Comparative nucleotide sequences encoding the immunity proteins and the carboxyl-terminal peptides of colicins E2 and E3*

Peter C.K.Lau**, R.Wally Rowsome, Michael Zuker+ and Louis P.Visentin

Molecular Genetics Group and +Biomathematics, Division of Biological Sciences, National Research Council of Canada, 100 Sussex Drive, Ottawa, Ontario K1A OR6, Canada

Received 18 May 1984; Revised and Accepted 18 October 1984

ABSTRACT

Using the M13 dideoxy sequencing technique, we have established the DNA sequences of colicins E2 and E3 which encompass the receptor-binding and the catalytic domains of each of the nucleases, and their immunity (imm) genes. The imm gene of plasmid ColE2-P9 is 255 bp long and is separated from the end of the col gene by a dinucleotide. This gene pair is arranged similarly in plasmid ColE3-CA38 except that the intergenic space is 9 bp and the E3 imm gene is one codon shorter than its E2 counterpart. Comparisons of the E2 and E3 imm sequences indicate considerable divergence whereas the receptor-binding domains of both colicins are highly conserved. The two nuclease domains appear to share some sequence homology. A possible evolutionary relationship between colicin E3 and other microbial extracellular ribonucleases is also suggested from the sequence alignment analysis.

INTRODUCTION

The plasmid-coded bactericidal proteins, colicins E2 and E3 and the E3-related cloacin DF13, constitute an interesting group of toxins in that each protein with an approximate M_r of 60,000 contains information for i) translocation across cell membranes, ii) receptor binding on the surface of sensitive E. coli cells and iii) nucleolytic activities (for recent reviews, see ref. 1-3). Colicin E2 is an endonuclease active on both single and double stranded DNA but with undefined specificity (4,5) whereas colicin E3 and cloacin DF13 cleave the 16S ribosomal RNA specifically near the 3' terminus (6-8). Although colicins E2 and E3 have different biochemical targets, they share a common receptor protein (9), a btu B gene product utilized also by vitamin B12 and phage BF23 (10,11). Each bacteriocin is released from the producing cells as an equimolar complex with another plasmid-coded product known as the 'immunity' protein whose primary function is to protect the colicinogenic cells from the lethal action of its own toxin (4,12,13). The basis for the immunity system has been attributed to the specific electrostatic interactions of the acidic immunity protein with the basic nuclease domain localized at the C-terminus of each colicin (5,14-18).

Because the neutralization of the toxin action requires specific interactions between the colicin and an immunity protein of a 'homologous' but not that of a 'heterologous' system (1,2,13), the various colicin-immunity proteins and their genes (19,20) offer an excellent system to study protein-protein interactions and protein/gene evolution.

Previous studies from this laboratory have localized the genes for the immunity proteins and the 3'-ends of the neighbouring colicin E2 and E3 genes to a seemingly non-homologous DNA region on the two producing plasmids, ColE2-P9 and ColE3-CA38 respectively (21-23). In this report we present a comparison of the nucleotide sequences and the deduced amino acid sequences for the C-terminal peptides of colicins E2 and E3 and for their immunity proteins. After the completion of this work, two reports (24,25) came to our attention establishing part of the E3 nucleotide sequence presented here. In addition to confirming their sequences we extend the sequence information towards the receptor binding domain of the colicin E3.

MATERIALS AND METHODS

Plasmids and bacterial strains

Plasmids ColE2-P9 and ColE3-CA38 carried in E. coli WA802 strain were used and purified as previously described (21). Bacteriophage M13 mp7, mp8 and mp9 replicative form DNAs (26) were purchased from PL Biochemicals. E. coli strain JM103 was used as recipient for the M13 phages (26).

Enzymes and reagents

Restriction endonucleases and T4 DNA ligase were obtained either from New England Biolabs or Bethesda Research Labs. Inc. (BRL). Dideoxynucleoside triphosphates were from PL Biochemicals and deoxynucleoside triphosphates and DNA polymerase I (Klenow fragment) were from Boehringer Mannheim. [α-^{32}P]-dATP was from New England Nuclear or Amersham corporation.

DNA sequencing

The M13-dideoxy procedure was used (27). The cloned templates of ColE2-P9 restriction fragments in M13 mp8 and mp9, and ColE3-CA38 restriction fragments in M13 mp7 and mp8 (Fig. 1) were prepared and sequenced using the universal synthetic 14-mer primer as previously described (28). To eliminate the 'smile' effect along the 80 cm-long sequencing gels (BRL Model S1) during high voltage electrophoresis, an 1/8 inch aluminum sheet was clamped along the length of one side of the sequencing gel plate. Sequencing data were assembled with the aid of the BRL Nucleic Acid Analyzer 2. The amino acid sequence of the nuclease-active T2A fragment of colicin E3 previously

determined by Suzuki and Imahori (29) aided in positioning and orientation of one of the Alu I clones (Fig. 1). Otherwise the sequences across all those restriction cleavage sites used for cloning were determined. In all cases, the nucleotide sequences were obtained from at least 2 independently generated ladders.

RESULTS AND DISCUSSION

Locations of the col and imm genes

The approximate positions of the col and imm genes of plasmids ColE2-P9 and ColE3-CA38 have been defined previously (22,23). The colicin E3 gene is localized within the 2.1 kb HincII-KpnI fragment adjacent to the imm gene which contains the unique KpnI restriction site (Fig. 1). In ColE2-P9, these two genes are within the 2.5 kb HincII-PvuII restriction fragment which also encompasses the region of the plasmid that is apparently not homologous to the E3 DNA (23,33). In this study and independently established by Masaki and Ohta (24), an unique ClaI site was found within the HincII-PvuII fragment of both plasmids so that the presumptive non-homologous regions are further confined to the ClaI-PvuII-cleaved fragments. Using the strategy depicted in Fig. 1 we have sequenced these latter fragments which are 1058 bp and 1395 bp in length for the E2 and E3 plasmids respectively. Portions of these sequences containing the structural genes for the specific immunity proteins and the 3'-ends of the respective colicins are presented in Fig. 2. The sequences near the PvuII ends have been reported elsewhere (30,32).

Sequence comparison of the 3'-coding regions of colicins E2 and E3

i) The nuclease domains. The nucleotide sequences and the deduced amino acid sequences for the C-terminal nuclease domains of colicins E2 and E3 are compared in Fig. 2. Identification of the reading frame for colicin E2 (nts 1-617) was established with the aid of the previously determined amino acid composition of (his,arg,gly)-lys-COOH for the colicin E2 C-terminus (17). In addition, the amino acid composition of the peptide segment spanning from gly85 to lys204 are in good agreement with that determined for the C-terminal tryptic fragment of colicin E2 (designated E2-T2A, Table 1) which retains the DNAase activity of the intact protein (14). This tryptic fragment was also found to contain a glycine residue at its amino-terminus (17).

Similar to the RNAase domain of colicin E3 (amino acid residues 86-182 in Fig. 2; also ref. 24,25,29) the DNAase domain of colicin E2 is characterized by high asp+glu (17.5%) and lys+arg (24%) contents resulting in a net positively-charged polypeptide. The lysine and arginine residues are

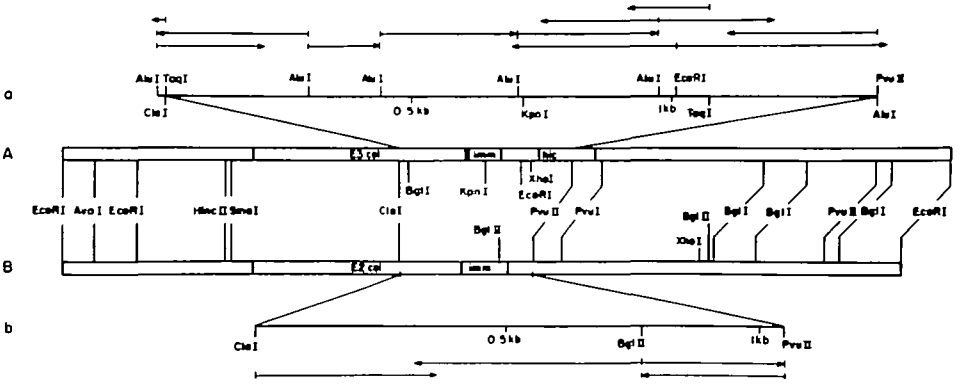

Figure 1. Summary of sequencing strategy. The gene maps of the 7.2 kb ColE3-CA38 plasmid (A) and the 6.8 kb ColE2-P9 plasmid (B) are aligned by their common restriction sites (21). The approximate locations of the colicin (col) and immunity (imm) gene pairs and the hic gene which has been shown to be responsible for high production of colicin E3 (22,23,30) are as indicated. Not shown in the E3 map is the location of the newly-characterized ColE8 imm gene (31) which contains the unique XhoI site of the E3 plasmid (32). The distribution, strandedness and extent of sequenced templates relative to the ClaI-PvuII fragments of both plasmids are shown in a and b.

also often found in pairs or its multiples. In both nucleases, the proline contents are relatively high (6-8%) and these are often associated with lysine and/or aspartate. These occurrences in the E2 sequence are at positions 86-87, 105-107, 133-136, 174-176 and 194-196. In the E3 sequence, three (pro-lys) and one (lys-pro) dipeptides are clustered between residues 99 and 119. Further occurrences of this kind are towards the C-terminus, at positions 162-164 and 172-176. Clusters of basic amino acids and proline have been found in the Semliki forest virus nucleocapsid protein (35) and other nucleic acid-associated protein sequences (36 and ref. therein). In all these cases, it was assumed that interaction of the nucleoprotein with RNA or DNA may act via clusters of basic amino acids and possibly with proline.

Despite the different catalytic activities of the E2 and E3 nucleases, by careful alignment of their domain sequences, two striking features are revealed (Fig. 2). One is the apparent homology between the two nucleases and the other, an insertion or deletion of some 30 amino acid residues within the E2 and E3 sequences respectively. About 20% of the two amino acid sequences are identical, 6 residues are chemically similar and an additional

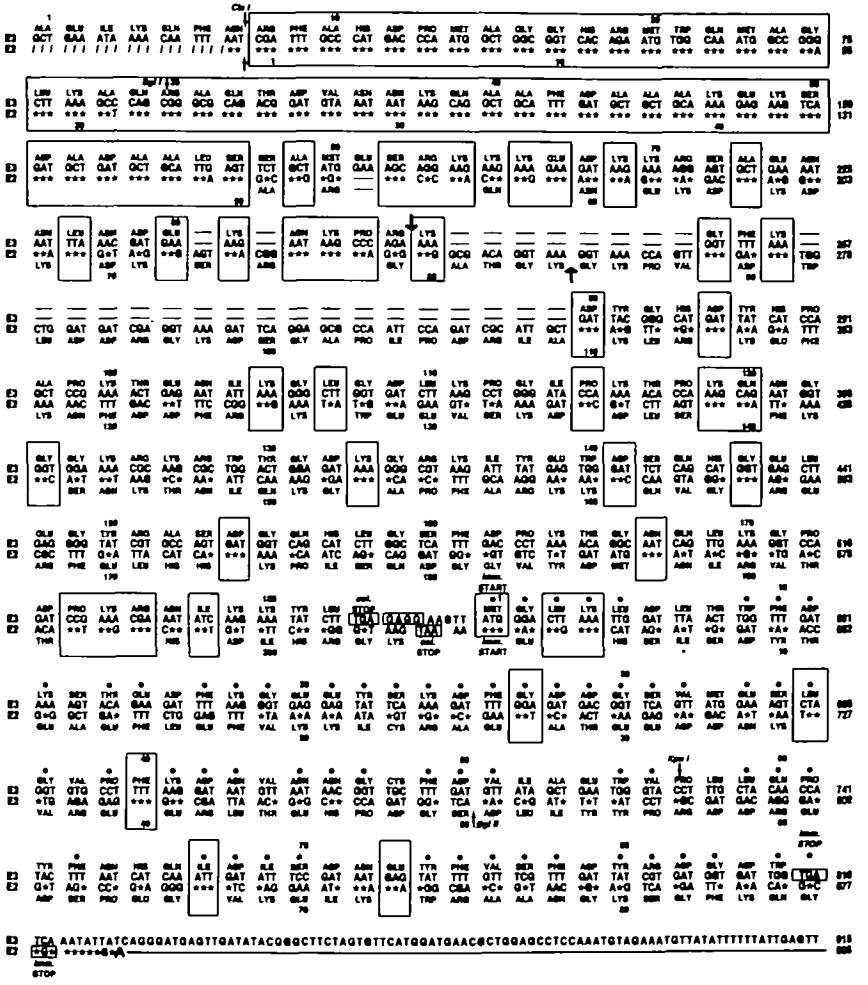

Figure 2. Nucleotide and predicted amino acid sequences of the immunity proteins and the C-terminal peptides of colicins E2 and E3. Identical nucleotides between the E2 and E3 sequences are marked by asterisks and conserved amino acid residues are boxed, maximum alignment of the two sequences being achieved by allowing various deletions indicated by double solid lines. Nucleotide numberings are indicated alongside the sequences and the amino acid residues are numbered from the coding open reading frame or the start codon of the relevant polypeptide. Termination codons and potential ribosome-binding sequences are also boxed. The solid circles above the E3 immunity protein sequence indicate those amino acid residues which are homologous to the plasmid ClODF13 immunity protein (34). Throughout the sequences, several unique restriction sites, indicated by arrows, are given as landmarks. The predicted cleavage site of colicin E2 by trypsin (see text) and the known tryptic fragmentation site of colicin E3 (29) are indicated by the thick arrows.

Table 1. Amino acid compositions of predicted colicin E2 nuclease domain and its immunity protein. The compositions predicted from nucleotide sequencing are compared to those determined for the colicin E2-T2A fragment (16) and for the purified immunity protein by Jakes (1) and Watson *et al.* (23).

	Nuclease domain		Immunity protein		
Amino acid	Trypsin T2A fragment	DNA sequence gly_{88}-lys_{204}	Proposed Jakes	Proposed Watson *et al.*	DNA sequence
Ala	6	4	5	7	5
Arg	10	10	6	6	6
Asp	] 26	16	]10	]11	8]10
Asn		5] 21			2
Cys	-	0	1	1	1
Glu	] 11	5	]18	]15	14]15
Gln		4]9			1
Gly	13	11	7	6	6
His	5	4	2	2	2
Ile	8	7	4	5	5
Leu	5	5	6	5	5
Lys	20	19	8	7	8
Met	1	1	1	1	1
Phe	8	6	4	3	4
Pro	9	7	1	3	3
Ser	6	5	5	5	5
Thr	2	3	3	3	3
Trp	-	2	0	0	1
Tyr	1	1	2	2	3
Val	6	5	3	5	3
TOTAL	137	120	86	87	86

6 residues are the result of single nucleotide changes. Although the level of similarity is low, the match is certainly better than random and therefore may be significant. Among the identical amino acids, half of them contain silent third base replacements. These observations indicate a possible divergent evolutionary relationship of the two colicins. Whether the conserved amino acids may be essential to the respective colicin function or have some vital function at the nucleotide level remains unanswered.

ii) Potential receptor-binding domain and identification of a common hydrophilic peptide. Located immediately 5' to the coding regions of the colicins E2 and E3 nuclease domains and extending towards the unique ClaI restriction site of either *col* gene, is a DNA segment that is highly conserved (Fig. 2). Within these sequences (nts 1-254 of E2 and 22-255 of E3) two regions of homology are evident, first an almost identical stretch of

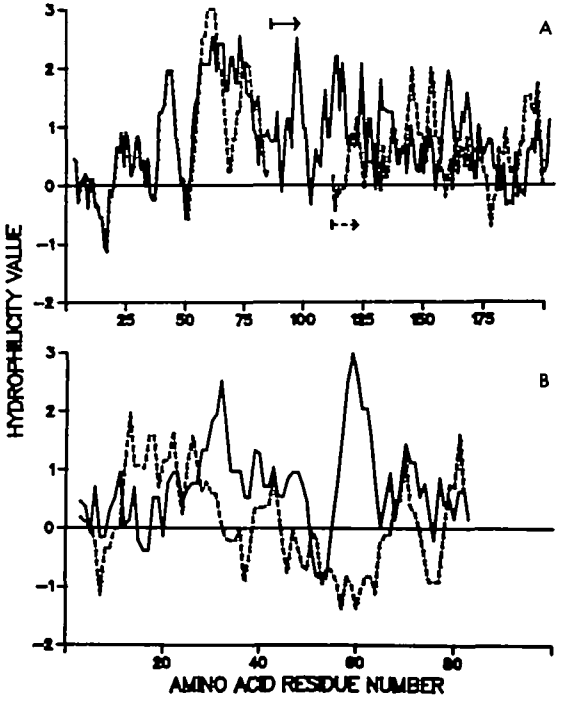

Figure 3. Hydrophilicity plots for (A) the C-terminal peptides of colicins E2 and E3 and (B) their immunity protein sequences. The E2 and E3 profiles are shown in solid and broken lines respectively. The first 6 amino acids of the colicin E3 sequence are not included in the hydrophilicity calculations so that amino acid residue 1 is the arg residue at the ClaI site (Fig 2). The hydrophilicity profiles of the nuclease domains are in the directions of the arrows. A line discontinuity was introduced in the E3 profile in order to delineate the position of the insertion or deletion sequence as aligned in Fig. 2. In all cases, hydrophilicity is indicated by the positive values above the midlines.

nucleotides which code for some 50 identical amino acid residues (1-50 of E2 and 8-57 of E3) and second, a less conserved region spanning amino acid residues 51-84 of E2 and 58-85 of E3. The fact that the two colicins recognize a common receptor protein (9-11) leads us to conclude that this common conserved region constitutes part of the receptor-binding domain, a function which has been assigned to the relatively large central portions of both colicins (17).

When the nuclease domains and the putative receptor-binding domains are computer-analyzed using the method of Hoop and Woods (37), a marked hydrophilic profile is predicted (Fig. 3A). This is expected because of the

preponderance of hydrophilic residues in these sequences especially within the less conserved peptide segments between residues 51-84 of E2 and 58-85 of E3 (Fig. 2). Among the amino acid changes within these less conserved regions, it is curious that the substitutions are mainly between polar residues. In addition, there are several silent mutations which do not result in amino acid changes. Perhaps, there is an evolutionary pressure to conserve the extreme hydrophilic nature of this part of the colicin molecule. According to the criteria of Hoop and Woods (37), the point of highest local hydrophilicity value may represent an antigenic determinant. Accordingly, it is tempting to assign such a role to the above-mentioned hydrophilic domains. On the other hand, the general hydrophilic profiles shown in Fig. 3A indicate several other possibilities, although some of the hydrophilic peaks may simply represent inter-domain structures (38) of the colicin molecule.

Locations and comparison of E2 and E3 immunity structures

Nine bases downstream from the opal stop codon (nts 547-549) of the colicin E3 gene (Fig. 2) begins the 255 nucleotide-long structural gene which codes for the 9290-dalton E3 immunity protein (24,25,39; this study). In contrast, the stop (ochre) codon of colicin E2 is separated from the beginning of its specific *imm* gene by only a dinucleotide (Fig. 2). The open reading frame with an initiation codon ATG (nts 620-622), capable of coding a protein of 86 residues (M_r: 9400) is identified as the E2-*imm* gene for the following reasons: i) the N-terminal sequence predicted from nucleotide sequence is identical to the amino acid sequence of the first 20 residues of the E2 immunity protein (23), and ii) the amino acid composition (Table 1) of the predicted protein is in excellent agreement with previous studies of the purified protein (1,23).

The E2 immunity protein contains 25.6% acidic (D+E), 18.6% basic (K+R+H), 9.3% aromatic (F+Y+W) and 20.9% hydrophobic (L+V+I+F+M) amino acid residues. [The groupings of amino acids in their one-letter codes are according to Dayhoff *et al.* (40)]. This distribution is markedly different from the immunity proteins of E3 (24,25,39; this study) and ColDF13 (34). The E2 and E3 immunity proteins only share 8 common residues including the N-terminal methionine which is absent in the purified E3 protein (23). However, each of these immunity proteins maintains a high % of charged residues. Although the E3 protein appears to be more acidic than that of E2, the latter is generally more hydrophilic. This is evidenced by the hydrophilicity plot (Fig. 3B) which further indicates that the hydrophilic

domains, which may be antigentically relevant (37,38), are significantly different for the two proteins. The E3 protein appears to have a central hydrophobic core whereas the corresponding region of the E2 protein is quite hydrophilic. This central hydrophobic domain of the E3 sequence which stretches from residues 45-67 may contribute to the hydrophobic (41), in addition to the hydrophilic interactions with the respective colicin E3 molecule.

Further structural differences of the two immunity proteins are revealed by an analysis of their secondary structures by the method of Chou and Fasman (42). This empirical method predicts a high α-helical (58.1%) content concentrated at the N-terminal half of the E2 protein. The β-pleated sheet, turn and coil contents are 17.4%, 8.1% and 16.3% respectively. On the contrary, the E3 protein is rich in β-sheet (41.1%) and turn (27.1) but low in α-helixes (20%).

The different hydrophilicity profiles and predicted secondary structures may explain the inability of either immunity protein to complement each other in the neutralization of the 'heterologous' colicin even when these primary structures are highly homologous. For example, the ClODF13 immunity protein which is 70% homologous to the E3 immunity protein has been shown to interact only weakly with colicin E3 (13,43).

Putative regulatory sequences for imm gene expressions

It has been suggested that in either ColE2-P9 and ColE3-CA38 plasmids the imm gene may be regulated independently of colicin expression depending on the mitomycin C-induced or non-induced states of the host cells (44-46). A putative promoter sequence 5'-TTCTCA (nts 423-428) and 5'-TATCGTG (nts 448-454) (Fig. 2) located within the colicin E3 gene for immunity gene expression has been proposed (24). Similarly, in the E2 sequence, the nucleotides 5'-GTGACC (nts 573-578) and 5'-CATATTG (nts 591-597) are in good agreement with the consensus bacterial promoter sequences of TTGACA and TATAATG also known as the '-35' and '-10' regions respectively (47). In addition, a CAT sequence (nts 603-605), a potential transcriptional startsite was found appropriately spaced from the '-10' consensus sequence (47).

Alternatively, the expressions of both of the col-imm gene pairs may be 'translationally coupled' as described for the E. coli tryptophan and galactose operons (48,49). First observed in the trpED gene pair in the trp operon (48), the phenomenon of translational coupling is characterized by an overlap of regulatory signals on a polycistronic message, and for a successful translation initiation of a distal gene, an efficient translation

Table 2. Sequence alignment of the catalytic residues and the surrounding amino acids of various microbial RNAases with colicin E3. The sequences of RNAases T1 (*A. oryzae*), C2 (*A. clavatus*), U2 (*U. sphaerogena*), Ms (*A. saitoi*), Ba (*B. amyloliquefaciens*), Bi (*B. intermedius*) and St (*S. erythreus*) are from ref. 53. The amino acid residues that are homologous in 4 or more sequences are also boxed according to ref. 53. The catalytic residues are discussed in the text. Selected segments of the colicin E3 sequence are from Fig. 2. The number of amino acid residues separating each sequence segment is also indicated.

Fungal	T1	Y P H^{40} K Y	14	Y - E^{58} W P I	13	A D R^{77} V V	6	L A G V I T H^{92}
	C2	Y P H^{40} Q Y	14	Y - E^{58} W P I	13	A D R^{77} V V	6	L A G L I T H^{92}
	U2	Y P H^{40} Q Y	16	W S E^{61} F P L	17	P D R^{86} V I	7	F C A T V T H^{100}
	Ms	Y P H^{39} E Y	14	Y - E^{57} Y P I	13	A D R^{76} V I	6	L A G V I T H^{91}
Colicin	E3	Y G H^{93} D Y	14	L - K^{111} P G I	11	R K R^{128} W T	7	Y E W D S Q H^{144}
Bacterial	St	D G^{41} T V F	14	Y H E^{61} Y T V	9	A R R^{76} F V	6	Y F Y T E D H^{91}
	Bi	G G^{52} D V F	14	W R E^{72} A D I	8	A D R^{86} L V	6	I Y K T T D H^{101}
	Ba	G G^{53} D I F	14	W R E^{73} A D I	8	S D R^{87} I L	6	I Y K T T D H^{102}

of the preceding gene is necessary. Examination of the sequence characteristics of the col-imm intergenic boundaries (Fig. 2) shows two potential overlapping ribosome-binding sequences (5'-GAGAGG, nts 548-553), one of which also overlaps the stop codon of the colicin E3 gene. A reasonable candidate (5'-TAAG) for a ribosomal-binding sequence (50,51) is also found for the E2 imm gene within the colicin E2 coding sequence. These overlaps of regulatory sequence signals in conjunction with the fact that both of the col-imm gene pairs code for products that form equimolar complexes (4,12,13), and the possibility that in either plasmid the colicin and immunity proteins are synthesized from a polycistronic message (22,24,25,45,46) make the model of translational coupling very attractive.

Active-site sequence homology between colicin E3 and other microbial RNAases

Several extracellular ribonucleases of bacterial and fungal origins have been shown to share considerable structural and sequence homology (52,53). Among the conserved sequences are those amino acid residues which are involved in catalysis (eg. E58, R77 and H92 of RNAase T1) or implicated in substrate binding (eg. H40 of RNAase T1) (Table 2). When the sequence of the colicin E3 RNAase domain is compared to that of fungal and bacterial nucleases, an important feature emerges. Although limited identity or similarity is observed (Table 2), the amino acid residues, H93, R128 and H144 of colicin E3 are found to be conserved and spaced in a similar manner as

those of H40, R77 and H92 (T1 numbering). The invariant catalytic or active-site E58 residue (T1 numbering) is replaced by lysine at the equivalent position (residue 111) in the E3 sequence. This substitution could be due to a single base change from GAG to AAG[333] (Fig. 2). Nevertheless, the tripeptide Y-E-W in the T1 and C2 sequences containing the active E residue also exists in the colicin E3 sequence at positions 138-140, in the vicinity of H144. This misplacement may explain the failure of an earlier attempt (29) to identify the extent of relatedness as reported here.

Based on the observed identity (Table 2), it appears that the RNAase domain of colicin E3 is more related to the fungal enzymes and that they possibly share a common ancestry (cf. 54). On the other hand, a similar relationship with the bacterial enzymes cannot be ruled out. It is possible that during divergent evolution of the 'nuclease family' gene(s) the active-site or substrate binding residues may have been conserved as have been reported for other proteins (see ref. 55 for a review). Interestingly, the sizes of the colicin E3 nuclease fragment (97 residues) and the various RNAases (101-113 residues, ref. 53) are similar. The larger size of the colicin E3 molecule (582 residues, ref. 17) compared to the various RNAases may be accounted by a gene fusion event involving a 'primordial' nuclease domain and a second gene domain in which the translated product has the intrinsic property of cell entry, i.e. the receptor-binding and transmembrane portion of the colicin molecule.

CONCLUDING REMARKS

The primary sequences presented in this study are necessary prerequisites for the long-term goal of understanding the interaction between colicin and its specific immunity protein as well as with its nucleic acid target. A preliminary study of the colicin E3-immunity protein complex by small angle X-ray diffraction has been reported (41).

The E2 and E3 immunity structures reported here represent another case of divergent evolution whereby a specific function is preserved for two proteins with remarkedly dissimilar primary structures. The elucidation of the sequences for the nuclease domains of colicins E2 and E3 also indicates that these structures, despite having different substrate specificities, are more closely related to each other than was previously realized.

The tentative identification of the catalytic or active-site residues for colicin E3 provides us a working hypothesis for current site-directed mutagenesis experiments to yield further information on the structure and

function of this molecule. Current work is also directed towards an investigation of gene expression and regulation of the colicin operon in both ColE2-P9 and ColE3-CA38 plasmids.

ACKNOWLEDGMENTS

We thank L. Bramall for the hydrophilicity plots; H. Turner and B. Kane for photography; Drs. A. Vaisius, T. Vernet and M. Yaguchi for helpful discussions and Dr. P. Davies for a critical reading of the manuscript. The excellent secretarial assistance of J. Strang is appreciated.

*NRCC Publication No. 23865

**To whom correspondence should be addressed.

ABBREVIATIONS

E2 and E3 refer to plasmids ColE2-P9 and ColE3-CA38 respectively unless otherwise qualified; DNAase, deoxyribonuclease; RNAase, ribonuclease; bp, base pairs; kb, kilobase pair; nts, nucleotides.

REFERENCES

1. Jakes, K. (1982) in Molecular Action of Toxins and Viruses (Cohen, P. and van Heyningen, S. eds.), pp. 131-167, Elsevier Biomedical Press.
2. Konisky, J. (1982) Ann. Rev. Microbiol. 36, 125-144.
3. Cramer, W.A., Dankert, J.R. and Uratani, Y. (1983) Biochim. et Biophys. Acta. 737, 173-193.
4. Schaller, K. and Nomura, M. (1976) Proc. Natl. Acad. Sci. U.S.A. 73, 3989-3993.
5. Yamamoto, H., Nishida, K., Beppu, T. and Arima, K. (1978) J. Biochem. 83, 827-834.
6. Bowman, C.M., Dahlberg, J.E., Ikemura, T., Konisky, J. and Nomura, M. (1971) Proc. Natl. Acad. Sci. U.S.A. 68, 964-968.
7. Ohno, S. and Imahori, K. (1978) J. Biochem. 84, 1637-1640.
8. de Graaf, F.K., Niekus, H.G.D. and Klootwijk, J. (1973) FEBS Lett. 35, 161-165.
9. Imajoh, S., Ohno-Iwashita, Y. and Imahori, K. (1982) J. Biol. Chem. 257, 6481-6487.
10. Di Masi, D.R., White, J.C., Schnaitman, C.A. and Bradbeer, C. (1973) J. Bacteriol. 115, 506-513.
11. Konisky, J. (1979) in Bacterial Outer Membranes Biogenesis and Functions (Inouye, M. ed.), pp. 319-359. John Wiley and Sons, New York.
12. Jakes, K., Zinder, N.D. and Boon, T. (1974) J. Biol. Chem. 249, 438-444.
13. Oudega, B., Klaasen-Boor, P. and de Graaf, F.K. (1975) Biochim. et Biophys. Acta. 392, 184-195.
14. Ohno, S., Ohno-Iwashita, Y., Suzuki, K. and Imahori, K. (1977) J. Biochem. 82, 1045-1053.
15. Suzuki, K. and Imahori, K. (1978) J. Biochem. 84, 1021-1029.
16. Ohno-Iwashita, Y. and Imahori, K. (1979) FEBS Lett. 100, 249-252.
17. Ohno-Iwashita, Y. and Imahori, K. (1980) Biochemistry 19, 652-659.
18. Suzuki, K., Ohno, S. and Imahori, K. (1980) J. Biochem. 87, 761-769.

19. Watson, R., Rowsome, W., Tsao, J. and Visentin, L.P. (1981) J. Bacteriol. 147, 569-577.
20. Mock, M. and Pugsley, A. (1982) J. Bacteriol. 150, 1069-1076.
21. Watson, R. and Visentin, L.P. (1980) Gene 10, 307-318.
22. Watson, R. and Visentin, L.P. (1982) Gene 19, 191-200.
23. Watson, R., Konarska-Kozlowska, M., Iyer, V.N., Yaguchi, M. and Visentin, L.P. (1983) J. Bacteriol. 153, 1552-1557.
24. Masaki, H. and Ohta, T. (1982) FEBS Lett. 149, 129-132.
25. Mock, M., Miyada, C.G. and Gunsalus, R.P. (1983) Nucleic Acids Res. 11, 3547-3557.
26. Messing, J. (1981) in Recombinant DNA (Walton, A.G. ed.), pp. 143-153, Elsevier Scientific Publishing Co., Amsterdam.
27. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
28. Lau, P.C.K. and Spencer, J.H. (1982) Biosc. Rep. 2, 687-696.
29. Suzuki, K. and Imahori, K. (1978) J. Biochem. 84, 1031-1039.
30. Watson, R.J., Lau, P.C.K., Vernet, T. and Visentin, L.P. (1984) Gene 29, 175-184.
31. Chak, K.-F. and James, R. (1984) J. Gen. Microbiol. 130, 701-710.
32. Lau, P.C.K., Rowsome, R.W., Watson, R.J. and Visentin, L.P. (1984) Biosc. Rep. 4, 565-572.
33. Inselburg, J. (1973) Nature New Biol. 241, 234-237.
34. Van den Elzen, P.J.M., Gaastra, W., Spelt, C., de Graaf, F.K., Veltkamp, E. and Nijkamp, H.J.J. (1980) Nucleic Acids Res. 8, 4349-4363.
35. Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H. and Delius, H. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 6376-6380.
36. van Rompuy, L., MinJou, W., Huylebroeck, D., Devos, R. and Fiers, W. (1981) Eur. J. Biochem. 116, 347-353.
37. Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 3824-3828.
38. Juillerat, M.A., Barkas, T. and Tzartos, S.J. (1984) FEBS Lett. 168, 143-148.
39. Mochitate, K., Suzuki, K. and Imahori, K. (1981) J. Biochem. 89, 1609-1618.
40. Dayhoff, M.O., Dayhoff, R.E. and Hunt, L.T. (1972) in Atlas of Protein Sequence and Structure (Dayhoff, M. ed.), Vol. 5, pp. 301-310.
41. Levison, B.L., Pickover, C.A. and Richards, F.M. (1983) J. Biol. Chem. 258, 10967-10972.
42. Chou, P.Y. and Fasman, G.D. (1978) Adv. in Enzymology 47, 45-148.
43. Kool, A.J., Pols, C. and Nijkamp, H.J.J. (1975) Antimicrob. Agents Chemother. 8, 67-75.
44. Tyler, J. and Sherratt, D.J. (1975) Mol. Gen. Genet. 140, 349-353.
45. Pugsley, A.P. and Schwartz, M. (1983) Mol. Gen. Genet. 190, 366-372.
46. Pugsley, A.P. (1983) Mol. Gen. Genet. 190, 379-383.
47. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
48. Oppenheim, D.S. and Yanofsky, C. (1980) Genetics 95, 785-795.
49. Schumperli, D., McKenney, K., Sobieski, D.A. and Rosenberg, M. (1982) Cell 30, 865-871.
50. Atkins, J.F. (1979) Nucleic Acids Res. 7, 1035-1041.
51. Kozak, M. (1983) Microbiol. Rev. 47, 1-45.
52. Hartley, R.W. (1980) J. Mol. Evol. 15, 355-358.
53. Hill, C., Dodson, G., Heinemann, U., Saenger, W., Mitsui, Y., Nakamura, K., Borisov, S., Tischenko, G., Polyakov, K. and Pavlovsky, S. (1983) Trends Biochem. Sci. 8, 364-369.
54. Doolittle, R.F. (1981) Science 214, 149-159.
55. Bajaj, M. and Blundell, T. (1984) Ann. Rev. Biophys. Bioeng. 13, 453-492.