lipopolysaccharide (Hasman et al., 1999, 2000). However, the relatively small number of amino acid differences in the surface-exposed α3 subunits does result in changes in molecular size and apparent epitope expression. It is also conceivable that the principal differences observed, A(α3B)226T(α3A) and A(α3B)230T(α3A), result in changes in O-glycosylation patterns. Certainly, there is growing evidence that bacterial surface structures can be glycosylated (Stimson et al., 1995; Brimer & Montie, 1998; Forest et al., 1999; Lindenthal & Elsinghorst, 1999; Sleytr & Beveridge, 1999). Additional experimentation, involving refined chemical and immunological analysis, is required to fully resolve these issues.

Size-variable α3 subunits suggestive of antigenic variation have been identified in this study of *E. coli* strain ML308-225 and also in previous studies of EPEC isolates (Owen et al., 1996). Our present studies confirm and extend these latter observations and demonstrate the probable presence of multiple agr43 alleles in certain EPEC strains. It is also clear from sequence analysis that,
whereas the agn43 genes of strains ML308-225 and K-12 show very high homology (indeed the primary sequence of α43 from ML308-225 and α43 from K-12 are identical), those of other strains (e.g. agn43<sub>KNC1911</sub>) can encode α43 subunits which differ more substantially, especially towards the N terminus, and can lack putative functional motifs. Ongoing sequence analysis of the genome of the uropathogenic <i>Escherichia coli</i> isolate CFT073 confirms the thrust of our data and indicates the presence in this strain of two ORFs showing strong homology with agn43 and predicted to encode proteins lacking RGD motifs (Blattner et al., 2000).

ACKNOWLEDGEMENTS

This work was supported by a research grant, SC/97/302, from Enterprise Ireland. The authors would like to thank Mary Meehan, Ian Henderson, Stephen Smith and Richard Fitzpatrick for help, expertise and scientific discussion.

REFERENCES


Owen, P. & Kaback, H. R. (1978). Molecular structure of mem-


Received 15 June 2000; revised 10 August 2000; accepted 23 August 2000.