Volume 12 Number 12 1984 Nucleic Acids Research

Nucleotide sequence of the repressor gene of the TN*10* tetracycline resistance determinant

Kathleen Postle, Toai T.Nguyen and Kevin P.Bertrand

Department of Microbiology and Molecular Genetics, University of California, Irvine, CA 92717, USA

Received 23 April 1984; Accepted 21 May 1984

ABSTRACT

The Tn10 tetR gene encodes the repressor that regulates transcription of the Tn10 tetracycline resistance determinant. We have determined the DNA sequence of the tetR gene and a 905 base pair region immediately 3' to tetR. The tetR gene is located on a 701 base pair HincII restriction fragment. Deletions at either end of this region eliminate synthesis of the wild-type TetR protein in E. coli minicells, and eliminate TetR activity as measured by repression of β-galactosidase synthesis in tetA-lacZ operon fusion strains. Taken together, the DNA sequence and the genetic data indicate that tetR encodes a 207 amino acid protein with a calculated molecular weight of 23,328. This value is in good agreement with estimates of 23,000-25,000 based on electrophoretic mobility in SDS-polyacrylamide gels. There is 47% amino acid sequence homology between the deduced sequences of the Tn10 and RP1/Tn1721 TetR proteins. There is, in addition, significant amino acid sequence homology between an NH_2-terminal region of the Tn10 TetR repressor and the DNA recognition regions of other DNA-binding proteins.

INTRODUCTION

The transposon Tn10 confers tetracycline resistance in E. coli and other enteric bacteria [1,2]. The Tn10 tetracycline resistance determinant (tet) resembles other plasmid-encoded tetracycline resistance determinants [3,4] in that (i) the mechanism of resistance involves the active efflux of tetracycline [5,6], and (ii) the expression of resistance is regulated; prior exposure to low levels of tetracycline increases the level of resistance [7].

As shown in Figure 1, the Tn10 tet region consists of two genes, the resistance gene (tetA) and the repressor gene (tetR), that are transcribed from divergent promoters located in the regulatory region between the two genes [8-12]. The resistance gene encodes a membrane-associated protein (TetA) that is necessary and, it appears, sufficient for resistance [8,9,13-17]. Based on its electrophoretic mobility in SDS-polyacrylamide gels, the TetA protein appears to have a molecular weight of about 36,000; however, the molecular weight deduced from the tetA DNA sequence is 43,200 [16,18]. The repressor gene encodes a regulatory protein (TetR) that, based on its electrophoretic mobility in SDS-polyacrylamide gels, has a molecular weight of 23,000-25,000 [9,10,19,20].

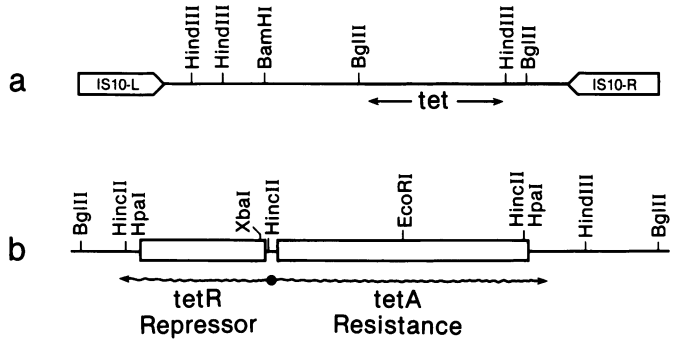


Figure 1. Physical and genetic maps of the transposon Tn10 and the Tn10 tet region. (a) Tn10 is 9,300 base pairs (bp) in length [27]. The central region is flanked by 1,330 bp IS10 sequences, designated IS10-Left and IS10-Right, which are in opposite orientations [28]. (b) The central 2,790 bp BglII fragment spans the structural genes for the TetR repressor and TetA resistance proteins (open bars); tetA and tetR are transcribed from divergent promoters located between the structural genes.

The regulation of the Tn10 tet genes has been examined both in vivo [8-10,14,20-23] and in vitro [12,24]. Hillen and his colleagues [12,19,25,26] have purified the TetR repressor and characterized its interaction with tetracycline and tet operator DNA in vitro. All of the data support a model of negative regulation at the level of transcription initiation: (i) the TetR repressor inhibits transcription of both tetA and tetR by binding to tandem operator sites that overlap the tetA and tetR promoters, and (ii) tetracycline induces transcription of both tetA and tetR by binding to the repressor and reducing its affinity for the tet operators. In this paper, we report the DNA sequence of the Tn10 tetR structural gene and the deduced amino acid sequence of the repressor protein.

METHODS

Plasmid Constructions

Restriction enzymes were purchased from Bethesda Research Laboratories or New England Biolabs. Plasmids were constructed and propagated in the E. coli K-12 strain JA221 (lacY1 leuB6 trpΔE5 thi-1 tonA2 supE44 recA1 hsr) [29], except plasmids to be digested with the restriction enzyme BclI, which were propagated in the adenine methylase deficient strain GM33 (dam-3) [30]. pBT1010 ($Ap^rNm^rTc^r$) [31] consists of Tn10 transposed from λNK370 (λ b221 cI857 cI171::Tn10 Ouga261, N. Kleckner) into the 3,670 bp plasmid pACYC177 (Ap^rNm^r) [32]. pBT107 (Ap^r Nm^r Tc^r) [31] consists of the 2,790 bp $tetR^+$ $tetA^+$ BglII fragment of λ::$Tn10_1$ [27], inserted into the BamHI site of pACYC177; tetR and bla are transcribed in the same orientation. pBT201 (Ap^r Tc^r) consists of the 4,800 bp $tetR^+$ $tetA^+$ HindIII fragment of λ::$Tn10_1$, inserted into the

HindIII site of pACYC177; tetR and bla are transcribed in the same orientation. pBT211 is identical to pBT201, except that the HindIII fragment is in the opposite orientation. pBT402 (Nm^r; Figure 2) consists of the 701 bp $tetR^+$ HincII fragment of pBT107 inserted into the HincII site of pACYC177; tetR and bla are transcribed in the same orientation. pBT401 (Nm^r) is identical to pBT402, except that the 701 bp fragment is in the opposite orientation. pBT4022 (Nm^r; Figure 2) is identical to pBT402, except that it contains a deletion of the 44 bp BclI fragment within the 3' end of tetR. pBT4022 was constructed by digesting pBT402 DNA with BclI and religating; there are no BclI sites in pACYC177. pBT4025 (Nm^r; Figure 2) is identical to pBT402, except that it contains a deletion of the 83 bp HincII-HinfI fragment that spans the 5' end of tetR. pBT4025 was constructed by digesting the 701 bp HincII fragment of pBT402 with HinfI, incubating with deoxynucleoside triphosphates and the large fragment of DNA polymerase I (Bethesda Research Laboratories) and ligating with HincII-digested pACYC177 DNA. pBT4028 (Nm^r; Figure 2) is identical to pBT402, except that it contains a deletion of the 25 bp HincII-XbaI fragment that spans the 5' end of tetR. pBT4028 was constructed in a manner analogous to the construction of pBT4025.

β-Galactosidase Assays

B2550(λRStet158-43) lysogens synthesize β-galactosidase under the control of the Tn10 tetA promoter-operator region [23]. Plasmid-containing derivatives of B2550(λRStet158-43) were incubated at 32°C in LB broth containing neomycin sulfate (40 μg/ml). Saturated cultures were diluted 1:100 into fresh medium and incubated until the optical density at 550 nm reached 0.3-0.6. Cultures to be induced contained the indicated amounts of 5a,6-anhydrotetracycline (kindly provided by N. Belcher, Pfizer) during both the pre-growth and subculturing steps [23,29]. β-Galactosidase activities were determined as described by Miller [33].

Analysis of Plasmid-Encoded Proteins in Minicells

The procedures for minicell preparation, labeling, and gel electrophoresis of protein samples were described previously [29]. In brief, plasmids were introduced into the minicell strain P678-54 and minicells were prepared and then labeled for 15 min with [^{35}S]methionine (20 μCi/ml; 1,200 Ci/mmole; Amersham). Samples to be induced received anhydrotetracycline (1 μg/ml) immediately prior to the addition of [^{35}S]methionine. Labeled minicell proteins were fractionated in a 12.5% polyacrylamide-SDS slab gel. The molecular weight standards (Bio-Rad Laboratories) were: lysozyme (14,400); soybean trypsin inhibitor (21,500); carbonic anhydrase (31,000); ovalbumin (45,000); bovine serum albumin (66,200); and phosphorylase B (92,500).

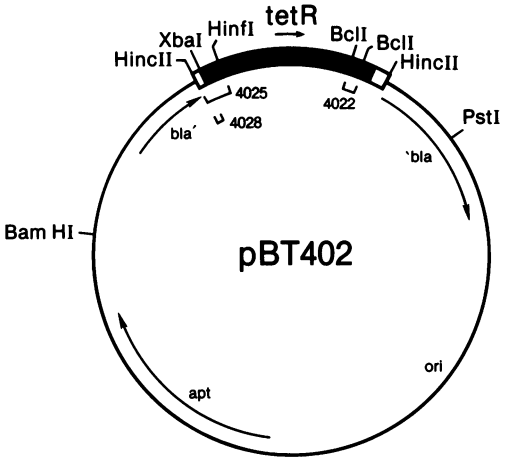


Figure 2. Physical and genetic map of the tetR$^+$ plasmid pBT402. The 701 bp Tn10 HincII fragment was inserted into the HincII site in the β-lactamase (bla) gene of pACYC177. The 701 bp fragment spans the tetR structural gene (closed bar); however, it does not include the tetR promoter-operator region. Arrows indicate the extent of the bla and aminoglycoside phosphotransferase (apt) genes and the direction of their transcription. Brackets indicate the extent of the deletions in the tetR$^-$ plasmids pBT4022, pBT4025 and pBT4028.

DNA Sequence Analysis

The DNA sequence was determined by the methods of Maxam and Gilbert [34]. DNA fragments were either labeled at their 5' termini by treatment with alkaline phosphatase (Bethesda Research Laboratories) and incubation with [γ-^{32}P]ATP (3,000 Ci/mmole, Amersham) and polynucleotide kinase (Bethesda Research Laboratories), or labeled at their 3' termini by incubation with [α-^{32}P]dGTP (3,000 Ci/mmole, Amersham) and the large fragment of DNA polymerase I. DNA fragments for sequencing were isolated from the plasmids pBT107, pBT201, and pBT211. Computer programs from Staden [35] were used to analyze the DNA sequence data.

RESULTS AND DISCUSSION

Localization of the tetR Gene

The capacity of tetR$^+$ plasmids to repress β-galactosidase synthesis in tetA-lacZ fusion strains provides a sensitive assay for functional TetR repressor [23]. Previous genetic and biochemical studies localized the tetR structural gene within a "695" bp HincII fragment (actually a 701 bp fragment) and determined the direction of tetR transcription [9-11,14,19,23]. Thus the plasmids pBT402 (Figure 2) and pBT401, which contain the 701 bp HincII fragment inserted into the pACYC177 bla gene in opposite orientations, both repress β-galactosidase synthesis in the tetA-lacZ fusion strain

Table 1. Repression of β-galactosidase synthesis by tetR plasmids in the tetA-lacZ fusion strain B2550(λRStet158-43)

Plasmid	tetR region	β-Galactosidase activity*
pACYC177	—	5,820
pBT1010	Tn10	330
pBT402	HincII 701	50
pBT401	HincII 701	170
pBT4022	HincII 701 Δ(BclI 44)	5,480
pBT4025	HincII 701 Δ(HincII-HinfI 83)	5,350
pBT4028	HincII 701 Δ(HincII-XbaI 25)	5,050

*β-Galactosidase activity is expressed in units as defined by Miller [33].

B2550(λRStet158-43) (Table 1). Since the 701 bp HincII fragment does not include the tetR promoter(s) [9,11,12], expression of tetR in pBT402 must reflect the activity of the bla promoter, whereas expression of tetR in pBT401 must reflect the activity of a pACYC177 promoter that transcribes the bla gene in the opposite (anti-sense) orientation. The DNA sequence across the junction of the bla gene [36] and the 701 bp tetR fragment in pBT402 predicts the synthesis of a truncated Bla protein as opposed to a Bla-TetR fusion protein. As expected, minicells containing pBT402 synthesize at least 10-fold more repressor than do minicells containing pBT401 (Figure 3). Anhydrotetracycline does not induce repressor synthesis in minicells containing either pBT402 or pBT401, as it does in minicells containing pBT1010 (pACYC177::Tn10). In contrast to tetR in pBT402 and pBT401, tetR in pBT1010 is transcribed from the autoregulated tetR promoter-operator region. In keeping with the minicell data for pBT402 and pBT401, we find that a 10- to 20-fold higher concentration of anhydrotetracycline is required to induce half-maximal β-galactosidase synthesis in a tetA-lacZ fusion strain containing pBT402 as opposed to pBT401 (Figure 4).

We constructed deletions at or near the ends of the 701 bp $tetR^+$ HincII fragment in order to examine the extent of the tetR gene within this region (Figure 2). Deletion of either the 25 bp HincII-XbaI fragment (pBT4028) or the 83 bp HincII-HinfI fragment (pBT4025) at the promoter-proximal end of the 701 bp HincII fragment eliminates the synthesis of functional repressor as judged by the inability of these plasmids to repress β-galactosidase synthesis in a tetA-lacZ fusion strain (Table 1).

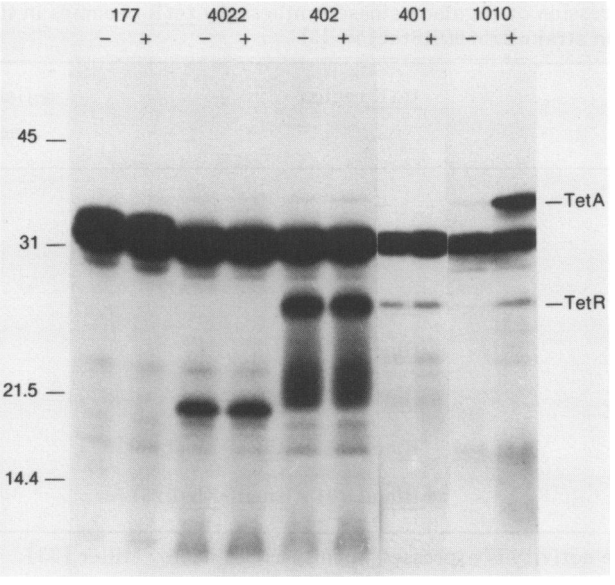


Figure 3. Minicell analysis of the Tn10 TetR repressor. [^{35}S]methionine-labeled proteins synthesized in minicells containing the plasmids pACYC177, pBT4022, pBT402, pBT401, and pBT1010 were electrophoresed in 12.5% polyacrylamide-SDS gels. Minicells were labeled either in the absence (-) or presence (+) of the inducer anhydrotetracycline. The positions of the wild-type TetA resistance protein and TetR repressor are indicated on the right; molecular size standards (in kilodaltons) are indicated on the left.

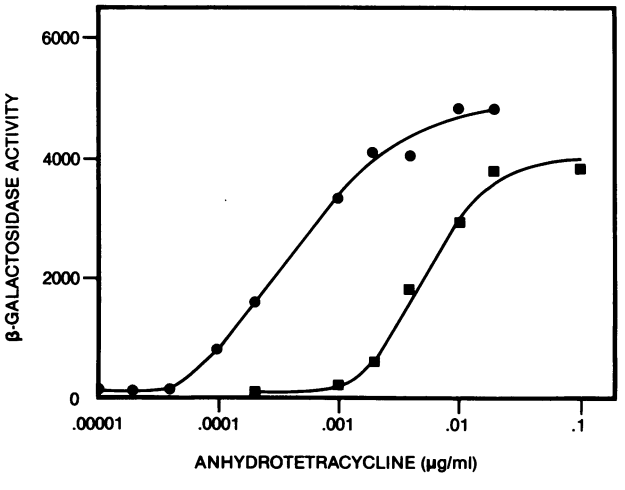


Figure 4. Induction of β-galactosidase synthesis in tetA-lacZ operon fusion strains harboring tetR^{+} plasmids. (●) B2550(λRStet158-43)(pBT401). (■) B2550 (λRStet158-43)(pBT402).

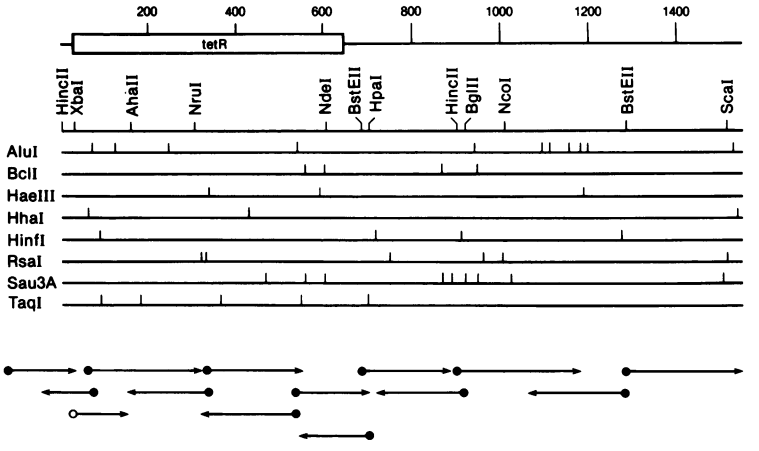


Figure 5. Restriction map and sequencing strategy for Tn10 tetR and its 3' flanking region. (top) Position of tetR with respect to the sequenced region. Distances are given in base pairs from the HincII site on the left. (middle) Partial restriction map, including the restriction sites used to generate DNA fragments for sequencing. (bottom) Sequencing strategy. The closed circles indicate 5' end-labeled DNA and the open circle indicates 3' end-labeled DNA. The arrows indicate the direction of sequencing and the length of the sequence determined.

Deletion of the 44 bp BclI fragment (pBT4022) near the promoter-distal end of the 701 bp HincII fragment similarly eliminates repressor activity, and minicells containing pBT4022 synthesize a truncated repressor protein (Figure 3). Thus sequences necessary for the expression of functional repressor span the 527 bp region between the XbaI and BclI sites within the 701 bp tetR$^+$ HincII fragment.

DNA Sequence of the tetR Gene

The restriction map of tetR and the sequencing strategy are shown in Figure 5. The complete sequence of the 701 bp tetR$^+$ HincII fragment was determined on both strands by the Maxam and Gilbert [34] technique (Figure 6). We have also sequenced several mutant tetR genes by the dideoxy sequencing technique of Sanger [37] and, in the process, confirmed the wild-type tetR sequence (L. Smith and K. Bertrand, in preparation).

There is only one significant open translational reading frame in the direction of tetR transcription; it extends from the ATG codon at bp 22-24 to the TAA stop codon at bp 646-648. There are three potential initiation codons near the 5' end of this open reading frame: the ATG at bp 22-24, a second ATG at bp 25-27, and a GTG at bp 49-51. Of these, the second ATG at bp 25-27 is the most probable initiation codon. We constructed a tetR-lacZ gene fusion by inserting the 451 bp AluI fragment that spans the tetR-tetA promoter-operator region into the lacZ fusion plasmid pMC1403 [38]

Figure 6. The DNA sequence of the Tn10 tetR gene and its 3' flanking region and the predicted amino acid sequence of the TetR repressor. The DNA sequence is numbered from the HincII site (GTCAAC) as in Figure 5. The 5' end of tetR RNA [11,12] is shown by an arrow at nucleotide 7. The Shine-Dalgarno sequence for tetR is underlined.

(L. Smith and K. Bertrand, unpublished observation). The synthesis of a TetR-LacZ fusion protein in strains harboring this plasmid indicates that tetR translation is initiated on the promoter-proximal side of the AluI site at bp 68-71 (corresponding to TetR amino acids 15 and 16 in Figure 6). It is unlikely that the GTG codon at bp 49-51 is the initiation codon, since deletion of the 25 bp HincII-XbaI fragment (bp 4-28) eliminates expression of functional repressor (Table 1). It is also unlikley that the first ATG codon at bp 22-24 is the initiation codon. The tandem ATG codons at bp 22-27 are preceded by a characteristic Shine-Dalgarno sequence (T-AGGA at bp 12-17); however, there are few if any examples of prokaryotic translation initiation signals in which the spacing between the last nucleotide of the Shine-Dalgarno sequence and the initiation codon is less than five nucleotides [39]. We conclude that the coding region for the repressor almost certainly consists of the 207 codons extending from the ATG at bp 25-27 to the TAA at bp 646-648.

The molecular weight of the repressor calculated from the predicted amino acid sequence is 23,328. This value is somewhat lower than our estimate of 25,500 $\pm$ 500 from SDS-polyacrylamide gels such as those shown in Figure 3. Other investigators have estimated the molecular weight of the repressor to be either 23,000 [9,10,19] or 25,000 [20]. The plasmid pBT4022 contains a deletion of the 44 bp BclI fragment in the 3' end of tetR; the tetR DNA sequence predicts that pBT4022 will encode a protein consisting of the NH_2-terminal 179 amino acids of TetR joined to a frame-shifted 7 amino acid COOH-terminal segment. The predicted molecular weight shift of 2,100 (23,300 versus 21,200) is significantly less than the molecular weight shift of 4,500 (25,500 versus 21,000) calculated from gels such as the one shown in Figure 3. Interestingly, the predicted and observed molecular weights of the mutant TetR protein are in good agreement. Perhaps the properties of the COOH-terminus of wild-type TetR account for its disproportionately lower mobility in the gel system we use. In addition, minicells that synthesize high levels of wild-type TetR (pBT402) also appear to accumulate lower molecular weight protein species (20,000-22,000). This latter difference between pBT402- and pBT4022-containing minicells may reflect a unique susceptability of the COOH-terminus of wild-type TetR to proteolytic degradation.

The amino acid composition of the repressor as deduced from the DNA sequence (Table 2) indicates that the repressor contains a slightly higher percent of charged amino acids (29.0%) than the "average" protein (25.1%) [40]. This includes 14.0% acidic residues and 15.0% basic residues. The repressor is somewhat rich in glutamic acid (10.1% versus 6.0% for the "average" protein) and histidine (4.4% versus 2.0% for the "average" protein). There are two tryptophan residues (positions 43 and 75) and six cysteine residues (positions 68, 88, 121, 144, 195, and 203).

Beck et al. [10] reported that certain TetR-LacZ fusion proteins fractionate

Table 2. Predicted amino acid composition of Tn10 TetR and TetA proteins

	TetR	TetA*		TetR	TetA
Ala	16	41	Leu	33	60
Arg	10	10	Lys	12	9
Asn	7	10	Met	3	13
Asp	8	9	Phe	9	28
Cys	6	1	Pro	7	14
Gln	12	14	Ser	11	31
Glu	21	10	Thr	11	25
Gly	13	39	Trp	2	12
His	9	5	Tyr	3	7
Ile	6	37	Val	8	26
				207	401

*The predicted amino acid composition of the TetA protein is based on the sequence reported by Nguyen et al. [16].

with the cytoplasmic membrane of lysed cells. They concluded that the repressor is a membrane-associated protein, and suggested that it may play a role in the mechanism of tetracycline resistance apart from its role in regulating tet transcription. We note that the deduced NH_2-terminal sequence of the repressor does not resemble a secretory signal sequence [41], nor does the repressor appear to contain obvious transmembrane segments as predicted by the hydropathy analysis of Kyte and Doolittle [42] (Figure 7). In fact, the hydrophilic character of the repressor is in marked contrast to the hydrophobic character of the TetA resistance protein [16,18] (Table 2). One possibility is that the repressor is a peripheral membrane protein [43]. Alternatively, the apparent repressor-membrane interaction may be indirect; repressor may, for example, bind non-specifically to plasmid DNA, which is known to

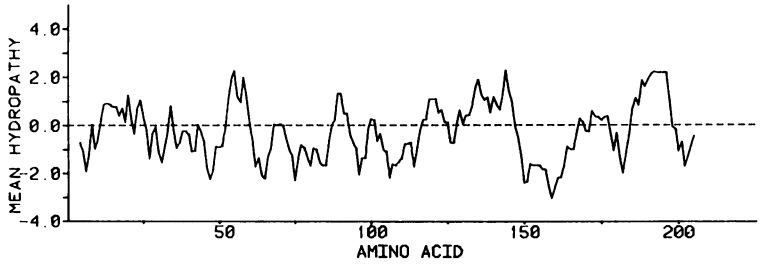


Figure 7. Hydropathy profile of the Tn10 TetR protein. A version of the SOAP program described by Kyte and Doolittle [42] was used to determine the mean hydropathy (hydrophilicity or hydrophobicity) of a moving segment of seven amino acid residues, proceeding from the NH_2-terminus on the left to the COOH-terminus on the right. Relatively hydrophobic regions appear above the midline.

fractionate with the cytoplasmic membrane under some experimental conditions [44].

Waters et al. [45] have determined the DNA sequence of the tetracycline resistance determinant encoded by the plasmid RP1 and the transposon Tn1721. There is 47% amino acid sequence homology between the predicted sequences of the Tn10 and RP1/Tn1721 TetR repressors. This homology will be discussed in more detail in conjunction with the deduced amino acid sequence of the pSC101 TetR repressor (T. Nguyen and K. Bertrand, in preparation).

DNA Sequence of the Region 3' to tetR

Tait and Boyer [46] identified a tetracycline-inducible 14 kilodalton (kDa) protein encoded by the plasmid pSC101, and suggested that this protein reduces the initial ATP-dependent adsorption of tetracycline, thereby contributing to the pSC101 tetracycline resistance phenotype. We recently determined that the structural gene for this 14 kDa protein is immediately 3' to, and in the same orientation as, the pSC101 tetR gene (T. Nguyen and K. Bertrand, in preparation). We were interested in whether the region 3' to the Tn10 tetR gene encodes an analogous protein. The DNA sequence in Figure 6 includes 905 bp immediately 3' to Tn10 tetR. The longest open reading frames in the same orientation as tetR could encode 47 amino acid (bp 966-1106) and 44 amino acid (bp 736-867) proteins. There is, in addition, an open reading frame in the opposite orientation that could encode a 101 amino acid protein (bp 1251-946). However, there is no detectable sequence homology between the predicted amino acid sequences of any of these hypothetical Tn10 proteins and the pSC101 14 kDa protein. We conclude that the region 3' to the Tn10 tetR gene does not encode a protein analogous to the pSC101 14 kDa protein.

Upon inspection, the 905 bp region 3' to tetR does not appear to contain a characteristic rho-independent transcription termination signal [47], as was noted in the DNA sequence immediately 3' to the Tn10 tetA resistance gene [18]. There is, in addition, genetic evidence that a significant level of tetracycline-inducible transcription proceeds across the BglII site 3' to tetR (bp 919-924). When the 2,790 bp Tn10 BglII fragment is inserted, in the appropriate orientation, between the promoter of the Tn5 aminoglycoside phosphotransferase (apt) gene and the apt structural gene, tetR-apt operon fusions are generated, resulting in the tetracycline-inducible expression of kanamycin resistance [48]. Experiments to define the 3' end(s) of tetR RNA are in progress.

Codon Usage and Base Composition

The codon usage in the tetR gene is given in Table 3, where it is compared to the codon usage in the tetA resistance gene. Most E. coli genes show a preference for synonomous codons corresponding to the most abundant tRNA isoacceptors in E. coli [49]. Using the assignments of optimal and non-optimal E. coli codons developed by

Table 3. Codon usage in Tn10 tetR and tetA

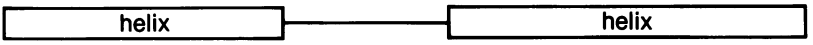
	tetR	tetA*		tetR	tetA		tetR	tetA		tetR	tetA
UUU Phe	7	22	UCU Ser	2	4	UAU Tyr	3	4	UGU Cys	2	1
UUC Phe	2	6	UCC Ser	0	0	UAC Tyr	0	3	UGC Cys	4	0
UUA Leu	18	23	UCA Ser	1	9	UAA	1	0	UGA	0	0
UUG Leu	4	13	UCG Ser	0	7	UAG	0	1	UGG Trp	2	12
CUU Leu	3	11	CCU Pro	4	5	CAU His	6	3	CGU Arg	2	5
CUC Leu	4	2	CCC Pro	0	3	CAC His	3	2	CGC Arg	1	2
CUA Leu	3	4	CCA Pro	2	5	CAA Gln	9	8	CGA Arg	1	1
CUG Leu	1	7	CCG Pro	1	1	CAG Gln	3	6	CGG Arg	2	1
AUU Ile	2	22	ACU Thr	5	4	AAU Asn	5	9	AGU Ser	5	9
AUC Ile	3	10	ACC Thr	1	9	AAC Asn	2	1	AGC Ser	3	2
AUA Ile	1	5	ACA Thr	5	6	AAA Lys	10	6	AGA Arg	2	1
AUG Met	3	13	ACG Thr	0	6	AAG Lys	2	3	AGG Arg	2	0
GUU Val	0	8	GCU Ala	6	14	GAU Asp	7	8	GGU Gly	6	13
GUC Val	2	4	GCC Ala	5	5	GAC Asp	1	1	GGC Gly	1	9
GUA Val	4	6	GCA Ala	5	13	GAA Glu	14	7	GGA Gly	3	8
GUG Val	2	8	GCG Ala	0	9	GAG Glu	7	3	GGG Gly	3	9

*The codon usage in tetA is based on the sequence reported by Nguyen et al. [16].

Ikemura [49], tetR and tetA contain only 45.2% and 51.2% optimal codons, respectively. The pattern of leucine codon usage is particularly striking; CUG, the optimal leucine codon in E. coli, comprises only 3.3% (1/33) and 11.7% (7/60) of the leucine codons in tetR and tetA, respectively. Konigsberg and Godson [50] noted a high frequency of rare codons in several E. coli regulatory genes, and suggested that this may be one of the mechanisms that determine the low level of expression of the corresponding regulatory proteins. We do not believe that the high frequency of rare codons in tetR reflects such a regulatory feature. The pattern of codon usage in tetR is very similar to that in tetA, and in both cases the deviation from optimal E. coli codon usage appears to reflect the low GC content of Tn10. The GC contents of Tn10 tetR and tetA are 40.3% and 43.3%, respectively, as compared to 51% for the total E. coli genome [51]. The GC content of the 905 bp region immediately 3' to tetR is 41.8%. In contrast, the GC contents of the pSC101 tetR and tetA genes are 59.1% and 61.5%, respectively, and the codon usage in these genes is more similar to that of moderately expressed E. coli genes (T. Nguyen and K. Bertrand, in preparation). The significantly different GC contents of the Tn10 and pSC101 tet genes presumably reflect the GC contents of the different bacterial hosts in which these genes evolved.

Sequence Homology Between the TetR Repressor and Other DNA-Binding Proteins

Recent crystallographic studies suggest that the lambda repressor, lambda Cro, and E. coli CAP proteins recognize their DNA target sites in a similar fashion [52-



Gln-Glu-Ser-Val-Ala-Asp-Lys-Met-Gly-Met-Gly-Gln-Ser-Gly-Val-Gly-Ala-Leu-Phe-Asn **λ cI** (33-52)
Gln-Thr-Lys-Thr-Ala-Lys-Asp-Leu-Gly-Val-Tyr-Gln-Ser-Ala-Ile-Asn-Lys-Ala-Ile-His **λ Cro** (16-35)
Arg-Gln-Glu-Ile-Gly-Gln-Ile-Val-Gly-Cys-Ser-Arg-Glu-Thr-Val-Gly-Arg-Ile-Leu-Lys **CAP** (169-188)
Leu-Tyr-Asp-Val-Ala-Glu-Tyr-Ala-Gly-Val-Ser-Tyr-Gln-Thr-Val-Ser-Arg-Val-Val-Asn **LacI** (6-25)
Ile-Lys-Asp-Val-Ala-Arg-Leu-Ala-Gly-Val-Ser-Val-Ala-Thr-Val-Ser-Arg-Val-Ile-Asn **GalR** (4-23)
Gln-Arg-Glu-Leu-Lys-Asn-Glu-Leu-Gly-Ala-Gly-Ile-Ala-Thr-Ile-Thr-Arg-Gly-Ser-Asn **TrpR** (66-85)
Arg-Ala-Glu-Ile-Ala-Gln-Arg-Leu-Gly-Phe-Arg-Ser-Pro-Asn-Ala-Ala-Glu-Glu-His-Leu **LexR** (28-47)
Thr-Arg-Lys-Leu-Ala-Gln-Lys-Leu-Gly-Val-Glu-Gln-Pro-Thr-Leu-Tyr-Trp-His-Val-Lys **TetR** (27-46)

Figure 8. Alignment of the Tn10 TetR repressor and other DNA-binding proteins based on sequence homology and the correspondence of the three-dimensional structures of the lambda repressor (cI), the lambda Cro protein, and the E. coli CAP protein. The helix 2 and helix 3 regions of lambda repressor are indicated. Residues that are identical in Tn10 TetR and in one or more of the other sequences are underlined. The alignments of lambda repressor, Cro, CAP, LacI, GalR, TrpR and LexR are from Pabo and Sauer [58].

54]. The helix 2-turn-helix 3 regions of lambda repressor and Cro and the helix D-turn-helix F region of CAP, the regions which are implicated in DNA binding, have very similar three-dimensional structures. Moreover, amino acid sequence homologies between these regions of lambda repressor, Cro, and CAP and other DNA-binding proteins suggest that many DNA-binding proteins use similar α-helical structures for DNA recognition [55-58].

As shown in Figure 8, there are significant sequence homologies between an NH_2-terminal region of the TetR repressor (positions 27-46) and the regions of lambda repressor, Cro, and CAP that have been implicated in DNA-binding. There are 4-6 identities (20-30% sequence homology) between the TetR sequence and each of the other sequences in Figure 8. Comparison of any of the sequences in Figure 8 to the remaining sequences yields a total of 22-38 identities; there are 33 identities between the TetR sequence and the remaining sequences. The TetR sequence resembles the other sequences in several respects: (i) the presence of highly conserved Ala and Gly residues at positions 31 and 35, respectively; (ii) the presence of hydrophilic residues at positions 27, 28, 29, 32, 33, 37, 38, and 40; and (iii) the presence of hydrophobic residues at positions 30, 34, 36, and 41. Based on these sequence comparisons, it seems very likely that positions 27-46 of the TetR repressor form a helix-turn-helix structure which is involved in binding to tet operator DNA. If this is in fact the case, the TetR repressor resembles the lambda repressor [59] and the LacI repressor [60] in that an NH_2-terminal region of the protein mediates DNA binding. Lastly, the observation that functional TetR repressor is a dimer [26] adds further support to the notion that TetR repressor interacts with tet operator DNA in a manner that is basically similar to the interaction of the lambda repressor, Cro, and CAP proteins

with their DNA target sites.

The DNA sequence of the Tn10 tetR gene and the deduced amino acid sequence of the repressor protein provide a framework for further biochemical and genetic analysis of the structure of the repressor and its interactions with tetracycline and operator DNA.

ACKNOWLEDGEMENTS

We thank Robert Sauer for his critical comments on repressor sequence homologies, M. Kathleen Toth and Stephanie Broderick for their excellent technical assistance, and David Farrell for his contribution to the plasmid constructions. This work was supported by Research Fellowship GM06702 (K.P.), Research Grant AI16735 (K.P.B.), and Research Career Development Award AI00470 (K.P.B.) from the National Institutes of Health.

REFERENCES

1. Foster, T. J., Howe, T. G. B. and Richmond, K. M. V. (1975) J. Bacteriol. 124, 1153-1158.
2. Kleckner, N., Chan, R. K., Tye, B.-K. and Botstein, D. (1975) J. Mol. Biol. 97, 561-575.
3. Mendez, B., Tachibana, C. and Levy, S. B. (1980) Plasmid 3, 99-108.
4. Levy, S. B. (1981) In: New Trends in Antibiotics: Research and Therapy, G. G. Grassi and L. D. Sabath, eds., Elsevier/North-Holland Biomedical Press, Amsterdam, 27-44.
5. McMurry, L., Petrucci Jr., R. E. and Levy, S. B. (1980) Proc. Natl. Acad. Sci. USA 77, 3974-3977
6. Ball, P. R., Shales, S. W. and Chopra, I. (1980) Biochem. Biophys. Res. Commun. 93, 74-81.
7. Izaki, K., Kiuchi, K. and Arima, K. (1966) J. Bacteriol. 91, 628-633.
8. Jorgensen, R. A. and Reznikoff, W. S. (1979) J. Bacteriol. 138, 705-714.
9. Wray Jr., L. V., Jorgensen, R. A. and Reznikoff, W. S. (1981) J. Bacteriol. 147, 297-304.
10. Beck, C. F., Mutzel, R., Barbe, J. and Muller, W. (1982) J. Bacteriol. 150, 633-642.
11. Bertrand, K. P., Postle, K., Wray Jr., L. V. and Reznikoff, W. S. (1983) Gene 23, 149-156.
12. Hillen, W., Schollmeier, K. and Gatz, C. (1984) J. Mol. Biol. 172, 185-201.
13. Levy, S. B. and McMurry, L. (1974) Biochem. Biophys. Res. Commun. 56, 1060-1068.
14. Coleman, D. C. and Foster, T. J. (1981). Mol. Gen. Genet. 182, 171-177.
15. Coleman, D. C., Chopra, I., Shales, S. W., Howe, T. G. B. and Foster, T. J. (1983) J. Bacteriol. 153, 921-929.
16. Nguyen, T. T., Postle, K. and Bertrand, K. P. (1983) Gene 25, 83-92.
17. Curiale, M. S., McMurry, L. M. and Levy, S. B. (1984) J. Bacteriol. 157, 211-217.
18. Hillen, W. and Schollmeier, K. (1983) Nuc. Acids Res. 11, 525-539.
19. Hillen, W., Klock, G., Kaffenberger, I., Wray Jr., L. V. and Reznikoff, W. S. (1982) J. Biol. Chem. 257, 6605-6613.
20. Zupancic, T. J., King, S. R., Pogue-Geile, K. L. and Jaskunas, S. R. (1980) J. Bacteriol. 144, 346-355.
21. Beck, C. F. (1979) Proc. Natl. Acad. Sci. USA 76, 2376-2380.
22. Wray Jr., L. V. and Reznikoff, W. S. (1983). J. Bacteriol. 156, 1188-1191.

23. Bertrand, K. P., Postle, K., Wray Jr., L. V. and Reznikoff, W. S. (1984) J. Bacteriol. 158, in press.
24. Yang, H.-L., Zubay, G. and Levy, S. B. (1976) Proc. Natl. Acad. Sci. USA 73, 1509-1512.
25. Hillen, W. and Unger, B. (1982) Nature 297, 700-702.
26. Hillen, W., Gatz, C., Altschmied, L., Schollmeier, K. and Meier, I. (1983) J. Mol. Biol. 169, 707-721.
27. Jorgensen, R. A., Berg, D. E., Allet, B. and Reznikoff, W. S. (1979) J. Bacteriol. 137, 681-685.
28. Halling, S. M., Simons, R. W., Way, J. C., Walsh, R. B. and Kleckner, N. (1982) Proc. Natl. Acad. Sci. USA 79, 2608-2612.
29. Clarke, L. and Carbon, J. (1978) J. Mol. Biol. 120, 517-532.
30. Marinus, M. G. (1973) Mol. Gen. Genet. 127, 47-55.
31. Moyed, H. S., Nguyen, T. T. and Bertrand, K. P. (1983) J. Bacteriol. 155, 549-556.
32. Chang, A. C. Y. and Cohen, S. N. (1978) J. Bacteriol. 134, 1141-1156.
33. Miller, J. H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
34. Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
35. Staden, R. (1977) Nuc. Acids Res. 4, 4037-4051.
36. Sutcliffe, J. G. (1978) Proc. Natl. Acad. Sci. USA 75, 3737-3741.
37. Sanger, F., Nicklen, S. and Coulsen, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
38. Casadaban, M. J., Chou, J. and Cohen, S. (1980) J. Bacteriol. 143, 971-980.
39. Gold, L., Pribnow, D., Schneider, T., Schinedling, S., Singer, B. S. and Stormo, G. (1981) Ann. Rev. Microbiol. 35, 365-403.
40. Dayhoff, M. O., Hunt, L. T. and Hurst-Calderone, S. (1978) In: Atlas of Protein Sequence and Structure, Vol. 5, Suppl. 3, M. Dayhoff, ed., National Biomedical Research Foundation, Washington, 363-369.
41. Michaelis, S. and Beckwith, J. (1982) Ann. Rev. Microbiol. 36, 435-465.
42. Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105-132.
43. Oliver, D. B. and Beckwith, J. (1982) Cell 30, 311-319.
44. Sparks Jr., R. B. and Helinski, D. R. (1979) Nature 277, 572-575.
45. Waters, S. H., Rogowsky, P., Grinsted, J., Altenbuchner, J. and Schmitt, R. (1983) Nuc. Acids Res. 11, 6089-6105.
46. Tait, R. C. and Boyer, H. W. (1978) Cell 13, 73-81.
47. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
48. Berg, D. E., Egner, C., Hirschel, B. J., Howard, J., Johnsrud, L., Jorgensen, R. A. and Tlsty, T. D. (1981) Cold Spring Harbor Symp. Quant. Biol. 45, 115-123.
49. Ikemura, T. (1981) J. Mol. Biol. 151, 389-409.
50. Konigsberg, W. and Godson, G. N. (1983) Proc. Natl. Acad. Sci. USA 80, 687-691.
51. Normore, W. and Brown, J. R. (1970) In: Handbook of Biochemistry, H. A. Sober, ed., CRC Press, Cleveland, H24-H74.
52. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. and Mathews, B. W. (1981) Nature 290, 754-758.
53. Pabo, C. O. and Lewis, M. (1982) Nature 298, 443-447.
54. McKay, D. B. and Steitz, T. A. (1981) Nature 290, 744-749.
55. Mathews, B. W., Ohlendorf, D. H., Anderson, W. F. and Takeda, Y. (1982) Proc. Natl. Acad. Sci. USA 79, 1428-1432.
56. Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M. and Pabo, C. O. (1982) Nature 298, 447-451.
57. Takeda, Y., Ohlendorf, D. H., Anderson, W. F. and Mathews, B. W. (1983) Science 221, 1020-1026.
58. Pabo, C. O. and Sauer, R. T. (1984) Ann. Rev. Biochem. 53, in press.
59. Sauer, R. T., Pabo, C. O., Meyer, B. J., Ptashne, M. and Backman, K. C. (1979) Nature 279, 396-400.
60. Ogata, R. T. and Gilbert, W. (1979) J. Mol. Biol. 132, 709-728.