Chloramphenicol acetyltransferase gene of staphylococcal plasmid pC221

Nucleotide sequence analysis and expression studies

William V. Shaw, Daniel G. Brenner, Stuart F.J. LeGrice*, Sarah E. Skinner+ and Alastair R. Hawkins†

Department of Biochemistry, University of Leicester, University Road, Leicester LE1 7RH, England and *Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, MA 02111, USA*

Received 22 October 1984

The nucleotide sequence of the inducible chloramphenicol acetyltransferase gene *(cat)* of *Staphylococcus aureus* plasmid pC221 has been determined. The deduced primary structure for the 215 residue polypeptide (25.9 kDa) is in agreement with partial amino acid sequence data on the purified protein, previously designated as the type C variant of CAT. In common with the inducible *cat* elements of pC194 and *B. pumilus*, the 5′ non-coding region of the *cat* of pC221 contains an inverted complementary repeat ('stem-loop' or 'hairpin') which may sequester the predicted ribosome bonding site of the mRNA. The likely transcription initiation site has been determined in vitro using purified *B. subtilis* RNA polymerase. Recombinant plasmids carrying the *cat* of pC221 on a 1156 bp *TaqI* fragment are expressed inefficiently in *Escherichia coli*, wherein induction is both poor and orientation-specific.

Chloramphenicol acetyltransferase *DNA sequence* *Plasmid* *Antibiotic resistance*

1. INTRODUCTION

Chloramphenicol acetyltransferase (EC 2.3.1.28) commonly serves as the effector of chloramphenicol resistance in bacteria by catalysing the acetyl-CoA dependent acetylation and inactivation of the antibiotic. The active form of CAT is a tetramer consisting of identical subunits of 215 to 220 amino acids, and various *cat* determinants may be found on plasmids in both Gram-positive and Gram-negative bacteria (review [1,2]). Plasmid pC221 is a 4.5 kb multicopy replicon (Inc 4) of *Staphylococcus aureus* which replicates in *Bacillus subtilis* [3] and yeasts [4] and which specifies the synthesis of the type C variant of CAT [5,6]. The synthesis of CAT by pC221 in *S. aureus* is induced by the presence of chloramphenicol [7], and the inducible phenotype is observed for the *cat* determinants of *S. aureus* plasmid pC194 [8] and the chromosomal *cat*-86 of *Bacillus pumilus* [9]. The nucleotide sequences of the latter genes have been determined [10,11]. Each contains a striking inverted complementary repeat in the 5′ non-coding region, yielded a predicted 'stem-loop' which sequesters the probable ribosomal binding sites in the deduced mRNA [12,14]. Although the precise mechanism by which chloramphenicol promotes expression is not yet clear, there is evidence that control is mainly exerted following the initiation of transcription and requires an intact stem-loop for inducibility.

* Present address: Zentrale Forschungs Einheit, Hoffman-La Roche A.G., CH 4057 Basel, Switzerland

+ Present address: Division of Biology and Medicine, Brown University, Providence, RI 02912, USA

† Present address: Department of Genetics, The University of Newcastle, Newcastle upon Tyne, NE1 7RU, England

Abbreviations: CAT, chloramphenicol acetyltransferase; kb, kilobase pairs

Published by Elsevier Science Publishers B.V.
00145793/85/$3.30

The nucleotide sequence of the *cat* gene of plasmid pC221 was examined in order to: (a) provide structural data for a gene which has been studied functionally in some detail; (b) confirm and extend the existing partial amino acid sequence of the type C variant of CAT (Fitton, J.E., unpublished); and (c) provide information which may be useful for the construction of prokaryotic vectors and expression systems. A study of the *cat* of pC221 was also a preliminary step in the larger task of understanding the structure and functions of pC221 and part of an ongoing analysis of the evolution of the CAT 'family' of genes and their products [1], the best-known example of which is the gene for the type I enzyme variant encoded by Tn*9* [15] and pBR325 [16]. Although subject to catabolite repression mediated at the transcriptional level by cyclic AMP [17,18], expression of the *cat* of Tn*9* is constitutive and is not responsive to chloramphenicol.

2. MATERIALS AND METHODS

The properties of plasmids pC221 and pCW41 and the methods for their isolation and purification from *S. aureus* 8325-4 have been described in detail elsewhere [19]. The *cat* determinant (see fig.1) and its flanking regions are contained within a 1156 bp *Taq*I fragment of pC221 which was inserted into the *Cla*I site of pBR322 to generate plasmids pSES15 and pSES16. The latter constructions contain the *cat* gene in the same and opposite

Fig.1. The nucleotide sequence of the *cat* gene of pC221 and flanking regions displayed as the 1156 base pair *Taq*I fragment corresponding to the approximate coordinates 2 kb to 3.2 kb on the map in [19]. The rightward arrow from residue 127 indicates the start of the *cat* transcript. The underlined nucleotide clusters marked −35 and −10 are the putative RNA polymerase recognition sites [29], and SD indicates the likely ribosome binding site from the observations of Shine and Dalgarno [33]. The opposing arrows correspond to the ascending and descending limbs of the proposed stem-loop structure for the corresponding mRNA (see text and fig..2). Commonly used restriction sites are marked. The vertical arrow above the His at nucleotide residues 772–774 calls attention to the amino acid likely to play a critical role in catalysis [1]. The underlined nucleotides 921–940 and 973–9 constitute a pair of inverted complementary sequences, the latter of which is followed immediately by five T residues, the composite structure being the mosst likely termination site for the mRNA [28].

orientations, respectively, to the direction of transcription of the *tet* gene of pBR322 [20] and were used to study CAT expression in *E. coli*. Plasmid pCW41 is a 2-kb replicon which contains *cat* and replication functions and which was generated by *Mbo*I cleavage of pC221 [19].

The nucleotide sequence data were obtained by the combined use of two independent strategies. The detailed restriction map of pC221 (and pCW 41) [19] allowed the use of selected restriction fragments to provide partial sequence data by the dideoxy chain termination method, using phages M13 mp8 and mp9 and a universal primer external to the cloned segment [21]. An unambiguous sequence on both strands was then obtained for the *Taq*I fragment shown in fig.1 by the use of a 'universal template' consisting of pC221 (cloned in both orientations via its unique *Hin*dIII site (in M13 mp8) and synthetic oligonucleotides as primers [22].

The determination of the start of the in vitro transcript for CAT involved the use of highly purified RNA polymerase from *B. subtilis,* dinucleotides as primers, and pCW41 as template (supercoiled or linearized by cleavage at the unique site for *Bst*EII endonuclease). The reaction conditions and polyacrylamide gel methods for the sizing of transcripts were as in [23].

The CAT induction experiments were performed essentially as in prior studies using 3-deoxychloramphenicol [7] or the 3-fluoro analogue of chloramphenicol [24], a gift of Dr G. Miller, Schering-Plough, Bloomfield, NJ.

3. RESULTS AND DISCUSSION

The nucleotide sequence data shown in fig.1 yield a deduced amino acid sequence for CAT which corresponds to a polypeptide of 25.9 kDa and which is in substantial agreement with directly determined partial sequence data on the purified protein [6,25]. The 5′ non-coding sequence contains the stem-loop structure expected by analogy with the corresponding regions of the *cat* genes of pC194 and *B. pumilus* [10,11]. The likely stability of each of these secondary structures may be estimated by calculations [26] which yield the data given in fig.2. The deduced initiation codons for *cat* of pC221 and pC194 are both AUG rather than UUG, the corresponding start codon for *cat*-86 of

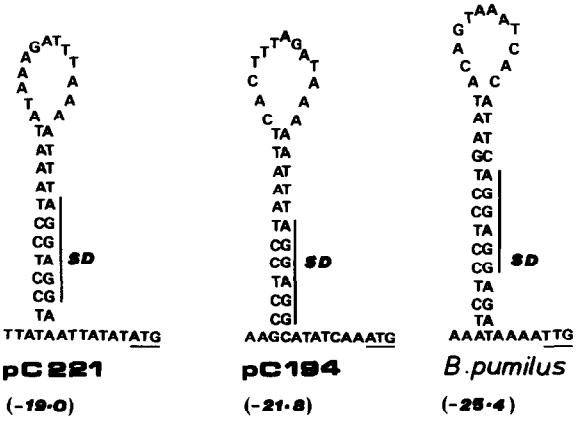

Fig.2. Predicted secondary structures of non-coding regions of the mRNA for three chloramphenicol-inducible *cat* genes with estimates of their likely relative stabilities (ΔG) in kcal · mol^{-1} [26]. For comparison with fig.1, the secondary structure diagrams show the antisense strand of the DNA duplex rather than the mRNA. The sequence for pC221 is from this study and those of pC194 *cat* and *cat*-86 pf *B. pumilus* from published data [10,11]. Nuclease digestion results for the pC194 *cat* mRNA [13] are in keeping with the structure shown, and deletion at the DNA level of the leading arm leads to constitutive expressions in *B. subtilis* [14].

B. pumilus [11] and the staphylococcal β-lactamase [27].

The likely start of the *cat* mRNA has been deduced from the results of in vitro transcription with *B. subtilis* RNA polymerase (see section 2); The 'run off' transcript observed with *Bst*EII cut plasmid was estimated to be 650–675 bases in length, localizing the likely start site to the 115–140 region of the sequence in fig.1. Dinucleotide priming of the in vitro system yielded stimulation with UpU, UpA and ApC, thus favoring the sequence -TTAC- at the start of the transcript with the A at 127 as the likely first base of the mRNA. Upstream sequences (boxed in fig.1) which seem likely to be involved in recognition by RNA polymerase are TATAAT and TTTATA. The former is identical with the '−10' site favored by both the *E. coli* holoenzyme [28] and by the *B. subtilis* vegetative (σ^{55}) transcription system [29]. The proposed '−35' sequence for the *cat* of pC221 differs at only two positions from the consensus sequence (TTGACA) deduced for *E. coli* and the *B. subtilis* σ^{55} polymerases. The corresponding −10 and −35

sequences deduced for the *cat* of pC194 are TCTAAT and TTGACT, respectively [13]. No studies were done to define the termination site, but two possible signature sequences are marked in fig.1. It should be noted that both lie distal to the *Mbo*I site (897) which is one of the two cleavage points in pC221 that were used to generate pCW41 [19]. The latter plasmid may therefore lack a normal termination site for the *cat* transcript such that it extends into the replication region of pCW41.

The chloramphenicol-inducible synthesis of CAT in Gram-positive bacteria remains a puzzle. The mRNA for the *cat* of pC194 [13] and that of the *cat*-86 of *B. pumilus* [12] seem certain to assume secondary structures that involve stem-loop ('hairpin') configurations which sequester the putative ribosome binding sites in each case. Fig.2 indicates that the *cat* of pC221 is no exception. Studies of the synthesis in *S. aureus* of CAT encoded by pC221 suggest that specific induction occurs at a concentration of chloramphenicol which is below that which inhibits growth completely, but at which effects on protein synthesis are noted [7]. Since chloramphenicol (the inducer) is acetylated following induction of CAT, it has been necessary to turn to non-substrate congeners of the parent antibiotic such as 3-deoxychloramphenicol [7] and the 3-fluoro analogue [24]. Use was made of both of these compounds here to test the inducibility of *cat* in *E. coli* in a hybrid vector consisting of the *Taq*I 'cartridge' shown in fig.1 cloned (in both orientations) into the *Cla*I site of pBR322. The

```
                1   5   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80
Tn9             MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAFLKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMK D
pC221             MTFNIIKLENWDRKEYFEHY FNQQTTYSITKEIDITLFKDMIKKKGYEIYPSLIYAIMEVVNKNKVFRTGINSE
pC194             MNFNKIDLDNWKRKEIFNHY LNQQTTFSITTEIDISVLYRNIKQEGYLFYPAFIFLVTRVINSNTAFRTGYNSD
B. pumilus        M FKQID ENYLRKEHFHHYMTLTRCSYSLVINLDITKLHAILKEKKLKVYPVQIYLLARAVQKIPEFRMDQVND
                                                                      *

                   85   90   95  100  105  110  115  120  125  130  135  140  145  150  155  160
Tn9             GELVIWDSVHPCYTVFHEQTETFSSLWSEYHDDFRQFLHIYSQDVACYGDNLAYFPKGFI ENMFFVSANPWVSFTSFDL
pC221           NKLGYWDKLNPLYTVFNKQTEKFTNIWTESDNNFTSFYNNYKNDLLEYKDKEEMFPKKPIPENTIPISMIPWIDFSSFNL
pC194           GDLGYWDKLEPLYTIFDGVSKTFSGIWTPVKNDFKEFYDLYLSDVDKYNGSGKLFPKTPIPENAFSLSIIPWTSFTGFNL
B. pumilus      ELGYWEILHPSYTILNKDTKTFSSIWTPFDENFAQFYKSCVADIETFSKSSNLFPKPHMPENMFNISSLPWIDFTSFNL
                          *                                                  *    *         *

                  165  170  175  180  185  190  195  200  205  210  215  220  225  230  235  240
Tn9             NVANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGGA
pC221           NIGNNSNFLLPIITIGKFYSENNKIYIPVALQLHHAVCDGYHASLFMNEFQDIIHKVDDWI
pC194           NINNNSNYLLPIITAGKFINKGNSIYLPLSLQVHHSVCDGYHAGLFMNSIQDLSDRPNDWLL
B. pumilus      NVSTDEAYLLPIFTIGKFKVEEGKIILPVAIQVHHAVCDGYHAGQYVEYLRWLIEHCDEWLNDSLHIT
                          *                  *        ▲
```

Fig.3. Comparison of the deduced amino acid sequence of the CAT encoded by pC221 with those pC194 [10], *cat*-86 of *B. pumilus* [11], and Tn*9* [15,16]. The sequences have been aligned to maximize homologies which are emphasized in bold type. Sites of identity are indicated by underlining, and the positions where prolines are conserved are marked with asterisks. The arrow at His-195 of the sequence for the CAT of Tn*9* marks the residue implicated as the catalytic center [1]. The deduced sequence for the CAT of pC221 is identical with that of pC194 at 118 residues and with those of *cat*-86 *(B. pumilus)* and Tn*9* at 94 and 80 residues, respectively.

basal levels of CAT were more than an order of magnitude lower than those observed with pC221 in *S. aureus,* and inducibility by 3-deoxychloramphenicol occurred only when the *TaqI cat* cartridge was in the *opposite* orientation to that of the *tet* gene of pBR322. A 5-fold increase in the specific activity of CAT was observed following exposure to 3-deoxychloramphenicol (7 μM), whereas a 50-fold increase is regularly observed for pC221 in *S. aureus* under similar conditions or when the 3-fluoro analogue of chloramphenicol (3 μM) was used (not shown). Circumstantial evidence favors the view that a sub-population of ribosomes to which chloramphenicol is bound are involved in translating the *cat* mRNA. The efficiency of this phenomenon may be related not only to recognition of the 16 S ribosomal RNA of the 30 S subunit by the Shine-Dalgarno sequence but also to a chloramphenicol-responsive change in the structure of the 50 S ribosomal subunit, allowing the latter to bind the complement of the Shine-Dalgarno sequence and destabilize the hairpin [13]. Theoretical considerations [26] argue that the stability of the stem-loop for the *cat* of pC221 is likely to be less than those of pC194 and *B. pumilus* (see fig.2).

A comparison of the deduced amino acid sequence for the CAT of pC221 with those of other variants for which published data are available is shown in fig.3. Pair-wise comparisons and a scan for over-all identities reveal patterns which are likely to be important for folding the subunits (proline residues at 7 positions) and catalytic activity (cluster centered at His-196). The conserved arginines and lysines seem likely to be involved in binding the anionic backbone of CoA [30]. The deduced sequence of the CAT encoded by pC221 is identical with that of pC194 for 118 residues and with those of *cat*-86 *(B. pumilus)* and Tn*9* at 94 and 80 positions, respectively.

The codon usage for the pC221 *cat* is shown in fig.4 and reveals a pattern not unlike that observed for the other two *cat* genes from Gram-positive bacteria [10,11]. The predominance of codons using A or U (70% overall; 77 and 78%, respectively, for the second and third bases) reflects the known content (65–68%) of A+T for *S. aureus* [31] and an even greater abundance (70%) in pC221 [22]. In the fully induced state, *S. aureus* harboring multiple copies ($\geq$20) of pC221 produces approx. 0.4% of the soluble protein as CAT [25]. Use of the rules in [32] to relate the level of gene expression of pC221 to codon usage yields an ambiguous result because of a consistent preference for U over C at the third position of both 'quartet' and 'duet' codons. Thus, in high A+T genomes base com-

Table 1

Codon usage for the *cat* gene of pC221

UUU	Phe	15	UCU	Ser	3	UAU	Tyr	13	UGU	Cys	1
UUC	Phe	1	UCC	Ser	0	UAC	Tyr	1	UGC	Cys	0
UUA	Leu	6	UCA	Ser	3	UAA	End	0	UGA	End	0
UUG	Leu	6	UCG	Ser	0	UAG	End	1	UGG	Trp	5
CUU	Leu	2	CCU	Pro	6	CAU	His	4	CGU	Arg	0
CUC	Leu	0	CCC	Pro	0	CAC	His	1	CGC	Arg	0
CUA	Leu	0	CCA	Pro	1	CAA	Gln	3	CGA	Arg	0
CUG	Leu	0	CCG	Pro	2	CAG	Gln	2	CGG	Arg	0
AUU	Ile	18	ACU	Thr	8	AAU	Asn	16	AGU	Ser	3
AUC	Ile	1	ACC	Thr	1	AAC	Asn	8	AGC	Ser	2
AUA	Ile	6	ACA	Thr	2	AAA	Lys	17	AGA	Arg	2
AUG	Met	6	ACG	Thr	2	AAG	Lys	4	AGG	Arg	0
GUU	Val	3	GCU	Ala	3	GAU	Asp	11	GGU	Gly	4
GUC	Val	0	GCC	Ala	0	GAC	Asp	1	GGC	Gly	0
GUA	Val	3	GCA	Ala	1	GAA	Glu	13	GGA	Gly	2
GUG	Val	1	GCG	Ala	0	GAG	Glu	2	GGG	Gly	0

position may over-ride a tendency towards the optimization of codon–anticodon interaction energies for highly or moderately expressed genes.

ACKNOWLEDGEMENTS

This work was supported by grants from the Medical Research Council and the Science and Engineering Research Council.

REFERENCES

[1] Shaw, W.V. (1983) Crit. Rev. Biochem. 14, 1–46.
[2] Shaw, W.V. (1984) Br. Med. Bull. 40, 36–41.
[3] Ehrlich, S.D. (1977) Proc. Natl. Acad. Sci. USA 74, 1680–1682.
[4] Gourst, R., Goze, A., Niaudet, B. and Ehrlich, S.D. (1982) Nature 298, 488–490.
[5] Shaw, W.V. and Brodsky, R.F. (1968) J. Bacteriol. 95, 28–36.
[6] Fitton, J.E. and Shaw, W.V. (1979) Biochem. J. 177, 575–582.
[7] Winshell, E. and Shaw, W.V. (1969) J. Bacteriol. 98, 1248–1257.
[8] Iordanescu, S. (1975) J. Bacteriol. 124, 597–601.
[9] Williams, D.M., Duvall, E.J. and Lovett, P.S. (1981) J. Bacteriol. 146, 1162–1165.
[10] Horinouchi, S. and Weisblum, B. (1982) J. Bacteriol. 150, 815–825.
[11] Harwood, C.R., Williams, D.M. and Lovett, P.S. (1983) Gene 24, 163–169.
[12] Duvall, E.J., Williams, D.M., Lovett, P.S., Rudolph, C., Vasantha, N. and Guyer, M. (1983) Gene 24, 171–177.
[13] Byeon, W.-H. and Weisblum, B. (1984) J. Bacteriol. 158, 543–550.
[14] Ambulos, N.P. jr, Chow, J.H., Mongkolsuk, S., Preis, L.H., Volmar II, W.R. and Lovett, P.S. (1984) Gene 28, 171–176.
[15] Alton, N.K. and Vapnek, D. (1979) Nature, 864–866.
[16] Marcoli, R., Iida, S. and Bickle, T. (1980) FEBS Lett. 110, 11–14.
[17] De Crombrugghe, B., Pastan, I., Shaw, W.V. and Rosner, J.L. (1973) Nature New Biol. 241, 237–239.
[18] Le Grice, S.F.J. and Matzura, H. (1981) J. Mol. Biol. 150, 185–196.
[19] Wilson, C.R., Skinner, S.E. and Shaw, W.V. (1981) Plasmid 5, 245–258.
[20] Stuber, D. and Bujard, H. (1981) Proc. Natl. Acad. Sci. USA 78, 167–171.
[21] Messing, J. (1981) in: Methods in Enzymology (Wu, R., Grossman, L. and Moldane, K. eds) pp. 20–78, Academic Press, New York.
[22] Brenner, D.G. and Shaw, W.V. (1984) EMBO J., submitted.
[23] Moran, C.P. jr, Lang, N., Le Grice, S.F.G., Lee, G., Stephens, M., Sonenshein, A.L., Pero, J. and Losick, R. (1982) Mol. Gen. Genet. 186, 339–346.
[24] Syriopoulou, V.P., Harding, A.L., Goldman, D.A. and Smith, A.L. (1981) Antimicrob. Agents Chemother. 19, 294–297.
[25] Fitton, J.E. (1977) Ph.D dissertation, University of Leicester.
[26] Tinoco, I. jr, Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40–41.
[27] McLaughlin, J.R., Murray, C.L. and Rabinowitz, J.C. (1981) J. Biol. Chem. 256, 11283–11291.
[28] Rosenberg, M. and Court, D. (1979) Annu. Rev. Genet. 13, 319–353.
[29] Johnson, W.C., Moran, C.P. jr and Losick, R. (1983) Nature 302, 800–804.
[30] Remington, S., Wiegand, G. and Huber, R. (1982) J. Mol. Biol. 158, 111–152.
[31] Garrity, F.L., Detrick, B. and Kennedy, E.R. (1969) J. Bacteriol. 97, 557–560.
[32] Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) Nucleic Acids Res. 9, r43–r74.
[33] Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342–1346.