The immunity genes of colicins E2 and E8 are closely related*

Peter C. K. LAU1,*, R. Wally ROWSOME1,
Robert J. WATSON2 and Louis P. VISENTIN1

1Molecular Genetics Group, Division of Biological Sciences,
National Research Council of Canada, 100 Sussex Drive,
Ottawa, Ontario, Canada K1A OR6; and 2Chemistry and Biology
Research Institute, Agriculture Canada,
Ottawa, Ontario, Canada K1A 0C6

(Received 3 July 1984)

We have determined the nucleotide sequence of the
newly characterized colicin E8 imm gene which exists in
tandem with the colicin E3 imm gene in the
ColE3-CA38 plasmid. Comparison of these immunity
structures reveals considerable sequence divergence, but
the ColE8 imm gene is markedly homologous to the
coli E2 imm gene from the ColE2-P9 plasmid.

Colicins E2 and E3 are nuclease coded by the col genes present in
the non-conjugative plasmids, ColE2-P9 (6.8 kb) and ColE3-CA38 (7.2
kb) which are native to Shigella sonnei P9 and Escherichia coli
CA38 cells respectively (Watson & Visentin, 1980; Watson et al.,
1981). The two plasmids are 80-90% homologous by heteroduplex
analyses (Inselburg, 1973) and their restriction maps are very similar
(Inselburg & Johns, 1975; Watson & Visentin, 1980). By molecular
cloning and characterization of recombinant plasmids carrying restric-
tion fragments of the ColE2 or the ColE3 DNA, it was found that the
entire imm gene coding for the immunity protein and one end of the
col gene are contained within a short DNA region of the two plasmids
which is not homologous to each other (Watson & Visentin, 1982;
Watson et al., 1983). The immunity protein is a 10-kilodalton acidic
polypeptide which protects the specific colicin-producing cells from its
own toxin by binding to the basic carboxyl-terminal nuclease-active
domain of the colicin as an equimolar complex (Schaller & Nomura,
1976; Jakes et al., 1974; Ohno et al., 1977; Yamamoto et al., 1978).
Because the interaction of the colicin and immunity protein is highly
specific (for recent reviews, see Jakes, 1982; Konisky, 1982),
knowledge of the primary structures of these polypeptides is essential
for an understanding of the protein-protein interactions involved.
Toward this end, we have sequenced the nucleotides coding for the
DNAase and the RNAase domains of colicins E2 and E3 respectively
and for the entire immunity structural genes (Lau et al., manuscript
submitted).

*Issued as NRCC no. 23586 and as CBRI no. 1480.
**To whom correspondence should be addressed.
Recently, in addition to the existing list of ColE plasmids, designated E1 to E7 (Watson et al., 1981; Mock & Pugsley, 1982) two new colicins, E8 and E9, were reported (Cooper & James, 1984) and interestingly, by transposon mutagenesis it was further shown that the \textit{imm} gene for colicin E8 is present in the ColE3-CA38 plasmid distinct from the ColE3 \textit{imm} gene (Chak & James, 1984). The presence of two immunity genes in a single ColE plasmid exemplifies a new phenomenon relevant to the evolution of the ColE plasmids. In this report, we present the nucleotide sequence of the ColE8 \textit{imm} gene and its deduced amino acid sequence and these sequences are compared with the corresponding gene conferring resistance to colicin E2 (Lau et al., manuscript submitted) or colicin E3 (Masaki & Ohta, 1982; Mock et al., 1983). Possible mechanisms responsible for the existence of the ColE8 immunity gene in the ColE3 plasmid are discussed.

\textbf{Materials and Methods}

\textbf{Bacterial strains, plasmid and phage DNA}

Plasmid ColE3-CA38 carried in bacterial strain WA802 was purified by the 'non-amplifiable' technique as previously described (Watson & Visentin, 1980). Replicative form DNA of bacteriophage M13 mp7 and mp8 derivatives (Messing, 1981) used for cloning and sequencing purposes were obtained from PL Biochemicals Inc. The host strain for the M13 phages was JM103 (Messing, 1981).

\textbf{M13 clones and DNA sequencing}

The unfractionated DNA fragments of the ColE3-CA38 plasmid generated by either AluI, EcoRI, PvuII, or TaqI restriction digestions were cloned into the appropriate sites of the M13 mp7 or mp8 DNA by the shotgun approach (Messing et al., 1981) as previously described (Lau & Spencer, 1982). The phage supernatants (3- replicate aliquot from a 1.2-ml culture) were then dotted onto nitrocellulose papers and probed with \textsuperscript{32}P-nicked translated plasmid pBB56, a derivative of pBR322 containing the \textit{col} gene, the \textit{imm} gene, and the DNA region designated \textit{his} (Watson & Visentin, 1982) so that clones to be sequenced are restricted to the pertinent segment (Fig. 1). Sequencing was performed by the dideoxy procedure (Sanger et al., 1980) using the synthetic 14-mer universal primer as described by Lau & Spencer (1982) with the following modification: To extend sequences up to 450 nucleotides from a single DNA template, a second reaction was routinely carried out with lower dideoxyribonucleotide concentrations than those used in the first reactions for the short run (5-7 h). The whole reaction mixtures were then applied on to the 80-cm-long 6% polyacrylamide - 7M urea gel and high-voltage electrophoresis was allowed to proceed overnight (16-20 h).

\textbf{Results and Discussion}

\textbf{Location and identification of the ColE8 imm gene}

Fig. 1 shows a summary of the gene arrangements within the 1.0-kb and the 1.4-kb ClaI-PvuII restriction fragments of the ColE2 and ColE3 plasmids containing the non-homologous DNA segments of the two plasmids (Lau et al., submitted; Watson et al., 1984, and this
Fig. 1. Strategy for sequencing ColE8 imm gene (hatched area). The distribution, strandedness, and extent of the sequenced M13 templates relative to the ClaI-PvuII restriction fragment of the ColE3-CA38 (E3) plasmid are indicated by the horizontal arrows. For comparison, the gene organization of the ColE2-P9 (E2) plasmid within the 1.0kb ClaI-PvuII fragment is provided. The locations and nucleotide lengths of the imm genes, col-3'-ends of the colicin genes, and hic-5'-ends of the hic genes and the various intergenic spaces are as indicated. These sequences are presented in Fig. 2. The dotted lines delineate the relative positions of the imm genes. The locations of several endonuclease restriction sites are given as landmarks.

study). The ColE2 and ColE3 maps differ mainly in that the latter contains an additional 424-bp region (166+258) between the imm gene and a portion of the hic gene shown in the case of ColE3 plasmid responsible for the release of colicin from the colicin-producing cells (Watson & Visentin, 1982; Watson et al., 1984). The 258-bp region within the 424 insertional sequence in the ColE3 plasmid is designated the imm gene for the newly described colicin E8 (Cooper & James, 1984) considering the biological observations described by Chak and James (1984) and the sequence characteristics described below.

Located downstream from the ColE3 imm gene (Fig. 2) are two potential initiation starts, both GTG codons at positions 977 and 9839, each preceded by a potential ribosome-binding sequence (5'-AGGAG or 5'-GAGGTG) and both of the open reading frames terminating at position 1240 by the opal codon, TGA. Arbitrarily we have chosen the second GTG triplet as the ColE8 imm initiator codon because of the correspondence of the ensuing 3 codons (viz. Glu, Leu, and Lys) with the ColE2 imm sequence (Lau et al., submitted; Fig. 2, nucleotides 620-880); the remainder of the two sequences are also highly homologous. Our assignment of the ColE8 imm coding sequence is also consistent with the results obtained by transposon mutagenesis which localized the gene between the EcoRI and PvuII restriction sites of the ColE3 plasmid (Chak & James, 1984; see also Fig. 2).

Comparison with ColE2 and ColE3 imm structures

At the nucleotide level, the ColE2 and ColE8 imm genes are 67% conserved (Fig. 2) and among the 83 nucleotide changes there are more transitions than transversions, indicating that the nucleotide substitution is not random (Li, 1983). At the amino acid level, the
Fig. 2. Localization and nucleotide sequence of the ColE8 \textit{imm} gene in the ColE3-CA38 plasmid and comparison with the ColE2 \textit{imm} gene. Nucleotide numberings for both the ColE2 and ColE3 plasmids are indicated alongside the sequences and amino acid numberings are shown above or below the coding sequences. Identical nucleotides are marked by asterisks and conserved amino acid residues are boxed. Termination codons and potential ribosome-binding sites are also boxed. For the purpose of direct comparison, the sequences of the ColE2 \textit{imm} gene and its protein from nucleotide numbers 620-880 (in brackets) are duplicated directly underneath the ColE8 \textit{imm} sequence. Note that for maximum alignment, the threonine codon at position 29 of the ColE2 \textit{imm} sequence was deleted, as indicated by the symbol Δ.
two immunity proteins are 60% identical, notably the 24 residues at the carboxyl-terminus. An additional 18% of the amino acids are chemically similar by the criteria of Dayhoff et al. (1972). That there may be an evolutionary pressure to keep the amino acid sequence preserved is indicated by the presence of various silent mutations among the conserved residues. Furthermore, many of the amino acid changes between the two immunity proteins are of a conservative nature although they involve charge differences. For example, many of the lysine changes involve the conversions into glutamate or aspartate, and among the 5 arginine changes, 3 are converted to asparagine and 2 to glutamate. The resultant net charges at physiological pH of both the ColE2 and E8 imm proteins are calculated to be -8. This is in contrast to the net charge of -13 of the ColE3 imm protein, whose sequence is considerably different from the former two structures (Fig. 2). Secondary structure predictions using the empirical rules of Chou and Fasman (1978) estimate that the ColE2 imm protein contains 58% α-helix and 17% β-sheet whereas these structural contents are 51% and 27% respectively for the ColE8 imm protein. On the other hand, the ColE3 imm protein is predicted to be 20% α-helix and 41% β-sheet.

The homology of the ColE2 and ColE8 imm sequences parallels the relatedness of the ColE3 and the cloacin DF13 (van den Elzen, 1980) imm proteins which are 70% homologous to each other (Lau et al., submitted) but are insufficient to complement each other in the neutralization of the exogenously added colicin (Oudega et al., 1975; Jakes, 1982). The observation of homology of the ColE2 and ColE8 immunity proteins also predicts that the structure of the colicin E8 carboxyl-terminus will be closely related to that of colicin E2 (Lau et al., submitted) in an analogous manner as those displayed by the RNAase domains of colicin E3 (Masaki & Ohta, 1982; Mock et al., 1983) and cloacin DF13 (van den Elzen et al., 1983), which are 82% homologous.

Potential regulatory sequences for ColE8 imm gene

A possible promoter sequence is found on the 5'-side of the ColE8 imm coding region (Fig. 3a). The sequence 5'-GTTATAATT at nucleotides 920-928 is extremely homologous with the consensus bacterial promoter sequence of gTATAATg at the -10 region (Rosenberg & Court, 1979). At nucleotides 907-914, the sequence 5'-TATTGAGT is also in agreement with the consensus RNA polymerase recognition sequence of tGTTGACA at the -35 region (Rosenberg & Court, 1979). However, unlike the most frequently occurring spacer length of 17 bp (Stefano & Gralla, 1979), i.e. the distance between the -10 and the -35 sequences, the spacer length in
this case is only 8 bp. Whether this aberrant feature affects the RNA polymerase-promoter interactions for an efficient transcription (Stefano & Gralla, 1982; Ackerson & Gralla, 1982) of the ColE8 imm message remains to be seen. If the assignment of the promoter is correct, mRNA transcription would be expected to initiate at one of the A residues at positions 936 or 937. As shown in Fig. 3a, the nucleotide sequence in this region could be folded into an AT-rich hairpin loop structure.

Another interesting structural feature is shown in Fig. 3b. The sequence between nucleotides 1246 and 1281 (886-921 of the ColE2 numbering) can be folded into a reasonably stable hairpin loop and is followed by the sequence 5'-TTCTTT. This conformation resembles a bacterial transcription terminator (Rosenberg & Court, 1979) but again this needs to be demonstrated biochemically. It will be of considerable interest to see how the ColE8 imm gene is regulated relative to the expression of the ColE3 imm gene which is constitutive as well as under the control of the colicin operon, the latter being inducible by DNA-damaging agents such as mitomycin C (Tyler & Sherratt, 1975; Jakes, 1982; Mock et al., 1983; Watson et al., 1984).

Fig. 3. Possible regulatory sequences at the (a) 5'- and (b) 3'- regions flanking the ColE8 imm coding sequence. See text for explanation. The energies of the hypothetical hairpin structures are calculated according to Salser (1977). The numbers in brackets in (b) refer to the ColE2 sequence numbering, the sequences in this DNA region being identical for both plasmids (see Fig. 2).
Molecular basis for a double immunity gene plasmid

The manner in which the second copy of the \textit{imm} gene may have come about in the ColE3 plasmid is a matter of speculation. One possibility is gene duplication followed by different rates of mutations or divergence and to consider deletion of one of the copies in the case of the 'one immunity' plasmids such as ColE2 and ColE8. The possibility of a deletion formation (Albertini et al., 1982; Jones et al., 1982; N. Kleckner, pers.comm.) is suggested by the presence of the unique short direct repeated sequence of 6 bp, 5'-AAATAT-3', flanking the ColE2, E3, and E8 \textit{imm} genes (Fig. 2). On the other hand, two inverted repeats (I.R.1 and I.R.2) are found to border the ColE2 \textit{imm} gene (Fig. 2). While these features are reminiscent of a transposon-like insertion mechanism (Kleckner, 1981), no evidence is available in this study to support this hypothesis, although the role of translocating genetic elements and DNA sequence insertions in the structural evolution of bacterial plasmids has been well documented (Cohen, 1976; Kopecko, 1980). Luria and Suit (1982) have also suggested that the colicin gene and its whole regulatory apparatus may have originated from the bacterial chromosome.

Concluding Remarks

Although the type of recombination or gene exchange events which might generate the second immunity gene in the ColE3-CA38 plasmid remains unknown, this may represent a unique case of a plasmid 'caught' in the evolutionary process. The origin of the intergenic space between the ColE3 and ColE8 \textit{imm} genes also remains to be clarified. Nonetheless, the exchange of genetic information seems to have occurred between two plasmids native to two different species, \textit{E. coli} and \textit{Shigella sonnei}. Considering the origins of these plasmids and the recent evidence provided by Olsson et al. (1983) that the naturally occurring chromosomal ampicillin resistance in \textit{E. coli} may have evolved by horizontal transfer of the \textit{ampC} DNA from \textit{Shigella} species, genetic exchange between \textit{Escherichia} and \textit{Shigella} may be greater than had been previously realized and the traditional classification system for these genera may need questioning.

Acknowledgements

We thank Dr. A. Vaisius for a critical reading of the manuscript and Dr. M. Zuker and his colleagues for help in computer analyses. We are also grateful to H. Turner and B. Kane for photography and L. Boucher for excellent secretarial assistance.

References