Transcription initiation at the *tet* promoter and effect of mutations

Calvin B. Harley^{2*}, Jonathon Lawrie³, Mary Betlach, Roberto Crea^{1,4}, Herbert W. Boyer and Joe Hedgpeth⁵

Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143-0448 and ¹Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080, USA

Received June 4, 1988; Revised and Accepted June 27, 1988

ABSTRACT

We have identified the startpoint for transcription *in vitro* of the tetracycline resistance gene (*tet*) of pBR322 and several deletion and insertion mutations which alter *tet* promoter structure. Tetracycline resistance in host bacteria correlates qualitatively with the efficiency of DNA fragments from these plasmids to promote *tet* transcription *in vitro*. Only in active promoters could we find by computer analysis promoter structures in which the -10 and -35 sequences and the relative spacing of the two regions agree with consensus sequence determinants. These data support the current model of the *E. coli* promoter sequence. Two promoter mutants gave heterogeneous 5' termini with additional A residues not encoded by the DNA sequence.

INTRODUCTION

Promoters are DNA sequences recognized by RNA polymerases which signal the start of RNA synthesis (1). Analysis of these sequences has contributed to our understanding of genetic expression and regulation (2-6). In prokaryotes, the characteristics of a DNA sequence which distinguish it as a promoter have been determined by three approaches. First, mutations that affect transcription have been isolated, their DNA sequences have been determined, and differences in the ways that mutant and wild-type sequences interact with RNA polymerase *in vitro* have been determined (3,7,8). Second, by comparing DNA sequences "upstream" from the startpoints of RNA transcripts, "consensus sequences" have been identified in which bases AT given distances from the startpoint are conserved to a high degree (2,3,9). Finally, points of contact between RNA polymerase and DNA during initiation of transcription have been suggested by identifying sites which are protected from or made more susceptible to chemical or enzymatic attack in the presence of RNA polymerase (3,10). The results of these diverse experiments tend to agree and lead to the following conclusion. Two regions of DNA upstream from the start of the RNA transcript are important in initiation of transcription: (i) a 6-7 basepair (bp) sequence centered about -10 with respect to the transcriptional startpoint (the -10 region or

"Pribnow box"), and (ii) a 6-7 bp sequence centered about -35 (the -35 region). The effect of insertions and deletions between the -35 and -10 regions suggest that the distance between these two regions is also a critical part of promoter structure (11-13).

We have identified by analysis of RNA transcripts *in vitro* the startpoint for transcription of the gene encoding tetracycline resistance, *tet*, of pBR322. Previous references to this startpoint were our unpublished data (6,14,15). We then generated several deletion and insertion mutations that alter the distance between the -35 and -10 regions and studied their effects on the *in vivo* level of *tet* resistance in host cells and the *in vitro* level of transcription from the *tet* promoter. We determined the startpoint for transcription in the active promoters which allowed us to locate putative -35 and -10 sequences and to compare these with sequences predicted by a statistical analysis of known promoters (6). This analysis explains the inactivity of certain mutants according to the promoter sequence model.

MATERIALS AND METHODS

Construction of Mutated Plasmids. Deletion mutations were generated by treating *Cla*I-digested pBR322 derivatives with nuclease S1, followed by DNA polymerase (Klenow fragment) and T4 DNA ligase (16). DNA isolated from ampicillin resistant transformants was sequenced in the region originally containing the *Cla*I site. Some plasmids contained deletions of 1, 3, 4 and 5 bp as well as the 2 bp deletion expected if S1 nuclease removed only the 2 bases of the *Cla*I single-stranded ends (Fig. 5).

One type of insertion mutation was generated by treating plasmid DNA with *Cla*I or *Hind*III, "filling-in" the single-stranded termini with DNA polymerase (Klenow fragment), and rejoining blunt ends with T4 DNA ligase (17). A second type of insertion mutation was made by introducing synthetic DNA fragments into the *Hind*III site of pBR322. Two synthetic DNA fragments (5' HO-AGCTAAATTT [1], and 5' pTATATAAAATTT [2]) with partial complementarity were ligated then labeled with T4 polynucleotide kinase in the presence of γ -[³²P]ATP after heat inactivation of the ligase. Products were subjected to electrophoresis in a polyacrylamide gel and radioactive bands located by autoradiography. Components with the size expected for ligation products 5'[1][2] and 5'[1][2][2] were excised, eluted and ligated with *Hind*III-treated pBR322. Strain RR1 was transformed with the DNA, and transformants selected on LB plates containing ampicillin (50 μ g/ml). Colonies from each transformation were pooled, plasmid DNA was purified and treated successively with *Hind*III and

exonuclease III to eliminate non-mutant plasmids. Strain RR1 was transformed with DNA from each pool. Plasmid DNAs from ampicillin resistant transformants were screened for inserts at the *Hind*III site, and one plasmid from each original pool was subjected to DNA sequence analysis. Fig. 1 shows the sequence of plasmids pTA22 and pTA33. Plasmid pTA33 unexpectedly contained one fewer T:A bp in the series of T:A following the *Hind*III-complementary end.

DNA Sequence Analysis. The DNA sequence of the pertinent region of all mutant plasmids was determined by a modified dideoxynucleoside triphosphate method (18). A synthetic single-stranded oligodeoxynucleotide complementary to positions 37 to 48 of the top strand of *tet* was used to prime plasmid DNA or a *Hae*III fragment extending from positions 4345 to 174 on the pBR322 map (19). For some plasmids the opposite strand of *tet* was sequenced using exonuclease III treatment of *Bam*HI-linearized plasmid DNA to generate a single-stranded template (20) which was primed with a 70 bp *Dde*I-*Eco*RI fragment (positions 4290-4360) from pBR322.

In vitro Transcription and Fingerprinting. Transcription reactions were done in 30 μ l of a mixture containing 20 mM Tris-acetate pH 7.5, 0.1 mM EDTA, 2 mM magnesium acetate, 100 mM KCl, 50-200 μ M of each ribonucleoside triphosphate including the labeled nucleoside triphosphate (150 μ Ci, 2000- 4000 Ci/mmol) and 0.2 mM dithiothreitol. Nucleotide concentration in this range had no effect on the results. About 0.5 pmoles template were added to the mixture and after 2 min at 37°C, 2 pmoles of RNA polymerase were added. Reactions in which magnesium or ribonucleoside triphosphates were used to start the reaction gave identical results to those when RNA polymerase was added first. After 30 min at 37°C, an equal volume of stop mix (90% formamide, 5 μ g tRNA, 0.1% xylene cyanol, 0.1% bromophenol blue and 20 mM Na₂EDTA) was added and the solution boiled for 3 min. Separation of RNA transcripts by acrylamide gel electrophoresis, elution and fingerprint analysis have been described (21). Analysis of nuclease P₁ products (22) and synthesis of γ -[³²P]ATP and γ -[³²P]GTP (23) have been described.

RESULTS

Location of the Startpoint for *tet* mRNA Synthesis in pBR322.

Several facts support the notion that the pBR322 *tet* promoter occurs in the vicinity of the *Hind*III site (Fig. 1): (i) cells carrying pBR322 hybrid plasmids with DNA inserted at the *Hind*III site are more sensitive to tetracycline than cells carrying pBR322, yet significant resistance is sometimes retained, suggesting that the structural gene was not affected (24);

Fig. 1. DNA sequence of the *tet* coding strand of the *Hae*III fragment of pBR322 containing the *tet* promoter and amino-terminal region of *tet*. Spaces following G delineate T1 products which are labeled above the sequence (Table 1 and Figs. 3, 5). Numbers below the sequence represent pBR322 positions (19). Boxes represent -35 and -10 regions of pBR322. The 4, 22 and 33 bp fragments inserted at the *Hind*III site are shown at top. The underlined ATG is the probable start codon for the *tet* protein (39). The circled G indicates the startpoint for *tet* transcription in pBR322.

(ii) the *Hind*III site is protected from restriction if *E. coli* RNA polymerase is allowed to form a stable complex but not initiate transcription (17); (iii) electron microscopic analysis shows that RNA polymerase complexes pBR322 near the *Hind*III site (25); and (iv) the DNA sequence around the *Hind*III site is homologous with known promoter (-35 and -10) sequences (ref. 6; see also Fig. 1). These observations predict that the *tet* mRNA startpoint occurs 6-9 bases downstream of the putative -10, "Pribnow box" region (Fig. 1).

The exact *tet* startpoint was determined by analyzing the 5'-terminus of *tet* transcripts made *in vitro*. A 192 bp *Hae*III fragment containing this region of pBR322 was used as a template for *E. coli* RNA polymerase. The transcripts from separate reactions, each containing a different α -³²P-labeled ribonucleoside triphosphate were separated by acrylamide gel electrophoresis (see Fig. 2), subjected to fingerprinting (Fig. 3), and characterized (ref. 26,27 and Table 1). The transcript indicated by the arrow (Fig. 2, lane A) contained

Fig 2. Polyacrylamide gel electrophoresis of *in vitro* transcripts labeled with α -[32 P]GTP from *Hae*III fragments of (A) pBR322; (B) pBR313H; (C) pTA22; (D) pTA33; (E) pBR322 *Cla*I-5; (F) pTA33 *Cla*I-4; (G) pTA33 *Cla*I-3; (H) pBR322 *Cla*I-2; (I) pTA22 *Cla*I-1; (J) pTA22 *Cla*I+2; (K) pTA33 *Cla*I+2; (L) no DNA. Dots (XC) correspond to position of the xylene cyanol dye marker. Arrows denote major transcripts shown by fingerprint analysis to come from the *tet* promoter region extending into *tet*. Bands immediately below arrows were not always present but gave fingerprints similar to those from indicated bands minus several nucleotides corresponding to 3' termini (data not shown).

the T1 oligonucleotide expected for the region downstream from the *Hind*III site on the coding strand for *tet* (Table 1). It contained T1 products UUUAUCACAG and pppGp, but not ---AAUG (Fig. 3A, Table 1). Hence, the transcript initiated at one of four Gs between positions 39 and 45 (see Fig. 1). Position 45 was identified as the startpoint because the transcript contained pancreatic ribonuclease product pppGUp which was labeled only by α -[32 P]GTP or

Fig. 3. Fingerprint analyses of *tet* transcripts from pBR322 (A,B,C), pBR313H (D,E), pTA22 (F) and pTA33 (G,I,J). RNA isolated as in Fig. 2A was digested with T1 (A,D,E,F,G) or pancreatic RNase (B,I,J) and subjected to electrophoresis on cellulose acetate at pH 3.5 (first dimension) then homochromatography (second dimension). Fragments were identified by secondary analysis and comparison with the DNA sequence (see also Table 1). The labeled precursor was [32 P]GTP (A,B,C,D,F,G,I) or [32 P]UTP (E,I,J). The spot labeled pppGU (B) was digested with nuclease P_1 and chromatographed on PEI-cellulose in 0.75 M potassium phosphate buffer (pH 3.5) with ribonucleoside triphosphate markers (C). The spot labeled AGC (D) was a contaminant which occurred only in this [32 P]GTP fingerprint. The composite drawing (H) shows T1 oligonucleotides which are common (solid circles) or not common (dashed circles) among *tet* transcripts from the four plasmids (pBR322, pBR313H, pTA22 and pTA33) (see also Table 1). RNA used in fingerprint (J) was treated with alkaline phosphatase prior to nuclease digestion.

α -[32 P]UTP, with α -[32 P]GTP incorporated as the 5' triphosphate (Fig. 3B,C). No triphosphate-containing spot appeared when α -[32 P]ATP was used, and the transcript could be labeled by γ -[32 P]GTP but not γ -[32 P]ATP (data not shown). This startpoint agrees with that predicted by putative -35 and -10 sequences at pBR322 positions 10 to 15 and 33 to 38, respectively (Fig. 1; also ref. 6,15).

Effect of Insertions and Deletions on Expression of *tet*.

Two unique restriction sites within or near the -10 region of the *tet* promoter allowed us to make deletion and insertion mutations in this area

("Materials and Methods" and Fig. 1, 5). We investigated the effects of these mutations on expression of *tet* *in vivo* and transcription *in vitro*. Mutants generated by insertions at the *Hind*III sites are pBR313H (17), pTA22 and pTA33. Those generated by "filling-in" the *Cla*I site are indicated by *Cla*I+2 while those deleting one or more bp at the *Cla*I site are indicated by *Cla*I-1, etc.

Tetracycline Resistance. Tetracycline resistance as measured by colony formation on tetracycline-containing medium has been used as an indication of the transcription efficiency of the *tet* promoter (see ref. 17 and Table 2). On the basis of tetracycline resistance, mutant plasmids can be placed in three

TABLE 1
Pancreatic Ribonuclease Products of T₁ Oligonucleotides

Oligo	Label α -[³² P]-		Sequence ³
	GTP	UTP	
T1	Gp	Gp	Gp
T2	Cp	--	CGp
T3	Up	Gp	UGp
T4	Gp,Cp	--	CCGp(G)
T5	AGp	AGp	CAGp(U)
T6	Up	Gp,Cp	CUGp(U)
T7	AGp,AUp	AUp	AUGp
T8	AGp,AUp	AUp	UAGp
T9	AGp	--	UCAGp
T10	AUp,Gp,Up	AUp,Cp,Up	CUUGp(G)
T11	AUp,Gp,Up	AUp,Cp,Up	UAUGp
T12	Cp	Cp	CACCGp(U)
T13	AGp	AUp	CAUAGp
T14	Up	ACp	UACUGp
T15	AUp	AUp,Up	UUAUGp
T16	AAC	Cp	CUAACGp
T17	Cp	AUp,Gp,Cp	CUCAUCGp(U)
T18	Up,Gp	Cp	UCACCCUGp(G)
T19	Up	AAAAUp,Up	UUAAAAUUGp
T20	Gp,Cp	AUp,Cp	UCAUCCUCGp(G)
T21	AGp	AGp,AUp,Up	UUUAUCACAGp(U)
T22	AAUp	AAAUp,AAUp,Cp	AAAUUAACAAUGp
T23	AAUp	AAUp,Up,Cp	CUUUAAUG
T24	pppGp	pppGp	pppGp
T25	AGp	X,Up	ppp(Ap)UUUAG
T26	AGp	X,AAAAUp, AUp,Up	ppp(Ap) _n UUUUUAUAAAAAUUUAG

Assignment of sequence was based on characteristic pancreatic oligonucleotides, nearest neighbor labeling, size, composition, and comparison to expected T₁ products from DNA sequence (Fig. 1). In product T8, AGp was labeled by α -[³²P]UTP only in transcripts from pBR313H, pTA22 and pTA33. T23 was found only in transcripts from pTA22 and pTA33. T25 and T26 were found in transcripts from pTA22 and pTA33, respectively.

groups. The group least affected contains plasmids having insertions only at the *Hind*III site. Resistance conferred by these plasmids is reduced by 40% or less relative to pBR322. A second group is comprised of mutants containing a 2 bp insertion at the *Cla*I site of pTA22 or pTA33 or a single bp deletion at the *Cla*I site of pTA22. These mutations reduced tetracycline resistance more severely (about 70% or greater relative to pBR322 or about 50% relative to the parent plasmid). The group most severely affected consists of plasmids having 2 and 5 bp deletions or a 2 bp insertion at the *Cla*I site of pBR322; and those having a 3 or 4 bp deletion at the *Cla*I site of pTA33. These exhibit LC50s of less than 5 μ g tetracycline per ml.

Effect of Mutations on *tet* Promoter transcription *in vitro*. Relative tetracycline resistance was correlated with relative transcriptional efficiency by analyzing *in vitro* transcription from *Hae*III fragments containing the *tet* promoter region from each mutant. The startpoint of *tet* mRNA was determined by

TABLE 2
Tetracycline Resistance Conferred on Strain MM294 by Various Plasmids
and *in vitro* and Computer-Predicted Promoter Strengths

Plasmid	(a) <i>tet</i> LC50 ¹ (μ g/ml)	(b) Relative Promoter ₂ Strength <i>in vitro</i>	(c) Relative Promoter Homology Index ³
pBR322	70	++	100
pBR322 <i>Cla</i> I+2	2	-	6
pBR322 <i>Cla</i> I-2	4	-	6
pBR322 <i>Cla</i> I-5	5	-	6
pBR313H	50	+	30
pTA22	45	++++	200
pTA33	45	++++	200
pTA22 <i>Cla</i> I+2	12	-	10
pTA22 <i>Cla</i> I-1	12	-	100
pTA33 <i>Cla</i> I+2	24	+/-	10
pTA33 <i>Cla</i> I-3	2	-	5
pTA33 <i>Cla</i> I-4	2	-	4

¹Concentration of tetracycline which reduces plating efficiency by 50%.

²Qualitative measure of promoter strength from analysis of *in vitro* transcription products as described in Materials and Methods.

³Log promoter homology indices were calculated as described previously (6), converted to a linear scale, and normalized to the pBR322 *tet* promoter.

fingerprint analysis to allow identification of likely -10 and -35 promoter sequences.

Transcripts made from the *Hae*III *tet* promoter fragment in the presence of α -[³²P]GTP were separated by polyacrylamide gel electrophoresis, identified by autoradiography and relative intensities estimated by densitometry (Fig. 2, Table 2). RNA species indicated by arrows were identified as *tet* transcripts by sequence analysis (see below). The minor RNA species in lanes A through D varied in relative amounts in different experiments and consist of three types: (i) *tet* transcripts having the same 5'-termini as those indicated by arrows, but which are shorter than the major transcript (these include the bands immediately below the major species); (ii) minor *tet* transcripts starting at other positions; and (iii) transcripts coming from the strand opposite the *tet* coding strand. Characterization of the minor transcripts will be presented elsewhere (CBH, JL, HWB and JH). Major *tet* transcripts occurred only with template DNA from plasmids conferring significant tetracycline resistance (compare lanes A-D with lanes E-K). Specific *tet* transcripts from fragments containing alterations at the *Cla*I site are reduced at least ten-fold in comparison to their parent sequences. The major *tet*-specific transcript pTA33 *Cla*I+2 was reduced slightly less than other *Cla*I mutants (Fig. 2, lane K) and tetracycline resistance was reduced by only 50% compared to pTA33.

Transcripts from pTA22 and pTA33 consistently contained 4- to 5-fold more radioactivity than transcripts from pBR322 (Fig. 2, lanes A, C, D) regardless of the labeled ribonucleoside triphosphate used. Since pTA22 and pTA33 conferred lower tetracycline resistance than pBR322, the relative level of transcription seen *in vitro* does not always correlate with the relative tetracycline resistance. Although *tet* transcripts were somewhat heterogeneous in length, transcripts from *Hae*III fragments containing 4, 22 and 33 bp insertions were, respectively, about 4, 20 and 30 nucleotides longer than the pBR322 transcript, indicating that the transcripts initiated upstream from the pBR322 startpoint. This was verified by fingerprint analysis.

Startpoint of *Hind*III+4 (pBR313H) *tet* transcript. T1 fingerprint patterns of transcripts from pBR313H differed from the corresponding pBR322 fingerprint in containing the spot UAGp, where Gp was labeled by α -[³²P]UTP. Also, α -[³²P]UTP did not label the 5' triphosphate-containing T1 nucleotide of pBR313H (Table 1). The pancreatic RNase fingerprint of the transcript labeled with α -[³²P]GTP contained a single triphosphate-containing oligonucleotide

Fig. 4. Gel electrophoresis (20% acrylamide, 7M urea) of γ -[32 P]ATP-labeled pTA33 (A,B) and pTA22 (C,D) transcripts digested with pancreatic (A,C) or T1 (B,D) RNase. Labeled transcripts were generated and isolated as described in Materials and Methods. Marker oligonucleotides (homopolymers of ppp(A)_n-OH were isolated from abortive initiation products generated by transcribing pTA33 fragment with ATP as the only nucleoside triphosphate present. Pancreatic and T1 RNase digestion products of size N nucleotides have 3'-phosphates (20) and hence migrate with 3'-hydroxyl terminated oligomers of size N-1 nucleotides. Digestion products migrating with pppAAA-OH, for example, contained the sequence pppAAAU_p. Numbers next to the ladder represent the number of nucleotides in the nuclease digestion products. Bands at position 2 and 3 are distorted in migration by salt.

which in turn yielded only pppGp as a labeled T1 product (Fig. 3D). This pancreatic RNase oligonucleotide was labeled by α -[32 P]UTP, but then yielded only Gp as a labeled T1 product (Fig. 3E). These data indicate that the 5'-oligonucleotide is pppGGUp which is compatible with its position in the pancreatic RNase fingerprint (Fig. 3D,E). Hence, the startpoint for this transcript is position 41, 4 bp upstream from the startpoint in pBR322 (Figs. 1,5). This accounts for the T1 spot, UAGp, labeled by α -[32 P]UTP and is consistent with the slightly slower migration of this transcript in acrylamide gel electrophoresis (Fig. 2, lanes A and B).

Startpoints and heterogeneity of tet transcripts from pTA22 and pTA33.

Transcripts for pTA22 and pTA33 fragments (22 transcript and 33 transcript, respectively) contained all T1 oligonucleotides found in the pBR322 transcript except T24 (Fig. 3F,G, Table 1). In addition, both the 22 and 33 transcripts contained CUUAAUGp (T23) from position 32-39 on the pBR322 sequence, upstream from the pBR322 startpoint (Fig. 1). T1 oligonucleotides T25 and T26 containing the 5'-triphosphates were actually a series of products differing at the 5' termini as shown by the following. Both T25 and T26 yielded only AGp as a pancreatic RNase product when labeled with α -[32 P]GTP. When labeled with α -[32 P]UTP, T25 yielded the pancreatic RNase product Up and a product that remained at the origin (Table 1). Treatment of this second product with alkali yielded labeled Ap and Up. With T25 labeled with α -[32 P]ATP, this second pancreatic RNase product yielded pppAp and Ap when treated with alkali, or ATP and pA when treated with nuclease P1 (not shown). These data indicate that T25 contained the 5' terminus of the transcripts (pppA--) followed by one or more A's, then U. Since pancreatic RNase did not release AAAUp from T25, the startpoint for the 22 transcript lies within the A₄ series between positions 45' and 48' (Fig. 1).

Oligonucleotide T26 from α -[32 P]UTP-labeled 33 transcript yielded pancreatic RNase products Up, AUp, and A₄Up plus material at the origin (Table 1). This material contained ATP and gave the same analysis as that of T25, suggesting that this transcript also started within the A₄ series between 44' and 47' of pTA33 (Fig. 1). However, several experiments indicated that the 5'-oligonucleotides from the 22 and 33 transcripts were heterogeneous and contained more than the four A's predicted by the DNA sequence. First, α -[32 P]UTP-labeled 22 and 33 transcripts were subjected to pancreatic RNase fingerprinting with or without prior treatment with alkaline phosphatase. Since only terminal phosphates were removed, 5'-termini migrated more slowly in the first dimension (27). In the 33 transcript the 5' oligonucleotides aligned

Fig. 5. Structure of *tet* promoters. DNA sequences are aligned according to the *Hind*III site of pBR322. Boldface type represents DNA inserted at the *Hind*III site. CG represents the 2 bp insertions generated by "filling in" the *Cla*I site. Bars represent deletions generated by S1 nuclease. Shaded regions represent -35 and -10 regions. Brackets show sequences in the mutated plasmids which would have to function as -10 sequences if the -10 region retained a 17 bp spacing from the wild type -35 sequence. Initiation points for the four active promoters are indicated above the DNA sequence.

in a series expected for $(Ap)_nUp$; where $n > 4$ (Fig. 3I,J). The 22 transcript gave identical results (not shown), except that A_4U and AU were less heavily labeled. This is consistent with startpoints in both pTA22 and pTA33 occurring between 44' and 48' (see Fig. 1), where A_4U and two AUs occur in oligonucleotide T26 but not in T25.

In a second series of experiments, the 22 and 33 transcripts were labeled with γ - ^{32}P ATP, isolated, digested with T1 or pancreatic RNase, and subjected to electrophoresis (28). Pancreatic RNase products of the two transcripts gave identical patterns (Fig. 4, lanes A and C) consistent with $ppp(Ap)_nUp$ where n ranges from 4 to 10 or more with a mode of 5 or 6. T1 products were also heterogeneous with a model size of about 10 or 11 nucleotides (22 transcript) or 22 or 23 nucleotides (33 transcript). Hence transcripts from both insertions appear to "start" within a series of A's beginning around A45' (Fig. 1) and incorporate a variable number of A's generating heterogeneous 5' termini (Fig. 5).

Promoter Homology Index of Wild Type and Mutant *tet* Promoters

Both strands of the DNA sequence from about 1 to 50 (-45 and +5 with respect to *tet* transcription) of pBR322 and the mutants were analyzed for promoter signals as described previously (6). This analysis yields subsequences which are similar to known promoters on strictly statistical grounds. The major promoter identified by computer for all sequences which were active in *in vitro* assays was that predicted by the experimental data. Table 2c gives the relative weight of the homology index for these promoters as well as that for the best possible promoters at corresponding locations in the non-active sequences. The relative strength of the promoters determined in the *in vitro* and *in vivo* assays was in general agreement with the relative promoter homology indices for these sequences (Table 2a-c and Discussion).

DISCUSSION

The transcripts we have analyzed appear to be those of the tetracycline promoter because (i) they originate in the proper region of the DNA and extend into *tet*, (ii) they are the major *tet* coding-strand transcripts made from DNA of plasmids conferring tetracycline resistance, and (iii) mutations that reduced tetracycline resistance concomitantly reduced or abolished the transcripts. However, the level of transcription *in vitro* did not necessarily correlate with the level of tetracycline resistance *in vivo* since pTA22 and pTA33 plasmids gave about 60% of wild-type tetracycline resistance while increasing *in vitro* transcription about 4-5 fold compared to pBR322. Relative *tet* expression *in vivo* and *in vitro* could differ for a number of reasons. First, there may be differential effects of DNA supercoiling on transcription as seen in both *E. coli* (30,31) and eukaryotes (32). Second, 5'-oligo (A) or other differences in the mutant mRNAs could make them less stable or less efficiently translated. Third, if the *tet* transcript is induced by tetracycline *in vivo*, the mutant promoters may be induced less than the wild-type promoter. Finally, there may be greater interference of expression from competing promoters or anti-sense RNA (33,34) *in vivo* than *in vitro*, where polymerase initiation is expected to be more frequent. Our computer analysis predicts numerous weak promoters on both DNA strands near the *tet* promoter in pTA22 and pTA33, but not in pBR322 (not shown). In addition, the transcript of one of these promoters, Pamp1 (15), runs antiparallel to *tet* and overlaps it in pTA22 and pTA33, but not in pBR322.

The pTA22 and pTA33 *tet* promoters are more efficient *in vitro* than the pBR322 promoter. In other promoters, the region -10 to +8 is "melted" in

forming the "closed complex" (3,35) and the AT-rich insertions in pTA22 and pTA33 might facilitate this process. Moreover, the -10 sequence of pTA22 and pTA33 (TAAATT) is statistically a factor of 2-3 more like the consensus than the pBR322 -10 sequence (TTTAAT), while maintaining a 17 bp spacing from the same -35 sequence (TTGACA). Mulligan et al. (4) have shown that similarity to the consensus promoter correlates well with promoter strength (selectivity) *in vitro*. Analysis of most of the other mutations we generated show a similar correlation between computer-predicted promoter homology index and transcriptional strength *in vitro*. For example, in pBR322 *Cla*I+2, -2 and -5, the homology index is decreased at least 16-fold compared to wild-type, while in pBR313H, the index is decreased by only a factor of 3-4. These figures predict the relative level of major *tet* transcripts seen *in vitro* and *tet* resistance *in vivo* (Table 2). However, it is not clear why some mutants made by altering the *Cla*I site of pTA22 or pTA33 conferred measurable levels of *tet* resistance even though the major *in vitro* *tet* transcript was severely reduced. There is also weak correlation to some computer predictions: mutant pTA22 *Cla*I gave lower *tet* resistance *in vivo* than is predicted by the statistical analysis (Table 2a,c).

Finding heterogeneous, non-cognate termini for pTA22 and pTA33 transcripts was unexpected, although similar results have been reported with other promoters (36,37). A more detailed analysis of this phenomenon will be presented elsewhere (CBH, JL, HWB and JH, in preparation). We will assume for this discussion that the startpoint is position 45' in pTA22 and 44' in pTA33 (Fig. 1) and that the heterogeneity is caused by events subsequent to initiation.

Using experimentally determined startpoints as guides to locate the relevant promoter structures, the behaviour of mutant promoters can be explained as follows. Each insertion at the *Hind*III site, although lengthening the distances between original -10 and -35 regions, provides a new -10 the proper distance from the original -35 sequence (Fig. 5). This combination of -35 and -10 is responsible for the major *tet* transcript in these plasmids. Even though other -10 sequences are present, including the wild-type -10, computer searches showed that they scored poorly with -35 sequences.

The *Cla*I site occurs in the spacer region between the -35 and -10 regions. Deletions and insertions at this site severely reduced the major *tet* transcript. Similar data in other promoter systems have also shown that changes of one or two bp from the consensus 17 bp spacing between the -35 and -10 regions dramatically alter promoter strength (11,13). With exception of

pTA22 *Cla*I, our data on these spacer mutants are qualitatively predicted by computer analysis: no strong -35 and -10 regions were found with a spacing of 17 bp, and the gap penalty (6) for the new spacing of the original sequences severely reduced the homology index. Thus, although we cannot rule out some sequence dependent changes in polymerase interaction within the spacer region (38), it is likely that improper distance between -35 and -10 sequences accounts for most of the mutational effects at the *Cla*I site.

In conclusion, we have identified the startpoint and relative efficiency of transcription *in vitro* from wild-type and mutant *tet* promoters and have determined *tet* resistance *in vivo* of cells carrying the plasmids. Although exceptions remain, the general agreement between results presented here and statistical analysis of these and other promoter sequences provides further support for the current model of the *E. coli* promoter sequence.

ACKNOWLEDGMENTS

We thank P. Greene for helpful discussion. Supported by the National Institutes of Health grants GM28749 (HWB) and GM27472 (JH) and the Medical Research Council and Natural Sciences and Engineering Research Council of Canada (CBH).

Present addresses: ²Department of Biochemistry, McMaster University, Hamilton, Ontario, L8N 3Z5, Canada, ³Amoco, Five Oak Terrace, Milford, MA 01757, ⁴Creative Biomolecules, Inc., 35 South Street, Hopkinton, MA 01748 and ⁵CODON, 213 East Grand Avenue, South San Francisco, CA 94080, USA

*To whom correspondence should be addressed

REFERENCES

1. Jacob, F. and Monod, J. (1961) Cold Spring Harbor Symp. Quant. Biol. 25, 193-198.
2. Rosenberg, M. and Court, D. (1979) Ann. Rev. Gene. 13, 319-353.
3. Siebenlist, U., Simpson, R.B. and Gilbert, W. (1980) Cell 20, 269-281.
4. Mulligan, M., Hawley, D., Entriken, R. and McClure, W. (1984) Nucl. Acids Res. 12, 789-200.
5. McClure, W.R. (1985) Ann. Rev. Biochem. 54, 171-204.
6. Harley, C.B. and Reynolds, R.P. (1987) Nucl. Acids Res. 15, 2343-2359.
7. Miller, J., Ippen, K., Scaife, R. and Beckwith, J. (1969) J. Mol. Biol. 38, 413-423.
8. Stefano, J.E. and Gralla, J.D. (1980) J. Biol. Chem. 255, 10423-10430.
9. Probnow, D. (1975) Proc. Natl. Acad. Sci. USA 72, 784-789.
10. Johnsrud, L. (1979) Proc. Natl. Acad. Sci. USA 75, 5314-5318.
11. Stefano, J.E. and Gralla, J.D. (1982) Proc. Natl. Acad. Sci. USA 79, 1069-1072.
12. Mandecki, W. and Reznikoff, W. (1982) Nucl. Acids Res. 19, 903-912.
13. Aoyama, T., Takanami, M., Ohtsuka, E., Yanayama, Y., Marumoto, R., Sato, H. and Ikehara, M. (1983) Nucl. Acids Res. 11, 2943-2957.

14. Brosius, J., Cate, R.L. and Perlmutter, A.P. (1982) *J. Biol. Chem.* 257, 9205-9210.
15. Hawley, D.K. and McClure, W.R. (1983) *Nucl. Acids Res.* 11, 2237-2255.
16. Hedgpeth, J., Clement, J.M., Marchal, C., Perrin, D. and Hofnung, M. (1980) *Proc. Natl. Acad. Sci. USA* 77, 2621-2625.
17. Rodriguez, R.L., West, R.W., Heyneker, H.L., Bolivar, F. and Boyer, H.W. (1979) *Nucl. Acids Res.* 6, 3267-3287.
18. Sanger, F. and Coulson, R. (1978) *FEBS Lett.* 87, 107-110.
19. Sutcliffe, G. (1978) *Cold Spring Harbor Symp. Quant. Biol.* 43, 77-90.
20. Smith, A.J.H. (1980) *Methods Enzymol.* 65, 560-580.
21. Belin, D., Hedgpeth, J., Selzer, G. and Epstein, R. (1979) *Proc. Natl. Acad. Sci. USA* 76, 700-704.
22. Johnson, L.D. and Iazzarini, R.A. (1979) *Virology* 77, 863-866.
23. Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74, 560-564.
24. Tait, R.C. and Boyer, H.W. (1978) *Cell* 13, 78-81.
25. Stuber, D. and Bujard, H. (1981) *Proc. Natl. Acad. Sci. USA* 78, 167-171.
26. Barrell, B.G. (1971) In Cantoni, G. and Davies, D. (eds.) *Procedures in Nucleic Acids Research*, Harper and Row, New York, Vol. 2, pp. 751-779.
27. Brownlee, G.G. (1972) In *Determination of Sequences in RNA*, North Holland-Amsterdam.
28. Simoncsits, A., Brownlee, G.G., Brown, R.S., Rubin, J.R. and Bulley, H. (1977) *Nature* 269, 833-836.
29. Staden, R. (1984) *Nucl. Acids Res.* 12, 505-519.
30. Chan, P.T., Lebowitz, J. and Bastia, D. (1979) *Nucl. Acids Res.* 7, 1247-1261.
31. Bertrand-Burggraf, E., Schnarr, M., Lefevre, J.F. and Duane, M. (1984) *Nucl. Acids Res.* 12, 7741-7752.
32. Hirose, S. and Suzuki, Y. (1988) *Proc. Natl. Acad. Sci. USA* 85, 718-722.
33. Coleman, J., Green, P. and Inouye, M. (1984) *Cell* 37, 439-436.
34. Blake, K., Murakami, A. and Miller, P.S. (1985) *Biochem.* 24, 6139-6145.
35. Gamper, H.B. and Hearst, J.E. (1982) *Cell* 29, 81-90.
36. Heyden, B., Nusslein, C. and Schaller, H. (1975) *Eur. J. Biochem.* 55, 147-155.
37. Machida, C., Machida, Y. and Ohtsubo, E. (1984) *J. Mol. Biol.* 177, 247-267.
38. Youdarian, P., Bouvier, S. and Susskind, M.M. (1982) *Cell* 30, 843-853.
39. Peden, K.W.C. (1983) *Gene* 22, 277-280.