JOURNAL OF BACTERIOLOGY, Sept. 1985, p. 850–857
0021-9193/85/090850-08$02.00/0

Vol. 163, No. 3

Nucleotide Sequence of *Klebsiella pneumoniae lac* Genes

WILSON E. BUVINGER AND MONICA RILEY*

Biochemistry Department, State University of New York at Stony Brook, Stony Brook, New York 11794

Received 4 February 1985/Accepted 29 May 1985

The nucleotide sequences of the *Klebsiella pneumoniae lacI* and *lacZ* genes and part of the *lacY* gene were determined, and these genes were located and oriented relative to one another. The *K. pneumoniae lac* operon is divergent in that the *lacI* and *lacZ* genes are oriented head to head, and complementary strands are transcribed. Besides base substitutions, the *lacZ* genes of *K. pneumoniae* and *Escherichia coli* have suffered short distance shifts of reading frame caused by additions or deletions or both during evolutionary divergence from a common ancestral gene. Relative to corresponding *E. coli* sequences, the nucleotide sequences of the *lacZ* and *lacY* genes are 61 and 67% conserved, and the *lacI* genes are 49% conserved. A comparison of both nucleotide and amino acid sequences revealed that the *K. pneumoniae* and *E. coli lacI* genes and *lac* repressor proteins each are related to the *galR* gene and *gal* repressor of *E. coli* to about the same extent. In terms of evolutionary relationships, the divergence of the forerunner of the *galR* gene from an ancestral *lac* repressor gene preceded separation and differentiation of the *K. pneumoniae* and *E. coli lac* repressor genes.

The chromosomal *lac* operon in *Klebsiella pneumoniae* (formerly *Klebsiella aerogenes*) (20) has some characteristics that are similar to those of the *lac* system in *Escherichia coli*, as well as some that are dissimilar (12, 23). As in the *E. coli* system, isopropylthio-β-D-galactopyranoside acts as an active inducer of the *K. pneumoniae* operon. Expression of the *lac* operon in *K. pneumoniae* is subject to the glucose effect, suggesting that, as in *E. coli*, the complex of cyclic AMP and its receptor protein acts as a positive regulator. However, unlike *E. coli*, *K. pneumoniae* is only weakly lactose positive. Whether the low activity is a consequence of poor expression of *lac* genes or the production of a less active β-galactosidase is not known. *K. pneumoniae* is also different from *E. coli* in its response to melibiose, an α-galactoside. Melibiose induces β-galactosidase synthesis in *E. coli* (1), but in *K. pneumoniae* it potentiates repression and prevents induction of the *lac* operon (22).

It seems likely that in the course of evolution of enteric bacteria, the *lac* operons and *lacI* genes of *K. pneumoniae* and *E. coli* descended from a common ancestral source. A comparative analysis of the two contemporary *lac* systems could provide information on mechanisms of evolutionary change in these bacteria, as well as information on the molecular basis for the differences in the expression and properties of the *K. pneumoniae* and *E. coli lac* operons. The genes and proteins of the *E. coli lac* and *gal* repressors are related in both nucleotide and amino acid sequences (25). The two repressors share some recognition properties in that the respective operators show nucleotide sequence similarity (18) and both repressors appear to interact with the small effector molecule methylthio-β-D-galactopyranoside, although with different effects since methylthio-β-D-galactopyranoside acts as an inducer for the *lac* operon but as an anti-inducer for the *gal* operon (5). Sequence comparisons could establish the relationships of the *K. pneumoniae lac* repressor gene to the two *E. coli* repressor genes, *lacI* and *galR*.

As preparation for such analyses, previously a 4.8-kilobase (kb) *Hin*dIII fragment of *K. pneumoniae* DNA that contained the chromosomal *lac* genes was cloned, and a simple restriction map was generated (12). In this study we determined the complete nucleotide sequence of this *Hin*dIII fragment that contains the *lacI* gene, the *lac* regulatory region, the *lacZ* gene, and part of the *lacY* gene. The sequences of the structural genes are presented in this paper, and the regulatory sequence is presented in the accompanying paper (6).

MATERIALS AND METHODS

Source of *K. pneumoniae lac* DNA. A *Hin*dIII fragment of *K. pneumoniae lac* chromosomal DNA was previously isolated, and its size was estimated by electrophoretic mobility to be 4.95 kb (12). (In this work the true size was shown to be 4.8 kb.) This fragment was cloned into the pBR322 vector to create plasmid pCR100. Plasmid pCR100 DNA was isolated, and the *K. pneumoniae* fragment was excised by *Hin*dIII digestion and then was separated from vector DNA by preparative agarose electrophoresis, using standard procedures (13).

Cloning *Pst*I fragments in phage M13 vectors. The *Hin*dIII fragment of *K. pneumoniae* DNA was digested with *Pst*I, and the resulting fragments were ligated to the replicative form of one of the cloning derivatives of phage M13 (M13mp8, M13mp9, M13mp10, or M13mp11) that had been linearized by digestion with either *Pst*I or *Pst*I plus *Hin*dIII. After transfection of bacterial strain JM103, individual phages were isolated by picking colorless plaques from plates containing the indicator dye 5-bromo-4-chloro-3-indolyl-β-D-galactoside. The sizes of the insertions were estimated by gel electrophoresis of *Pst*I or *Pst*I-*Hin*dIII digests of small-scale replicative form preparations to be 1.7, 1.2, 1.0, 0.45, 0.4, and 0.15 kb. M13 derivatives with the fragment inserted in both orientations were identified by annealing single-stranded phage DNAs pairwise and determining mobility during gel electrophoresis.

Sequencing strategy. The nucleotide sequences of the three smallest fragments were determined without further subcloning. Both strands of two of the fragments were sequenced from both ends by using phage M13 containing the insertions in both orientations. The sequence of the smallest

* Corresponding author.

fragment was determined in one direction only. The three largest fragments were sequenced by using the kilobase approach (10). Each fragment was inserted into an appropriate M13 derivative in both orientations. A set of deletions was prepared for each, thus yielding two sets of deletions for each *Pst*I fragment in which one end of the fragment was unchanged while truncations were introduced at the other end. This approach provided clones with overlapping sequences for both complementary strands.

Preparation of sets of deletion derivatives. The method of Hong (10) was used to generate the sets of deletions. A low concentration of pancreatic DNase in the presence of manganous ion was used to introduce double-stranded cuts in random positions in the replicative forms of the M13 derivatives. The linearized DNA was then cut at a specific primer-proximal location (14) in the phage DNA by either *Hin*dIII (for M13mp8 or M13mp10) or *Eco*RI (for M13mp9 or M13mp11) in order to remove fragments of varying sizes from one end of the insertion DNA. Treated phage DNA then was cyclized with T4 DNA ligase. After transfection of host bacteria, the resulting phages were propagated from isolated plaques. DNA crudely derived from each phage culture was hybridized with a replicative form DNA in which a full-sized *Pst*I fragment was present but in the opposite orientation. The hybrid complex was subjected to S1 nuclease digestion to remove single-stranded DNA. The size of the resultant double-stranded insertion DNA was determined by agarose or polyacrylamide gel electrophoresis (13). From the collection of sized, deleted derivatives, a subset was chosen for sequencing that terminated at intervals of 100 to 200 base pairs (bp) throughout the original *Pst*I fragment, and this set was subjected to nucleotide sequencing procedures.

Nucleotide sequencing. The dideoxynucleotide chain termination method was used to determine the sequences (24). For most experiments, a kit consisting of chemicals, enzymes, and a 15-bp primer was used (supplied by the Bethesda Research Laboratories, Inc.), and the general features of the protocol of the manufacturer were followed. The sequence of each deletion fragment contained sequences that overlapped with the sequences of the adjacent deletion fragment, thus providing an independent check on the sequence in the overlap region. In addition, in almost all cases both strands of the DNA were sequenced, providing data for the complementary strand.

Computer analysis. The nucleotide and amino acid sequences were managed and analyzed by using the DNA Sequence Analysis Program of Bruce Conrad and David Mount (7) and the DNA Protein Sequence Analysis System of Pustell and Kofatos (21) (International Biotechnologies, Inc.). The *E. coli lac* operon sequences were obtained as the file "ecolac" from the DNA sequence data bank at Los Alamos National Laboratory.

RESULTS

The nucleotide sequence was determined for each *Pst*I fragment derived from the 4.8-kb *Hin*dIII *K. pneumoniae lac* DNA by using the strategy and methodology described above. The locations and limits of the individual genes of the *lac* operon within each fragment were detected by searching for similarities in nucleotide sequences with genes of the *E. coli lac* operon. Open reading frames were detected and amino acid sequences were deduced with the aid of computer programs. The geography of the *K. pneumoniae* operon was established in terms of the relative positions of the *lacZ*, *lacY*, and *lacI* genes (Fig. 1). In this operon, the

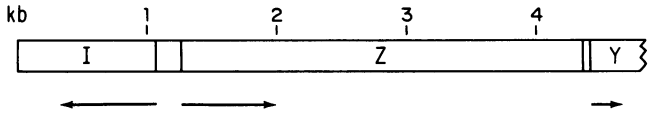


FIG. 1. Topography of the *K. pneumoniae lac* operon. The locations and directions of transcription of the genes in the 4.8-kb *Hin*dIII fragment of *K. pneumoniae* chromosomal DNA are shown. The scale (in kilobases) pertains to the DNA strand that encodes the *lacZ* and *lacY* genes.

lacZ and *lacY* genes are encoded by one strand of the DNA, and the *lacI* gene is encoded in the opposite direction by the complementary strand of the DNA. Consequently, the *K. pneumoniae lacZ* and *lacI* genes are situated in a head-to-head relationship.

The 3,102-bp sequence of the coding region of the *lacZ* gene and the deduced amino acid sequence of β-galactosidasewere analyzed relative to the corresponding *E. coli* sequences. The locations of conserved and differing residues were identified in addition to the loci at which one or the other sequence appeared to contain an excess of residues relative to the other. These features and the complete *K. pneumoniae lacZ* coding region are shown in Fig. 2. The C-terminal methionine of the *K. pneumoniae* amino acid sequence aligns with the second methionine (the third residue) of the *E. coli* amino acid sequence. The nucleotide and amino acid sequences of *K. pneumoniae lacZ* that were most poorly conserved relative to the *E. coli* sequences were found in the first 45 nucleotides, in a stretch of about 120 nucleotides located about 0.2 of the distance from the N-terminal end, in a stretch of about 350 nucleotides at about the 0.7 mark, in a segment of about 100 nucleotides at about the 0.8 mark, and finally in a group of 21 poorly conserved nucleotides at the C-terminal end. Several positions of unmatched nucleotides representing either additions or deletions in the *K. pneumoniae* and *E. coli lacZ* sequences were observed and are shown schematically in Fig. 3A. A cluster of discontinuities was found to lie in a 100-bp region near the 0.8 mark of the *lacZ* gene (Fig. 3B).

The 48-bp intercistronic region that lies between the stop codon at the end of the *lacZ* gene and the methionine codon at the beginning of the *lacY* gene was found to contain an inverted repeat sequence of eight bases (Fig. 4). Since no structures characteristic of promoters or operators were found in this region, the *K. pneumoniae lacZ* and *lacY* genes, like the *lacZ* and *lacY* genes of *E. coli*, appeared to be a single operon.

Only 401 nucleotides at the N-terminal end of the coding region of the *K. pneumoniae lacY* gene were sequenced (Fig. 5). This was due to the fact that the *lacY* gene is interrupted by a *Hin*dIII site and the gene lies at one end of the 4.8-kb *Hin*dIII fragment that was cloned from the *K. pneumoniae* genome (Fig. 1). By comparison with the corresponding *E. coli* sequences, the partial sequence of the *K. pneumoniae lacY* gene and the deduced amino acid sequence of the protein were found to be conserved or to differ at the residues indicated in Fig. 5.

The nucleotide sequence of the *K. pneumoniae lacI* gene was aligned and compared with the coding sequence of the *E. coli lacI* gene (Fig. 6). The deduced amino acid sequence of the *K. pneumoniae lac* repressor, comprising 352 amino acid residues, and the amino acid sequences of the *E. coli lac* and *gal* repressor proteins were aligned and compared for homology (Fig. 7). To optimize the similarities of the repressor amino acid sequences, four gaps of one amino acid each

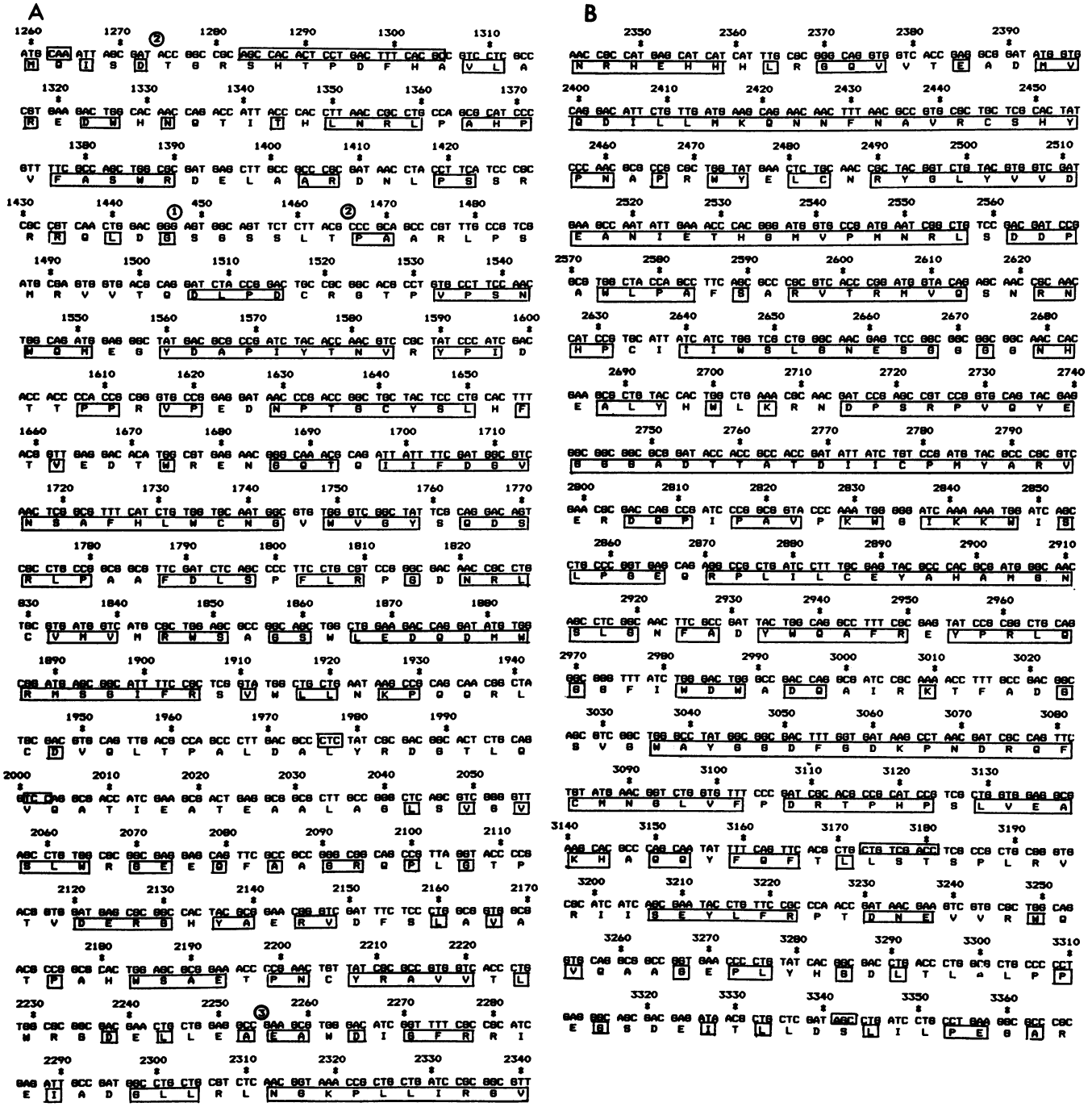
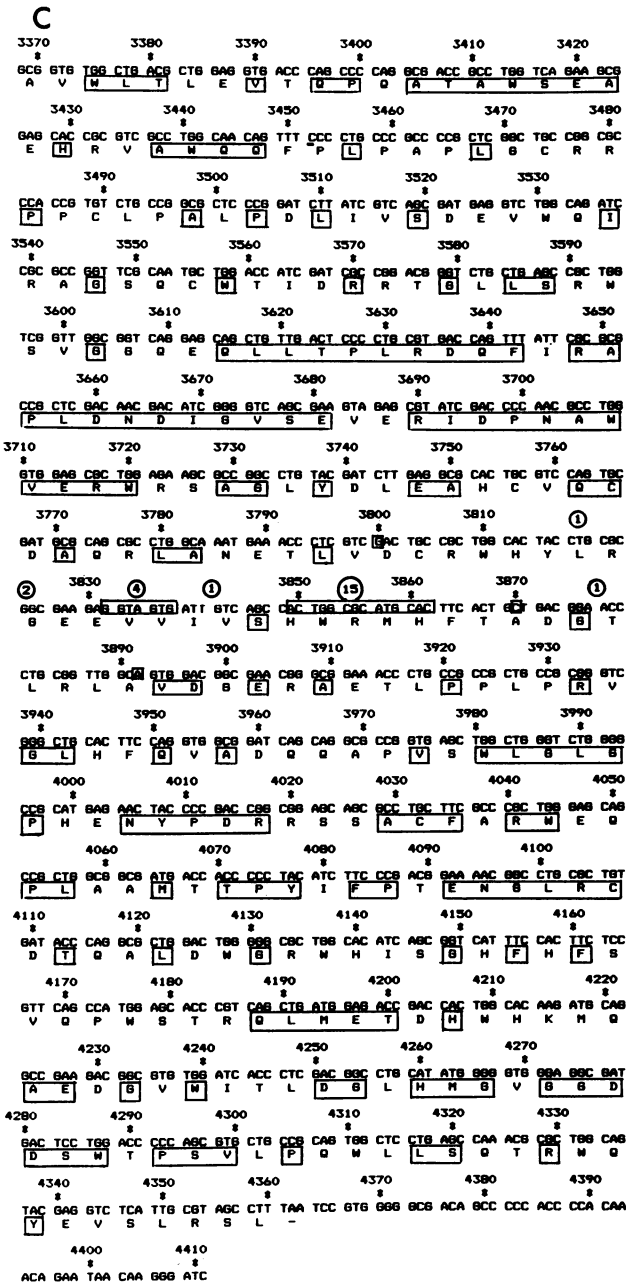


FIG. 2. Nucleotide sequence and deduced amino acid sequence of the *K. pneumoniae lacZ* gene. The numbers coincide with the kilobase coordinates in Fig. 1 and locate the nucleotides of the *lacZ* gene relative to the total sequence of the 4.8-kb *Hin*dIII fragment. The *K. pneumoniae lacZ* amino acid sequence was aligned with the *E. coli* sequence. The C-terminal methionine of the *K. pneumoniae* sequence corresponds to the second methionine in the *E. coli* sequence (the third residue from the C-terminal end). The *K. pneumoniae* amino acid residues that were conserved with respect to the amino acids of the *E. coli* sequence are shown in boxes. The excess nucleotides in the *K. pneumoniae* nucleotide sequence that have no counterpart in the *E. coli* sequence are also shown in boxes. The nucleotides in the *E. coli* sequence that have no counterpart in the *K. pneumoniae* sequence are shown as circled numbers that indicate the number of excess bases at each indicated position. The *E. coli lacZ* nucleotide sequence was determined by Kalnins et al. (11).

were introduced. Three of these occurred at the same locations in both of the *lacI* repressor sequences, and the other was in the *galR* sequence.

The degree of conservation of the nucleotide sequences and the amino acid sequences of the three repressors are summarized in Table 1. Pairwise comparisons of amino acid sequences, as well as nucleic acid sequences, revealed similar degrees of conservation of the *E. coli* and *K. pneumoniae lac* repressors and *lacI* genes. When the 5′ and 3′ overhangs were omitted from consideration, 49% of the



nucleotides of the *K. pneumoniae* and *E. coli lacI* genes were conserved in relation to each other, and 40% of the amino acid residues were conserved. The greatest concentration of conserved amino acids was found between amino acid residues 7 and 90, where 46 of 83 amino acids (55%) are the same, whereas the least conserved region was found between residues 91 and 149, where only 12 of 59 (20%) are the same. The extents of conservation of nucleotide sequences and amino acid sequences in the *lacZ*, *lacY*, and *lacI* genes are presented together with comparable data for the *trpA* gene (19) in Table 2. The relative frequencies of utilization of identical, synonymous, and replacement codons were determined for each of the four gene pairs. These data showed that the *lacI* gene is the least conserved of these genes, the *lacZ* and *lacY* genes are moderately conserved, and the *trpA* gene is the most highly conserved.

Approximately equal evolutionary distances were detected between the *E. coli lacI* and *K. pneumoniae lacI* sequences in relation to the *gal* repressor sequences. In both cases, 23% of the amino acids were conserved, and 30 to 31% of the nucleotides were conserved (Table 1), but the conserved amino acids were not identical in the two cases. Even so, a subset of 24 amino acids has been conserved in all three repressor molecules.

The patterns of codon usage were determined by using a computer program for the *lacI* and *lacZ* genes of *K. pneumoniae* and *E. coli*. A comparison revealed that the frequency of use of codons in the two *lac* genes is similar for *E. coli* and *K. pneumoniae*, except that utilization of synonymous codons with guanine or cytosine in the third position is greater in the *K. pneumoniae* genes (73% for *lacZ*, 78% for *lacI*) than it is in the *E. coli* genes (57% for *lacZ*, 59% for *lacI*).

DISCUSSION

Heterogeneity in conservation of the nucleotides and amino acids encoded within the *lac* genes of *K. pneumoniae* and *E. coli* undoubtedly reflects both stringent and loose relationships between primary structure and function. Fusion proteins with partial deletions of *E. coli* β-galactosidase have shown that the first 26 amino acids are dispensable for catalytic function of the enzyme (3, 17). Consistent with this observation, the first 16 amino acids of the *K. pneumoniae* enzyme are poorly conserved with respect to the *E. coli* sequence. It has been shown that the first eight amino acid residues of the *lac* permease of *E. coli* are dispensable (2). Consistent with this finding, the initial 14 N-terminal amino acid residues of the galactoside permease encoded by *lacY* genes of *K. pneumoniae* and *E. coli* are poorly conserved. Conservation is observed in the remaining sequence. The clusters of phenylalanine residues and the high proportion of other hydrophobic amino acid residues that characterize the *E. coli lac* permease protein (4) similarly characterize the deduced amino acid sequence of the *K. pneumoniae lac* permease (Fig. 5). The 48-bp nucleotide sequence that lies between the stop codon of the *K. pneumoniae lacZ* gene and the initiation codon of the *lacY* gene and contains an inverted repeat sequence which is high in G+C content (Fig. 4) is unlike the corresponding *E. coli* sequence, which contains an A+T-rich palindrome (4). Although this intercistronic sequence is not a good candidate for a rho-independent transcription termination sequence, since it lacks the characteristic run of thymine residues on the distal side of the hairpin loop, nevertheless the palindromic G+C-rich structure seems likely to play a role in regulating the rate of expression of the *lacY* gene relative to the *lacZ* gene.

The nucleotide sequences of the coding portions of the *K. pneumoniae* and *E. coli lacI* genes are related, with an overall level of conservation of the bases of 49% (Fig. 6). The deduced amino acid sequence of the *K. pneumoniae* repressor is similarly related to the *E. coli lac* repressor and is 40% conserved (Fig. 7). The regions that are most highly conserved among the *K. pneumoniae lac*, *E. coli lac*, and *gal* repressor proteins are the first 60 amino acids, constituting the headpiece that in the *E. coli lac* repressor is known to bind to DNA (3), and the residues between amino acids 180 and 300, in the body of the protein (Fig. 7). These commonly held residues presumably participate in establishing the general features of the three repressor proteins, the ability to bind to DNA, and the ability to interact with galactose or

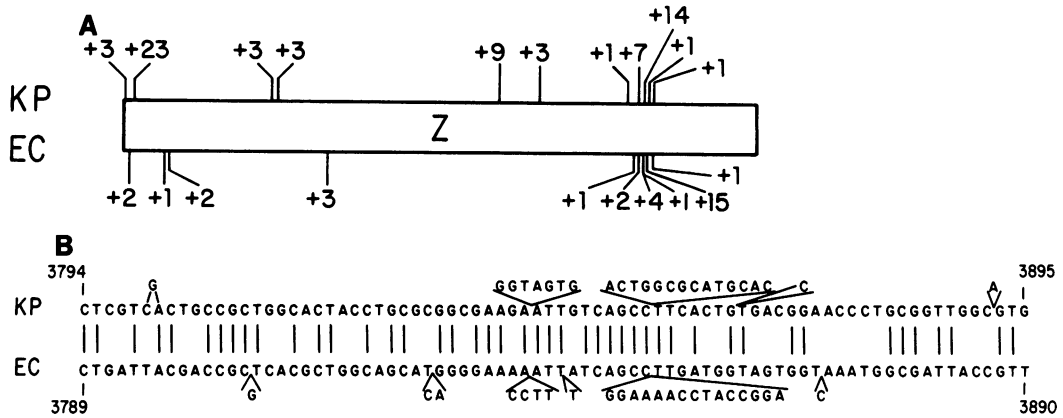


FIG. 3. Additions or deletions or both in the *K. pneumoniae lacZ* gene relative to the *E. coli lacZ* gene. (A) Schematic locations of unmatched nucleotides in the two *lacZ* genes. At each position indicated, the numbers indicate the number of nucleotides for which there is no counterpart in the other sequence. This is formally equivalent to gaps of the designated size in the opposite gene. (B) Mismatches for the aligned sequences of both the *K. pneumoniae* and *E. coli lacZ* genes in the region that contains the cluster of addition-deletion sites that are located near the 3′ ends of the genes. KP, *K. pneumoniae*; EC, *E. coli*.

small galactoside effector molecules. Presumably the unique specificities toward operator sequences and small effector molecules of the three repressor proteins are determined by some of the amino acids that differ among the three sequences. Detailed studies have been done on the effects of amino acid substitutions resulting from alterations within the *E. coli lacI* gene (15, 16). Alterations at amino acid residues 92 through 97 (i^{rc} mutants) were found to alter the interaction of the *E. coli lac* repressor with isopropylthio-β-D-galactopyranoside such that isopropylthio-β-D-galactopyranoside becomes an anti-inducer rather than an inducer (16). Of the corresponding six amino acids in the *K. pneumoniae* repressor sequence, only one is conserved, and none is conserved in the *E. coli galR* sequence. The dissimilarities in the primary sequences in this region could be related to the known differences in the physiological responses of the three operons to the small effector molecules. Reverse phenotypes are known to exist in that melibiose is an inducer for the *lac* operon in *E. coli* but is an anti-inducer with repression-potentiating action in *K. pneumoniae*, just as methylthio-β-D-galactopyranoside is an inducer for *lac* in *E. coli* but is an anti-inducer for the *gal* operon. In vitro mutagenesis experiments would address the proposition that amino acids 92 through 97 are critical in the specificity of interactions of repressors with galactosides.

Other differences among the amino acid sequences of the *lac* and *gal* repressors concern the lengths of the proteins and the existence of internal additions or deletions. At the amino termini of the proteins, the *K. pneumoniae* repressor is four amino acids longer than the *E. coli gal* repressor and two residues longer than the *E. coli lac* repressor (Fig. 7). At the C-terminal end, the *gal* repressor is 16 amino acids shorter than the *E. coli lac* repressor (26), and the *K. pneumoniae lac* repressor appears to be eight residues shorter than the *E. coli lac* repressor and eight residues longer than the *gal* repressor. However, the location of the C terminus of the *K. pneumoniae* repressor, as deduced from nucleotide sequence data, may be an artifact of cloning. The nucleotide sequence of the *lacI* gene is located at one end of the cloned 4.8-kb *Hin*dIII fragment. The cloned *lacI* gene was oriented so that the 3′ end of the gene abutted the pBR322 vector stop codon that follows the *Hin*dIII site, in frame with the coding sequence of the *lacI* gene. Since *Hin*dIII cut the *K. pneumoniae* DNA on the 5′ side of the genomic *lacI* gene stop codon, the position of the genomic stop has not yet been established. If in fact the *K. pneumoniae lacI* gene is artificially truncated in the cloned *Hin*dIII fragment, nevertheless the cloned sequence is adequate for function since expression of the associated *lacZ* gene in the plasmid is regulated, requiring action of the inducer isopropylthio-β-D-galactopyranoside to relieve repression (12).

A point-to-point comparison of the amino acid sequences of the three repressor genes showed that the two *lacI* genes are entirely congruent, but that shifts have occurred relative to the sequence of the *galR* gene (26). The four gaps that relate the two *lacI* repressors to the *galR* repressor (Fig. 7) indicate that during evolutionary divergence of the *gal* and

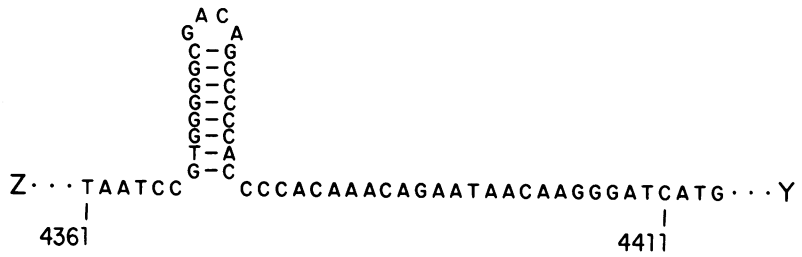


FIG. 4. Nucleotide sequence of the intercistronic DNA between the *lacZ* and *lacY* genes. The stop codon for the *lacZ* gene is shown at one end, and the ATG methionine codon of the *lacY* gene is shown at the other end. An inverted repeat sequence is shown in the form of a base-paired hairpin loop. The numerical coordinates coincide with those of Fig. 1.

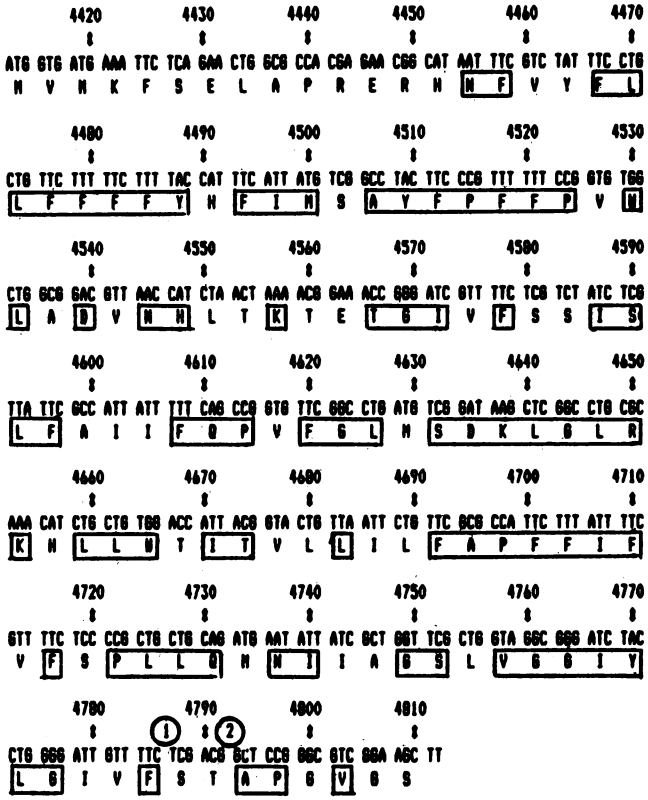


FIG. 5. Nucleotide sequence and deduced amino acid sequence of the N-terminal end of the *K. pneumoniae lacY* gene. The numerical coordinates coincide with those of Fig. 1. The amino acids that are conserved relative to the residues of the *E. coli* galactoside permease are shown in boxes. The locations and quantities of excess nucleotides in the *E. coli* sequence that have no counterpart in the *K. pneumoniae* sequence are shown as circled numbers. The nucleotide sequence of the *E. coli lacY* gene was determined by Buchel et al. (4).

lac repressor genes, four amino acid codons either were added or were deleted from either the precursor to the *lacI* genes or the precursor to the *galR* gene.

The evolutionary distance between the *galR* gene and either of the *lacI* genes is greater than the distance between the two *lacI* genes, both in terms of sequence homologies and in terms of the gaps (Table 1). These relationships indicate that an ancestral repressor gene gave rise to a *lac* type of repressor gene and a *gal* type of repressor gene at a time that preceded the divergence of the *K. pneumoniae* and *E. coli lac* repressor genes (Fig. 8). Therefore, the separation and differentiation of the enterobacterial *lac* and *gal* repressor genes appear to have occurred in an ancestral enterobacterium, preceding the subdivision and emergence of the two enteric genera *Escherichia* and *Klebsiella*.

TABLE 1. Pairwise comparisons of *lac* and *gal* repressor genes

Gene pair	Level of similarity (%)		No. of gaps
	Amino acids	Nucleotides	
E. coli lacI-K. pneumoniae lacI	40	49	0
E. coli lacI-E.coli galR	23	31	4
K. pneumoniae lacI-E. coli galR	23	30	4

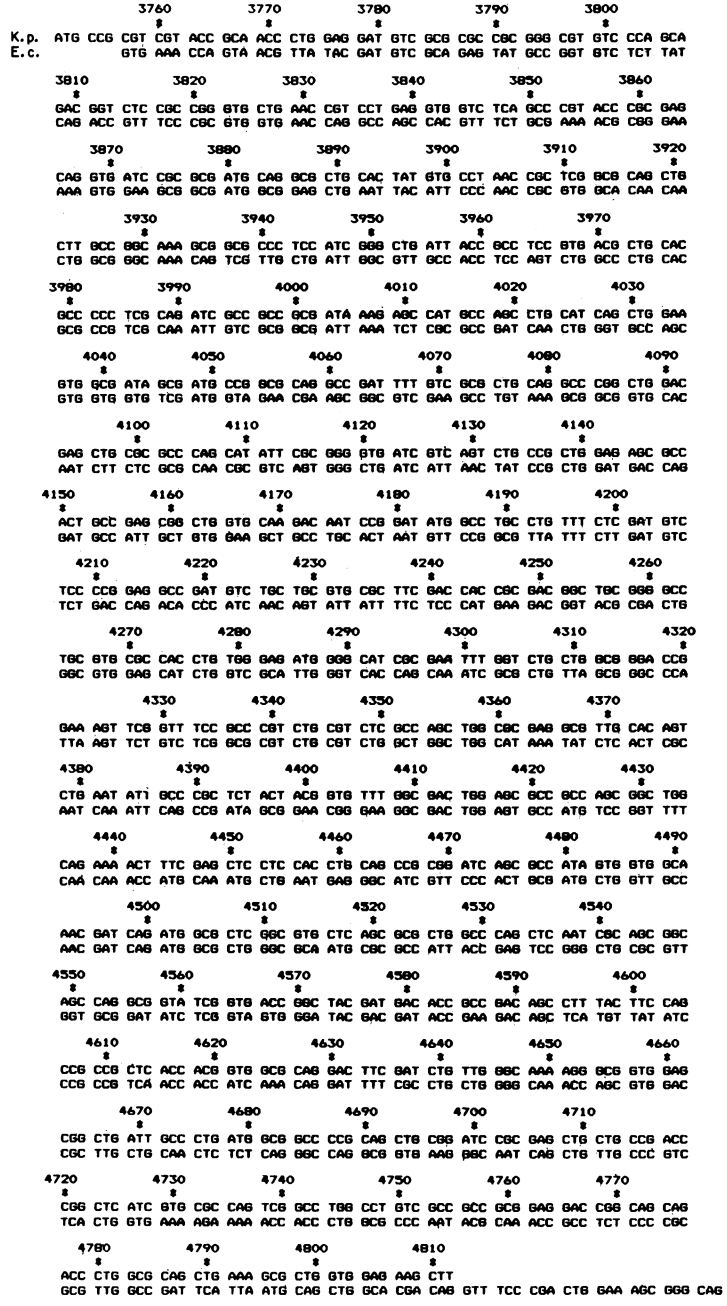


FIG. 6. Nucleotide sequences of the *K. pneumoniae* and *E. coli lacI* genes. The nucleotide numbers shown are those that were assigned to one of the strands of the 4,812-bp *Hin*dIII fragment of *K. pneumoniae* chromosomal DNA that was cloned in plasmid pCR100 (12). Numbering is clockwise relative to the clockwise numbering of vector pBR322. The *K. pneumoniae lacI* gene resides at the clockwise end of the *Hin*dIII fragment. The sequence of the *E. coli lacI* gene was determined by Farabaugh (8). K.p., *K. pneumoniae*; E.c., *E. coli*.

Codon usage in the *K. pneumoniae* and *E. coli lacZ* and *lacI* genes and in the N-terminal parts of the *lacY* genes were determined by a computer sequence analysis (data not shown). Comparisons of the codon usage data revealed that codons in the *K. pneumoniae* genes all exhibit a higher G+C content in the third position than is found in the corresponding codons of the *E. coli* genes. The same relationship was

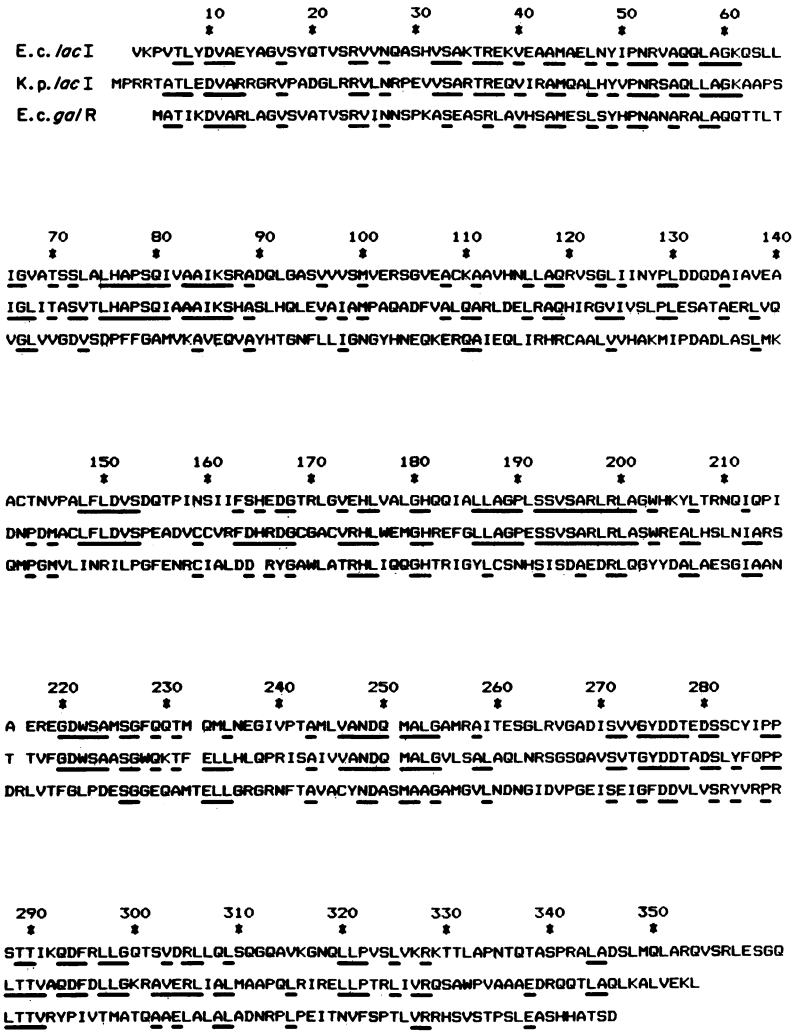


FIG. 7. Amino acid sequences encoded by the *K. pneumoniae* and *E. coli lacI* genes and the *E. coli galR* gene. See references 8 and 26. The numbers are the numbers deduced for the *K. pneumoniae lac* repressor amino acid sequence. Conserved amino acids relative to the *K. pneumoniae* sequence are underlined. Gaps were introduced at positions 165, 216, 233, and 251 to optimize sequence matching. K.p., *K. pneumoniae*; E.c., *E. coli*.

found for the *trpA* genes of *K. pneumoniae* and *E. coli* (19). These data are consistent with the greater G+C content (56%) of *K. pneumoniae* genomic DNA compared with the lesser content (51%) of *E. coli* DNA (25).

The frequency of use of certain codons has been shown to be characteristic of the level at which the genes are expressed (9). In *K. pneumoniae*, the *lac* genes utilize codons in the manner of moderately expressed genes, like the *trpA* gene of *K. pneumoniae* (19) and unlike the *lpp* gene that exhibits codon distributions that are characteristic of a high level of expression (27).

TABLE 2. Conservation of nucleotide and amino acid sequences of *K. pneumoniae* and *E. coli* genes

Gene	% of conserved nucleotides	% of conserved amino acids	% of identical codons	% of synonymous codons	% of replacement codons	No. of additions or deletions
lacZ	65	61	34.5	26	39.5	21
lacY[a]	67	65	31.2	33.8	35	2
lacI	49	40	18	22	60	0
trpA[b]	76	87	41	46	13	0

[a] Data for 400 nucleotides at the N-terminal end of the coding region of the *lacY* gene.
[b] Data from reference 19.

Point mutations are not the only kinds of changes that alter the coding regions of genes during evolutionary divergence. Additions or deletions can introduce gaps in the alignment of two related sequences that in molecular terms might reflect errors in replication or errors in sister strand recombination. The *lacZ* gene of *E. coli* (Fig. 2 and 3) embodies more interruptions than are observed in the *K. pneumoniae* and *E. coli lacY*, *lacI*, and *trpA* genes (Table 2). Clusters of the addition-deletion events within the *K. pneumoniae lacZ* gene or *E. coli lacZ* gene or both probably identify portions of the β-galactosidase polypeptide chain whose primary sequence is not critical for subunit interaction or catalytic activity. From an evolutionary standpoint, a single addition or deletion event can be substantially more powerful than a single base substitution event. As is the case for many loci in

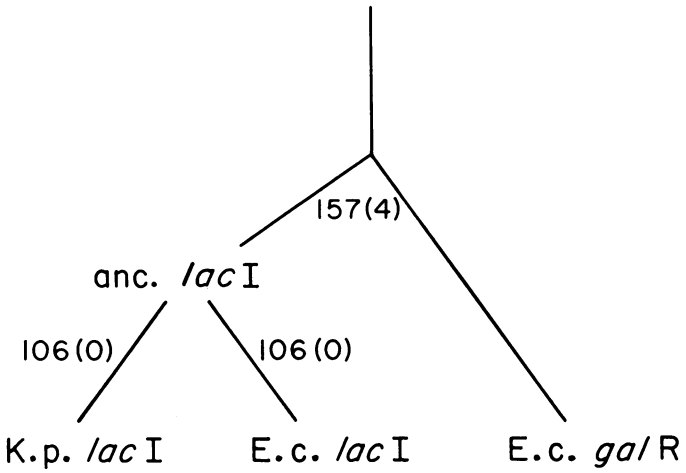


FIG. 8. Evolutionary branching order of the *K. pneumoniae* and *E. coli lacI* genes and the *E. coli galR* gene. anc, ancestral; E.c., *E. coli*; K.p., *K. pneumoniae*. The numbers of calculated nucleotide replacements are shown for each branch, and the numbers in parentheses indicate the numbers of gaps introduced.

the *K. pneumoniae* and *E. coli lacZ* genes, when additions and deletions introduce gaps that are not multiples of three into the colinear alignment of the genes, then the reading frame is changed so that the amino acid sequence downstream of the gap is altered. The new frameshift persists until similar events restore the initial reading frame as a consequence of achieving an algebraic sum of a multiple of three. Thus, a potential for major change exists as a consequence of a single molecular event.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant GM28926 from the National Institutes of Health.

LITERATURE CITED

1. **Barkeley, M. D., and S. Bourgeois.** 1978. Repressor recognition of operator and effectors, p. 177–220. *In* J. H. Miller and W. S. Reznikoff (ed.), The operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
2. **Bocklage, H., and B. Müller-Hill.** 1983. $lacZ^{-}$-Y^{+} fusions in *Escherichia coli*. Eur. J. Biochem. **137**:561–565.
3. **Brake, A. J., A. V. Flower, I. Zabin, J. Kania, and B. Müller-Hill.** 1978. β-Galactosidase chimeras: primary structure of a *lac* repressor-β-galactosidase protein. Proc. Natl. Acad. Sci. U.S.A. **75**:4824–4827.
4. **Buchel, D. E., B. Gronenborn, and B. Müller-Hill.** 1980. Sequence of the lactose permease gene. Nature (London) **283**:541–545.
5. **Buttin, G.** 1963. Méchanismes régulateurs dans la biosynthèse des enzymes du métabolisme du galactose chez *Escherichia coli* K12. J. Mol. Biol. **7**:164–182.
6. **Buvinger, W. E., and M. Riley.** 1985. Regulatory region of the divergent *Klebsiella pneumoniae lac* operon. J. Bacteriol. **163**:858–862.
7. **Conrad, B., and D. Mount.** 1982. Microcomputer programs for DNA sequence analysis. Nucleic Acids Res. **10**:31–37.
8. **Farabaugh, P. J.** 1978. Sequence of the *lacI* gene. Nature (London) **274**:765–769.
9. **Grantham, R., C. Gautier, M. Gouy, M. Jacobzone, and R. Mercier.** 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. **9**:r43–r74.
10. **Hong, G. F.** 1982. A systematic DNA sequencing strategy. J. Mol. Biol. **158**:539–549.
11. **Kalnins, A., K. Otto, U. Ruther, and B. Müller-Hill.** 1983. Sequence of the *lacZ* gene of *E. coli*. EMBO J. **2**:593–597.
12. **MacDonald, C., and M. Riley.** 1983. Cloning chromosomal *lac* genes of *K. pneumoniae*. Gene **24**:341–345.
13. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
14. **Messing, J., and J. Viera.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.
15. **Miller, J. H., C. Coulondre, M. Hofer, U. Schmeissner, H. Sommer, and A. Schmitz.** 1979. Genetic studies of the *lac* repressor. IX. Generation of altered proteins by the suppression of nonsense mutations. J. Mol. Biol. **131**:191–222.
16. **Miller, J. H., and U. Schmeissner.** 1980. Genetic studies of the *lac* repressor. X. Analysis of missense mutations in the *lacI* gene. J. Mol. Biol. **131**:223–248.
17. **Müller-Hill, B., and J. Kania.** 1974. *Lac* repressor can be fused to β-galactosidase. Nature (London) **249**:561–563.
18. **Musso, R. R. Di Lauro, M. Rosenberg, and B. deCrombrugghe.** 1977. Nucleotide sequence of the operator-promoter region of the galactose operon of *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **74**:106–110.
19. **Nichols, B. P., M. Blumenberg, and C. Yanofsky.** 1981. Comparison of the nucleotide sequence of *trpA* and sequences immediately beyond the *trp* operon of *K. pneumoniae*, *Salmonella typhimurium* and *E. coli*. Nucleic Acids Res. **7**:1743–1755.
20. **Ørskov, I.** 1974. Genus VI. *Klebsiella* Trevisan 1885, p. 321–324. *In* R. E. Buchanan and N. E. Gibbons (ed.), Bergey's manual of determinative bacteriology, 8th ed. The Williams & Wilkins Co., Baltimore.
21. **Pustell, J., and F. C. Kafatos.** 1984. A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determination. Nucleic Acids Res. **12**:643–655.
22. **Reeve, E. C. R., and J. A. Braithwaite.** 1973. The lactose system in *Klebsiella aerogenes* V9A. III. Specific repression of the *lac* operon by melibiose and raffinose. Genet. Res. **22**:217–221.
23. **Reeve, C. R., and J. A. Braithwaite.** 1974. The lactose system in *K. pneumoniae* V9A. IV. A comparison of the *lac* operons of *Klebsiella* and *E. coli*. Genet. Res. **24**:323–331.
24. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
25. **Shapiro, H. A.** 1970. Distribution of purines and pyrimidines in nucleic acids, p. H83–H85. *In* H. A. Sober (ed.), Handbook of biochemistry: selected data for molecular biology, 2nd ed. CRC Press, West Palm Beach, Fla.
26. **von Wilcken-Bergmann, B., and B. Müller-Hill.** 1982. Sequence of *galR* gene indicates a common evolutionary origin of *lac* and *gal* repressor in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **79**:2427–2431.
27. **Yamagata, H., K. Nakamura, and M. Inouye.** 1981. Comparison of the lipoprotein gene among the *Enterobacteriaceae*. J. Biol. Chem. **256**:2194–2198.