Nucleotide sequence for the catalytic domain of colicin E3 and its immunity protein. Evidence for a third gene overlapping colicin

Michèle Mock^{1,2}, C.Garrett Miyada and Robert P.Gunsalus

Department of Microbiology and Molecular Biology Institute, University of California, Los Angeles, CA 90024, USA

Received 30 March 1983; Accepted 3 May 1983

ABSTRACT

We have determined the nucleotide sequence of a segment of ColE3 DNA coding for the carboxyl-terminal, catalytic peptide of colicin E3 and for the immunity protein. The end of the colicin E3 gene is separated from the beginning of the immunity gene by a nine-basepair intercistronic region, suggesting the two genes are expressed as a single transcriptional unit. The immunity gene is expressed, however, in *E.coli* strains containing ColE3-pBR322 hybrid plasmids deleted for the 5'-end of the colicin gene. The DNA sequence also contains an unexpected open reading frame (ORF). This ORF is contained within the colicin gene and is in the +1 reading frame with respect to that gene. Plasmids containing the ORF, directed the synthesis of an 11 kilodalton protein in a cell-free, transcription-translation system.

INTRODUCTION

Plasmids that direct the production of bacteriocins also confer immunity against the action of those bacteriocins (1). The mechanism of immunity was first established in the case of colicin E3 (2,3), a 67 kilodalton (K) protein that kills sensitive bacteria by catalyzing a specific cleavage in 16S rRNA. A 9K "immunity protein" inhibits the ribonuclease activity of the colicin by binding to its carboxyl-terminal, catalytic domain (4). The same immunity mechanism prevails in cloacin DF-13, which has the same enzymatic activity as colicin E3 (5), and probably in several other colicins (6). A similar immunity mechanism prevails for colicin E2, which is a deoxyribonuclease (7). The availability of a series of catalytically similar bacteriocins, such as colicins E3, E5, and E6, and cloacin DF-13, each sensitive to a particular immunity protein, provides a fascinating system to study protein-protein interactions. In the case of colicin E3, the amino acid sequences of both the immunity protein and carboxyl-terminal catalytic peptide have been recently determined (8,9).

In E.coli cells containing the ColE3-CA38 plasmid, the immunity protein is expressed, even when colicin synthesis is repressed. Within the population, there are some cells, which are spontaneously induced for both colicin and immunity production. Activation of the SOS system of the cell (10) will also cause the induction of both colicin and its immunity protein. Even though the immunity gene is closely linked to the colicin gene (11, this work), it may be expressed independently as well as simultaneously with colicin. The differential expression of colicin E3 and its immunity protein may be explained either by the existence of two separate promoters or differential mRNA processing. Determining the nucleotide sequence of the colicin and immunity genes constitutes an initial step in unraveling this regulatory mechanism.

In this paper we have cloned restriction fragments of ColE3 DNA into pBR322 to identify sequences required for the expression of the colicin and its immunity protein. The nucleotide sequence was determined for a 650 base pair (bp) DNA segment that encodes both the catalytic domain of colicin E3 and its immunity protein. The organization of the two genes suggests regulatory mechanisms which control their expression. When ColE3-pBR322 hybrid plasmids were used to direct an in vitro protein synthesis system, bands corresponding to colicin E3 and its immunity protein were identified in autoradiograms of SDS polyacrylamide gels. We identified a third protein in this system, which is immunologically-unrelated to either colicin E3 or its immunity protein but is also encoded by this region of DNA.

An article by Masaki and Ohta came to our attention while we were preparing this manuscript (12). These authors also established the sequence of a DNA segment encoding the catalytic domain of colicin E3 and the immunity protein. In addition to confirming their sequence we provide new informations regarding the expression of immunity, and regarding the existence of a third gene overlapping with that of the colicin.

MATERIALS AND METHODS

Bacterial strains and plasmids

Escherichia coli strain W3110 was used as a host for wild type ColE3 plasmid (13). Strains PK1022 (14) and 71-18 (15) were used as hosts for transformation and transfection respectively. The recombinant plasmids constructed in this work are described in Fig.1.

Media and Chemicals

The bacteria were grown in L-broth or in minimal M9 medium as described (13). When required, antibiotics were added at the following final concentrations: ampicillin, 100 μ g/ml; tetracycline, 20 μ g/ml. The restriction enzymes EcoRI, KpnI and Sau3AI and T4 DNA ligase were obtained from New England Biolabs. Restriction enzyme ClaI and DNA polymerase I were from Boehringer Mannheim. Oligodeoxyribonucleotides used as primers in DNA sequencing were purchased from Collaborative Research, Inc.

Preparation and analysis of phage and plasmid DNA

Small amounts of phage M13 replicative form (RF) DNA and of plasmid DNA were prepared according to Birnboim and Doly (16). Large amounts of plasmid DNA were extracted by the SDS-NaCl technique of Guerry *et al* (17) and purified by isopycnic ultracentrifugation in cesium chloride and ethidium bromide. Large amounts of phage RF DNA were prepared as previously described (18). DNA analysis with restriction endonucleases was performed using the reaction mixtures recommended by the manufacturers.

Isolation of DNA restriction fragments and cloning in plasmid or phage vectors.

The methods describing isolation of DNA restriction fragments from agarose or acrylamide gels, ligation reactions conditions, transformation and transfection have been described (18).

Deletion of the ClaI-PvuII fragment of pLAX4 was performed as follows: the plasmid was cleaved with ClaI, the resulting ends were filled in using DNA polymerase I (19), a secondary cut was performed with PvuII, and blunt end ligation was performed.

Procedures for testing strains for colicin production or immunity have also been reported elsewhere (6).

DNA sequencing

DNA was sequenced using the dideoxy method (20) by cloning isolated restriction fragments into the single-stranded DNA cloning vectors M13mp8 and M13mp9 (21). Phage inserts were verified by using direct gel electrophoresis of culture supernatant, by C-tests (BRL DNA sequencing manual) or by analyzing restriction digests of phage RF DNA.

Phage DNA was prepared for sequencing as previously described (22). M13mp8 and M13mp9 RF's were the gift of Dr. David Sheppard. Chemically synthesized oligodeoxyribonucleotides, a dodecamer, 5'(TCACGACGTTGT) 3', and a septadecamer, 5' d(GTAAACGACGCCAGT)3' were used as primers. Poly-

acrylamide (5 and 6%) gels were prepared as described (23) and the gel matrix was covalently attached to one glass plate using γ -methacryloxypropyl-trimethoxysilane (24). After electrophoresis gels were acid fixed, dried, and autoradiographed at room temperature usually overnight.

In vitro protein synthesis

In vitro synthesis, with various plasmid DNAs as templates, was conducted as described by Gunsalus *et al* (25).

The [35 S] methionine labeled proteins were analyzed by gel electrophoresis on 14% polyacrylamide gels and autoradiography (13), either directly or after precipitation with an antiserum (26) against the colicin E3-immunity protein purified complex.

Figure 1. Structure of pBR322-derived plasmids containing different ColE3-CA38 DNA inserts.

The plasmid DNA sequences located between the unique PvuII and PvuII restriction sites of pBR322, containing the bla and tet determinants, are diagrammed. The remaining portions of the vectors are not represented since they are identical in all of the plasmids. The notation tet' in pLAX4 and pLAX4Δimm6 indicates that the tet determinant is partially deleted. The positions of restriction sites within the inserts are from Visentin and Watson (11) and the present work. The arrow indicates the direction of transcription for colicin E3 and the immunity gene. Plasmid pAPBZ101 was obtained by cloning a 3kb EcoRI fragment from ColE3-CA38 into the EcoRI site of pBR322 (6). Plasmid pLAX821 was derived from pAPBZ101 by deleting the 2kb Aval restriction fragment. Construction of pLAX4 involved the cloning of the 1kb ClaI-EcoRI restriction fragment of pAPBZ101 between the ClaI and EcoRI sites of pBR322. Plasmid pLAX4Δimm6 was obtained after limited Bal 31 digestion of pLAX4 opened at the EcoRI site. Deletion Δimm6 is 369 bp long and extends into the immunity gene (Mock, Miyada and Cottier, manuscript in preparation).

RESULTS

Colicin E3 and immunity protein are encoded by the ColE3-CA38 plasmid. A 3 kilobase pair (kb) EcoRI restriction fragment of this plasmid containing the determinants for colicin and immunity production has been cloned into pBR322 (6). We found a single ClaI site in this insert and this site was used for subcloning a 1000 bp ClaI-EcoRI restriction fragment into pBR322. The resulting plasmid was designated pLAX4 (fig.1). E.coli strains carrying pLAX4 were immune to the same amounts of colicin E3 as those carrying a wild type ColE3-CA38 plasmid (6). Evidence was obtained that this expression did not result from readthrough transcription from the "anti-tet promoter" called PI (28) present in pBR322. Indeed, the expression of immunity still occurred when pLAX4 was deleted for the pBR322 sequences present between the ClaI and PvuII sites. The positions of Sau3AI, KpnI and RsaI restriction sites within the 1000 bp EcoRI-ClaI insert were determined (Fig.2). These sites were utilized for subcloning into the single-stranded DNA vectors M13mp8 and M13mp9 (21). The entire EcoRI-ClaI segment was then sequenced (Fig.3) by the dideoxytechnique. The position of the immunity gene (ceiC) and the carboxyl-terminus of the colicin E3 gene (ceaC) were identified by correlation with known amino acid sequences (9). In comparing the amino acid and nucleotide sequences, we observed only one difference in each gene. Instead of aspartic acids,

Figure 2. Restriction map of the 1000 bp ClaI-EcoRI restriction fragment of ColE3-CA38 DNA.

Distances, given in bp from the center of the EcoRI restriction site, are numbered below the restriction map. Horizontal arrows below the map indicate the direction of sequencing as performed on restriction fragments cloned into M13mp8 or M13mp9. Regions coding for the colicin E3 gene (ceaC) and the immunity gene (ceiC) are shown above the restriction map.

asparagine codons were found at amino acid position 45 in the immunity protein (8) and at position 36 in the catalytic peptide (9).

By localizing the positions of the two genes within the nucleotide sequence, we showed that the immunity gene lies in the 3' direction with respect to colicin E3 as previously suggested (11). The stop codon of colicin and the start codon of immunity are separated by nine basepairs. This nine basepair segment contains the sequence, 5'-GAGG-3' (boxed sequence, Fig.3), which shows considerable homology with the Shine-Dalgarno sequence (27). The Shine-Dalgarno sequence is complementary to the 3'-OH end of 16S-rRNA and is thought to aid in the initiation of translation. The DNA sequence shown in Fig.3 also contains an open reading frame (ORF). This ORF overlaps with the sequence coding for the catalytic peptide of the colicin (nucleotides 789 to 513, Fig.3). If expressed, the ORF would code for a 91 residue polypeptide (italicized print, Fig.3). The ORF starts with an AUG codon but no identifiable Shine-Dalgarno

Figure 3. Nucleotide sequence of the ClaI-EcoRI restriction fragment between nucleotides 850 and 200.

The numbers in the left hand margin indicate the position of the first nucleotide in each line. The nucleotides are numbered as in Fig.2, with respect to the EcoRI restriction site. Only the "sense" strand of DNA is shown. The unique KpnI restriction site is underlined and a potential Shine-Dalgarno sequence for the immunity gene is boxed. The predicted amino acid sequences for colicin E3 and the immunity protein are shown above the coding sequence. The amino-terminal end of the carboxyl-terminal tryptic peptide of colicin E3(4) is indicated by an arrow (nucleotide 769). An additional open reading frame was found starting at nucleotide 789 and overlapping with the sequence encoding the catalytic peptide of colicin E3. The predicted amino acid sequence for this open reading frame is shown in italics above the nucleotide sequence.

sequence is found at the usual distance from the start codon (6 to 9 nucleotides).

Plasmid pLAX4 was found to direct the synthesis of two proteins that are not seen when the control vector pBR322 was used as a template (Fig.4, lanes 3 and 1, respectively). One of these proteins (9K) was identified as the immunity protein. The plasmid pLAX4 Δ imm6 which contains a 3'-deletion of the immunity gene, was unable to direct the synthesis of the same protein (Fig.4, lane 4). The 9K protein was also identified as immunity protein immunologically since it was precipitated with antiserum against the colicin E3-immunity protein complex (Fig.5, lane B').

The synthesis of the second protein (11K) was directed by pLAX4 as well as pLAX4 Δ imm6 and pLAX821 (fig.4, lanes 2,3 and 4). Since the

Figure 4. *In vitro* translation products from ColE3-pBR322 plasmids.

In vitro protein synthesis was performed as described in "Materials and Methods" using as a DNA template either pBR322 (lane 1), pLAX821 (lane 2), pLAX4 (lane 3) or pLAX4 Δ imm6 (lane 4). [³⁵S]Methionine-labeled polypeptides were run on an SDS polyacrylamide gel (14%) and autoradiographed.

inserts present in these three plasmids all have different junctions with pBR322 DNA, the 11K protein observed cannot be the product of a hybrid segment containing sequences from both the insert and the vector. Examination of these plasmids for common insert sequences shows that this protein must be encoded between nucleotides 200 and 1000 in the EcoRI-ClaI fragment (Fig.2). The 11K protein is also not immunoprecipitated with the anti-serum against the colicin E3 immunity complex (Fig.5, lanes A' and B'), even though this antiserum can precipitate trypsin-digestion products of colicin E3 (29 and data not shown). Therefore, the 11K protein is probably not a degradation or a processing product of colicin E3. According to its size, the 11K protein is the product of the 91 amino acid-ORF contained in the ClaI-EcoRI insert. The only other significant ORF within this region of DNA would code for a 68 amino acid polypeptide. Provisionally, we have named the product of this open reading frame ET (for colicin E3 Terminal). By comparing the intensities of the bands corresponding to ET and to the immunity protein (11K and 9K, respectively, Fig.4) and taking into account the methionine content of these two proteins (five and one,

Figure 5. Immunoprecipitation of *in vitro* translation products using anti-colicin immunity protein complex serum.

The proteins synthesized *in vitro* from plasmids pLAX821 or pLAX4 were analyzed by gel electrophoresis and autoradiography either directly (lanes A and B respectively) or after immunoprecipitation (lanes A' and B' respectively) with antiserum against the colicin-immunity protein complex.

respectively), we estimate that ET is synthesized at about one-tenth the level of the immunity protein in our in vitro system.

DISCUSSION

The sequence established in this work is identical with that determined, by using a different method of sequencing, by Masaki and Ohta (12). These authors also compare the sequence of the immunity genes present on the ColE3-CA38 and Clo DF13 (30) plasmids and therefore this does not require further comments.

The DNA sequence shown in Fig.3 is very A+T-rich (62%). This is reflected at the level of codon usage, which is somewhat unusual and shows a strong bias for the NNU vs the NNC codons (92 NNU vs 13 NNC in the DNA sequences coding for the immunity protein and the catalytic peptide of colicin, Fig.3). Grosjean and Fiers (31) proposed rules in the selection of the synonymous codons, NNU or NNC, which are usually recognized by the same tRNA species. In highly expressed genes, for each given amino acid, an NNU or NNC codon is preferentially used. This preference, which is correlated with the base composition of the NN dinucleotide, would allow an optimization of codon-anticodon interaction energy and consequently facilitate the efficiency of translation. These rules do not seem to apply to weakly expressed genes where codon usage is more random. Highly or weakly expressed E.coli genes (31) have an average ratio of NNU vs NNC codons of 1.4 or 1.2, respectively. We observe a ratio of 7.0 (92/13) for the catalytic peptide and immunity protein of colicin E3. The significance of this bias is not known, but it could be related to the apparent existence of pauses in the translation of several colicin genes including that of colicin E3 (32).

The organization of the genes coding for colicin E3 and its immunity protein is the same as reported for cloacin DF13 (30) and probably colicin E2 (A.Pugsley, personal communication). Indeed, the distance between colicin and immunity genes, nine basepairs, is too short to contain a transcription termination signal. The two genes may be part of a single transcriptional unit, controlled by the SOS system of the cell (10). We have also shown, however, that plasmid pLAX4, which does not contain the 5' end of colicin gene, still expresses the immunity gene at wild-type levels. This was shown in vivo, since E.coli strains containing pLAX4 were equally immune to exogenously added colicin E3 as strains containing the ColE3-CA38 plasmid (Mock, Miyada and Collier, manuscript in preparation),

and in vitro (Fig.4, compare lanes 2 and 3). The expression of immunity still occurred when the P1 promoter of pBR322 was deleted. The remaining endogenous pBR322 promoters (28) are then in the wrong orientation to cause immunity expression by initiating transcription into the DNA insert. This suggests the existence of a second promoter, located within the colicin gene, for immunity expression. The same promoter may be responsible for the expression of immunity in cells which are not induced for colicin production.

The most intriguing part of our results concerns the presence of a 91 codon ORF which overlaps with the sequence encoding the catalytic portion of the colicin. The lack of an identifiable Shine-Dalgarno sequence suggests that if this ORF is expressed, it would be at a low level. Moreover, several of the codons present in this ORF correspond to minor tRNA's and this would also be expected to lower its expression. For example, five of six arginine residues are coded by AGA or AGG, both of which are rarely found in E.coli genes (31). It should be noted, however, that the same rare codons are also found in the sequence encoding colicin E3 itself (see the stretch of basic residues upstream from the catalytic peptide, Fig.3) and in the sequence coding for colicin E1 (33). Experiments reported here show that the ORF is expressed in vitro. Whether expression occurs in vivo and whether the product has any function remains to be established.

ACKNOWLEDGEMENTS

C.G.M. thanks G.Wilcox for support during this work. This work was performed while one of us (M.M.) was working in the laboratory of J.Collier, who is gratefully acknowledged for his constant interest in this work and many helpful discussions. We thank J.Collier and M.Schwartz for help in preparing the manuscript.

This work was supported by National Institute of General Medical Science Grant GM 30896, National Science Foundation Grant PCM78-26255 (G.W.). National Institute of Health Grant AI-07877 (R.J.C.). National Institute of Health Grant GM 29456 (R.P.G).

¹Present address: Institut Pasteur, Département de Biologie Moléculaire, 25 rue du Dr.Roux, 75724 Paris Cedex 15, France

²To whom reprint requests should be addressed

REFERENCES

- (1) Fredericq, P. (1958) *Symp. Soc. Exp. Biol.* 12, 104-122.
- (2) Jakes, K., Zinder, N.D. and Boon, T. (1974) *J. Biol. Chem.* 249, 438-444.
- (3) Sidikaro, J. and Nomura, M. (1974) *J. Biol. Chem.* 249, 445-453.
- (4) Ohno, S., Ohno-Iwashita, Y., Suzuki, K. and Imahori, K. (1977) *J. Biochem.* 82, 1045-1053.
- (5) De Graaf, F.K. and Klaasen-Boor, P. (1974) *FEBS Lett.* 40, 293-296.
- (6) Mock, M. and Pugsley, A. (1982) *J. Bacteriol.* 150, 1069-1076.
- (7) Schaller, K. and Nomura, M. (1976) *Proc. Natl. Acad. Sci. USA* 73, 3989-3993.
- (8) Mochitate, K., Suzuki, K. and Imahori, K. (1981) *J. Biochem.* 89, 1609-1618.
- (9) Suzuki, K. and Imahori, K. (1978) *J. Biochem.* 84, 1031-1039.
- (10) Witkin, E.M. (1976) *Bacteriol. Rev.* 40, 869-907.
- (11) Watson, R.J. and Visentin, L.P. (1982) *Gene*, 19, 191-200.
- (12) Masaki, H., and Ohta, T. (1982) *FEBS Lett.* 149, 129-132.
- (13) Mock, M. and Schwartz, M. (1977) *FEMS Microbiol. Lett.* 2, 181-184.
- (14) Charnay, P., Louise, A., Fritsch, A., Perrin, D. and Tiollais, P. (1979) *Mol. Gen. Genet.* 170, 171-178.
- (15) Hofschneider, P.H. (1977) *Proc. Natl. Acad. Sci. USA* 74, 3642-3646.
- (16) Birnboim, H.C. and Doly, J. (1979) *Nucl. Acids Res.* 7, 1513-1523.
- (17) Guerry, P., Le Blanc, D.J. and Falkow, S. (1973) *J. Bacteriol.* 116, 1064-1066.
- (18) Miyada, C.G., Soberon, X., Itakura, K. and Wilcox, G. (1982) *Gene* 17, 167-177.
- (19) Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular cloning*, Cold Spring Harbor Laboratory, New York, N.Y.
- (20) Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
- (21) Messing, J. and Vieira, J. (1982) *Gene* 19, 269-276.
- (22) Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) *J. Mol. Biol.* 143, 161-178.
- (23) Anderson, S. (1981) *Nucl. Acids Res.* 9, 3015-3027.
- (24) Garoff, H. and Ansorge, W. (1981) *Anal. Biochem.* 115, 450-457.
- (25) Gunsalus, R.P., Zurawski, G. and Yanofsky, C. (1979) *J. Bacteriol.* 140, 106-113.
- (26) Mock, M. and Schwartz, M. (1978) *J. Bacteriol.* 136, 700-707.
- (27) Shine, J. and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* 71, 1342-1346.
- (28) Stüber, D. and Bujard, H. (1981) *Proc. Natl. Acad. Sci. USA* 78, 167-171.
- (29) Mock, M. and Schwartz, M. (1980) *J. Bacteriol.* 142, 384-390.
- (30) Van den Elzen, P.J.M., Gaastra, W., Spelt, C.E., de Graaf, F.K., Veltkamp, E. and Nijkamp, H.J.J. (1980) *Nucl. Acids Res.* 8, 4349-4363.
- (31) Grosjean, H. and Fiers, W. (1982) *Gene* 18, 199-209.
- (32) Varenne, S., Knibiehler, M., Cavard, D., Morlon, J. and Lazdunski, C. (1982) *J. Mol. Biol.* 159, 57-70.
- (33) Yamada, M., Ebina, Y., Miyata, T., Nakazawa, T. and Nakazawa, A. (1982) *Proc. Natl. Acad. Sci. USA* 79, 2827-2831.