ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Nov. 1992, p. 2548–2551
0066-4804/92/112548-04$02.00/0

Vol. 36, No. 11

Comparative Sequence Analysis of the *catB* Gene from *Clostridium butyricum*

ANDREA S. HUGGINS, TRUDI L. BANNAM, AND JULIAN I. ROOD*
Department of Microbiology, Monash University, Clayton 3168, Australia

Received 12 May 1992/Accepted 1 September 1992

Sequence analysis of the *Clostridium butyricum* chloramphenicol acetyltransferase (CAT) gene, *catB*, showed that it encoded a CAT monomer of 219 amino acids with a molecular weight of 26,114. Comparison of the deduced amino acid sequence of the CATB monomer to those of sixteen other CATs showed that it was most closely related to the CATQ monomer from *Clostridium perfringens*.

Bacterial resistance to chloramphenicol is most commonly mediated by production of the enzyme chloramphenicol acetyltransferase (CAT), which catalyzes the transfer of an acetyl group from acetyl coenzyme A to the primary hydroxyl group of chloramphenicol (O-acetylation) (25). The O-acetoxy derivatives of chloramphenicol do not bind to bacterial ribosomes and are consequently devoid of antimicrobial activity (29).

Numerous CATs have been isolated from a diverse range of gram-positive and gram-negative bacteria. All natural variants of CAT are trimers composed of identical subunits of 20 to 26 kDa (8, 14). The *cat* genes which encode the CAT monomers can be divided into two groups, those in which chloramphenicol resistance is constitutively expressed and those whose expression is induced by a subinhibitory level of chloramphenicol. Usually, the *cat* genes from gram-negative bacteria are constitutively expressed, whereas those occurring in gram-positive bacteria are inducible (26). However, there are exceptions. For example, the *cat* genes found in *Clostridium perfringens* (2) and *Clostridium butyricum* (7) are constitutively expressed.

The genus *Clostridium* consists of obligately anaerobic, gram-positive spore-forming rods. Resistance to chloramphenicol is uncommon among clostridial species; however, several resistant isolates have been identified, and resistance has been shown to be due to the production of CATs. The five distinct clostridial *cat* genes that have been cloned include *catP* (1) and *catQ* (20) from *C. perfringens*, *catD* from *Clostridium difficile* (33), and *catA* and *catB* from *C. butyricum* (7). The *C. perfringens* genes *catP* (30, 31) and *catQ* (2) and the *C. difficile* gene *catD* (34) have been sequenced.

Hybridization studies carried out on the chromosomally encoded *C. butyricum* genes showed that *catA* exhibited sequence similarity to the *Staphylococcus aureus* chloramphenicol resistance plasmid pC194. However, the *catB* gene did not hybridize to any of the *cat* genes against which it was tested (7). The *catB* gene has been cloned into pUC19 to form the recombinant plasmid pWD212 (7). In this paper, we report the nucleotide sequence of the *C. butyricum catB* gene and describe the relationship of the deduced amino acid sequence to the sequences of CAT monomers from other bacteria.

Recombinant strains were derivatives of *Escherichia coli* DH5α (Bethesda Research Laboratories, Inc.) and were grown in 2YT medium (15) supplemented with ampicillin (100 μg/ml) and chloramphenicol (30 μg/ml) where appropriate. Cloning experiments conducted in this study utilized the *E. coli* plasmid vector pUC18 (35). Molecular methods used were those previously described (21). Small-scale plasmid DNA isolation was done by using an alkaline lysis method (16). DNA was recovered from agarose gels by using the Geneclean Kit (Bio 101, Inc.), following the manufacturer's instructions. Labeling of DNA probes and hybridization to Southern filters were achieved by using the Nonradioactive (DIG) DNA Labelling and Detection Kit (Boehringer-Mannheim). DNA sequence analysis of supercoiled plasmids was carried out by using the ^{T7}Sequencing kit (Pharmacia). The nucleotide sequence was read from autoradiographs by using the Rodent program (Pharmacia). The nucleotide sequence was manipulated and assembled by using the Eyeball Sequence Editor version 1.09c (E. Cabot, Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia, Canada) and MELBDBSYS, developed by the Walter and Eliza Hall Institute, the Ludwig Institute for Cancer Research, and the Howard Florey Institute (Melbourne, Australia). Sequence alignments and dendograms were obtained by using the TREEALIGN program (10). To optimize the alignments of CAT monomers, the gap penalty function of $g_k = 8 + 3 \cdot k$ was utilized, where k is the length of the insertions minus the length of deletions. The percentages of CAT identity, as well as the branch lengths, were calculated on the basis of this alignment.

The double-stranded nucleotide sequence of 892 bp of the plasmid pWD212 was determined completely on both strands and across all restriction sites, and the sequence of the *catB* coding strand is shown in Fig. 1. Analysis of the nucleotide sequence revealed the presence of two open reading frames (ORFs). ORF1 (positions 80 to 106) appeared to encode a peptide of nine amino acids which corresponded to a putative leader peptide. ORF1 was preceded by a potential ribosome-binding site (SD1; positions 69 to 74). In addition, there were potential −35 (positions 7 to 12) and −10 (positions 35 to 40) promoter sequences, with an interval gap of 22 nucleotides. This nonoptimal spacing, between the −35 and −10 promoter sequences, suggested a weak promoter. The final six nucleotides of ORF1 were found to be part of an inverted repeat (positions 101 to 111), while a second potential ribosome-binding site (SD2; positions 131 to 136) was located within the second half of the inverted repeat (positions 127 to 137) (ΔG [25°C] = −8 kcal/mol [1 cal = 4.184 J]). The arrangement and positioning of the putative leader peptide and inverted repeats in relation to one another

* Corresponding author.

```
TGAAAATTGAATAGTTCGGTATTACAGAATGTGCTATAATAAACTAAAGCGTAAATATCA 60
      -35                          -10

TTGTAAAAAGGAGATTGAAATGGCTAGGTCACGGAAAAAAGCCTTCTAAAATAGAATTAC 120
        SD1        M  A  R  S  R  K  K  A  F  *

GAAAATTTTTAGGAGGCCCGAATTATGAATTTTAATTTAATAGATATTAATCATTGGAGT 180
          SD2           M  N  F  N  L  I  D  I  N  H  W  S

AGAAAGCCATACTTTGAACATTATTTAAACAATGTGAAATGTACTTACAGTATGACTGCC 240
 R  K  P  Y  F  E  H  Y  L  N  N  V  K  C  T  Y  S  M  T  A

AATATAGAAATAACTGATTTATTGTATGAAATTAAACTTAAAAATATTAAATTTTATCCT 300
 N  I  E  I  T  D  L  L  Y  E  I  K  L  K  N  I  K  F  Y  P

ACACTTATTTATATGATTGCAACTGTGGTTAATAATCATAAAGAATTCCGTATTTGTTTT 360
 T  L  I  Y  M  I  A  T  V  V  N  N  H  K  E  F  R  I  C  F

GATCATAAAGGTAGTTTAGGATATTGGGATAGCATGAATCCAAGCTATACTATTTTTCAT 420
 D  H  K  G  S  L  G  Y  W  D  S  M  N  P  S  Y  T  I  F  H

AAAGAAAACGAAACATTTTCAAGTATTTGGACGGAATATAACAAAAGTTTTTTACGTTTT 480
 K  E  N  E  T  F  S  S  I  W  T  E  Y  N  K  S  F  L  R  F

TATAGTGATTATCTTGACGATATAAAAAACTATGGAAATATCATGAAGTTTACTCCGAAA 540
 Y  S  D  Y  L  D  D  I  K  N  Y  G  N  I  M  K  F  T  P  K

TCAAATGAACCTGACAATACATTTTCTGTATCAAGTATTCCTTGGGTGAGTTTTACAGGA 600
 S  N  E  P  D  N  T  F  S  V  S  S  I  P  W  V  S  F  T  G

TTTAACTTGAATGTTTATAATGAAGGAACATATTTAATTCCTATTTTTACTGCAGGAAAG 660
 F  N  L  N  V  Y  N  E  G  T  Y  L  I  P  I  F  T  A  G  K

TATTTCAAACAAGAAAATAAAATATTTATTCCTATATCAATACAAGTACATCATGCTATC 720
 Y  F  K  Q  E  N  K  I  F  I  P  I  S  I  Q  V  H  H  A  I

TGTGACGGTTATCATGCTAGTAGATTTATTAATGAAATGCAAGAATTAGCATTTAGTTTT 780
 C  D  G  Y  H  A  S  R  F  I  N  E  M  Q  E  L  A  F  S  F

CAAGAATGGTTAGAAAATAAATAAATATATTTAGATAGTCATGTCTTGATTACATTATTT 840
 Q  E  W  L  E  N  K  *

ATGGAGTAATAAAAATAAAGAATAGAAAAAATACCGCATTATTCAATAAAAA          892
```

FIG. 1. Nucleotide sequence of the *catB* determinant. The ORF beginning at nucleotide position 145 corresponds to the predicted CATB monomer of 219 amino acids. The ORF beginning at nucleotide position 80 corresponds to a potential leader peptide of 9 amino acids. Symbols: SD, Shine-Dalgarno; *, stop codon; ., 10-bp interval. Putative −35 and −10 sequences are in boldface and are labeled. The inverted repeat sequences are underlined. Amino acids which are conserved among all CAT variants are also underlined. The GenBank accession number of the *catB* sequence is M93113.

and to ORF2, the *catB* structural gene, were similar to those observed with inducible *cat* genes such as the *S. aureus* pUB112 *cat* gene and *Bacillus pumilus cat-86*. It has been shown that leader peptides are involved in the regulation of *cat* gene expression by a process known as translational attenuation (6, 19). An analogous situation occurs with the *C. perfringens catQ* gene, in which the structural gene is preceded by a putative leader peptide and an appropriately positioned inverted repeat (ΔG [25°C] = −25 kcal/mol) (2). Studies on the inducible *cat-86* and pUB112 *cat* genes have shown that the actual amino acid sequence of the leader peptide is important for translational attenuation, particularly amino acids 2 through 5 (6, 19). The amino acid sequences of the putative leader peptides of *catB* and *catQ* show very little similarity to the amino acid sequences of the two functional leader peptides, presumably explaining why the *catB* and *catQ* genes are not inducible.

The start codon of *catB* was located 8 bp downstream from the second ribosome-binding site, SD2. The *catB* gene was 657 bp in length, extending from position 145 to position 801. The G+C content of *catB* was 26% which was consistent with the G+C content of the *C. butyricum* genome (27 to 28%) (4). Comparison of *catB* and the *C. perfringens catQ* gene (2) indicated that they had 78% nucleotide sequence homology.

Translation of the *catB* structural gene corresponds to a protein product of 219 amino acids with a molecular weight of 26,114. The size of the predicted CATB monomer was consistent with the sizes of previously identified CAT monomers, which range from 207 to 220 amino acids. Alignment of the amino acid sequences of CATB and the known bacterial CAT variants revealed that CATB had 23 amino acids which were completely conserved among all of the CAT monomers (Fig. 1). On the basis of X-ray crystallographic studies, 15 of the 23 conserved amino acids are believed to have structural or functional significance (12, 13). The most significant region of sequence conservation is that surrounding the catalytic His-198 residue, QVHHAVCDG. This region is highly conserved among all CAT monomers, including

TABLE 1. Amino acid sequence similarities of CAT variants

Variant[a]	% Amino acid sequence similarity to variant:																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. CATB	100	72	54	53	47	47	47	48	45	42	42	41	37	51	45	47	47
2. CATQ		100	55	54	51	51	51	48	45	44	41	41	39	50	48	51	51
3. CATP			100	97	46	47	48	46	46	42	47	46	40	61	47	46	47
4. CATD				100	45	45	47	47	44	41	45	44	39	59	46	45	47
5. PC221					100	98	80	54	40	42	38	37	31	45	78	98	80
6. PUB112						100	80	53	39	43	38	37	31	45	77	99	80
7. PC223							100	57	40	41	40	38	33	46	86	80	95
8. PC194								100	41	41	39	40	33	44	54	53	57
9. PMIR									100	38	76	47	35	44	39	39	40
10. CAT86										100	39	41	32	41	41	43	42
11. ECOI											100	47	38	44	37	38	40
12. ECOIII												100	40	41	37	37	37
13. SACR													100	40	30	31	33
14. CCOLI														100	44	45	46
15. PSCS5															100	77	85
16. PSCS1																100	80
17. PSCS7																	100

[a] CAT abbreviations and sources of CAT sequence data: CATQ, *C. perfringens* CATQ (2); CATP, *C. perfringens* CATP (30); CATD, *C. difficile* CATD (34); CCOLI, *Campylobacter coli* CAT (32); PC221, *S. aureus* pC221 CAT (28); PUB112, *S. aureus* pUB112 CAT (3); PC223, *S. aureus* pC223 CAT (31); PC194, *S. aureus* pC194 CAT (11); PSCS7, *S. aureus* pSCS7 CAT (22); PSCS1, *S. intermedius* pSCS1CAT (24); PSCS5, *S. haemolyticus* pSCS5 CAT (23); PMIR, *Proteus mirabilis* CAT (5); CAT86, *B. pumilus* CAT86 (9); ECOI, *E. coli* CAT type I (27); ECOIII, *E. coli* CAT type III (18); and SACR, *Streptomyces acrimycini* CAT (17).

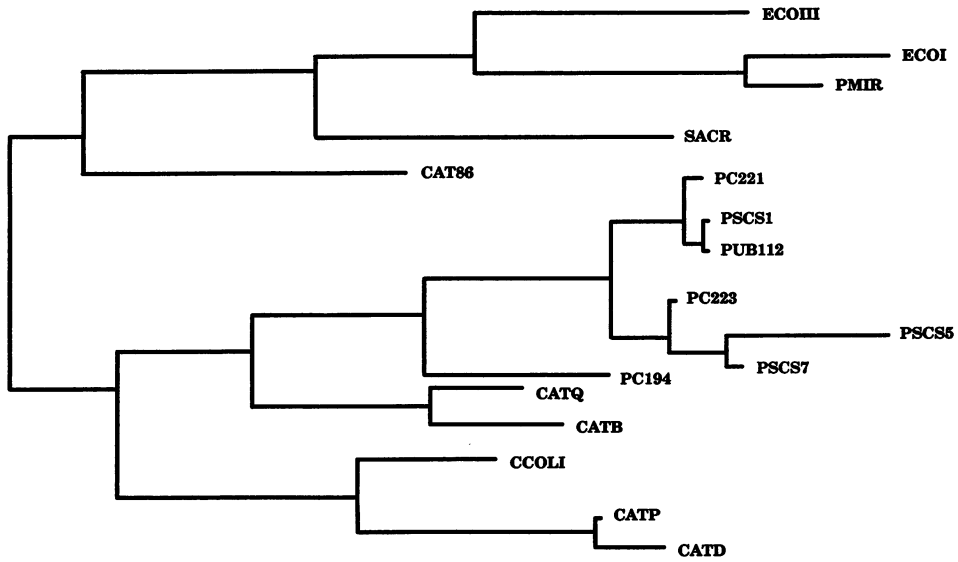

FIG. 2. Phylogenetic relationships between CAT monomers. A phylogenetic tree of CAT monomer amino acid sequences was produced by using the parsimony method in conjunction with pairwise alignment (10). For a definition of abbreviations, see Table 1, footnote *a*.

CATB, and is believed to represent the active site of CAT (12).

On the basis of sequence alignment, the amino acid sequence similarities between the CAT monomers were determined (Table 1). The CATB monomer had maximal sequence similarity (72%) with the CATQ protein. The CATB enzyme also exhibited significant levels of amino acid sequence similarity with the *C. perfringens* CATP (54%), *C. difficile* CATD (53%), and *Campylobacter coli* CAT monomers (51%) and several of the staphylococcal CAT determinants (47 to 48%). A less significant level of similarity existed between CATB and the CAT variants from *B. pumilus* (CAT-86), *Streptomyces acrimycini* (SACR), and other CAT monomers from gram-negative bacteria. Like the CATQ monomer, the CATB protein did not have the four-amino-acid (amino acids 38 to 41) deletion that is unique to the CATD, CATP, and *C. coli* monomers.

To study the potential evolutionary relationships between the various CAT monomers, a dendrogram was constructed (Fig. 2) by using the deduced amino acid sequences of the 17 sequenced CAT variants. The dendrogram suggested the presence of five main groups of CATs. The first group comprised the CAT variants of gram-negative bacteria, such as *E. coli* and *Proteus mirabilis*. The gram-positive *Streptomyces acrimycini* CAT monomer was also included within this group, though it appeared to have branched off early, as reflected by its relatively low degree of amino acid sequence similarity to all other CAT variants (30 to 40%). The second group consisted only of *B. pumilus* CAT-86, which diverged from the deepest branching point within the tree, suggesting that it was most closely related to the putative primordial CAT monomer (2). The third group consisted of the gram-positive staphylococcal monomers, all of which are plasmid encoded and inducibly expressed. The fifth group consisted of CATP, CATD, and the *C. coli* CAT monomer. The phylogenetic inconsistency within this group was the presence of the gram-negative *C. coli* CAT monomer, which was more closely related to the CATP and CATD enzymes than to the other clostridial monomers, CATB and CATQ.

The final two clostridial CAT monomers, *C. butyricum* CATB and *C. perfringens* CATQ, constitute the fourth group within the phylogenetic tree. Despite extensive nucleotide sequence similarity (78%), the G+C contents of the *catB* (26%) and the *catQ* (34%) genes are quite different and these genes did not hybridize under high-stringency conditions. The CATB and CATQ monomers occupied a position within the phylogenetic tree which was equidistant between the other clostridial CAT monomers and the staphylococcal CAT monomers. The relationship of the CATB and CATQ monomers to the staphylococcal monomers was further supported by the presence of nonfunctional leader sequences associated with *catB* and *catQ*, which were absent from *catP* and *catD*. It is proposed that the *catB*, *catQ*, and staphylococcal *cat* genes may have been derived from a common ancestral *cat* gene which was inducible. The induction mechanism has been maintained in relation to the staphylococcal *cat* genes but has presumably undergone mutation and been rendered nonfunctional in *catB* and *catQ*. Given the difference in their G+C contents, it also seems likely that the chromosomal *catB* and *catQ* genes have evolved independently from an ancestral gene rather than *catB* diverging from *catQ*, or vice versa.

Nucleotide sequence accession number. The GenBank accession number of the DNA sequence of the *catB* determinant is M93113.

We thank Walter Staudenbauer for providing pWD212 and the Australian Research Council for research support. T.L.B. was the recipient of an Australian Postgraduate Research Award.

REFERENCES

1. **Abraham, L. J., A. Wales, and J. I. Rood.** 1985. Worldwide distribution of the conjugative *Clostridium perfringens* tetracycline resistance plasmid, pCW3. Plasmid **14**:37–46.
2. **Bannam, T. L., and J. I. Rood.** 1991. Relationship between the *Clostridium perfringens catQ* gene product and chloramphenicol acetyltransferases from other bacteria. Antimicrob. Agents Chemother. **35**:471–476.
3. **Bruckner, R., and H. Matzura.** 1985. Regulation of the inducible chloramphenicol acetyltransferase gene of the *Staphylococcus aureus* plasmid pUB112. EMBO J. **4**:2295–2300.

4. **Cato, E. P., W. L. George, and S. M. Finegold.** 1986. Genus *Clostridium* Prazmowski 1880, 23^{AL}, p. 1141–1200. *In* P. H. A. Sneath, N. S. Mair, M. G. Sharpe, and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 2. The Williams & Wilkins Co., Baltimore.
5. **Charles, I. G., J. W. Keyte, and W. V. Shaw.** 1985. Nucleotide sequence analysis of the *cat* gene of *Proteus mirabilis*: comparison of the type I (Tn*9*) *cat* gene. J. Bacteriol. **164:**123–129.
6. **Dick, T., and H. Matzura.** 1988. Positioning ribosomes on leader mRNA for translational activation of the message of an inducible *Staphylococcus aureus* gene. Mol. Gen. Genet. **21:**108–111.
7. **Dubbert, W. D., H. Luczak, and W. L. Staudenbauer.** 1988. Cloning of two chloramphenicol acetyltransferase genes from *Clostridium butyricum* and their expression in *Escherichia coli* and *Bacillus subtilis*. Mol. Gen. Genet. **214:**328–332.
8. **Harding, S. E., A. J. Rowe, and W. V. Shaw.** 1987. The molecular mass and trimeric nature of chloramphenicol acetyltransferase. Biochem. Soc. Trans. **15:**513.
9. **Harwood, C. R., D. M. Williams, and P. S. Lovett.** 1983. Nucleotide sequence of a *Bacillus pumilus* gene specifying chloramphenicol acetyltransferase. Gene **24:**163–169.
10. **Hein, J.** 1990. Unified approach to alignments and phylogenies. Methods Enzymol. **183:**626–645.
11. **Horinouchi, S., and B. Weisblum.** 1982. Nucleotide sequence and functional map of pC194 that specifies inducible chloramphenicol resistance. J. Bacteriol. **150:**815–825.
12. **Kleanthous, C., C. Cullis, P. M. Cullis, and W. V. Shaw.** 1985. 3-(Bromoacetyl)chloramphenicol, an active site directed inhibitor for chloramphenicol acetyltransferase. Biochemistry **24:** 5307–5313.
13. **Leslie, A. G. W.** 1990. Refined crystal structure of type III chloramphenicol acetyltransferase at 1.75Å resolution. J. Mol. Biol. **213:**167–186.
14. **Leslie, A. G. W., P. C. E. Moody, and W. V. Shaw.** 1988. The structure of chloramphenicol acetyltransferase at 1.75Å resolution. Proc. Natl. Acad. Sci. USA **85:**4133–4137.
15. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Morelle, G.** 1989. A plasmid extraction procedure on a miniprep scale. Focus **11:**7–8.
17. **Murray, I. A., J. A. Gil, D. A. Hopwood, and W. V. Shaw.** 1989. Nucleotide sequence of the chloramphenicol acetyltransferase gene of *Streptomyces acrimycini*. Gene **85:**283–291.
18. **Murray, I. A., A. R. Hawkins, J. W. Keyte, and W. V. Shaw.** 1988. Nucleotide sequence analysis and overexpression of the gene encoding a type III chloramphenicol acetyltransferase. J. Biochem. **252:**173–179.
19. **Rogers, E. J., U. J. Kim, N. P. Ambulos, and P. S. Lovett.** 1989. Four codons in the *cat-86* leader define a chloramphenicol-sensitive ribosome stall sequence. J. Bacteriol. **172:**110–115.
20. **Rood, J. I., S. Jefferson, T. L. Bannam, J. M. Wilkie, P. Mullany, and B. W. Wren.** 1989. Hybridization analysis of three chloramphenicol resistance determinants from *Clostridium perfringens* and *Clostridium difficile*. Antimicrob. Agents Chemother. **33:**1569–1574.
21. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Schwarz, S., and M. Cardoso.** 1991. Nucleotide sequence and phylogeny of a chloramphenicol acetyltransferase encoded by the plasmid pSCS7 from *Staphylococcus aureus*. Antimicrob. Agents Chemother. **35:**1551–1556.
23. **Schwarz, S., and M. Cardoso.** 1991. Molecular cloning, purification, and properties of a plasmid-encoded chloramphenicol acetyltransferase from *Staphylococcus haemolyticus*. Antimicrob. Agents Chemother. **35:**1277–1283.
24. **Schwarz, S., U. Spies, and M. Cardoso.** 1991. Cloning and sequence analysis of a plasmid-encoded chloramphenicol acetyltransferase gene from *Staphylococcus intermedius*. J. Gen. Microbiol. **137:**977–981.
25. **Shaw, W. V.** 1967. The enzymatic acetylation of chloramphenicol by extracts of R factor-resistant *Escherichia coli*. J. Biol. Chem. **242:**687–693.
26. **Shaw, W. V.** 1975. Chloramphenicol acetyltransferase from chloramphenicol-resistant bacteria. Methods Enzymol. **43:**737–755.
27. **Shaw, W. V., D. G. Brenner, S. F. J. Legrice, S. E. Skinner, and A. R. Hawkins.** 1985. Chloramphenicol acetyltransferase of staphylococcal plasmid pC221: nucleotide sequence analysis and expression studies. FEBS Lett. **179:**101–106.
28. **Shaw, W. V., L. C. Packman, B. D. Burleigh, A. Dell, H. R. Morris, and B. S. Hartley.** 1979. Primary structure of a chloramphenicol acetyltransferase specified by R plasmids. Nature (London) **282:**870–872.
29. **Shaw, W. V., and J. Unowsky.** 1968. Mechanism of R factor-mediated chloramphenicol resistance. J. Bacteriol. **95:**1976–1978.
30. **Sloan, J., T. A. Warner, P. T. Scott, T. L. Bannam, D. I. Berryman, and J. I. Rood.** 1992. Construction of a sequenced *Clostridium perfringens-Escherichia coli* shuttle plasmid. Plasmid **27:**207–219.
31. **Steffen, C., and H. Matzura.** 1989. Nucleotide sequence analysis and expression studies of a chloramphenicol acetyltransferase coding gene from *Clostridium perfringens*. Gene **75:**349–354.
32. **Wang, Y., and D. E. Taylor.** 1991. Chloramphenicol resistance in *Campylobacter coli*: nucleotide sequence, expression, and cloning vector construction. Gene **98:**225–230.
33. **Wren, B. W., P. Mullany, C. Clayton, and S. Tabaqchali.** 1988. Molecular cloning and genetic analysis of a chloramphenicol acetyltransferase determinant from *Clostridium difficile*. Antimicrob. Agents Chemother. **32:**1213–1217.
34. **Wren, B. W., P. Mullany, C. Clayton, and S. Tabaqchali.** 1989. Nucleotide sequence of a chloramphenicol acetyl transferase gene from *Clostridium difficile*. Nucleic Acids Res. **17:**4877.
35. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.