JOURNAL OF BACTERIOLOGY, Feb. 1980, p. 973-976
0021-9193/80/02-0973/04$02.00/0

Vol. 141, No. 2

Promoter Sequence for Stringent Control of Bacterial Ribonucleic Acid Synthesis

ANDREW A. TRAVERS

MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, England

In bacterial promoters subject to stringent control the heptanucleotide sequence spanning nucleotide positions −5 to +2 (defining the −10 region [D. Pribnow, J. Mol. Biol. **99**:419-443, 1975] as positions −12 to −6) is highly conserved. This conserved region, for which the best consensus sequence is

```
−5      −1 +1
  c c
C g g C −  C C,
```

is absent from promoters which are not under stringent control.

When *rel*$^+$ strains of *Escherichia coli* are starved for an amino acid,the rate of both rRNA and tRNA synthesis is reduced by 10- to 20-fold (6, 19). Similarly, the synthesis of r-protein mRNA is strongly inhibited (4), whereas the synthesis of other mRNA species (13), for example *lac* (11) and *trp* (5, 7), continues at relatively high levels. This stringent response is accompanied by the rapid accumulation of the nucleotide guanosine 3′-diphosphate 5′-diphosphate (ppGpp) (1). In vitro ppGpp interacts directly with RNA polymerase (2, 22), altering the pattern of transcription initiation in a manner qualitatively similar to that observed during the accumulation of the nucleotide in vivo (2, 3, 23, 25). This alteration of transcriptional selectivity affects both the formation and the stability of the open polymerase-promoter complex in a promoter-specific manner (2, 14; J. Hamming, M. Gruber, and G. Ab, Nucleic Acids Res., in press).

Such a discriminatory process implies that a DNA sequence at or near a promoter site must encode the information specifying whether or not a particular promoter is under stringent control. Two regions within the RNA polymerase binding site are known to be important in promoter function and show strong sequence conservation. One, the −10 region, is a heptanucleotide sequence spanning positions −6 to −12 before the average start point for transcription (16). In both the *lac* UV5 and the $rrnE_1$ promoters, this region has the identical sequence, TATAATG. However, in vitro ppGpp stimulates the activity of the *lac* UV5 promoter and strongly inhibits that of $rrnE_1$ (Hamming et al., in press; P. G. Debenham, Ph.D. thesis, University of Cambridge, Cambridge, England, 1978). In general, the −10 regions of promoters under stringent control show a close fit to the "ideal" sequence for all promoters. In the second region conserved for all promoters, that at about 35 base pairs preceding the start point (17, 21), the sequence CTTTACA occurring in the region of greatest conservation is common to both the *lac* and tRNA^{Tyr} promoters, promoters which have opposite responses to ppGpp in vitro (3, 25). These arguments suggest that neither the −10 region nor the −35 region is a major signal for stringent control. Analysis of the sequences between the −10 region and the −35 region by sequence alignment with either conserved region in a fixed position also does not show any oligonucleotide common to promoters under stringent control. However, analysis of sequences spanning positions −5 to +8, defining the −10 region as positions −12 to −6 (20), reveals that there exists within this a region a heptanucleotide sequence which is conserved in six positions in 16 promoters known or believed to be under stringent control. This sequence, which in the antisense strand is

```
                −5      −1 +1     +4
−10 region  C G C C  −  C C −  −
```

is most highy conserved in positions +1 and +2 with C present in 15 of 16 examples in each case (Table 1). In positions −2 and −5, conservation is less pronounced, C being present in 13 of 16 examples. Considering all stringent promoters, C is favored for position −3 and G for −4.

This consensus sequence is derived from sequences which include five examples each of the proximal and distal promoters of the rRNA cistrons. To exclude bias due to this preponderance of rRNA promoters a second frequency distribution can be calculated from a set of stringently controlled promoters which includes one exam-

TABLE 1. *Sequences of E. coli promoters*[a]

Promoter	Position												
	−5	−4	−3	−2	−1	+1	+2	+3	+4	+5	+6	+7	+8
rrnA$_1$, *B*$_1$; −10 region	C	G	C	C	A	C	C	A	C	T	G	A	C
rrnD$_1$	C	G	C	C	T	C	C	G	T	T	G	A	G
rrnE$_1$, *X*$_1$	C	G	C	C	T	C	C	A	T	C	G	A	C
rrnA$_2$, *B*$_2$	C	A	C	A	C	C	C	C	G	C	G	C	C
rrnD$_2$, *E*$_2$, *X*$_2$	C	G	C	C	A	C	C	T	C	G	C	G	A
tRNA^{Tyr}	C	G	C	C	C	C	G	C	T	T	C	C	C
rpsL (Str)	C	G	G	C	G	T	C	C	T	C	A	T	A
rpsE (Spc)	C	C	G	C	G	C	C	C	T	C	G	A	T
rplKA	T	C	G	C	G	C	C	T	T	T	T	G	T
rpoA	G	C	C	A	G	C	C	A	A	T			
rpoB	T	T	A	C	C	C	C	C	A	C	G	T	A
Frequencies													
(i) A	–	2	1	3	5	–	–	5	2	–	1	6	5
C	13	3	12	13	4	15	15	6	5	7	4	3	7
G	1	10	3	–	4	–	1	1	2	3	9	4	1
T	2	1	–	–	3	1	–	4	7	6	1	2	2
(ii) A	–	–	1	1	1	–	–	2	2	–	1	2	3
C	5	3	4	7	2	7	7	4	1	4	1	1	2
G	1	4	3	–	4	–	1	–	–	1	4	2	–
T	2	1	–	–	1	1	–	2	5	3	1	2	2
(iii) A	10	12	11	8	7	20	8	13	12	12	10	8	14
C	4	11	10	9	13	9	5	5	4	9	6	7	6
G	14	8	5	12	8	8	10	10	7	7	4	9	8
T	11	8	13	10	11	2	16	11	15	10	14	14	10
Consensus sequence for stringency		c	c										
	C	g	g	C	–	C	C	–	–	–			

[a] Sequences of *E. coli* promoters known or believed to be under stringent control aligned with the −10 region in a fixed position. Known start points are underlined; for *rrnA*$_2$ the precise start point is uncertain, and for *rpoA* it is unknown. The frequency distributions are calculated for: all listed promoters under stringent control; *rrnX*, *rrnX*$_2$, tRNA^{Tyr}, *rpsL* (Str), *rpsE* (Spc), *rplKA*, *rpoA*, and *rpoB*; and all other known promoters listed in Table 2. Nomenclature and sequence data are taken from reference 20. The *rplKA* and *rpoB* sequences are from reference 27, the *rpoA* sequence is from Post and Nomura (personal communication), and the *rrnB*$_1$ and *rrnB*$_2$ sequences are from Earl and Cashel (personal communication). Sequences additional to those in reference 20 used for frequency distribution iii are those for ampicillinase (20) and ColE1-1 (10). Note that, for frequency distribution iii, 39 sequences are totaled for positions −5 to 13, and 38 are totaled for positions +4 to +7.

ple each of a proximal and distal rRNA promoter. Such an analysis supports the assignations at positions +2, +1, −2, and −5 but shows that the preference for a single base pair at positions −3 and −4 is a property of the rRNA promoters and not of stringent promoters as a group. Nevertheless, at these positions there remains a high frequency of guanosine plus cytosine base pairs. On this basis, the consensus sequence is refined to:

	−5		−1	+1	+4
		c	c		
−10 region	C	g	g	C	– C C – –

If this consensus sequence is functionally significant it should not occur in promoters which are not under stringent control. When 39 other promoters are scored to fit this sequence they fall in a Poisson distribution around the 2/6 fit expected from the random occurrence of any base at each position (Table 2). The promoter with the best fit, *lacI*, has the same extent of matching to the consensus sequence as three stringent promoters, *rrnA*$_2$, *rrnB*$_2$, and *rpoA*. The response of the *lacI* promoter to ppGpp is not known. Nevertheless, all promoters known not to be under stringent control in vivo or in vitro, e.g., *lac*, *galP*$_1$, *trp*, show a poor fit. A possible anomaly is the *rpoB* promoter which is proximal to the *rplJ*, *rplL*, and *rpoB* genes (8, 12, 24). In vivo the synthesis of the mRNA for these r-proteins is strongly inhibited on imposition of stringency while that of *rpoB* is not (9). Nor is the synthesis of the *rpoB* product, the β subunit of RNA polymerase, sensitive to ppGpp in vitro (18). At present it is unclear whether the

TABLE 2. *Fit of promoter sequences to consensus sequence* $C \overset{c}{g}\ g\ C - C\ C$ *in positions −5 to +2*[a]

Fit	Frequency	
	Stringent promoters	Others
6/6; ***rrnA*$_1$ *rrnB*$_1$ *rrnD*$_1$ *rrnE*$_1$ *rrnX*$_1$ *rrnD*$_2$ *rrnE*$_2$ *rrnX*$_2$ *rpsE*** **(Spc)**	9	—
5/6; **tRNA^{Tyr} *rplKA rpsL* (Str)**	3	—
4/6; ***rrnA*$_2$ *rrnB*$_2$ *rpoA lacI***	3	1
3/6; ***rpoB*** T7A3 φX174*A* φX174*B* φX174*D* *fd*IV Tet	1	6
2/6; λp_L, λp_O, λP_E T7A1 T7A2 *fd*II *lacP115 galP*$_2$ *araBAD trp bioA bioB bioP98*	—	13
1/6; λp_R, λp_M λ*cin* 434p_R *fd*III *fd*V *fd*VIII *fd*X T5$_{26}$ *lac araC* ColE1-1	—	12
0/6; λp_R λc$_{17}$ *fd*II′ T5$_{25}$ *galP*$_1$ *amp* SV40	—	7

[a] Promoters known or believed to be under stringent control are in boldface. —, Dashes indicate zero.

stringent control of *rplJ* and *rplL* is a consequence of regulation at the *rpoB* or the *rplKA* promoter.

The consensus sequence for stringent promoters clearly reflects the high proportion of C residues in the antisense strand and of guanosine plus cytosine base pairs in the region of sequence conservation. For such promoters these proportions are 67 and 86%, respectively, taking all stringent promoters, whereas for all other promoters these proportions, 22 and 46% respectively, are close to those expected from a random occurrence of any base pair at a particular position. Given the base composition of the −5 to +2 region, the probability of obtaining the consensus sequence in stringent promoters is 0.13, and in other promoters it is 0.0005. However, the probability of obtaining both the base composition and the consensus sequence of stringent promoters is ~0.0001 on the basis of random base distribution.

Promoter sequences may also be aligned with the transcriptional start point rather than the −10 region in a fixed position. When promoters under stringent control are scored for sequence conservation in this alignment the sequence

$$\begin{array}{cccccccccc} -5 & -4 & -3 & -2 & -1 & \text{sp} & +2 & +3 & +4 \\ - & C & - & - & C & - & \overset{c}{t} & C & - \end{array}$$

emerges as the best consensus sequence. The most highly conserved position is −1, which in every case is C. However, when other promoters are scored for fit to this sequence it is observed that both the T7A3 and *araBAD* promoters have a 75% match, identical to that of eight stringent promoters. The *araBAD* promoter should be available for activation in vivo in the presence of the high ppGpp levels accompanying the onset of glucose deprivation. Furthermore, for the *rpoA* promoter no possible start point within the usual limits, −1 to +3 relative to the −10 region (M. Rosenberg and D. Count, Annu. Rev. Genet., in press), has a >50% match to this sequence even though *rpoA* expression is sensitive to ppGpp in vitro (18). These exceptions to a correlation between sequence and function suggest that the consensus sequence derived from aligning promoter sequences with the start point in a fixed position is less likely to be a major determinant of the functional response to stringency.

In this paper I have identified a heptanucleotide sequence which with the exception of one position is highly conserved in promoters subject to stringent control and is in general absent from other promoters. This sequence is not totally conserved. Nor indeed is total conservation to be expected since promoters under stringent control vary in the apparent K_i for inhibition by ppGpp in vitro (3). The conserved nucleotides are exclusively G-C base pairs with a preponderance of C residues in the antisense strand. This base composition is fully consistent with the observation that ppGpp inhibits the transcription of poly d(I-C) but not of poly d(A-T) in vitro (2).

The consensus sequence occurs 11 times in the φX174 genome and thus by itself is insufficient for promoter recognition. More probably RNA polymerase is aligned by initial contacts at the −35 or −10 regions or both and then interacts with the sequence flanking the start point distal to the −10 region. This interaction is regulated by ppGpp. It is perhaps significant that this putative regulatory sequence lies within the 11 base pair region which is melted before the initiation of transcription (18), suggesting the possibility that this regulated interaction between a protein and DNA may require the recognition of single-stranded DNA.

ACKNOWLEDGMENTS

I thank L. Korn, J. D. Smith, and R. Staden for helpful discussions and R. Staden for determining the frequency of occurrence of the consensus sequence in the φX174 genome.

LITERATURE CITED

1. **Cashel, M., and J. Gallant.** 1969. Two compounds implicated in the function of the RC gene of *Escherichia coli*. Nature (London) **221**:838–841.
2. **Cashel, M., E. Hamel, P. Shapshak, and M. Boquet.** 1976. Interaction of ppGpp structural analogs with RNA polymerase. *In* N. O. Kjeldgaard and O. Maaløe (ed.), Control of ribosome synthesis, p. 279–290. Munksgaard, Copenhagen.

3. **Debenham, P. G., and A. Travers.** 1977. Selective inhibition of $tRNA^{Tyr}$ transcription by guanosine 3′ diphosphate 5′ diphosphate. Eur. J. Biochem. **72**:515–523.
4. **Dennis, P. P., and M. Nomura.** 1975. Stringent control of transcriptional activities of ribosomal protein genes in *E. coli*. Nature (London) **255**:460–465.
5. **Edlin, G., G. S. Stent, R. F. Baker, and C. Yanofsky.** 1968. Synthesis of a specific messenger RNA during amino acid starvation of *Escherichia coli*. J. Mol. Biol. **37**:257–268.
6. **Ikemura, T., and J. E. Dahlberg.** 1973. Small ribonucleic acids in *Escherichia coli*. II. Noncoordinate accumulation during stringent control. J. Biol. Chem. **248**: 5033–5041.
7. **Lavallé, R., and G. de Hauwer.** 1968. Messenger RNA synthesis during amino acid starvation in *Escherichia coli*. J. Mol. Biol. **37**:269–288.
8. **Linn, T., and J. Scaife.** 1978. Identification of a single promoter in E. coli for *rplJ*, *rplL* and *rpoBC*. Nature (London) **276**:33–37.
9. **Maher, D. L., and P. P. Dennis.** 1977. In vivo transcription of *E. coli* genes coding for rRNA, ribosomal proteins and subunits of RNA polymerase: influence of the stringent control system. Mol. Gen. Genet. **155**:203–211.
10. **Morita, M., and A. Oka.** 1979. The structure of a transcriptional unit on colicin E1 plasmid. Eur. J. Biochem. **97**:435–443.
11. **Morris, D. W., and N. O. Kjeldgaard.** 1968. Evidence for the non-coordinate regulation of ribonucleic acid synthesis in stringent strains of *Escherichia coli* J. Mol. Biol. **31**:145–148.
12. **Newman, A. J., T. G. Linn, and R. S. Hayward.** 1978. Evidence for cotranscription of the RNA polymerase genes *rpoBC* with a ribosomal protein gene of *Escherichia coli*. Mol. Gen. Genet. **169**:195–204.
13. **Nierlich, D. P.** 1968. Amino acid control over RNA synthesis: a re-evaluation. Proc. Natl. Acad. Sci. U.S.A. **60**: 1345–1352.
14. **Oostra, B. A., A. J. J. van Ooyen, and M. Gruber.** 1977. In vitro transcription of three different ribosomal RNA cistrons of *E. coli*; heterogeneity of control regions. Mol. Gen. Genet. **152**:1–6.
15. **Post, L. E., G. D. Strycharz, M. Nomura, H. Lewis, and P. P. Dennis.** 1979. Nucleotide sequence of the ribosomal gene cluster adjacent to the gene for the RNA polymerase subunit β in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **76**:1697–1701.
16. **Pribnow, D.** 1975. Bacteriophage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites. J. Mol. Biol. **99**:419–443.
17. **Seeburg, P., C. Nüsslein, and H. Schaller.** 1977. Interaction of RNA polymerase with promoters from bacteriophage fd. Eur. J. Biochem. **74**:107–113.
18. **Siebenlist, U.** 1979. RNA polymerase unwinds an 11 base-pair segment of a bacteriophage T7 promoter. Nature (London) **279**:651–652.
19. **Stomato, T. D., and D. E. Pettijohn.** 1971. Regulation of ribosomal RNA synthesis in stringent bacteria. Nature (London) **234**:99–102.
20. **Sutcliffe, J. G.** 1978. Nucleotide sequence of the ampicillin resistance gene of the *Escherichia coli* plasmid pBR322. Proc. Natl. Acad. Sci. U.S.A. **75**:3737–3741.
21. **Takanami, M., K. Sugimoto, H. Sugisaki, and T. Okamoto.** 1976. Sequence of promoter for coat protein gene of bacteriophage fd. Nature (London) **260**:297–302.
22. **Travers, A.** 1976. Modulation of RNA polymerase specificity by ppGpp. Mol. Gen. Genet. **147**:225–232.
23. **van Ooyen, A. J. J., M. Gruber, and P. Jorgensen.** 1976. The mechanism of action of ppGpp on rRNA synthesis *in vitro*. Cell **8**:123–128.
24. **Yamamoto, M., and M. Nomura.** 1978. Cotranscription of genes for RNA polymerase subunits β and β' with genes for ribosomal proteins in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **75**:3891–3895.
25. **Yang, H. L., G. Zubay, E. Urm, G. Reiness, and M. Cashel.** 1974. Effects of guanosine tetraphosphate, guanosine pentaphosphate and β-γ methylenyl-guanosine pentaphosphate on gene expression of *Escherichia coli in vitro*. Proc. Natl. Acad. Sci. U.S.A. **71**:63–67.