
The tetracycline resistance determinants of RP1 and Tn1721: nucleotide sequence analysis

Simon H. Waters, Peter Rogowsky*, John Grinstead, Josef Altenbuchner* and Rüdiger Schmitt*+

Department of Microbiology, Medical School, University of Bristol, Bristol BS8 1TD, UK, and
*Lehrstuhl für Genetik, Universität Regensburg, D-8400 Regensburg, FRG

Received 5 July 1983; Accepted 5 August 1983

ABSTRACT

Nucleotide sequences of the homologous tetracycline resistance (tet) determinants of plasmid RP1 and transposon Tn1721 have been determined. Two open reading frames of divergent polarity have been assigned to a regulatory gene (tetR) and a gene encoding a resistance protein (tetA). The intercistronic region contains appropriate regulatory and transcription signals. The tetR gene can code for a protein of 216 amino acids (deduced mol.wt. 23,288) and the tetA gene for a protein of 399 amino acids (deduced mol. wt. 42,205). Based on the deduced amino acid sequence, the tetA proteins of RP1/Tn1721 are 78% homologous with that of pBR322 and 45% homologous with that of Tn10. We conclude that a single tetA gene mediates resistance in each of these tet determinants.

INTRODUCTION

Tetracycline resistance (tet) determinants have been grouped into at least four classes by DNA-DNA hybridization [1]. The tet determinant of RP1 homologous to that of Tn1721 [2] belongs to class A; the tet determinants of Tn10, pSC101 (and the derivative pBR322 [3]) and RA1 have been assigned to classes B, C and D, respectively [1]. Despite the apparent lack of DNA homology between the three classes, their members share the same basic mechanism of resistance involving an active efflux of tetracycline [4]. A recent analysis of the Tn1721-borne tet region [5] has revealed a regulatory gene (tetR) encoding a 26 kDa repressor protein and a structural gene (tetA) encoding a membrane protein (34 kDa) responsible for resistance. The two adjacent genes have divergent transcriptional polarity suggesting that transcription signals are located in the intercistronic region. The arrangement of regulatory and structural tet genes on Tn10 [6-8] and their products are quite similar to those of Tn1721 and have been confirmed by a recent nucleotide sequence analysis [9]. On the other

hand, two genetic loci, tetA and tetB located in the tet structural portion of Tn10, have been deduced from complementation studies [10,11]. As a representative of class C, the recombinant plasmid pBR322 [12], the entire sequence of which is known [3], carries the tet structural portion of pSC101. Here the situation was even less clear, since three partially overlapping reading frames were found in the portion responsible for resistance [3].

We report here the nucleotide sequence of the homologous tet determinants of RP1 and Tn1721 containing two open reading frames of divergent polarity and appropriate transcription signals. We show, that extensive homology exists between these and the pBR322- and Tn10-borne tet sequences also pertaining to the presence of a single tetA gene responsible for resistance.

MATERIALS AND METHODS

Materials

Restriction enzymes, T4 DNA ligase, T4 DNA polymerase, Klenow subfragment of DNA polymerase I, alkaline phosphatase from calf intestine and T4 polynucleotide kinase were purchased from Boehringer Mannheim (FRG) and Bethesda Research Laboratories (Cambridge, UK). EcoRI and SalGI were generous gifts from Drs. S. Halford and T. Maxwell, respectively. [α -³²P]dATP and [γ -³²P]ATP were obtained either from New England Nuclear (Southampton, UK) or from Amersham Buchler (Braunschweig, FRG). Fuji RS (Fuji Photo Co., Tokyo, Japan) or KODAK XAR-5 films (Eastman Kodak, Rochester, NY, USA) were used for autoradiography.

Preparation of DNA

The low copy plasmid RP1 (and derivatives used here) were prepared from 500 ml overnight broth cultures. The high copy plasmid pJOE398 [5] used as source of the Tn1721 tet region was isolated from 400 ml broth cultures grown overnight at 37°C. Plasmid DNA was prepared as described previously [13] and purified twice by CsCl gradient centrifugation. DNA used for 5'end-labelling was separated from contaminating small RNA fragments by gel filtration on Sephadex G100-120 (1 ml column). Restriction fragments were labelled at their 5'ends using polynucleotide kinase and [γ -³²P]ATP. 3' termini of EcoRI and SalI fragments were labelled by use of Klenow DNA polymerase I and [α -³²P]dATP. For 3'end-la-

belling of SstII (or its isoschizomer SacII) fragments having 3' extensions, use was made of the 3' to 5' exonuclease activity of T4 DNA polymerase in the presence of dGTP, which leaves 5'GG extensions; these were taken as template to fill in the 3'ends with [α -³²P]dCTP using Klenow DNA polymerase I as before.

DNA Sequencing

Nucleotide sequences were determined essentially according to Maxam and Gilbert [14]. A modification concerning the preparation of labelled fragments eluted from acrylamide gels involves chromatography on DEAE-cellulose DE52 (0.3 ml columns) equilibrated with T/E buffer (10 mM Tris Cl, pH 7.5, 1 mM EDTA). After washing with 10 ml of T/E buffer, the DNA was eluted with 2 M NaCl in T/E buffer; 0.4 ml fractions were precipitated by adding 5 μ l of 10 ml/mg carrier RNA (Sigma Chemical Co., Poole, UK) and 1 ml ethanol, chilled on dry ice for 5 min and collected by centrifugation in an Eppendorf centrifuge. A combination of 6%, 8% and 20% sequencing gels allowed reading up to 460 nucleotides from the labelled end in the best cases.

RESULTS AND DISCUSSION

Sequence determination

The strategy used to determine the sequence of 2,709 base pairs (bp) of RP1 between coordinates 12 kb (SacII) and 14.8 kb (SmaI) [15], a portion encompassing the tet determinant, is shown in figure 1. Derivatives of pUB307 (RP1 with a deletion that has removed TnA) and of pUB367 (a pUB307::Tn801 recombinant with Tn801 inserted at RP1 coordinate 35.7 kb) containing Tn501 insertions in and around the tet determinant [16] were used. Since Tn501 has EcoRI sites within the 38-bp terminal repeats [17], these insertions introduced EcoRI sites at various loci in the region of interest, from which sequences of the neighbouring DNA could be derived. Eight Tn501 insertions and three of the restriction sites already present in the tet region of RP1 were used for sequencing. As can be seen from figure 1 (middle), most of the sequence of both strands of the RP1 tet determinant was obtained; where the sequence of only one strand was determined the gels were unambiguous.

A similar strategy was adopted for Tn1721 to sequence the

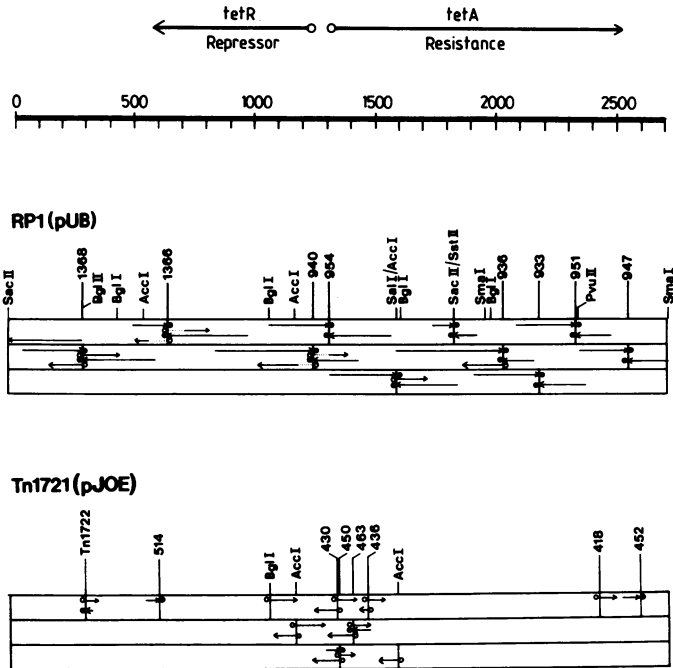


Figure 1 Top: Physical and genetic map of the 2.8-kb *SacII-SmaI* fragment of *RP1* used for sequencing (heavy single line). Numbers refer to sequenced nucleotides (Fig. 2); extent and polarities of genes *tetR* and *tetA* are indicated (arrows). Middle: Map of the *RP1* fragment showing restriction sites, positions of eight *Tn501* insertions used for sequencing (vertical lines with pUB numbers) and sequencing strategy (below). Horizontal arrows represent terminally labelled single strands pointing from 5' to 3' (o = 5'-labelling; ● = 3'-labelling). Bottom: Map of *Tn1721* portion corresponding to the superimposed *RP1* fragment with relevant restriction sites and positions of seven *Tn1725* insertions used for sequencing (vertical lines with pJOE numbers). *Tn1722* marks the transition between the *tet* region and the minor transposon (*Tn1722*) of *Tn1721* (see text). The sequencing strategy shown is represented as for *RP1*.

N- and C-terminal portions of the *tet* genes and the intercistronic region (Fig. 1, bottom). pJOE398 (an *in vitro* recombinant containing the *tet* determinant of *Tn1721*) and derivatives with *Tn1725* inserted in and around the determinant have been used previously to physically map two genes, *tetR* and *tetA*, involved in resistance [5]. *Tn1725* (a derivative of the "minor transposon" *Tn1722* with a chloramphenicol resistance marker [5]), like *Tn501*,

contains EcoRI sites within the 38-bp inverted repeats [18] thus generating new sites suitable for end-labelling. In addition to seven insertions, three restriction sites (AccI, BglI and HindII) were used for sequence determinations (Fig. 1, bottom). Here, in view of the already established homology between the RP1 and Tn1721 tet regions [1,2], the tet gene termini of Tn1721 and the extent of homology with RP1 rather than the entire sequence was probed.

Figure 2 displays the nucleotide sequence for 2,709 bp spanning the entire tet region of RP1 and table 1 lists the recognition sites within this region for various class II restriction endonucleases. It should be noted that the sequence derived here is contiguous with that previously determined for the RK2/RP1 origin of replication [19], and that the two sequences are identical for the 85 bp where they overlap. A stretch of 540 nucleotides of Tn1721 between BglI (at position 1041) and SalI (at position 1580; Fig. 2) shows identity to the corresponding sequence of RP1 apart from the six base changes marked in figure 2. None of the exchanges gives rise to a stop codon in the postulated coding sequence of tetA (see below). Five substitutions represent transitions, one is a transversion giving rise to the only non-conservative amino acid exchange (see below). Thus, the tet determinants of RP1 and Tn1721 are nearly identical with only 1% divergence in their nucleotide sequences and, therefore almost certainly have a common origin. We have previously proposed a model for the genealogy of Tn1721 [2] postulating an insertion of the "minor transposon" Tn1722 into RP1 next to the tet determinant [2,20]. The precise point of insertion between nucleotides 278 and 279 (Fig. 2) has now been identified by comparison with the sequence within Tn1721 adjacent to the minor transposon [18], thus supporting our previous notion. It is of interest to note that this is the precise insertion site of Tn501 into RP1 derivative pUB1368. It is possible that the transfer of the tet determinant of a low copy number plasmid like RP1 (RP4) onto a transposable element with the capacity, in a recombination-proficient (Rec^+) cell, for gene amplification [13] was promoted by the use of tetracycline in recent decades.

The evidence so far presented supports the model for the gene-

CTCGCACACGGCTTCGACGGCGTTTCTGGCGCTTTGCAGGGCCATAGACGGCCGCCAGC 60
GAGCGTGTGCCAAGCTGCCGCAAAGACCGCGCAAACGTCCCAGGTATCTGCCGGCGGTCC

CCAGCGGCGAGGGCAACCAGCCCGGTGAGCGTCGGAAGGGCGCTGGAAGCCCGTAGCGA 120
GGTCGCCGCTCCCCTGGTTCGGCCACTCGCAGCCTTCCGCGACCTTCGGGGCATCGCT

CGCGGAGAGGGGCGAGACAAGCCAAGGGCGCAGGCTCGAGTCGCAGCACGACATAGCCGG 180
GCGCCTCTCCCGCTCTGTTCCGTTCCCGCGTCCGAGCTCAGCGTCGTGCTGATCGGCC

TTCTCGCAAGGACGAGAATTTCCCTGCGGTGCCCTCAAGTGTCAATGAAAGTTTCCAAC 240
AAGAGCGTTCCTGCTCTTAAAGGGACGCCACGGGAGTTCACAGTTACTTTCAAAGGTTG
Tn1722
↓
GCGAGCCATTCGCGAGAGCCTTGAGTCCACGCTAGATCTATCTCATCTGCGCAAGGCAGA 300
CGCTCGGTAAGCGCTCTCGAACTCAGGTGCGATCTAGATAGAGTAGACGCTTCCGTCT
pUB1368

ACGTGAAGACGGCCGCCCTGGACCTCGCCCGCAGGCCAGCGCACGAGGCCGGCGCGCG 360
TGCACCTTCTGCCGGCGGGACCTGGAGCGGGCGCTCGCGGTGCGGTGCTCCGGCCGCGCG

GACCCGCCCGGGCCACGAGCGGACGCCCGCAGCAGGAGCGCCAGAAGGCCGCCAGAGAGG 420
CTGGGGCGCGCGGGTGTCTCGCTGCGGCGTCTCCTCGCGGTCTTCCGGCGGTCTCTCC

CCGAGCGCGGCGTGAGGCTTGGACGCTAGGGCAGGGCATGAAAAAGCCCGTAGCGGGCGC 480
GGCTCGCGCCGCACTCCGAACCTGCGATCCCGTCCCGTACTTTTTCGGGCATCGCCCGG

TACGGGGCTCTGACGCGGTGAAAGGGGGAGGGGATGTTGTCTACATGGCTCTGCTGTAG 540
ATGCCCCGAGACTGCGCCACCTTCCCCCTCCCCACAAACAGATGTACCGAGACGACATC

TGAGTGGGTTGCGCTCCGGCAGCGGTCCTGATCAATCGTCACCTTTTCTTCGGTCTTCAA 600
ACTCACCCAACGCGAGGCCGTCGCCAGGACTAGTTAGCAGTGGGAAAGAGCCAGGAAGTT
stop tetR pJOE514

CGTTCCTGACAACGAGCCTCCTTTTCGCGCAATCATCGACAATCACCGCGAGTCCCTGCT 660
GCAAGGACTGTTGCTCGGAGGAAAAGCGCTTAGGTAGCTGTTAGTGGCGCTCAGGGACGA
pUB1366

CGAACGCTGCGTCCGACCGGCTTCGTGCAAGGCGTCTATCGCGGCCCGCAACAGCGGCG 720
GCTTGCACGACAGGCCTGGCCGAAGCAGCTTCCGAGATAGCGCCGGGCGTGTGCGCCG

AGAGCGGAGCCTGTTCAACGGTGCCGCCGCGCTCGCCGGACTCGCTGTGCGCCGGCCTGCT 780
TCTCGCCTCGGACAAGTTGCCACGGCGGCGGAGCGGCTGAGCGACAGCGGCCGGACGA

CCTCAAGCACGGCCCAACAGTGAAGTAGCTGATTGTCATCAGCGCATTGACGGCGTCCC 840
GGAGTTCGTGCCGGGGTGTCACTTCATCGACTAACAGTAGTCGCGTAACTGCCGACGG

CGGCCGAAAAACCCGCCTCGCAGAGGAAGCGAAGCTGCGCGTCGGCCGTTTCCATCTGCG 900
GCCGGCTTTTTGGGCGAGCGTCTCCTTCGCTTCGACGCGCAGCCGGCAAAGGTAGACGC

GTGCGCCCGGTGCGGTGCCGGCATGGATGCGCGGCCATCGCGGTAGGCGAGCAGCGCCT 960
CACGCGGGCCAGCGCACGGCCGTACCTACGCGCGGGTAGCGCCATCCGCTCGTCGGGA

GCCTGAAGCTGCGGGCATTCCAGTCAGAAATGAGCGCCAGTCGTCGTGCGGCTCTCGGCA 1020
CGGACTTCGACGCCCCGTAAGGGTCAGTCTTTACTCGCGGTCAGCAGCAGCCGAGAGCCGT

CCGAAGTGCTATGATTCTCCGCCAGCATGGCTTCGGCCAGTGCGTCGAGCAGCGCCCGCT 1080
GGCTTCACGATACTAAGAGGGTTCGTACCGAAGCCGGTCACGCAGCTCGTCGCGGGCGA

TGTTCTGAAGTGCCAGTAAAGCGCCGGTCTGTAACCCCAACGGTTCGCCAGTTTGC 1140
ACAAGGACTTCACGGTCATTTGCGGGCCGACGACTTGGGGTTGGCAAGGCGCTCAAACG

GTGTCGTACAGACCGTCTACGCCGACCTCGTTCAACAGGTC TAGGGCGGCACGGATCACTG 1200
CACAGCAGTCTGGCAGATGCGGCTGGAGCAAGTTGTCCAGATCCCGCCGTGCCTAGTGAC

TATTCGGCTGCAACTTTGTCTATGCTTGACAGTTTATCACTGATAAACATAAATATGTCCAC 1260
ATAAGCCGACGTTGAAACAGTACGAACGTGAAATAGTGACTATTTGTATTATACAGGTG
← tetR start PUB940

CAACTTATCAGTGATAAAGAAATCCGCGCGTTC AATCGGACCAGCGGAGGCTGGTCCGGAG 1320
GTTGAATAGTCACTATTTCTTAGGCGCGCAAGTTAGCTGGTCGCTCCGACCAGGCCCTC
start tetA → G PUB954

GCCAGACGTGAAACCCAACATACCCCTGATCGTAATTCTGAGCACTGTGCGGCTCGACGC 1380
CGGTCTGCACTTTGGTTGTATGGGACTAGCATTAAAGACTCGTGACAGCGCGAGCTGCG
pJOE430 pJOE450

TGTCGGCATCGGCCTGATTATGCCGGTGTGCCGGGCCTCCTGCGCGATCTGGTTCACCTC 1440
ACAGCCGTAGCCCGACTAATAACGGCCACGACGGCCCCGGAGGACGCGCTAGACCAAGTGAG
pJOE463

GAACGACGTCAACCGCCACTATGGCATTCTGCTGGCGCTGTATGCGTTGGTGCAATTTGC 1500
CTTGCTGCAGTGGCGGCTGATACCGTAAGACGACCGCGACATACGCAACCACGTAAACG
pJOE435

CTGCGCACCTGTGCTGGGCGCGCTGTCCGATCGTTTCGGGCGGCGCCAATCTTGCTCGT 1560
GACGCGTGACACGACCCGCGGACAGCCTAGCAAAGCCCGCCGCGGTTAGAACGAGCA
GG

CTCGCTGGCCGGCGCCACTGTGACTACGCCATCATGGCGACAGCGCCTTTCTTTGGGT 1620
GAGCGACCCGGCCGCGGTGACAGCTGATGCGGTAGTACCGCTGTGCGGGAAAGGAAACCCA
TG

TCTCTATATCGGGCGGATCTGGCCGGCATCACCGGGGCGACTGGGGCGGTAGCCGGCGC 1680
AGAGATATAGCCCGCTAGCACCGGCCGTAGTGGCCCCGCTGACCCCGCCATCGGCCGCG

TTATATTGCCGATATCACTGATGGCGATGAGCGCGCGGCGCACTTCGGCTTCATGAGCGC 1740
 AATATAACGGCTATAGTGACTACCGCTACTCGCGCGCGCGTGAAGCCGAAGTACTCGCG

CTGTTTCGGGTTTCGGGATGGTTCGCGGGACCTGTGCTCGGTGGGCTGATGGGCGGTTTCTC 1800
 GACAAAGCCCAAGCCCTACCAGCGCCCTGGACACGAGCCACCCGACTACCCGCCAAAGAG

CCCCACGCTCCGTTCTTCGCGCGGCGAGCCTTGAACGGCCTCAATTTCTGACGGGCTG 1860
 GGGGGTTCGAGGCAAGAAGCGCGCCCTCGGAAGTTCGCGGAGTTAAAGGACTGCCCGAC

TTTCCTTTTTCGCGGAGTCGCACAAAGGCGAACCGCGCCGTTACGCCGGGAGGCTCTCAA 1920
 AAAGGAAAACGGCCTCAGCGTGTTCGCTTTCGCGCCGCAATGCGGCCCTCCGAGAGTT

CCCGCTCAGCTTCGTTTCGGTGGGCCCCGGGCATGACCGTCTGTCGCGCCCTGATGGCGGT 1980
 GGGCGAGTCGAAGCAAGCCACCCGGGCCCGTACTGGCAGCAGCGCGGGACTACCCCA

CTTCTTCATCATGCAACTTGTTCGGACAGGTGCCGCGCGCGTTTGGGTTCATTTTCGGCGA 2040
 GAAGAAGTAGTACGTTGAACAGCCTGTCCACGGCCGGCGCGAACCAGTAAAAGCCGCT
 pUB936

GGATCGCTTTTCACTGGGACCGCACCACGATCGGCATTTTCGCTTTCGCGCATTTGGCATTCT 2100
 CCTAGCGAAAGTGACCCTGCGCTGGTGTAGCCGTAAAGCGAACGGCGTAAACCCTAAGA

GCATTCACCTCGCCAGGCAATGATCACCGCCCTGTAGCCGCCCGGCTCGGCGAAAGCGC 2160
 CGTAAAGTGAGCGGGTCCGTTACTAGTGGCCGGGACATCGGCGGGCCGAGCCGCTTTCGCG

GGCACTCATGCTCGGAATGATTGCCAGCGCACAGGCTACATCCTGCTTGCCTTCGCGAC 2220
 CCGTTCAGTACGAGCCTTACTAACGGCTGCCGTGTCCGATGTAGGACGAACGGAAGCGCTG
 pUB933

ACGGGGATGGATGGCGTTCCCGATCATGGTCTTGCTTCGGGTGGCAGTTCGGAATGCC 2280
 TGCCCTACCTACCGCAAGGGCTAGTACCAGGACGAACGAAACCCACCGTAGCCTTACGG

GGCGCTGCAAGCAATGTTGTCCAGGCAGGTGGATGAGGAACGCCAGGGGCGAGTGCAGG 2340
 CCGCGACGTTTCGTTACAACAGGTCCGTCCACCTACTCTTCGGGTCCCCGTGACGCTTCC
 pUB951

CTCACTGGCGGCGCTCACCAGCCTGACCTCGATCGTTCGGACCCCTCCTCTTCACGGCGAT 2400
 GAGTGACCGCCGCGAGTGGTTCGACTGGAGCTAGCAGCCTGGGGAGGAGAAGTGCCGCTA

CTATGGCGGCTTCTATAACAACGTTGGAACGGGTGGGCATGGATTGCAGGCGCTGCCCTCTA 2460
 GATACGCCGAAGCATATTTGTCACCTTGCCACCCGTACCTAACGTCGCGACGGGAGAT
 pJOE418

CTTGCTCTGCCTGCCGGCGCTGCCGTGCGGGCTTTGGAGCGGCGCAGGGCAACGAGCCGA 2520
 GAACGAGACGGACGGCCGCGACGCAGCGCCGAAACCTCGCCGCTCCCGTTGCTCGGCT


```

      stop tetA           stop (tetA) .
TCGCTGATCGTGGA AACGATAGGCCTATGCCATGCGGGTCAAGGCGACTTCCGGCAAGGT 2580
AGCGACTAGCACCTTTGCTATCGGATACGGTACGCCCCAGTTCCGCTGAAGGCCGTTCCA
      pUB947
ATACGCGCCCTAGGAGTGC GGTTGGAACGTTGGCCCAGCCAGACACTCCCGATCACGAGC 2640
TATCGCGGGGATCCTCACGCCAACCTTGCAACCGGGTGGTCTGTGAGGGCTAGTGCTCG
pJOE452
AGGACGCCGATGATTTGAAGCGCACTCAGCGTCTGATCCAAGAACAACCATCCTAGCAAC 2700
TCCTGCGGCTACTAAACTTCGCGTGAGTCGCAGACTAGGTTCTTGTGGTAGGATCGTTG

AGCGCGGCG
TCGCGCCCG

```

Figure 2 Sequence of 2,709 nucleotides of the 2.8-kb RP1 fragment shown in figure 1 (middle) and of portions of Tn1721 as diagrammed in figure 1 (bottom). Seven nucleotide substitutions, that distinguish Tn1721 from RP1, are shown above the corresponding base pairs in the RP1 sequence. The transition from the tet region to the minor transposon (Tn1722) in Tn1721 and the positions of eight Tn501 insertions (boxes with pUB numbers) and of seven Tn1725 insertions (boxes with pJOE numbers) used for mapping [5,16] and sequencing are marked. The presumptive translation start and stop signals of genes tetR and tetA are underlined and marked accordingly.

alogy of Tn1721 [2,20] and the near identity of the tet determinants of RP1 and Tn1721. In subsequent discussions results pertaining to the tet genes of RP1 and Tn1721 will be taken to be equally applicable to either determinant.

Open reading frames

Bennett and Shales [21] used insertional inactivation by Tn501 and Tn801 to define regulatory and structural tet genes located on 2.2-kb fragment of RP1. By employing deletions, insertions and a protein synthesizing minicell system, Altenbuchner et al. [5] mapped two tet genes within a 2.1-kb portion of Tn1721 and characterized their products. Gene tetA, encoding a 34-kDa resistance protein, comprises approximately 1,250 bp and gene tetR encoding a 26-kDa repressor protein, comprises about 650 bp. The location and divergent transcriptional polarity of the two genes are indicated in figure 1 (top).

The nucleotide sequence (Fig. 2) contains seven potential open reading frames, but only two are non-overlapping. These are read

Table 1 Position of restriction endonuclease cleavage sites

Restriction endonuclease	Position
AccI	520, 1154, 1580
AvaI	155
BclI	569, 2121
BglI	411, 1041, 1589, 1967
BglII	274
EcoRII	317, 338, 2114, 2301, 2323
MboII	305
MstI	288, 1502
PvuI	2067, 2370, 2375
PvuII	2330
SacII (SstII)	1821, (-6)*
SalI	1580
SmaI	1944, (2750) ⁺

*Sequence overlaps that derived by Stalker et al. [19]

⁺Estimated from restriction analysis

from different strands, one running from position 1222 through 574, the other from position 1328 through 2525 (with a second stop codon in frame at position 2540). It was surmised, that they correspond to the genes tetR and tetA, respectively. This assumption is supported by the phenotypes conferred by Tn1725 insertions [5], whose loci have been sequenced (Fig. 2): insertions pJOE430, pJOE450, pJOE463, pJOE436 and pJOE418, but not pJOE462, were tetracycline-sensitive (inactivation of tetA), insertion pJOE514 expressed tetracycline resistance constitutively (inactivation of tetR). RP1 derivatives with Tn501 insertions behaved correspondingly (Fig. 1): pUB940, pUB954, pUB936, pUB933 and pUB951 did not confer tetracycline resistance, whereas pUB1368 and pUB947 did not affect the resistant phenotype [16]. The insertion pUB947 nicely defines the end of the presumptive tetA open reading frame since it lies only 11 nucleotides beyond the TGA stop codon. Taken together these data suggest, that the two open read-

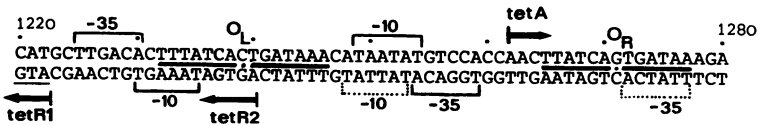


Figure 3 Portion of the intercistronic region (position 1220 to 1280, Fig. 2) containing regulatory and transcription signals of the divergently oriented *tetA* and *tetR* genes. The suggested palindromic operator sequences (O_L and O_R), promoter boxes and the initiation of the *tetA* transcript are marked above the double strand; primary and secondary (dotted) promoters and the corresponding transcription starts (*tetR1* and *tetR2*) are shown below the sequenced portion. Note that the primary transcript *tetR1* begins with AUG the presumptive *tetR* translational start codon (underlined).

ing frames, the translational starts and stops of which have been marked in figure 2, represent the regulatory (*tetR*) and structural (*tetA*) genes responsible for the inducible tetracycline resistance. The coding capacities inferred from this analysis correspond to 216 (*tetR*) and 399 (*tetA*) amino acids, respectively.

Transcription controls

The expression of both *tetR* and *tetA* is inducible by tetracycline [5]. Hence, the intercistronic region should contain transcription signals (promoters) and a repressor binding site(s), which control the two genes. Features consistent with such control elements were identified in the nucleotide sequence separating the two reading frames. The relevant portion between bp 1220 and 1280 (Fig. 2) is displayed in figure 3. This region contains two overlapping promoters for *tetR* (left) and *tetA* (right), respectively, represented by fairly well-conserved -10 and -35 boxes [22]. *In vitro* transcription studies from a fragment that contains the entire intercistronic region support this arrangement of the promoters (data not shown). There is possibly a secondary *tetR* promoter (Fig. 3, dotted) giving rise to a second, longer *in vitro* transcript (W. Hillen, personal communication). In this case, the superimposition of the -10 sequences of two divergently oriented promoter signals is a curious feature and a possible requisite of the autoregulatory system. The overlapping -10 boxes are flanked by two nearly identical sequences of dyad symmetry comprising 15 bp and 13 bp, respectively. These sequence

Table 2 Predicted amino acid composition of TetR and TetA polypeptides of RP1/Tn1721

	TetR	TetA
□ Lys (K)*	5	2
□ His (H)	3	7
□ Arg (R)	19	24
△ Cys (C)	1	4
□ Asp (D)	16	13
△ Asn (N)	7	6
△ Thr (T)	11	19
△ Ser (S)	9	17
□ Glu (E)	14	8
△ Gln (Q)	9	10
○ Pro (P)	9	19
△ Gly (G)	19	46
○ Ala (A)	33	61
○ Val (V)	14	26
○ Met (M)	4	16
○ Ile (I)	5	27
○ Leu (L)	25	56
○ Tyr (Y)	3	8
○ Phe (F)	8	21
○ Trp (W)	<u>2</u>	<u>9</u>
	216	399

*Letter code used in amino acid sequence presentation (Figs. 4 and 5). Residues classified as charged (□), uncharged polar (△) and nonpolar (○) are marked accordingly [24].

elements were shown to efficiently bind the purified Tet repressor (W. Hillen, personal communication) and were, therefore, designated as operators left (O_L) and right (O_R). The primary tetR transcript begins with AUG, the presumptive translational start implying that this message contains no leader for ribosome binding. In contrast, the second transcript (tetR2 in Fig.3) thought to be transcribed upon induction with tetracycline, carries the potential binding sequence GTGAT. This situation is reminiscent

MTKLQPN TVI	RAALDLLNEV	GVDGLTTRKL	AERLGVQQA	LYWHFRNKRA	50
LLDALAEAML	AENHSTSVPR	ADDDWRSFLT	GNARSRQAL	LAYRDGARIH	100
AGTRPGAPQM	ETADAQLRFL	CEAGFSAGDA	VNALMTISYF	TVGAVLEEQA	150
GDESEGERGG	TVEQAPLSPL	LRAAIDAFDE	AGPDAAFEQG	LAVIVDGLAK	200
RRLVVRNVEG	PRKGDD				

Figure 4 One letter code (Table 2) amino acid sequence of the tetR gene product consisting of 216 residues deduced from the nucleotide sequence (Fig. 2).

of phage lambda cI gene transcription from two promoters, Pre and Prm. The Prm-promoted mRNA also begins with the initiator codon [23]. Similar to cI gene expression in a lambda lysogen the low level of TetR protein found in uninduced (repressed) cells [5] may be due to inefficient translation from a leaderless mRNA.

The arrangement of three promoters and two operators shows striking resemblance to the tet regulatory region of Tn10 analysed by Hillen and Schollmeier [9]. Together with these investigators we have, therefore, tested the in vivo and in vitro complementation of repressor proteins and operator sequences derived from the Tn10 and RP1/Tn1721 systems. These experiments (to be presented elsewhere) show complementation of the regulatory elements of different origin, thus lending strong support to the concept depicted in figure 3.

Primary structure of the TetR protein

The amino acid composition of the putative TetR and TetA polypeptides derived from the corresponding nucleotide sequences are summarized in table 2. Figure 4 shows the predicted primary structure of the TetR repressor protein consisting of 216 amino acid residues (Table 2). This size (deduced mol. wt. 23,288) is in fair agreement with the 26 kDa TetR monomer identified by polyacrylamide gel electrophoresis [5]. The polypeptide has a 53% content of charged and polar residues [24] in accord with its location in the cytoplasm [5]. With regard to function it will be of interest to compare the primary structure to the sequence of the related TET repressors, encoded by Tn10 and pSC101, which are

now being determined (K. Bertrand and W.S. Reznikoff; W. Hillen, personal communication). Conserved regions are likely to have functional significance. Domains involved in DNA-binding and in protein-protein interactions are expected to come from crosslinking experiments, from mutant analysis and possibly from comparisons with well-studied repressor proteins like the CI- and Cro-Proteins of bacteriophage lambda [25-28].

Primary structure of the TetA proteins by RP1/Tn1721, pBR322 and Tn10

The existence of a single open reading frame (between nucleotides 1328 and 2525; Fig. 2) that encodes a TetA polypeptide chain of 399 amino acids (Table 2) is consistent with the observation of a single resistance protein in the minicell system [5]. This protein, responsible for the active efflux of tetracycline from the cell [4], is membrane-bound, a notion in agreement with the preponderance of nonpolar (hydrophobic) residues (61%) in the protein (Table 2). The calculated mol. wt. of 42,205 exceeds the size of the TetA polypeptide determined by gel electrophoresis by 20%. As discussed previously [5], this apparent discrepancy may be due to the reported higher mobility of hydrophobic proteins in SDS gel electrophoresis [29]. Another interpretation, however less likely, postulates a large precursor protein, which by rapid processing is reduced to 80% of its original size.

A comparison of the primary structures of TetA proteins derived from the corresponding nucleotide sequences of RP1/Tn1721 (Fig. 2), pBR322 [3] and Tn10 [9] indicates, that the terminal portions are largely conserved and are thus unlikely to be lost by processing. As shown in figure 5, the TetA polypeptides originating from tet determinants of three different classes [1] show striking resemblance in size and composition. The three proteins consist of approximately 400 amino acids and have a similar content and distribution of hydrophobic residues. The observed homologies are sufficient to account for an evolution from a common origin. However, amino acid and nucleotide homologies are far more obvious between RP1/Tn1721- and pBR322-borne genes than between these tet determinants and that of Tn10 (Table 3). The discrete differences suggest a pedigree with an early divergence of tetA (Tn10) from the other three determinants and a much later divergence of

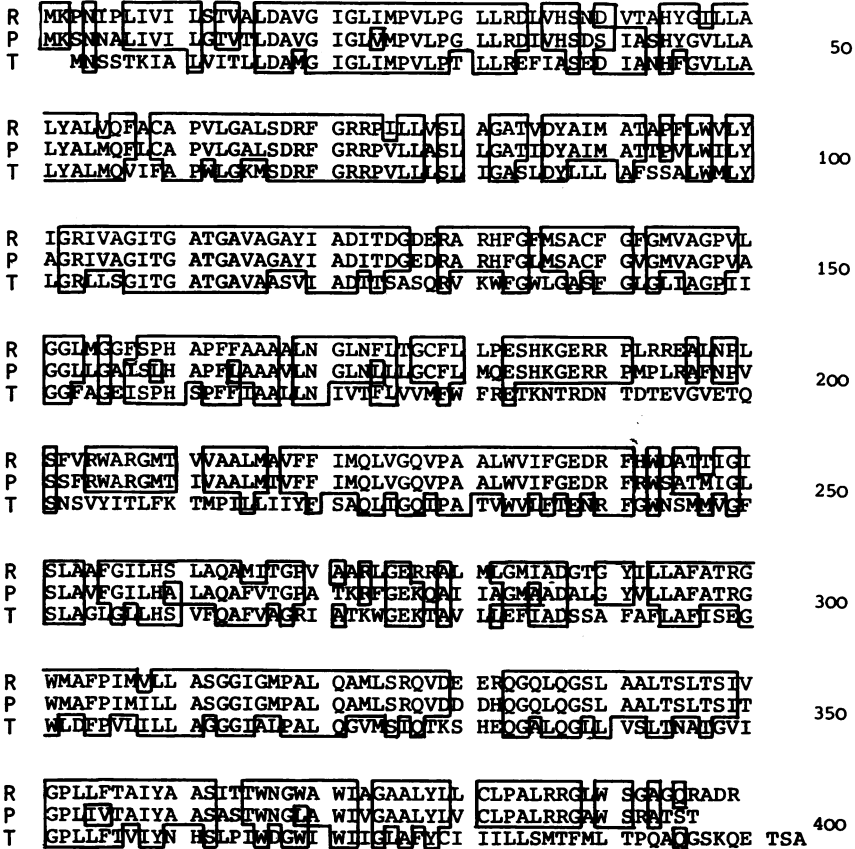


Figure 5 Comparison of the nucleotide sequence-derived primary structures of the TetA proteins of RP1/Tn1721 (R), pBR322 (p)[3] and Tn10 (T)[9]. The three amino acid sequences are superimposed, homologous residues are indicated by boxes.

Table 3 Homologies between the tetA genes borne by RP1/Tn1721, pBR322 and Tn10

Compared <u>tetA</u> sequences	Nucleotide homology	Amino acid homology	
		identical residues	polarity changes
RP1/Tn1721 : pBR322	74%	78.3%	6.6%
RP1/Tn1721 : Tn10	48%	45.1%	22.4%
pBR322 : Tn10	49%	44.9%	22.1%

the pBR322- and RP1/Tn1721-borne tet sequences with tetA (Tn1721) having evolved very recently [2,20].

This comparison was feasible due to a recent correction of the pBR322 nucleotide sequence [30] that inserts an additional G/C pair at position 526, thus generating a large open reading frame between positions 88 and 1275 [3] coding for 396 amino acid residues. It is, therefore, not necessary to account for overlapping reading frames in the analysis of tetA [9]. Similarly, complementation experiments in Tn10 [10,11] indicating the existence of two tet structural genes in the region comprised by the tetA open reading frame are thought to reflect intracistronic complementation between domains of one large membrane protein. This does not exclude the participation of an additional gene in the Tn10-mediated resistance, since another reading frame adjacent to the 5'end of tetA, but of opposite polarity, has been detected (C. Beck, personal communication). It may be possible that this extra gene will account for the extended substrate range observed in Tn10-borne tetracycline resistance [1].

ACKNOWLEDGEMENTS

We are indebted to Christoph Beck and Wolfgang Hillen for communicating their results before publication, to Sabine Unsin for technical assistance and to Helene Beier for artwork. This investigation was supported by grants from MRC (K) to JG and from the Deutsche Forschungsgemeinschaft and the Fonds der Chemie (FRG) to RS.

+To whom correspondence should be addressed

REFERENCES

- [1] Mendez, B.D., Tachibana, C. and Levy, S.B. (1980). Plasmid 3, 99-108
- [2] Schmitt, R., Altenbuchner, J., Wiebauer, K., Arnold, W., Pühler, A. and Schöffl, F. (1981) Cold Spring Harbor Symp. Quant. Biol. 45, 53-65
- [3] Sutcliffe, J.G. (1978) Cold Spring Harbor Symp. Quant. Biol. 43, 77-90
- [4] McMurry, L., Petrucci, E. Jr. and Levy, S.B. (1980). Proc. Natl. Acad. Sci. USA 77, 3974-3977
- [5] Altenbuchner, J., Schmid, K. and Schmitt, R. (1983) J. Bacteriol. 153, 116-123

-
- [6] Jorgensen, R.A. and Reznikoff, W.S. (1979). *J. Bacteriol.* 138, 705-714
- [7] Wray, L.V. Jr., Jorgensen, R.A. and Reznikoff, W.S. (1981) *J. Bacteriol.* 147, 297-304
- [8] Beck, C.F., Mutzel, R., Barbé, J. and Müller, W. (1982) *J. Bacteriol.* 150, 633-642
- [9] Hillen, W. and Schollmeier, K. (1983) *Nucl. Acids Res.* 11, 525-539
- [10] Curiale, M.S. and Levy, S.B. (1982) *J. Bacteriol.* 151, 209-215
- [11] Coleman, D.C., Chopra, I., Shales, S.W., Howe, T.G.B. and Foster, T.J. (1983) *J. Bacteriol.* 153, 921-929
- [12] Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heynecker, H.L. and Boyer, H.W. (1977) *Gene* 2, 95-113
- [13] Wiebauer, K., Schraml, S., Shales, S.W. and Schmitt, R. (1981) *J. Bacteriol.* 147, 851-859
- [14] Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-559
- [15] Ward, J.M. and Grinsted, J. (1982) *Plasmid* 8, 244-252
- [16] Grinsted, J., Bennett, P.H., Higginson, S. and Richmond, M.H. (1978) *Mol. Gen. Genet.* 166, 313-320
- [17] Brown, N.L., Choi, C.-L., Grinsted, J., Richmond, M.H. and Whitehead, P.R. (1980) *Nucl. Acids Res.* 8, 1933-1945
- [18] Schöffl, F., Arnold, W., Pühler, A., Altenbuchner, J. and Schmitt, R. (1981) *Mol. Gen. Genet.* 172, 53-65
- [19] Stalker, D.M., Thomas, C.M. and Helinski, D.R. (1981) *Mol. Gen. Genet.* 181, 8-12
- [20] Altenbuchner, J., Choi, C.-L., Grinsted, J., Schmitt, R. and Richmond, M.H. (1981) *Genet. Res., Camb.* 37, 285-289
- [21] Bennett, P.M. and Shales, S.W. (1981) in *Molecular Biology, Pathogenicity and Ecology of Bacterial Plasmids*, Levy, S.B., Clowes, R.C. and Koenig, E.L., Eds., p. 581 Plenum Press, New York and London
- [22] Hawley, D.K. and McClure, W.R. (1983) *Nucl. Acids Res.* 11, 2237-2255
- [23] Walz, A., Pirrotta, V. and Ineichen, K. (1976) *Nature* 262, 665-669
- [24] Capaldi, R.A. (1982) *TIBS* 7, 292-295
- [25] Pabo, C.O. and Mitchell, L. (1982) *Nature* 298, 443-447
- [26] Sauer, R.T., Yocum, R.R., Doolittle, R.F., Lewis, M. and Pabo, C.O. (1982) *Nature* 298, 447-451
- [27] Anderson, W.F., Ohlendorf, D.H., Takeda, Y. and Matthews, B.W. (1981) *Nature* 290, 754-758
- [28] Steitz, T.A., Ohlendorf, D.H., McKay, D.B., Anderson, W.F. and Matthews, B.W. (1982) *Proc. Natl. Acad. Sci. USA* 79, 3097-3100
- [29] Ehring, R., Beyreuther, K., Wright, J.K. and Overath, P. (1980) *Nature* 283, 537-540
- [30] Livneh, Z. (1983) *Proc. Natl. Acad. Sci. USA* 80, 238-241
-