DNA sequence analysis of the *imp* UV protection and mutation operon of the plasmid TP110: identification of a third gene

David Lodwick, Darerca Owen and Peter Strike*
Department of Genetics and Microbiology, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK

Received June 19, 1990; Revised and Accepted July 31, 1990 EMBL accession no. X53528

ABSTRACT

The sequence of the *imp* operon of the plasmid TP110 (which belongs to the $IncI_1$ incompatibility group) has been determined, and is shown to contain three open reading frames. This operon, involved in UV protection and mutation, is functionally analogous to the *umu*DC operon of *E.coli* and the *muc*AB operon of the plasmid pKM101, which belongs to the quite unrelated IncN incompatibility group. The *umu* and *muc* operons however contain only two open reading frames, coding for proteins of approximately 16kD and 46kD. The high degree of homology between the two 16kD proteins (UmuD and MucA) and between the two 46kD proteins (UmuC and MucB) clearly shows their relatedness. This is shown also to extend to the *imp* gene products, with ImpA sharing homology with UmuD and MucA, and ImpB sharing homology with UmuC and MucB. However, the two *imp* genes are preceded in the operon by a third gene, *impC*, which encodes a small protein of 9.5kD and which has no equivalent in the *umu* and *muc* operons.

INTRODUCTION

In *Escherichia coli*, the mutagenic responses to UV light and to a large number of chemical mutagens are largely dependent upon the inducible pathway commonly referred to as 'error-prone repair' (see (1) for a review). The essential components of this pathway include both the RecA protein and the products of the *umu* operon, UmuC and UmuD. These proteins, in some as yet uncharacterised way, may modify the DNA replication machinery to allow bypass synthesis across a damaged template. Several lines of evidence (2,3,4) suggest that the *umu* gene products may facilitate the prompt resumption of chain elongation at sites where the replication forks have been stalled by DNA damage.

Phenotypically, *umu*CD mutants show a greatly reduced mutagenic response to UV and many chemical mutagens, and they are also slightly UV sensitive. Although their gene products play an important role in mutation in *E.coli*, it has become apparent that many bacterial species do not carry equivalent genes (5). Even in the closely related species *Salmonella typhimurium*, the *umu* gene equivalents appear to be ineffective, and mutation in this species depends to a considerable extent on the presence of plasmids from a variety of incompatibility (Inc) groups, carrying functional analogues of the *umu* genes. Thus the sensitivity of the strains employed in the Ames mutagenicity test is greatly increased by the presence of the IncN plasmid pKM101, which carries the *muc* analogue of the *umu* operon (6). Sequence analysis of both the *umu* and *muc* operons has revealed that although there is a wide divergence between these systems at the level of DNA sequence, they are closely related in terms of their gene products. Both operons contain two genes encoding proteins of MW15,064 and 47,681 in the case of *umu* (*umu*D and *umu*C respectively), and MW16,371 and 46,362 in the case of *muc* (*muc*A and *muc*B respectively) (7,8).

In addition to being very similar in size, the proteins of these operons show considerable amino acid homology. UmuD and MucA proteins are 41% homologous, UmuC and MucB proteins are 55% homologous, these areas of homology being largely confined to blocks of amino acids which are almost completely conserved. The implication of these observations is clearly that the IncN plasmid *muc* genes and the chromosomal *umu* genes are evolutionarily related, and that they have probably evolved from a common ancestor. The similarities in the gene products leave little doubt that these two operons perform similar if not identical functions.

In this paper, we describe the sequence analysis of a second plasmid-encoded *umu*-analogue, encoded in this case by the *imp* genes of the $IncI_1$ plasmid TP110 (9). Although both this plasmid and pKM101 were originally derived from clinical isolates of *Salmonella typhimurium*, they can both be transferred very efficiently by conjugation to *E.coli*, where their UV protection and mutation properties are again manifested. We wished to determine whether the UV protection and mutation functions of this plasmid, which is completely unrelated to the IncN plasmids such as pKM101, were also evolutionarily related to the *umu* system, or whether this was a functionally analogous system with no structural similiarities. The results clearly indicate that all three systems, *umu*, *muc*, and *imp*, share a common

* To whom correspondence should be addressed

ancestry, but that there are some major differences between *imp* and the other two operons. The most obvious of these is that the *imp* operon contains a third gene, encoding a small protein of MW9,800.

MATERIALS AND METHODS

Bacterial Strains used included TG1−δ(*lac-pro*), *supE*, *thi*, *hsdD5* [F′ *traD36 proA*$^+$*B*$^+$ *lacI*^{q} *lacZ*δM15] (10); and JL652−as AB1157, F$^-$ *thi-1*, *thr-1*, *leu-6*, *proA2*, *argE3*, *his4*, *lacY1*, *galK2*, *xyl-5*, *ara-14*, *rpsL*, *sup37*, but containing also F′ *lacI*^{q} and pJL59 (J.Little, pers.comm)

Plasmids: pKG10 (5.9kb (11)) contains the 3.2kb *Eco*R1−*Bgl*II fragment of TP110 carrying the *imp* genes, cloned into the vector pT7-1. pDL2 (4.8kb; this work) contains the 2.4kb *Cla*I−*Xba*I fragment of pKG10 carrying the *imp*B gene, cloned into *Acc*I−*Xba*I digested pT7-2. pGP1-2, pT7-1, and pT7-2 make up the T7 RNA polymerase / promoter controlled expression system of Tabor and Richardson (12). pJL59 (Amp^r *lexA*$^+$) is a LexA over-producing plasmid kindly provided by Dr J. Little.

DNA Sequencing

Sequencing of the *imp* genes was achieved by directed subcloning of fragments from pKG10 into M13 mp18 and mp19 as appropriate (13), the clones used being shown in Figure 1. The subclones were sequenced by the dideoxy chain termination method of Sanger et al (14), initially using the standard Klenow (*E.coli* DNA polymerase 1 large fragment) method and latterly the modified T7 DNA polymerase (Sequenase) method (United States Biochemical Corporation), (15). C band compressions were resolved by the substitution of dITP for dGTP as necessary. The products of the reactions were labelled with [α-^{35}S]dATP (Amersham) and were subsequently analysed on buffer gradient gels (16). Single stranded DNA templates were prepared from PEG precipitated phage particles from the supernatent of an infected culture of strain TG1 (10).

LexA−Binding

LexA protein was purified from the overproducing strain JL652 (the generous gift of Dr John Little) using the method of Schnarr et al (17). Cells were lysed by sonication, cleared by centrifugation, and nucleic acid in the supernatent precipitated with Polymin P. Total protein was then precipitated with ammonium sulphate, dissolved, and subjected to chromatography on phosphocellulose. The peak fractions contained >95% pure LexA protein, as judged by stained SDS-PAGE. These fractions were pooled and used for subsequent analysis.

LexA−DNA binding reactions contained the end-labelled *imp* DNA fragment, prepared as described below (20−50,000 cpm per reaction), and purified LexA protein to give the required final concentration of between 10^{-5} and 10^{-11}M. Final reaction volumes of 120μl contained DNA, protein, and reaction buffer (100mM Tris-HCl, pH7.4; 1.6mM NaEDTA; 0.04mg/ml bovine serum albumen; 20% glycerol; 25mM NaCl). The reaction was incubated at 37°C for 5mins, and then on ice for 15mins. 100μl samples were loaded onto non-denaturing 5% polyacrylamide gels, made up in TBE buffer (0.09M Tris-borate, 0.002M EDTA, pH 8.0). The gel was run in 1× TBE at a constant current of 50mA for 2−3 hours, dried and then exposed to X-ray film at −70°C overnight or longer if required.

Labelling of DNA fragments

DNA fragments for use in gel retardation experiments were end-labelled using [α-^{35}S]dATP and T4 DNA polymerase in a replacement synthesis reaction, as described by Maniatis et al (18). Unincorporated material was removed by passage through Sephadex G50 (nucleic acid grade, Pharmacia-LKB).

Identification of Plasmid-Coded Gene Products

Proteins encoded by the *imp* genes were identified using the T7 RNA polymerase / promoter system as devised by Tabor and Richardson (12, and personal communication). Cells containing both the bacteriophage T7 RNA polymerase plasmid pGP1-2, and a pT7 recombinant plasmid, were grown in Luria broth to mid-log phase. A 0.2ml sample was taken, washed twice in M9 buffer, and resuspended in M9 buffer supplemented with 20μg/ml thiamine and 0.01% (w/v) 18 amino acids (excluding cysteine and methionine). The cells were grown with shaking at 30°C for 60 mins, then the temperature shifted to 43°C for 15 mins. Rifampicin was then added to a final concentration of 200μg/ml, to inhibit host directed transcription, and the cells were incubated at 42°C for a further 10 mins. The temperature was then shifted down to 30°C for 20 mins, when 10μCi ^{35}S-methionine (Amersham) was added. Incubation was continued for 5 mins, the cells were pelleted, and labelled proteins analysed by SDS-PAGE as described previously (9).

RESULTS

Sequencing the *imp* Operon

We have previously described the location of the *imp* genes of TP110 within a 3.2kb *Eco*R1−*Bgl*II fragment which is capable of fully expressing UV protection and mutation functions when cloned in high copy number vectors (9). Further subcloning defined the *imp* coding region as being completely contained on a ~2.6kb *Eco*R1−*Eco*RV fragment (11), and this fragment was chosen as the starting material for the determination of the DNA sequence. The strategy adopted for the sequence determination is shown in Figure 1. The complete sequence was determined on both strands using the dideoxy chain termination method of Sanger et al (14), as described above. The resulting sequence

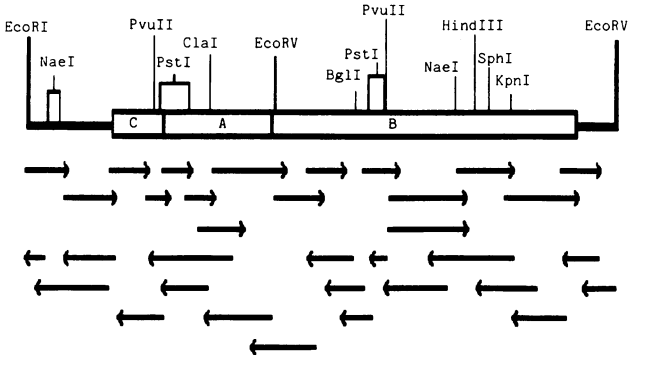

Figure 1. Gene organisation, restriction map and sequencing strategy of the *imp* operon of TP110. Sub-clones derived from the *Eco*RI−*Eco*RV fragment (~2.6kb) were used to determine the sequence on both strands for a distance of 2527 bp. from the *Eco*RI site. The extent of the sequenced fragments is shown as arrows. The positions of a number of six base pair restriction sites are shown, as are the positions of the three *imp* open reading frames, C, A, and B. Transcription across the operon proceeds in the left to right direction.

is shown in Figure 2. The restriction sites predicted from the DNA sequence were in complete agreement with those found earlier by restriction analysis (9), with the exception of an additional *Pst*I site at position 1685 (numbered from the *Eco*RI site). This site, together with the site at 1607, generates a small *Pst*I fragment of 78 bp. which had not previously been detected: a further pair of *Pst*I sites (at bases 629 and 790 in the sequence) also generate a small *Pst*I fragment. As the sub-clone carrying the *imp* operon was originally isolated as a *Pst*I partial digestion product, we felt it important to ensure that these small fragments were true components of the *imp* operon, and were not artefacts of the initial cloning strategy. *Eco*RI/*Eco*RV restriction digests of TP110 DNA were therefore hybridised with labelled M13 clones containing these two small fragments, cloned individually. In each case, hybridisation was observed to the appropriate band, confirming that the sequenced fragment represented an uninterrupted region of TP110.

Analysis of the sequence shown in Figure 2 reveals three major

```
1     GAATTCTCCT CTCACATGCG GCCTCCATCA GGCCGTGCCG CACCTGTTAC CGGGCGACCG CATACGAATG CAGGTGCGGC CCCGAAGCCG
91    CAGCGCGTGG AGGAGGAACG ACGAAACGTG ACCGGGTGAG GAGGCCGGCG ACAGCGACAC ACTTGCATCG GATGCAGCCC GGTTAAGTGC
181   CGGCACGGCC TGGGTAACCA GGTATTTTGT CCACAGAACC GTGCACAAAA TGCTGTGGAT AAGCTGGACA CAGCAGCCCA CAGCAGGCAT

271   ACAGCCACAC ACTCAGAGAA TTGAGCGTTT TTTAGTCACA ACAATGACAT TTGACTGTGT GCGGATTCAG TTATTAGAAT ACTGTATATA

361   CATACAGCAA AAGAAAAGGG AGATGAGAAC 390

391   ATG ATC CGC ATT GAA ATT CTT TTC GAC CGC CAG AGC ACA AAA AAC CTC AAA TCA GGC ACA CTT CAG GCG CTA CAG AAT
        M   I   R   I   E   I   L   F   D   R   Q   S   T   K   N   L   K   S   G   T   L   Q   A   L   Q   N
469   GAA ATC GAA CAA CGT CTG AAA CCA CAC TAC CCG GAA ATC TGG CTG CGT ATA GAC CAG GGC AGC GCC CCG TCT GTA TCT
        E   I   W   Q   R   L   K   P   H   Y   P   E   I   W   L   R   I   D   Q   G   S   A   P   S   V   S
547   GTC ACA GGA GCC CGC AAC GAC AAG GAT AAA GAA CGT ATC CTG TCT CTG CTG GAA GAA ATC TGG CAG GAC GAC AGC TGG
        V   T   G   A   R   N   D   K   D   K   E   R   I   L   S   L   L   E   E   I   W   Q   D   D   S   W
625   CTG CCT GCA GCA TGA  639
        L   P   A   A   Z
              636  ATG AGT ACC GTT TAT CAC CGT CCG GCT GAC CCT TCA GGC GAT GAC AGT TAC GTG CGA CCG TTG TTT
                     M   S   T   V   Y   H   R   P   A   D   P   S   G   D   D   S   Y   V   R   P   L   F
702   GCC GAT CGT TGC CAG GCC GGT TTT CCT TCG CCT GCC ACT GAT TAT GCT GAG CAG GAA CTG GAT CTG AAC AGC TAT TGC
        A   D   R   C   Q   A   G   F   P   S   P   A   T   D   Y   A   E   Q   E   L   D   L   N   S   Y   C
780   ATC AGC AGA CCT GCA GCC ACC TTC TTT CTG CGC GCC AGC GGT GAA TCG ATG AAC CAG GCT GGC GTG CAG AAT GGT GAT
        I   S   R   P   A   A   T   F   F   L   R   A   S   G   E   S   M   N   Q   A   G   V   Q   N   G   D
858   CTG CTG GTA GTG GAC AGG GCC GAA AAA CCA CAA CAC GGG GAC ATC GTT ATC GCT GAG ATC GAC GGT GAG TTC ACC GTC
        L   L   V   V   D   R   A   E   K   P   Q   H   G   D   I   V   I   A   E   I   D   G   E   F   T   V
936   AAA CGA CTG CTG TTG CGC CCA CGC CCG GCA CTG GAG CCG GTT TCA GAC AGC CCG GAG TTC CGC ACA CTG TAT CCG GAA
        K   R   L   L   L   R   P   R   P   A   L   E   P   V   S   D   S   P   E   F   R   T   L   Y   P   E
1014  AAC ATC TGT ATT TTT GGT GTT GTC ACT CAC GTG ATA CAC AGG ACG CGG GAG TTA CGC TGA  1073
        N   I   C   I   F   G   V   V   T   H   V   I   H   R   T   R   E   L   R   Z
                                                                                1073  ATG TTT GCA CTG GCT GAT
                                                                                        M   F   A   L   A   D
1091  ATC AAC AGT TTC TAC GCC TCA TGT GAA AAA GTT TTC CGC CCG GAC CTT CGC AAC GAA CCG GTC ATC GTA CTC AGC AAT
        I   N   S   F   Y   A   S   C   E   K   V   F   R   P   D   L   R   N   E   P   V   I   V   L   S   N
1169  AAC GAT GGC TGT GTG ATC GCG CGC AGC CCG GAG GCA AAA GCC CTT GGC ATC AGA ATG GGG CAG CCC TGG TTT CAG GTG
        N   D   G   C   V   I   A   R   S   P   E   A   K   A   L   G   I   R   M   G   Q   P   W   F   Q   V
1247  AGA CAA ATG CGC CTG GAG AAG AAA ATA CAT GTA TTT TCC AGC AAT TAT GCG CTG TAC CAC AGC ATG AGC CAA CGG GTT
        R   Q   M   R   L   E   K   K   I   H   V   F   S   S   N   Y   A   L   Y   H   S   M   S   Q   R   V
1325  ATG GCT GTT CTG GAG TCG CTT TCT CCC GCA GTT GAG CCC TAC TCA ATT GAT GAA ATG TTC ATT GAT TTG CGG GGG ATA
        M   A   V   L   E   S   L   S   P   A   V   E   P   Y   S   I   D   E   M   F   I   D   L   R   G   I
1403  AAT CAT TGC ATC TCT CCG GAG TTT TTT GGT CAT CAG CTC AGG GAA CAG GTA AAG AGC TGG ACA GGA CTC ACC ATG GGG
        N   H   C   I   S   P   E   F   F   G   H   Q   L   R   E   Q   V   K   S   W   T   G   L   T   M   G
1481  GTG GGC ATT GCG CCT ACA AAA ACG CTG GCT AAA AGT GCA CAG TGG GCA ACA AAG CAA TGG CCA CAG TTT TCC GGA GTG
        V   G   I   A   P   T   K   T   L   A   K   S   A   Q   W   A   T   K   Q   W   P   Q   F   S   G   V
1559  GTC GCG CTG ACG GCA GAA AAC CGT AAT CGG ATC TTG AAG CTA CTG GGG CTG CAG CCA GTT GGT GAG GTC TGG GGA GTA
        V   A   L   T   A   E   N   R   N   R   I   L   K   L   L   G   L   Q   P   V   G   E   V   W   G   V
1637  GGA CAC AGA CTG ACG GAA AAG CTG AAT GCG CTG GGT ATT AAC ACA GCA CTG CAG CTG GCG CAG GCT AAC ACG GCA TTC
        G   H   R   L   T   E   K   L   N   A   L   G   I   N   T   A   L   Q   L   A   Q   A   N   T   A   F
1715  ATC CGG AAA AAC TTC AGC GTC ATT CTT GAG CGT ACG GTA CGC GAA CTC AAC GGC GAG TCC TGC ATA TCC CTG GAA GAA
        I   R   K   N   F   S   V   I   L   E   R   T   V   R   E   L   N   G   E   S   C   I   S   L   E   E
1793  GCA CCA CCG GCA AAA CAG CAG ATT GTC TGT AGT CGC AGT TTT GGT GAA CGA ATC ACA GAC AAA GAT GCC ATG CAC CAG
        A   P   P   A   K   Q   Q   I   V   C   S   R   S   F   G   E   R   I   T   D   K   D   A   M   H   Q
1871  GCT GTT GTT CAG TAT GCA GAG CGG GCC GCA GAG AAA CTA CGT GGG GAG CGT CAG TAT TGC CGG CAG GTG ACG ACA TTT
        A   V   V   Q   Y   A   E   R   A   A   E   K   L   R   G   E   R   Q   Y   C   R   Q   V   T   T   F
1949  GTA CGG ACA TCA CCC TTT GCA GTA AAA GAA CCC TGT TAC AGC AAT GCC GCT GTG GAA AAG CTT CCA TTG CCC ACA CAG
        V   R   T   S   P   F   A   V   K   E   P   C   Y   S   N   A   A   V   E   K   L   P   L   P   T   Q
2027  GAC AGC CGG GAC ATT ATT GCC GCC GCA TGC AGA GCC TTA AAC CAT GTC TGG CGT GAA GGG TAC CGC TAT ATG AAG GCA
        D   S   R   D   I   I   A   A   A   C   R   A   L   N   H   V   W   R   E   G   Y   R   Y   M   K   A
2105  GGT GTC ATG CTG GCT GAT TTC ACA CCA TCG GGT ATA GCG CAG CCG GGA TTA TTT GAT GAA ATC CAG CCC CGT AAA AAC
        G   V   M   L   A   D   F   T   P   S   G   I   A   Q   P   G   L   F   D   E   I   Q   P   R   K   N
2183  AGT GAA AAG TTA ATG AAA ACA CTC GAT GAA CTG AAC CAG TCG GGA AAA GGG AAA GTG TGG TTT GCG GGG CGA GGA ACC
        S   E   K   L   M   K   T   L   D   E   L   N   Q   S   G   K   G   K   V   W   F   A   G   R   G   T
2261  GCC CCT GAA TGG CAA ATG AAA CGG GAA ATG TTG AGT CAG TGT TAT ACA ACT AAA TGG CGA GAT ATT CCC CTG GCC AGG
        A   P   E   W   Q   M   K   R   E   M   L   S   Q   C   Y   T   T   K   W   R   D   I   P   L   A   R
2339  CTG GGT TAG    2347
        L   G   Z
2348  TTCAGTCATC TCCGAACATA TTTTCAGCGT TTCTTCTGGT CTGGTCTTCA CCGTCATTTC TCGACAGACT CTGCTCTGTA AGAGGAGTAC
2438  TGCAATCCAC AGACGTATGG AGACCAGATT GTTTTTCCAG TAATTCAAGC AATTTCTTTT CAACCGAAAA AGCACCTGGT ATTACGCTAA
```

Figure 2. DNA sequence of the *imp* operon and predicted amino acid sequence of the gene products. The sequence starts at the *Eco*RI site shown in Figure 1. Structural features indicated in bold type include the likely promoter sequence (−35 and −10), the ribosome binding site (SD), and the two LexA binding sites (LexA1 and LexA2). The translation products are shown for each of the three *imp* open reading frames (*imp*A, 391−639; *imp*B, 636 1073; *imp*C, 1073−2347).

overlapping reading frames, downstream of a good consensus promoter sequence which is overlapped by two potential LexA binding sites. One of these overlaps the promoter −35 sequence, while the other overlaps the −10 region. Gel retardation studies using purified LexA protein and the 392 bp. *Eco*R1−*Mbo*1 *imp* fragment revealed only a single-step binding reaction (Figure 3), and it appears likely that only one of these potential binding sites is strongly bound by LexA protein. This is most likely to be the site overlapping the promoter −10 sequence, since the central (N_{10}) of the CTG (N_{10}) CAG motif is in this instance very AT rich (90%), whilst the site which overlaps the −35 box contains only 50% AT base pairs. It is however possible that the −35 Lex box does play a role, particularly as a similar box with a 50% GC content also exists in front of the *umu*DC operon (7). If LexA binding to such sites is weak, it is likely not to be detected under the conditions of the gel retardation assay.

Open Reading Frames Within The Sequence

The first open reading frame of the *imp* operon, designated *imp*C, potentially encodes a protein of MW 9,491. The predicted amino acid sequence (Figure 2) shows no homology to the products of either the *muc* or *umu* operons.

The second open reading frame, designated *imp*A, potentially encodes a protein of MW16,201. This is similar in size to both UmuD (MW 15,064) and MucA (MW 16,371). Although there is little nucleotide sequence homology between these three genes, the degree of amino acid homology is striking (Figure 4). ImpA is 43% homologous to MucA, and 49% homologous to UmuD. The homology between MucA and UmuD is slightly lower at 40%. Thirty one per cent of amino acids are common to all three proteins, this figure rising to fifty one per cent if conservative substitutions are allowed. As might be expected, this high degree of homology is achieved by the retention of blocks of amino acids which are essentially 100% conserved, these blocks being separated by more variable regions. It seems reasonable to assume that these blocks of conserved sequence represent sites essential for the mutagenic function of these proteins, while the variable sequences have other roles. Perry et al (7) demonstrated considerable amino acid homology between UmuD, MucA and the region of LexA protein surrounding its Ala−Gly (residues 84/85) cleavage site. This homology, as shown in Figure 4, extends to the amino acid sequence predicted for ImpA, and includes the residues Serine and Lysine (residues 119 and 156 in LexA; 61 and 98 in ImpA), as well as the Ala and Gly (residues 27/28 in ImpA). In this latter respect, the ImpA protein resembles the MucA protein rather than UmuD in having an Ala−Gly as opposed to a Cys−Gly cleavage site. The serine and lysine residues in LexA have been shown to be important for the autolytic cleavage of the protein (19), and this conservation of sequence indicates that ImpA may undergo RecA mediated self cleavage in a manner similar to that demonstrated for LexA and UmuD (20,21).

The third *imp* open reading frame, designated *imp*B, potentially encodes a protein of MW 47,731, similar in size to UmuC (47,681) and MucB (46,362). Again, although there is little nucleotide homology between the coding regions, there is significant homology between the predicted amino acid sequences of the gene products (Figure 5). ImpB is 51% homologous to MucB, and 54% homologous to UmuC. Homology between MucB and UmuC is slightly higher at 56%. In total, 43% of the amino acids are common to all three proteins, and if substitutions of similar amino acids is allowed, the figure rises to 57%.

The *imp*A and *imp*B coding regions overlap by one base pair (Figure 2), in the same way as those of *umu*D and *umu*C (7), which may have implications for translational coupling. There is an excellent match to the Shine-Dalgarno consensus sequence in the appropriate position (within *imp*A) to promote efficient translation of *imp*B. The *imp*C and *imp*A coding regions also overlap, but there is no obvious Shine-Dalgarno sequence upstream of the *imp*A initiation codon. This may account for the apparent low level of translation of *imp*A which has previously made it difficult to identify its product.

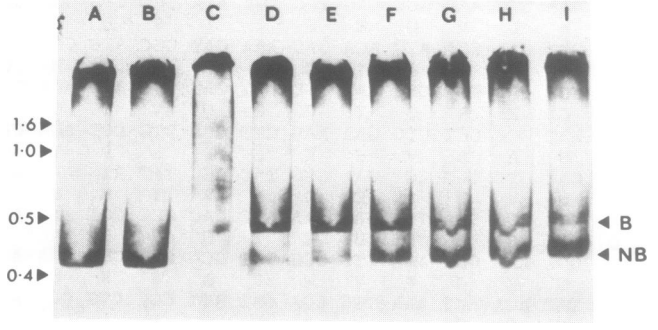

Figure 3. Electrophoretic mobility of the 413 bp *Eco*RI−*Taq*I fragment of the imp operon in the presence of purified LexA protein. Lanes A and B contain the end-labelled DNA fragment, and no LexA protein. Lanes C−I contain decreasing amounts of LexA, the concentration reducing ten-fold in each track from 10^{-5}M in track C to 10^{-11}M in track I. Arrows adjacent to track A indicate the positions of the molecular weight markers, and arrows labelled B and NB indicate the positions of the DNA band when the protein is either bound or not bound respectively. Electrophoresis conditions are described in the main text.

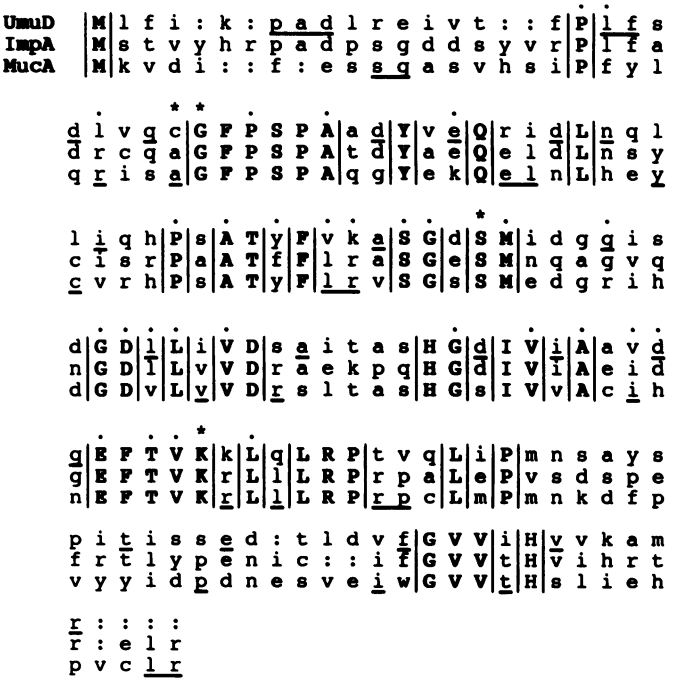

Figure 4. Sequence homology comparisons of the ImpA protein with UmuD and MucA. Amino acid residues which are conserved in all three equivalent gene products are shown in bold type and are flanked by bars. Sequences common to either UmuD and ImpA, or MucA and ImpA, are underlined. The Ala−Gly (Cys−Gly in UmuD)−Ser−Lys residues involved in the autocleavage reaction are starred (*). Other residues common to ImpA and LexA are indicated by points (·). Sequence data for UmuD and MucA are taken from Perry et al (7)

Identification of Gene Products

In order to identify the products of the *imp* open reading frames, the coding region was inserted into suitable pT7 vectors and expression was induced by providing T7 RNA polymerase *in trans* (12). Figure 6 shows the results of an analysis using two plasmid constructs, namely pKG10, in which the 3.2kb *Eco*RI−*Bgl*II fragment is cloned in pT7-1, and pDL2, in which the *Cla*I−*Eco*RV fragment encoding the 3′ end of *imp*A and the entire sequence of *imp*B (Figure 1), is cloned in pT7-2.

Comparison of the lanes reveals plasmid coded proteins of approximately 10kD, 16kD and 50kD in the pKG10 track, but suprisingly no plasmid coded products in the pDL2 track. We have previously identified proteins of approximately 11kD and 51kD as *imp* products (9), which were at that time suggested to be the products of the *imp*A and *imp* genes respectively. It is clear from the sequence data that this is unlikely to be the case, and that while the assignment of the ~50kD protein to *imp* is still valid, the ~10kD protein is almost certainly the product of the *imp*C gene. The use of the pT7 system allows the visualisation of a ~16kD protein not previously seen, which is consistent in size with the potential product of the *imp*A gene. Additional experiments using *imp*C$^+$ δ(*imp*AB) derivatives (containing either the *Eco*RI−*Cla*I or *Eco*RI−*Pst*I fragments (Figure 1) confirmed unequivocally that the ~10kD protein is produced not by *imp*A, but by *imp*C (data not shown).

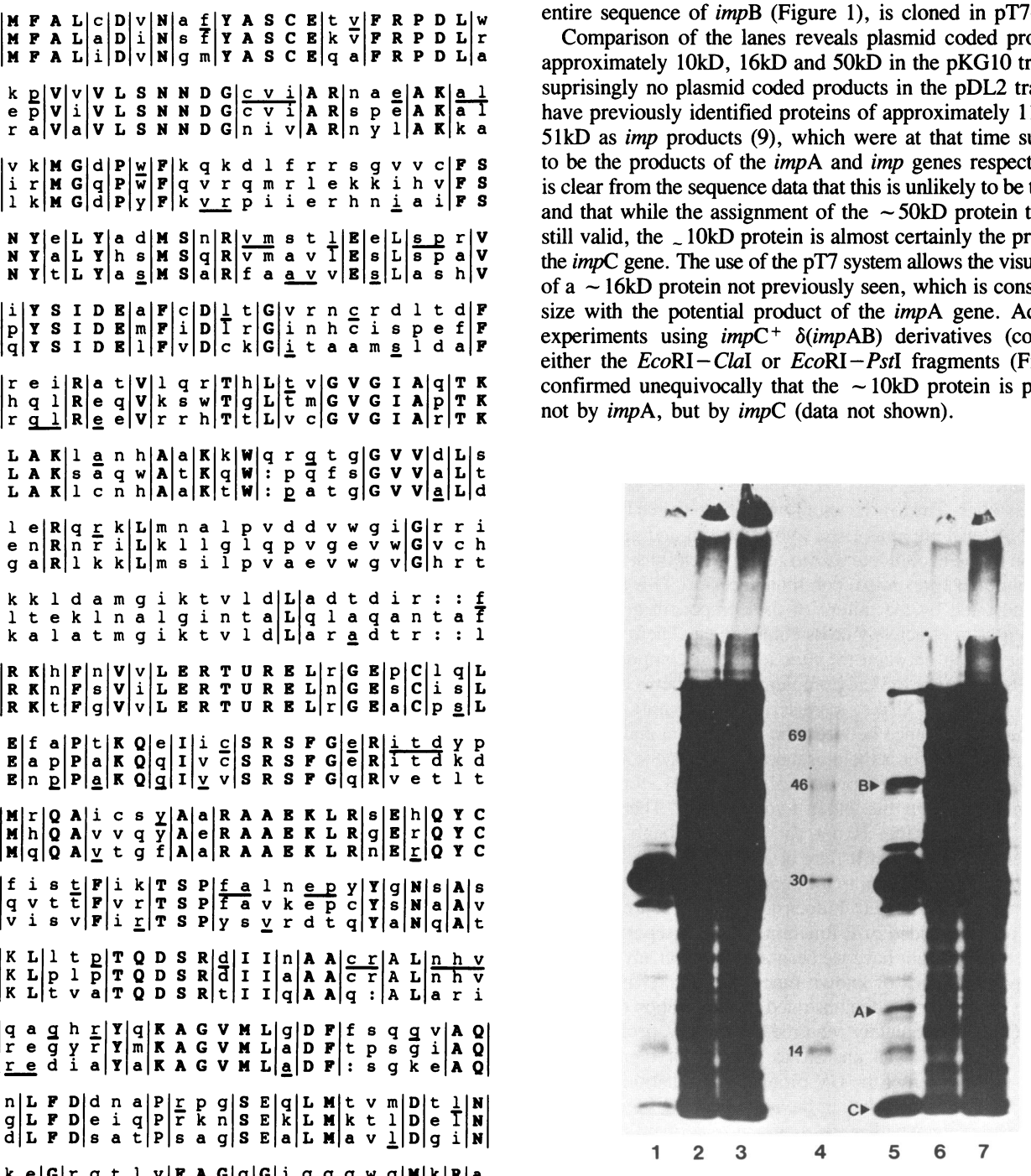

Figure 5. Sequence homology comparisons of the ImpB protein with UmuC and MucB. Amino acid residues which are conserved in all three equivalent genes are shown in bold type and are flanked by bars. Sequences common to either UmuC and ImpB, or MucB and ImpB, are underlined. Sequence data for UmuC and MucB are taken from Perry et al (7).

Figure 6. Polyacrylamide gel electrophoresis analysis of *imp* encoded proteins produced from T7 expression plasmids. Lanes 1, 2, and 3 show proteins produced by the plasmid pDL2 (*imp*C$^-$A$^-$B$^+$), lanes 5, 6, and 7 show proteins produced by pKG10 (*imp*C$^+$A$^+$B$^+$). The construction of the two plasmids is described in the main text. Lanes 3 and 7 show proteins prepared from cultures which have not been induced for T7 RNA polymerase. Lanes 2 and 6 contain samples prepared from induced cells, and lanes 1 and 5 contain samples from induced cells to which rifampicin (200μg/ml) has been added to inhibit host mediated transcription. Lane 4 contains molecular weight markers, the sizes of which are indicated in kilodaltons. Arrows indicate the three *imp* gene products in lane 5, with approximate sizes of 10kD (ImpC), 16kD (ImpA) and 47kD (ImpB).

The lack of visible *imp*-coded gene products in the pDL2 track was an unexpected observation, since this plasmid contains an intact *imp*B gene which is known from genetic complementation to be completely functional (Lodwick and Strike, manuscript in preparation). It seems likely that the high level expression of this gene cannot be achieved in this construct, perhaps due to the lack of translational coupling which normally occurs when the *imp*C and *imp*A genes are present.

DISCUSSION

The DNA sequence of the *imp* operon reveals a close relationship to the two previously characterised UV protection / mutation operons *muc* and *umu*. However, despite the fact that both *imp* and *muc* are found on plasmids originally isolated from *Salmonella typhimurium*, while *umu* occupies a chromosomal location in *E.coli*, *imp* and *muc* show no greater homology to each other than to *umu*. Each of these three operons codes for proteins of ~16kD and ~47kD, which are essential for the UV protection / mutation effect. The high degree of conservation of blocks of amino acids amongst the three ~16kD proteins and amongst the three ~47kD proteins clearly indicates those regions of the protein essential for function. However, it is also clear that while these systems are closely related, they are quite distinct from each other. The low level of homology between the operons at the DNA level prevents cross-hybridisation, and gives a clear indication of the evolutionary divergence which must have taken place from their (presumed) common ancestor. This divergence is also indicated by the failure of the component genes of the three operons to effectively cross complement. There is no cross complementation between the *muc* and *umu* component genes (1), and only a very limited complementation between *imp* and *umu* (Lodwick and Strike, manuscript in preparation).

The major difference between the *imp* operon and the *muc* / *umu* operons however is the presence of a third gene, designated *imp*C, coding for a small protein of MW 9,491, for which no homologue exists in the other two operons. The biological function of this protein is not yet clear, although our initial experiments indicate that it may be involved in the restoration of the induced *imp* operon to the repressed state, once repair of DNA damage is completed. It does not however share homology with the plasmid-coded *psi*B function, which is reported to have a similar role (22), nor have we been able to detect any significant homology to proteins of known function in the SwissProt data base. We have however demonstrated by the creation of deletions that *imp*C is not absolutely required for the UV protection and mutation effects (Lodwick and Strike, manuscript in preparation). In terms of evolution of the UV protection / mutation functions, it is not possible to say whether the *imp* operon has acquired this extra sequence, or whether the *umu* and *muc* operons have lost it. It is perhaps relevant however that *imp*-like sequences are widely distributed amongst most $IncI_1$, IncB and IncFIV plasmids (Lodwick and Strike, manuscript in preparation), both in plasmids derived from recent clinical infections, and from plasmids isolated before the general use of antibiotics (11). In each of these cases, sequences equivalent to *imp*C are present and have not been lost. The ImpC protein does appear to be toxic to the cell when produced constitutively in large quantities, and also appears to interfere with the SOS induction process (Lodwick and Strike, manuscript in preparation). Experiments are currently in progress to further characterise its biological properties.

ACKNOWLEDGEMENTS

This work was supported by a project grant from the Medical Research Council to P.S. The excellent technical assistance of Jenni Jones is gratefully acknowledged. We are particularly grateful to Dr John Little for the supply of LexA overproducing strains, for the supply or purified LexA protein, and for help and advice on purification procedures. Thanks are also due to Dr Stan Tabor for kindly supplying the pT7 expression plasmids.

REFERENCES

1. Walker,G.C. (1984) *Microbiol. Rev.*, **48**, 60–93
2. Bridges, B.A. and Woodgate, R. (1984) *Mol. Gen. Genet.*, **196**, 364–366
3. Tessman, I. (1985) *Proc. Natl. Acad. Sci. USA*, **83**, 6614–6618
4. Christensen, J.R., Le Clerc, J.E., Tata, P.V., Christensen, R.B. and Lawrence, C.W. (1988) *J. Mol. Biol.*, **203**, 635–641
5. Sedgwick, S.G. and Goodwin, P.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4172–4176
6. McCann, J., Spingarn, N.E., Kobori, J. and Ames, B.N. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 979–983
7. Perry,K.L., Elledge, S.J., Mitchell, B.B., Marsh, L. and Walker, G.C. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4331–4335
8. Kitagawa, Y., Akaboshi, E., Shinagawa, H., Horii, T., Ogawa, H. and Kato, T. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4336–4340
9. Glazebrook, J.A., Grewal, K.K. and Strike, P. (1986) *J. Bacteriol.*, **168**, 251–256
10. Carter, P., Bedouelle, H. and Winter, G. (1985) *Nucleic Acid Res.*, **13**, 4431–4443
11. Sedgwick, S., Thomas, S.M., Hughes, V.M., Lodwick, D. and Strike, P. (1989) *Mol. Gen. Genet.*, **218**, 323–329
12. Tabor, S. and Richardson, C.C. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1074–1078
13. Messing, J. and Viera, J. (1982) *Gene*, **19**, 269–276
14. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **71**, 5463–5467
15. Tabor, S. and Richardson, C.C. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4767–4771
16. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965
17. Schnarr, M., Pouyet, J., Granger-Schnarr, M. and Daune, M. (1985) *Biochemistry*, **24**, 2812–2818
18. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor University Press, Cold Spring Harbor.
19. Slilaty, S.N. and Little, J.W. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 3987–3991
20. Burckhardt, S.E., Woodgate, R., Scheuermann, R.H. and Echols, H. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 1811–1815
21. Shinagawa, H., Iwasaki, H., Kato, T. and Nakata, A. (1988) *Proc. Natl. Acad. Sci. USA*, **82**, 1074–1078
22. Dutreix,M., Backman,A., Celerier,J., Bagdasarian,M.M., Sommer,S., Bailone,A., Devoret,R., and Bagdasarian,M. *Nucleic Acids Res.*, **16**, 10669–10679