Structural Relatedness of Lysis Proteins from Colicinogenic Plasmids and Icosahedral Coliphages[1]

Peter C. K. Lau, Mary A. Hefford,* and Petr Klein†*

*Molecular Genetics and †Biomathematics Sections, Division of Biological Sciences, National Research Council of Canada

The host-lysis-inducing functions of ϕX174 protein E and MS2 protein L were recently shown to reside on the N-terminal and C-terminal halves of the two respective lysis proteins. In the present study it is shown that the small lysis proteins encoded in various colicinogenic plasmids share local sequence similarities and certain structural characteristics with the essential peptides of their coliphage-coded counterparts. Despite their dissimilar sizes and origins, it is suggested that the colicinogenic lysis proteins are functionally analogous and evolutionarily related to those of icosahedral single-stranded DNA and RNA phages.

Introduction

Colicinogenic or Col plasmids carried by various species of bacteria are factors that are responsible for the synthesis of both a bactericidal protein (colicin in *Escherichia coli* and cloacin in *Enterobacter cloacae*) and an immunity protein that protects the host cell from the lethal action of its own toxin (Reeves 1972). Superficially, colicins resemble bacteriophages in that, on solid media layered with sensitive bacterial cells, they produce areas of inhibition of growth known as lacunae or plaques. In addition, the synthesis of colicins may be induced, for example, by ultraviolet irradiation; inducibility is also an inherent property of temperate phages (Lewin 1977). Recently it was shown that Col plasmids, like most phages, contain a gene encoding a lysis protein (Pollock et al. 1978; Beremand and Blumenthal 1979; Hakkaart et al. 1981*a*; Altman et al. 1983; Sabik et al. 1983; Jakes and Zinder 1984; Lawrence and James 1984; Pugsley 1984; Watson et al. 1984; Cavard et al. 1985; Cole et al. 1985; Rennell and Poteete 1985). In *E. coli* infected by such phages as lambda or ϕX174, lysis of the host cells leads to the release of the replicated phage particles (Denhardt and Sinsheimer 1965; Tsugita 1971; Garrett et al. 1981), whereas with the Col plasmids lysis results in the export and release of colicins (Hakkaart et al. 1981*a;* Pugsley and Schwartz 1984; Watson et al. 1984; Cavard et al. 1985; Cole et al. 1985).

Prompted by these observations and the notion of coevolution of bacteria and their plasmids and phages (Levin and Lenski 1983), we have compared the amino acid sequences of the Col plasmid–coded and the phage-coded lysis proteins. In addition, it is of considerable interest to compare the structures of these proteins, which are devoid of enzymatic or lysozymic activities (Eigner et al. 1963; Cole et al. 1985), with that of phage T4 lysozyme, which is also known to act on the peptidoglycan cell

1. Key words: lysis protein, computer analysis, coliphage, Col plasmids.

Address for correspondence and reprints: Dr. Peter C. K. Lau, Molecular Genetics Section, Division of Biological Sciences, National Research Council of Canada, 100 Sussex Drive, Ottawa, Ontario, Canada K1A 0R6.

Mol. Biol. Evol. 4(5):544–556. 1987.

0737-4038/87/0405-0008$02.00

wall from the cytoplasmic side (Tsugita 1971). Although no similarity can be detected in the primary structures, the tertiary structure of phage T4 lysozyme corresponds to those of goose-egg-white lysozyme (GEWL) and hen-egg-white lysozyme (HEWL) (Weaver et al. 1985).

Here we report the structural correspondence of Col plasmid–coded lysis proteins with the essential portions of lysis proteins E and L, which are products of single-stranded DNA (ϕX174, G4) and single-stranded RNA phages, respectively (Berkhout et al. 1985; Bläsi and Lubitz 1985). This structural relatedness implies a mode of action for Col lysis proteins that is similar to that of the icosahedral coliphage systems ϕ174 and MS2, which require a functional host autolytic enzyme system (Holtje and van Duin 1984; Lubitz et al. 1984).

Material and Methods

The name of the deduced lysis-protein sequences used in the present study and the bacteriocinogenic plasmids from which they were obtained are as follows: kil of ColE1-K30 (Hakkaart et al. 1981*a*; Chan et al. 1985), celB of ColE2-P9 (Cole et al. 1985), hic or lys of ColE3-CA38 (Jakes and Zinder 1984; Watson et al. 1986), cal of ColA-CA31 (Cavard et al. 1985), and H of CloDF13 (Hakkaart et al. 1981*a*). All plasmids are from *E. coli* except the last, which is from *Enterobacter cloacae*. (For simplicity, the various lysis proteins are referred to in the text as lys but are differentiated by their carrier plasmids.) E proteins of ϕX174, G4 (Godson et al. 1978), and S13 (Lau and Spencer 1985) were from single-stranded DNA phages; L protein of MS2 was from single-stranded RNA phages (Beremand and Blumenthal 1979); S, R, and Rz were from phage lambda (Sanger et al. 1982); and L13 and L19 were from *Salmonella* phage P22 (Rennell and Poteete 1985). The phage T4 lysozyme sequence (T4L), GEWL, and HEWL are from the National Biomedical Research Foundation (NBRF) Protein Sequence Database (Orcutt et al. 1983).

Computer analysis of local sequence similarities utilized the Goad-Kanehisa program (Goad and Kanehisa 1982; Kanehisa 1982, 1986). Similarity between two aligned sequences is measured by a score that depends on the values of penalties for amino acid substitutions and deletions. Amino acid similarity is defined as in Dayhoff et al.'s (1979) mutation data matrix. The penalty for the deletion used was eight, which is equal to the highest penalty for substitution. To assess the statistical significance of the score, we compared it with the distribution of similarity scores for 50 random permutations of the sequences. The significance is expressed either as the difference (in SD) between the score and the mean score of the permuted sequences or as the probability that the observed or higher score would arise by chance (this assumes that the distribution of scores is Gaussian).

The program of Garnier et al. (1978) was used for secondary-structure predictions. In this program certain parameters—i.e., the decision constants—can be set to favor a particular state (helix or sheet) if there is some evidence on percentages of ordered structures (e.g., from circular dichroism) or folding type. We used the method of Klein and DeLisi (1986) for a preliminary folding-type prediction. Transmembrane segments were analyzed by means of the program of Klein et al. (1985), which locates all segments with significant hydrophobicity. Hydrophobicity profiles were obtained by averaging the hydrophobicity values per six amino acids according to Hopp and Woods (1981). The amino acid variability plot was constructed according to the method of Wu and Kabat (1970).

Results and Discussion

With the exception of the closely related colicin Ia and Ib plasmids (Mankovich et al. 1986), the lysis gene of all bacteriocinogenic plasmids characterized thus far is located downstream of the colicin or bacteriocin gene on one transcriptional unit such that the two genes are expressed from a common inducible promoter under the SOS regulatory system (Hakkaart et al. 1981*b*; Sabik et al. 1983; Watson et al. 1984; Cavard et al. 1985; Cole et al. 1985). An alignment of four representative lysis-protein sequences and their variability plot are shown in figure 1. Extensive sequence similarity is evident among the various lysis proteins, the highest percentage of matches being between the aligned CloDF13-coded and ColE1-coded proteins. The sequence from ColA has the least similarity to the other lysis proteins. In each plasmid, the small lysis protein is characterized by a relatively variable hydrophobic N-terminal segment (17–21 amino acids) that contains a typical signal-peptide sequence (Perlman and Halvorson 1983; Vlasuk et al. 1983) potentially cleavable either at the alanyl-cysteinyl bonds of ColE3, CloDF13, and ColA or at the glycyl-cysteinyl bond of ColE1. Some evidence has been provided that the lysis proteins of CloDF13, ColE2, and ColA are synthesized as precursor molecules from which these amino extensions are proteolytically removed (Oudega et al. 1984; Cavard et al. 1985; Cole et al. 1985). Interpretation of these data, however, is complicated by the inherent difficulties associated with accurate sizing of small proteins by means of conventional methods, as illustrated by the anomalous migration of the lambda S gene product in gel electrophoresis (Altman et al. 1983). Therefore, the "mature" forms of the lysis proteins of 28–33 residues having a common N-cysteine residue remain to be characterized. If the signal peptide is in fact cleaved,

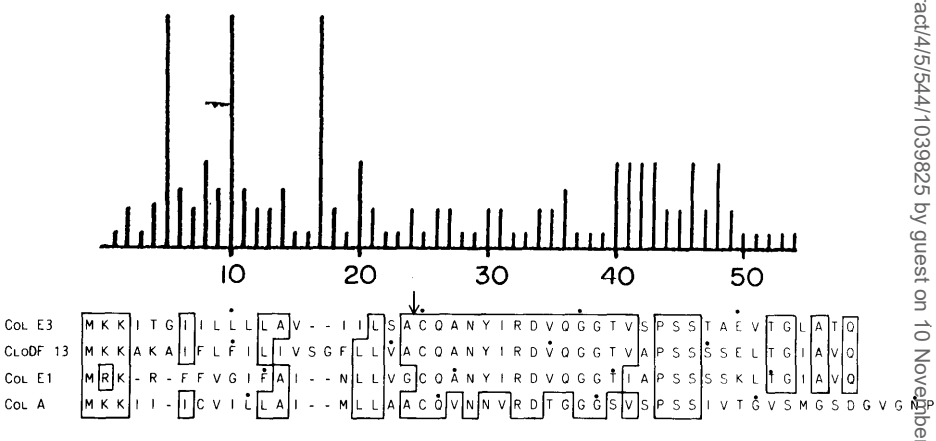

FIG. 1.—Alignment of ColA, ColE, and CloDF13 plasmid-coded lysis-protein sequences and their variability plot. Identities of three or four amino acid residues at a given position are boxed. The arrow indicates the site of potential cleavage by bacterial signal peptidase II. The fourth and last amino acids of ColE1 protein are given here as R (codon AGA) and Q (codon CAG) instead of as K and V, respectively, as elsewhere reported (Hakkaart et al. 1981*a*). The percent matches of the various pairs of proteins with (and without) the putative signal peptide are as follows: CloDF, E1 75.0 (92.6); CloDF, E3 63.3 (78.6); E1, E3 60.0 (73.1); E1, A 47.2 (55.6); E3, A 46.8 (42.9); and CloDF, A 45.7 (61.1). The lys sequence of ColE2 is identical to that of ColE3 (Cole et al. 1985; Watson et al. 1986). In the variability plot the lengths of the vertical lines indicate the extent of variability at the given position, measured as the number of different amino acids at that position divided by the frequency of the most common one (Wu and Kabat 1970; Kanehisa 1986).

the N-terminal 13 amino acids of ColE1, E3, and CloDF13 lysis proteins (fig. 1) and that of ColE2 (Cole et al. 1985; not shown) are identical.

Figure 2 summarizes local similarities found in the lysis-protein sequences of phage and plasmid origins. HEWL and GEWL were included in the comparison because of their three-dimensional similarity to phage T4L protein (Weaver et al. 1985). As in earlier reports, extensive similarities were observed between the following sequences: E proteins of phage ϕX174 and G4 (Godson et al. 1978) (the E protein sequence of the closely related phage S13 is identical to that of ϕX174 except for the penultimate C-terminal residue [Lau and Spencer 1985]); the *Salmonella* P22 gene L13 protein and the lambda gene S protein; and, to a lesser extent, P22 L19 and T4L (Rennell and Poteete 1985). Our analysis showed additional local sequence similarities.

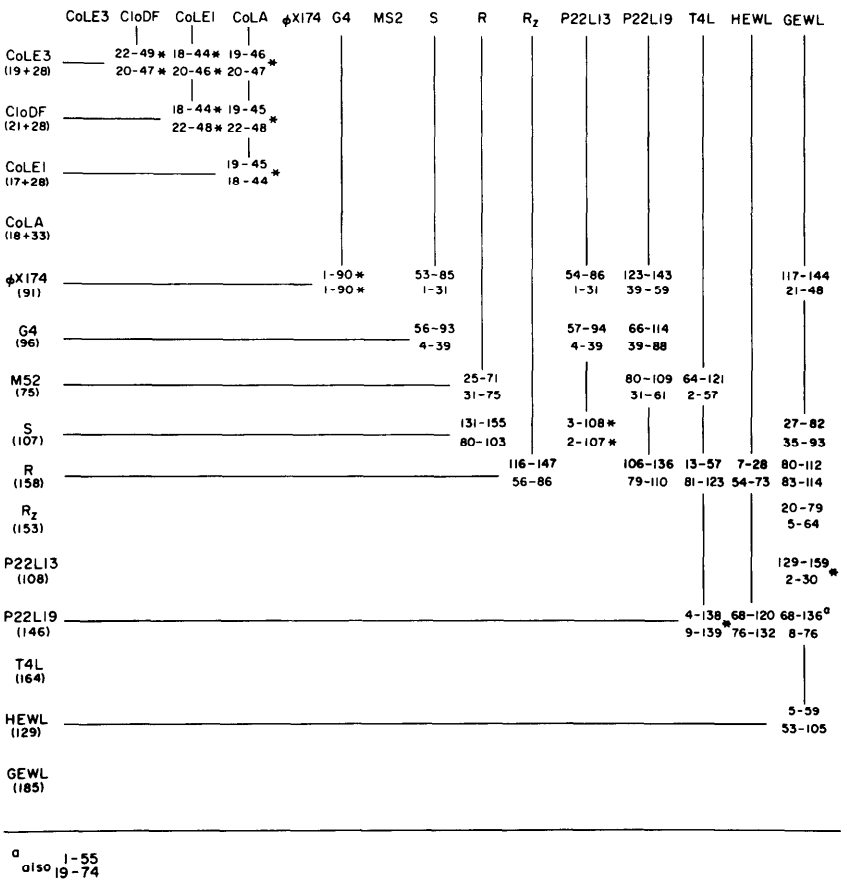

FIG. 2.—Local similarities among the various lysis proteins. The number in parentheses under the protein name represents the length of the protein. Where there are two numbers (the Col proteins), the first is the length of the putative signal-peptide sequence that is excluded from the homology search. The numbers written in the matrix refer to the segments of proteins where local similarities were found (the upper segment refers to proteins listed horizontally, the lower to proteins listed vertically). The asterisks alongside the peptide segments indicate the SD from the mean scores for permuted sequences (see Material and Methods): ** = difference >10 SD (significant); * = difference >6 SD (probably significant); and no asterisk = difference >3 SD (possibly significant) (Lipman and Pearson 1985).

Although some of the similarities found—such as those of lambda S and P22 L13 with ϕX174 or G4—can perhaps be attributed to the presence of a similar hydrophobic transmembrane segment, others cannot. The significance of the latter finding is not obvious at this time. In the absence of the putative signal-peptide sequences, we did not find any significant similarity (>20 residues and >3 SD) between the Col plasmid–coded and the phage-coded proteins. It is not unexpected to see sequence disparity between the Col lysis proteins and the lysozymes or the R and Rz endolysin proteins of lambda, which have transglycosylase or endopeptidase activities, respectively (Bienkowska-Szewczyk et al. 1981; Altman et al. 1983).

If, however, the putative leader sequences of the various Col proteins are included in the analysis, significant local similarities in at least three gene pairs are found (fig. 3). Especially noteworthy is the similarity between almost the complete lys sequence of ColA and the portion of G4 E protein that is very similar to the N-terminal 51 amino acids of ϕX174 E protein that has been shown to be essential for lysis function (Bläsi and Lubitz 1985). (A significance of 4.8 SD indicates that the probability that this similarity occurred by chance is $\sim 10^{-6}$). We found that in the whole NBRF Protein Sequence Database of >3,300 sequences there are only 16 sequences (0.48%) ≥30 residues long that have local sequence similarity with ColA lys at significance

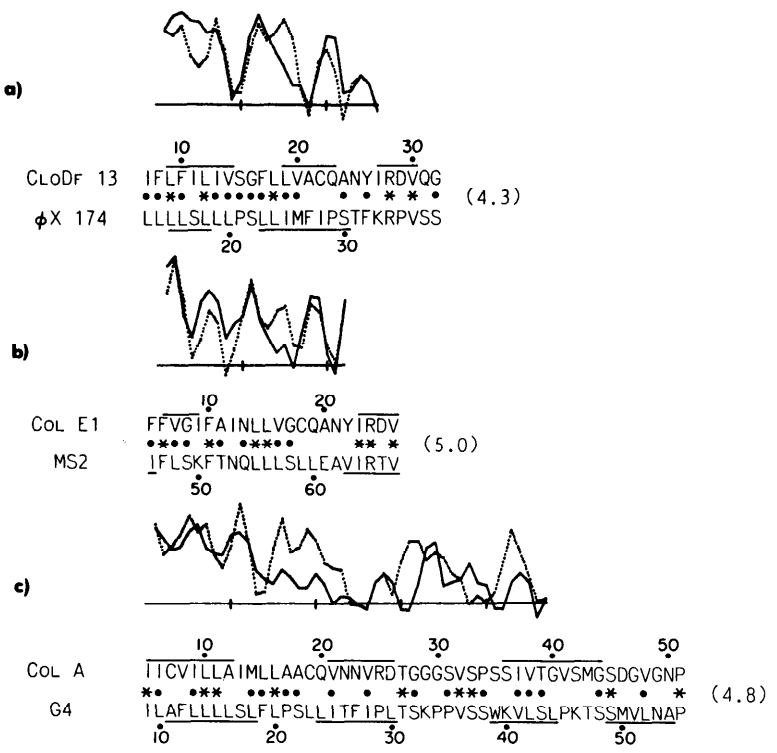

FIG. 3.—Alignment of the various lysis-protein sequences and their structural characteristics. Identical and chemically similar amino acid residues are indicated by asterisks and dots, respectively. The numbers in brackets alongside the sequence pairs represent the significance in SD units. Underlined and overlined residues have been predicted as a beta strand by means of the method of Garnier et al. (1978). Hydropathy profiles have been obtained by smoothing hydrophobicity values (Kanehisa 1986). The plots for CloDF13, Col E1, and Col A are represented by solid lines and those of ϕX174, MS2, and G4 by dotted lines.

values higher than those characterizing the similarity between ColA and G4 (table 1). However, in contrast to the alignment between the latter proteins, which covers virtually the whole ColA sequence and a substantial part of the G4 sequence, most of these 16 sequences are much longer and the parts similar to ColA lysis are just small portions of the relatively large proteins (table 1). The highly significant similarity found with a portion of human apolipoprotein is interesting in the context of reported partial similarities between *E. coli* outer-membrane lipoprotein and ColE1 lys (Kobayashi et al. 1986). Nonetheless, none of the 16 listed proteins is known to exhibit any cell lytic function.

The similarities seen between CloDF and ϕX174 and between ColE1 and MS2 as shown in figure 3 do not cover the whole sequences. Substantial parts of the aligned sequences are highly hydrophobic, with hydrophobicity values typical of signal peptides and transmembrane segments. To ascertain that the relationship found cannot be attributed only to the presence of the hydrophobic N-terminal sequences, we searched 383 NBRF-database proteins having N-terminal signal sequences for similarities both with CloDF and with ColE1 lys proteins. We found only three sequences (beta lactamase precursor, interferon gamma precursor, and the immunoglobulin heavy-chain precursor V region) that showed (1) local similarities with CloDF >25 residues and (2) higher significance values than that found between CloDF and ϕX174. Although the alignment found between CloDF13 and ϕX174 (fig. 3a) covers 29% of the latter sequence (and 51% of the part shown to be sufficient for lysis [Bläsi and Lubitz 1985]), the portions of the three unrelated sequences locally similar to CloDF represent only 15%–27% of the lengths of these three proteins. Similarly, there are only four sequences (atrial gland peptide A precursor, the immunoglobulin heavy-chain precursor V region,

Table 1
Sequence Similarity Between ColA lys and Other Proteins

ColA lys (Residues)	Protein (Source)	Length	Residues	Significance (SD)
5–51	E protein (phage G4)	96	9–55	4.8
1–38	Ig heavy-chain precursor V region (rat)	142	1–37	5.0
1–37	Hypothetical BZLF2 (Epstein-Barr virus)	223	7–42	5.0
6–43	Actin 1 (soybean)	378	135–174	5.1
5–50	Coat protein A precursor (phages fd, M13, and f1)	424	217–262	5.1
7–44	Lactose permease (*E. coli*)	417	176–213	5.2
1–50	Outer-membrane-protein F precursor (*E. coli*)	362	2–51	5.2
8–49	32 K Thylakoid membrane protein precursor (spinach)	353	47–90	5.3
11–44	DNA-directed DNA polymerase I (*E. coli*)	928	586–620	5.7
4–48	50S ribosomal protein L4 (*E. coli*)	201	147–193	5.7
4–45	Immediate-early protein (frog virus 3)	157	98–138	5.9
3–50	16.5 K Submaxillary gland protein (mouse)	138	2–51	6.0
10–51	Glycophorin A (human)	131	84–126	6.3
4–47	Seminal vesicle secretory protein IV precursor (rat)	112	6–49	6.3
1–48	Gag-pol polyprotein (Rous sarcoma virus)	1596	871–918	6.9
9–58	Probable early 17 K protein 1 (adenovirus 2)	153	41–82	9.5
8–42	Proapoliprotein A-1 precursor (human)	267	5–38	10.6

SOURCE.—NBRF Protein Sequence Database.

apolipoprotein E precursor, and murein-lipoprotein precursor) showing (1) alignments with ColE1 lys ≥20 residues and (2) significance values higher than those characterizing the alignment between ColE1 and MS2 lys proteins. Here the portion aligned (fig. 2b) covers 31% of the whole MS2 L protein sequence and 63% of that 35-residue portion at the C-terminus that is sufficient for lysis (Berkhout et al. 1985). The alignments with the unrelated proteins represent 23%, 24%, 9%, and 55% of their lengths. The lack of similarity with unrelated hydrophobic sequences is in agreement with observations that there is no homology between various signal peptides (von Heijne 1981*a*) and that homologous membrane proteins are more similar in the regions outside the membrane than they are in those within it (von Heijne 1981*b*).

The amino acid similarity matrix that we used for sequence comparisons is based on amino acid mutation data (Dayhoff et al. 1979). Recently, Bacon and Anderson (1986) derived a similarity matrix that is based on amino acid structural preferences and physico-chemical properties. Another matrix was derived by Levin et al. (1986) for secondary-structural predictions. To check the consistency of our results, we repeated sequence-similarity searches and significance evaluations by means of these matrices. Using the matrix of Bacon and Anderson (1986), we were able to confirm the alignment between ColA and G4 lysis proteins: the alignment between G4 residues 4–55 and the whole sequence of ColA lys had a significance of 3.7 SD and contained the alignment shown in figure 3. Using the matrix of Levin et al. (1986), we found significant alignments between ColE3 residues 5–39 and ϕX174 E protein residues 9–41 (3.8 SD), between ColE3 lys residues 1–22 and MS2 L protein residues 48–72 (4.2 SD), and between ColE1 residues 4–28 and MS2 residues 45–69 (5.2 SD).

To further verify the possibility of evolutionary relationships, we have analyzed secondary structures of colicin, MS2, ϕX174, and G4 lysis proteins. Since a preliminary analysis of folding class indicated beta-rich structures for all of these proteins, the decision constants in the method of Garnier et al. (1978) were set to favor beta sheets. Figure 3 shows that among the three aligned sequence pairs there is a striking correspondence of those residues that were predicted to be in beta structure.

The hydrophobic effect plays an important role in protein folding (Kyte and Doolittle 1982; Dill 1985); and residues with hydrophilic side chains tend to be nearer the protein surface in the folded protein than do the residues with hydrophobic side chains, the latter of which are thus shielded from the unfavorable interactions with water (Chothia 1976). Plotting averaged or otherwise smoothed hydrophobicity values for amino acid residues in the sequence helps identify exterior and interior regions in the folded protein (Rose and Roy 1980; Hopp and Woods 1981; Kyte and Doolittle 1982). The rationale for such plots lies in the observed correlation between hydrophobicity profiles and packing or solvent accessibility profiles derived from X-ray-determined protein structures (Rose and Roy 1980), although important differences also have been reported (Lipman et al. 1987). Similar accessibility profiles would presumably indicate similar folding, and this could imply structural relationship. Hydrophobicity profiles are used here only to further illustrate similarity found by alignment (see, e.g., Norcross and Kanehisa 1985). In figure 3 hydrophobicity profiles show regions where local similarity is to be found. In figure 4 the plots are extended to show them in the context of the whole molecules.

The hydrophobicity profiles of Col-coded lysis proteins are represented here by ColE3 and ColA (fig. 4a). The homologous G4 ϕX174 E proteins are represented by a ϕX174 profile (fig. 4b). It is apparent that the profile shows correspondence of the

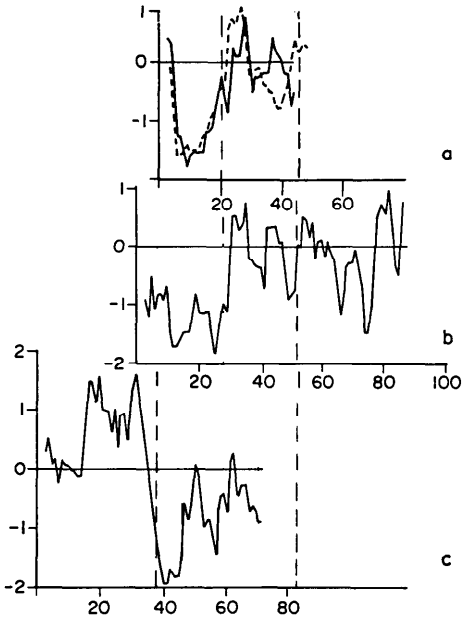

FIG. 4.—Hydropathic profiles of lysis proteins of (a) ColE3 (unbroken line) and ColA (broken line), (b) ϕX174, and (c) MS2. Hydrophilic and hydrophobic regions are represented by the positive and negative values relative to the zero baseline, respectively. The alignment of the three profiles is consistent with the sequence alignments in fig. 2. The essential portions of ϕX174 E protein and MS2 L protein and the "mature" form of the ColE3 protein are approximately within the dashed lines.

"mature" form of ColE3 lys (residues 20–47) with the profile of ϕX174 E protein (fig. 4b) in the region that has been shown to be essential for lytic activity; Bläsi and Lubitz (1985) showed that the lytic function of the E protein is within the first 51 amino acids of the 91-residue polypeptide. Figures 4b and 4c show that there is also some correspondence (although with vertical shifts) between ϕX174 and MS2 lysis proteins, roughly in the region of residues 40–75 of MS2, a region shown to be sufficient for lysis (Berkhout et al. 1985). Note that the last peak in which colicin E3 and A lysis proteins differ is missing in the MS2 L protein.

Although these data do not constitute proof of a relationship, we reckon that it is more than a coincidence that the essential portions of ϕX-E and MS2-L proteins should display secondary-structure and hydropathic characteristics similar to those of bacteriocinogenic-coded proteins. Indeed, several biochemical studies have documented similarities in these systems. Before onset of lysis, accumulation of either the ϕX gene E product or of the ColE2 or CloDF13 lysis protein apparently causes an induction of the cellular phospholipase activity (Pugsley and Schwartz 1984; Lubitz and Pugsley 1985; Luirink et al. 1986). In ϕX174, during the first 20 min of cell lysis, lyso-phosphatidyl ethanolamine was found to increase by 40%–50%, together with an increase in free fatty acid content but a drop in phosphatidyl glycerol and phosphatidyl ethanolamine contents (Pugsley and Schwartz 1984). Induction of phospholipase activity was viewed as a consequence of membrane perturbation owing to the integration of the lysis protein into the cytoplasmic membrane of *E. coli*. In ϕX174 and MS2 systems, it was further shown that the phage-induced lysis requires the activity of the host autolysins (Holtje and van Duin 1984; Lubitz et al. 1984). It is tempting to

speculate that there is a similar requirement for the bacteriocinogenic-coded proteins. The exact mechanism triggering the autolysins remains to be seen, however.

Another intriguing relationship between the bacteriocinogenic-coded and the phage-coded proteins is their gene arrangement. Either the genes are overlapping or their expression is translationally coupled in an operon unit. Whereas the MS2 L gene overlaps with the coat-protein and replicase genes (Beremand and Blumenthal 1979; Kastelein et al. 1982), the genes E of ϕX174, S13, and G4 are completely within the genes D that specify a scaffold protein (Godson et al. 1978; Lau and Spencer 1985). In MS2, no host lysis occurs unless enough coat protein has been made (Kastelein et al. 1982). In the Col or CloDF13 plasmids (where the *lys* and the *col* genes are organized in an operon), there is no lysis unless colicins are made (Jakes and Zinder 1984; Watson et al. 1984; Cavard et al. 1985; Colc et al. 1985; Luirink et al. 1986).

If coliphages, by virtue of their intrinsic ability to synthesize coat proteins and to exist outside the host cell, represent a more evolved form than plasmids (coatless phage?), it is not surprising that phage lysis proteins are bigger than those of plasmid origin. (One could argue that simpler plasmid functions might be derived from more complex phage functions or vice versa [Levin and Lenski 1983]. It is beyond the scope of this paper to resolve the progenitor issue.) In general, the size of the gene of homologous proteins increases with evolution of increasing complexity of function (Bajaj and Blundell 1984; Neurath 1985). For the ϕX174-type and MS2 proteins, it is not difficult to envision a gene-fusion event whereby the remaining sequences (N-terminal and C-terminal, respectively) have evolved a regulatory function (Berkhout et al. 1985; Schuller et al. 1985). It is worth noting that the N-termini of ϕX174 and G4 E products are highly conserved whereas the C-terminal portions account for most of the sequence differences, which are 19% and 44% at the nucleic acid and amino acid levels, respectively (Godson et al. 1978).

On one hand, the present study shows that there could well be a common structural basis for the observed similarity in lytic function of various lysis proteins from different origins. On the other hand, one may extrapolate from structural correspondence to a common mechanism of action for other lysis proteins that are less characterized. For example, one might predict a phospholipase activation and a requirement for a functional host-autolytic system for the lambda S protein, which shares some structural features with the ϕX174 E protein (Altman et al. 1985). Certainly, this hypothesis remains to be tested experimentally.

Because plasmids and phages depend on their bacterial hosts for reproduction—and because, in a reciprocal manner, the host's survival and reproduction can be influenced by phage or plasmid replication—it has been proposed that "coevolution" must contribute significantly to the overall evoluton of bacteria and their plasmids and phages. (Readers are referred to Levin and Lenski [1983] for mathematical models.) It is also known that plasmids and phages, as well as their bacterial hosts, can accumulate genetic variations through processes of recombination, mutation, and acquisition or loss of transposable elements. Recently, Ollis et al. (1985) reported a domain structure common to *E. coli* polymerase I and phage T7 DNA polymerase. Toh (1986) has extended this relationship to other replication-related gene products of T7 and *E. coli,* such as the DNA exonuclease, DNA primase, and helix-destabilizing protein. Our findings of structural similarities between bacteriocinogenic and phage proteins lend additional support to the notion of coevolution of bacteria and their phages and plasmids.

Acknowledgments

We thank L. Bramall for the hydropathy plots, H. Turner for photography, and J. Strang for excellent secretarial assistance. M.A.H. is a Canadian Government Laboratory Visiting Fellow. This paper is National Research Council of Canada publication 27439.

LITERATURE CITED

ALTMAN, E., R. K. ALTMAN, J. M. GARRETT, R. J. GRIMAILA, and R. YOUNG. 1983. S gene product: identification and membrane localization of a lysis control protein. J. Bacteriol. **155**:1130–1137.

ALTMAN, E., K. YOUNG, J. GARRETT, R. ALTMAN, and R. YOUNG. 1985. Subcellular localization of lethal lysis proteins of bacteriophages λ and φX174. J. Virol. **53**:1008–1011.

BACON, D. J., and W. F. ANDERSON. 1986. Multiple sequence alignment. J. Mol. Biol. **191**: 153–161.

BAJAJ, M., and T. BLUNDELL. 1984. Evolution and the tertiary structure of proteins. Annu. Rev. Biophys. Bioeng. **13**:453–492.

BEREMAND, M. N., and T. BLUMENTHAL. 1979. Overlapping genes in RNA phage: a new protein implicated in lysis. Cell **18**:257–266.

BERKHOUT, B., M. H. DE SMIT, R. A. SPANJAARD, T. BLOM, and J. VAN DUIN. 1985. The amino terminal half of the MS2-coded lysis protein is dispensable for function: implications for our understanding of coding region overlaps. EMBO J. **4**:3315–3320.

BIENKOWSKA-SZEWCZYK, K., B. LIPINSKA, and A. TAYLOR. 1981. The R gene product of bacteriophage λ is the murein transglycosylase. Mol. Gen. Genet. **184**:111–114.

BLÄSI, U., and W. LUBITZ. 1985. Influence of C-terminal modifications of φX174 lysis gene E on its lysis-inducing properties. J. Gen. Virol. **66**:1209–1213.

CAVARD, D., R. LLOUBES, J. MORLON, M. CHARTIER, and C. LAZDUNSKI. 1985. Lysis protein encoded by plasmid ColA-CA31: gene sequence and export. Mol. Gen. Genet. **199**:95–100.

CHAN, P. T., H. OHMORI, J. TOMIZAWA, and J. LEBOWITZ. 1985. Nucleotide sequence and gene organization of ColE1 DNA. J. Biol. Chem. **260**:8925–8935.

CHOTHIA, C. 1976. The nature of the accessible and buried surfaces in proteins. J. Mol. Biol. **105**:1–14.

COLE, S. T., B. SAINT-JOANIS, and A. P. PUGSLEY. 1985. Molecular characterization of the colicin E2 operon and identification of its products. Mol. Gen. Genet. **198**:465–472.

DAYHOFF, M. O., R. M. SCHWARTZ, and B. C. ORCUTT. 1979. A model of evolutionary change in proteins. Pp. 345–352 *in* Atlas of protein sequence and structure. Vol. **5**, suppl. 3. National Biomedical Research Foundation, Washington, D.C.

DENHARDT, D. T., and R. L. SINSHEIMER. 1965. The process of infection with bacteriophage φX174. III. Phage maturation and lysis after synchronized infection. J. Mol. Biol. **12**:641–646.

DILL, K. A. 1985. Theory for the folding and stability of globular proteins. Biochemistry **24**: 1501–1509.

EIGNER, J., A. H. STOUTHAMER, I. VAN DER SLUYS, and J. A. COHEN. 1963. A study of the 70S component of bacteriophage φX174. J. Mol. Biol. **6**:61–84.

GARNIER, J., D. J. OSGUTHORPE, and B. ROBSON. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**:97–120.

GARRETT, J., R. FUSSELMAN, J. HISE, L. CHIOU, D. SMITH-GRILLO, J. SCHULTZ, and R. YOUNG. 1981. Cell lysis by induction of cloned lambda lysis genes. Mol. Gen. Genet. **182**:326–331.

GOAD, W. B., and M. I. KANEHISA. 1982. Pattern recognition in nucleic acid sequences. I. A general method for finding local homologies and symmetries. Nucleic Acids Res. **10**:247–263.

GODSON, G. N., J. C. FIDDES, B. G. BARRELL, and F. SANGER. 1978. Comparative DNA sequence analysis of the G4 and ϕX174 genomes. Pp. 51–86 in D. T. DENHARDT, D. DRESSLER, and D. S. RAY, eds. The single-stranded DNA phages. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

HAKKAART, M. J. J., E. VELTKAMP, and H. J. J. NIJKAMP. 1981*a*. Protein H encoded by plasmid CloDF13 involved in lysis of the bacterial host. I. Location of the gene and identification and subcellular location of the gene H product. Mol. Gen. Genet. **183**:318–325.

———. 1981*b*. Protein H encoded by plasmid CloDF13 involved in lysis of the bacterial host. II. Functions and regulation of synthesis of the gene H product. Mol. Gen. Genet. **183**:326–332.

HOLTJE, J. V., and J. VAN DUIN. 1984. MS2-phage induced lysis of *E. coli* depends upon the activity of the bacterial autolysins. Pp. 195–199 in C. NOMBELA, ed. Microbial cell wall synthesis and autolysis. Elsevier Science,

HOPP, T. P., and K. R. WOODS. 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78**:3824–3828.

JAKES, K., and N. D. ZINDER. 1984. Plasmid ColE3 specifies a lysis protein. J. Bacteriol. **157**: 582–590.

KANEHISA, M. I. 1982. Los Alamos sequence analysis package for nucleic acids and proteins. Nucleic Acids Res. **10**:183–196.

———. 1986. IDEAS: integrated database and extended analysis system for nucleic acids and proteins. User manual. NCI Frederick Cancer Research Facility, Frederick, Md.

KASTELEIN, R. A., E. REMAUT, W. FIERS, and J. VAN DUIN. 1982. Lysis gene expression of RNA phage MS2 depends on a frame shift during translation of the overlapping coat protein gene. Nature **295**:35–41.

KLEIN, P., and C. DELISI. 1986. Prediction of protein structural class from the amino acid sequence. Biopolymers **25**:1659–1672.

KLEIN, P., M. KANEHISA, and C. DELISI. 1985. The detection and classification of membrane-spanning proteins. Biochim. Biophys. Acta **815**:468–476.

KOBAYASHI, T., C. KATO, T. KUDO, and K. HORIKOSHI. 1986. Excretion of the penicillinase of an alkalophilic *Bacillus* sp. through the *Escherichia coli* outer membrane is caused by insertional activation of the *kil* gene in plasmid pMB9. J. Bacteriol. **166**:728–732.

KYTE, J., and R. F. DOOLITTLE. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.

LAU, P. C. K., and J. H. SPENCER. 1985. Nucleotide sequence and genome organization of bacteriophage S13 DNA. Gene **40**:273–284.

LAWRENCE, G. M. P., and R. JAMES. 1984. Characterization of the ColE8 plasmid, a new member of the group E colicin plasmids. Gene **29**:145–155.

LEVIN, B. R., and R. E. LENSKI. 1983. Coevolution in bacteria and their viruses and plasmids. Pp. 99–127 in D. J. FUTUYMA and M. SLATKIN, eds. Coevolution. Sinauer, Sunderland, Mass.

LEVIN, J. M., B. ROBSON, and J. GARNIER. 1986. An algorithm for secondary structure determination in proteins based on sequence similarity. FEBS Lett. **205**:303–308.

LEWIN, B. 1977. Gene expression. Vol. **3**. Plasmids and phages. Wiley, New York.

LIPMAN, D. J., R. W. PASTOR, and B. LEE. 1987. Local sequence patterns of hydrophobicity and solvent accessibility in soluble globular proteins. Biopolymers **26**:17–26.

LIPMAN, D. J., and W. R. PEARSON. 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.

LUBITZ, W., R. E. HARKNESS, and E. E. ISHIGURO. 1984. Requirement for a functional host cell autolytic enzyme system for lysis of *Escherichia coli* by bacteriophage ϕX174. J. Bacteriol. **159**:385–387.

LUBITZ, W., and A. P. PUGSLEY. 1985. Changes in host cell phospholipid composition of ϕX124 gene E product. FEMS Microbiol. Lett. **30**:171–175.

LUIRINK, J., C. VAN DER SANDE, J. TOMMASSEN, E. VELTKAMP, F. K. DE GRAAF, and B. OUDEGA. 1986. Effects of divalent cations and of phospholipase A activity on excretion of cloacin DF13 and lysis of host cells. J. Gen. Microbiol. **132**:825–834.

MANKOVICH, J. A., C-H. HSU, and J. KONISKY. 1986. DNA and amino acid sequence analysis of structural and immunity genes of colicins Ia and Ib. J. Bacteriol. **168**:228–236.

NEURATH, H. 1985. Proteolytic enzymes, past and present. Fed. Proc. **44**:2907–2913.

NORCROSS, M. A., and M. KANEHISA. 1985. The predicted structure of the Ia beta-1 domain: a hypothesis for the structural basis of major histocompatibility complex–restricted T cell recognition of antigens. Scand. J. Immunol. **21**:511–523.

OLLIS, D. L., C. KLINE, and T. A. STEITZ. 1985. Domain of *E. coli* DNA polymerase I showing sequence homology to T7 DNA polymerase. Nature **313**:818–819.

ORCUTT, B. C., D. G. GEORGE, and M. O. DAYHOFF. 1983. Protein and nucleic acid sequence database systems. Annu. Rev. Biophys. Bioeng. **12**:419–441.

OUDEGA, B., A. YKEMA, F. STEGEHUIS, and F. K. DE GRAAF. 1984. Detection and subcellular localization of mature protein H involved in excretion of cloacin DF13. FEMS Microbiol. Lett. **22**:101–108.

PERLMAN, D., and H. O. HALVORSON. 1983. A putative signal peptidase recognition site and sequence in eukaryotic and signal peptides. J. Mol. Biol. **167**:391–409.

POLLOCK, T. J., E. S. TESSMAN, and I. TESSMAN. 1978. Identification of lysis protein E of bacteriophage ϕX174. J. Virol. **28**:408–410.

PUGSLEY, A. P. 1984. Genetic analysis of the Col N plasmid determinants for colicin production, release and immunity. J. Bacteriol. **158**:523–529.

PUGSLEY, A. P., and M. SCHWARTZ. 1984. Colicin E2 release: lysis, leakage or secretion? possible role of a phospholipase. EMBO J. **3**:2393–2397.

REEVES, P., ed. 1972. Molecular biology, biochemistry and biophysics. Vol. **11**. The bacteriocins. Springer, New York.

RENNELL, D., and A. R. POTEETE. 1985. Phage P22 lysis genes: nucleotide sequences and functional relationships with T4 and λ genes. Virology **143**:280–289.

ROSE, G. D., and S. ROY. 1980. Hydrophobic basis of packing in globular proteins. Proc. Natl. Acad. Sci. USA **77**:4643–4647.

SABIK, J. F., J. F. SUIT, and S. E. LURIA. 1983. *cea-kil* Operon of the ColE1 plasmid. J. Bacteriol. **153**:1479–1485.

SANGER, F., A. R. COULSON, G. F. HONG, D. F. HILL, and G. B. PETERSEN. 1982. The nucleotide sequence of bacteriophage λ DNA. J. Mol. Biol. **162**:729–773.

SCHULLER, A., R. E. HARKNESS, U. RUTHER, and W. LUBITZ. 1985. Deletion of C-terminal amino acid codons of PhiX174 gene E: effect on its lysis inducing properties. Nucleic Acids Res. **13**:4143–4153.

TOH, T. 1986. T7 and *E. coli* share homology for replication-related gene products. FEBS Lett. **194**:245–248.

TSUGITA, A. 1971. Phage lysozyme and other lytic enzymes. Pp. 343–411 in P. D. BOYER, ed. The enzymes. Academic Press, New York.

VLASUK, G. P., S. INOUYE, H. ITO, K. ITAKURA, and M. INOUYE. 1983. Effects of the complete removal of basic amino acid residues from the signal peptide on secretion of lipoprotein in *Escherichia coli*. J. Biol. Chem. **258**:7141–7148.

VON HEIJNE, G. 1981*a*. Membrane proteins: the amino acid composition of membrane-penetrating segments. Eur. J. Biochem. **120**:275–278.

———. 1981*b*. On the hydrophobic nature of signal sequences. Eur. J. Biochem. **116**:419–422.

WATSON, R. J., P. C. K. LAU, T. VERNET, and L. P. VISENTIN. 1984. Characterization and nucleotide sequence of a colicin-release gene in the *hic* region of plasmid ColE3-CA38. Gene **29**:175–184.

———. 1986. Corrigenda. Gene **42**:351–355.

WEAVER, L. H., M. G. GRUTTER, S. J. REMINGTON, T. M. GRAY, N. W. ISAACS, and B. W. MATTHEWS. 1985. Comparison of goose-type, chicken-type, and phage-type lysozymes illustrates the changes that occur in both amino acid sequence and three-dimensional structure during evolution. J. Mol. Evol. **21**:97–111.

WU, T. T., and E. A. KABAT. 1970. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarities. J. Exp. Med. **132**:211–250.

ROBERT K. SELANDER, reviewing editor

Received December 4, 1986; revision received March 5, 1987