Structure of the DNA distal to the gene for ribosomal protein S20 in *Escherichia coli* K12: presence of a strong terminator and an IS1 element

George A.Mackie

Department of Biochemistry, University of Western Ontario, London, Ontario, Canada N6A 5C1

ABSTRACT

The sequence of nucleotides extending over 2.3 kb distal to the gene for ribosomal protein S20 of E. coli has been determined. Included in the sequence is an efficient rho-independent terminator 50 b.p. distal to the coding sequence for S20, a complete copy of IS1 which lacks, however, flanking direct repeats, and finally, an open reading frame capable of encoding a 28 kDa polypeptide of unknown function. Several lines of evidence suggest that the IS1 sequence described here must represent one of the copies resident in the bacterial chromosome rather than a newly transposed copy. Northern blotting experiments show that the gene for S20 is functionally monocistronic under all conditions tested in several genetic backgrounds. Thus it seems unlikely that the distal copy of IS1 plays any role in the termination or stability of mRNA transcribed from the gene for S20.

INTRODUCTION

Two general features of the organization of the genes for ribosomal proteins in E. coli have emerged from investigations of their structure and regulation (reviewed in ref. 1, 2, 3). Virtually all such genes are clustered into transcriptional units ("operons") containing from 2 to 17 cistrons. Moreover, with two exceptions (the S10 and L28 operons), ribosomal protein operons also contain genes which encode non-ribosomal functions. Co-transcribed genes include those for elongation factors, RNA polymerase subunits, RNA modifying enzymes, and others (3). The reasons for such clustering remain obscure although two explanations seem most reasonable: first, adjacent genes may have evolved by tandem duplication from their neighbours followed by subsequent sequence divergence and secondly, cotranscription of functionally related genes would simplify global regulatory mechanisms.

The gene for ribosomal protein S20/L26 (rpsT) stands out as an exception to the generalizations mentioned above. It is unlinked to other ribosomal genes but is close to the genes for isoleucyl tRNA synthetase (ileS) (4, 5.) and lipoprotein signal peptidase (lsp) (6). It is, however, divergently transcribed from them. The presence of a putative rho-independent terminator 50 b.p. beyond the stop codon for S20 (7) is reminiscent of the organization of the S21 operon where the gene for ribosomal protein S21 is separated from the genes for primase (dnaG) and sigma factor (rpoD) by a similar functional site (8, 9). I have, therefore, determined the DNA sequence distal to that published for S20 (7) and have examined whether transcription originating at the S20 promoters is capable of driving the expression of genes distal to that for S20. The results show that the distal DNA sequence contains an intact IS element, IS1, and that the gene for S20 behaves as if it were functionally monocistronic.

MATERIALS AND METHODS

Strains

Recombinant plasmids and phage are ultimately derived from lambda dapB2 (10) and the vectors pBR322 (11), pK01 (12), pSP64 (13), pSP65 (13), and M13mp9 (14) using methods described previously (15). Host strains for these constructions include: MM294 (from M. Meselson), C600 galK (12), N100 (12), and JM103 (14). Other strains are referenced in the appropriate legend.

DNA Sequencing

Residues 1-690 in Fig. 2 have been sequenced previously by the Maxam and Gilbert method (7). Residues 552-2882 were sequenced by the modified dideoxy technique of Biggin et al. (16), using a 17 residue primer obtained from P-L Biochemicals. DNA ligase and most restriction enzymes were obtained from Boehringer-Mannheim; the Klenow fragment of DNA polymerase I was purchased from P-L Biochemicals or from BRL Inc. The computer programs of Larson and Messing (17) were used to facilitate the analysis of sequence data.

In vitro Transcription

I employed the conditions described previously (18) using 0.25 pmoles linearized plasmid DNA as template and $[a^{32}P]$-UTP as

labelled substrate in a volume of 0.025 ml. The concentration of KCl was 50 mM where both S20 promoters are active. Products were analyzed by electrophoresis in a 6% polyacrylamide gel containing 8 M urea and were subsequently visualized by autoradiography.

Extraction and Analysis of RNA

RNA was extracted from exponential cultures of E. coli by a modification of published methods (19, 20) which together substantially eliminate DNA. Total RNA (0.004 mg) was denatured with glyoxal (21), electrophoresed on 1.5% agarose gels, and blotted to Biodyne A nylon membranes (Pall, Inc.). Hybridizations were performed at 43° in the following buffer: 0.75 M NaCl, 0.075 M Na-citrate (pH 7.0), 0.04 M Pipes (pH 6.4), 50% formamide (previously deionized), 0.1% Na-dodecylsulfate, 0.25 mg/ml crude yeast RNA, 0.020 mg/ml sonicated calf thymus DNA, 0.05% Ficoll (Pharmacia), 0.05% polyvinylpyrolidone, and 0.05% bovine serum albumin (Fraction V) (22). Radiolabelled probes were generated by transcription of the appropriate EcoRI-cleaved plasmid (see below) with SP6-encoded RNA polymerase in the presence of [a^{32}P]-UTP and non-radioactive ATP, CTP, and GTP as recommended by the supplier of the enzyme (Promega Biotec). No attempt was made to remove template DNA or to denature the probe prior to use. Following annealing for 30-48 h, excess probe was removed by washing the membrane with 0.3 M NaCl, 0.03 M Na-citrate (pH 7.0) containing 0.1% Na-dodecylsulfate at 36-37°. The following plasmids were constructed by subcloning portions of the sequence in Fig. 2 (indicated in parentheses) between the HindIII and EcoRI sites, respectively, of pSP64 (13) obtained from Promega Biotec: pGM49 (residues 551-232; 49E probe), pGM59 (residues 2538-2094; 59E probe), pGM60 (residues 1021-552; 60E probe which includes the 5' end of IS1), and pGM61 (residues 1793-1236; 61E probe which includes the 3' end of IS1). In all cases, the RNA transcript of the linearized template is complementary to the upper strand of the sequence in Fig. 2. See also Fig. 1c.

Enzyme Assays

Cultures of strains of interest were grown in supplemented M9 medium containing 0.2% fructose, 0.35% casamino acids, and 0.03 mg/ml ampicillin where appropriate. Toluenized extracts

were prepared as described by McKenney et al. (12). The assays
for beta-lactamase (23) and galactokinase (12) have been
described. All assay results are the average of duplicates or
triplicates performed in the linear range of the assay.

RESULTS AND DISCUSSION

DNA Sequence distal to the gene for S20

The diagram in Fig. 1a illustrates the structure of the DNA
and the location of potential coding sequences around the gene
for ribosomal protein S20 based on data from this and previous
work (4, 7,). The plasmids pGM7 and pGM8 were used as sources of
DNA. Inserts from these plasmids were purified, digested with
one of AluI, HaeIII, HinfI, HpaII, Sau3AI, or TaqI, and subcloned
into M13mp8 (see the legend to Fig. 1d). Fig. 1d diagrams the
subclones so obtained and the extent to which the sequence was
determined from each. The complete sequence of the region,
including portions already published for S20 (7,24) is tabulated
in Fig. 2. The most interesting feature of the sequence is the
presence of a complete (768 b.p.) copy of IS1 between residues
789-1557. This element was identified after the sequence in Fig.
2 was substantially complete by a comparison of the open reading
frames with amino acid sequences in a protein sequence data bank
using the fastp program of Lipman and Pearson (25). A more
detailed comparison has shown that the IS1 element in Fig. 2 is
virtually identical to that first sequenced by Ohtsubo and
Ohtsubo (26) except at residues 1050, 1086, 1089, and 1092
(residues 262, 298, 301, and 304 of IS1) where the nucleotides
agree with the sequence determined by Johnsrud (27). Unlike
transposed copies of IS1 (28, 29), this copy lacks flanking
direct repeats. Since transposition of IS1 is known to generate
deletions (30), it is conceivable that the original transposition
of IS1 to this site was accompanied by deletion(s) on one (or
both) side(s) of the IS1 element.

Two lines of evidence indicate that the IS1 element
identified here is almost certainly an endogenous copy in the E.
coli K12 chromosome, and is unlikely to be an artifact of
cloning. First, restriction maps of lambda dapB2 (4, 15) and of
the Clarke-Carbon plasmid pLC3-13 (31, 32) which were constructed
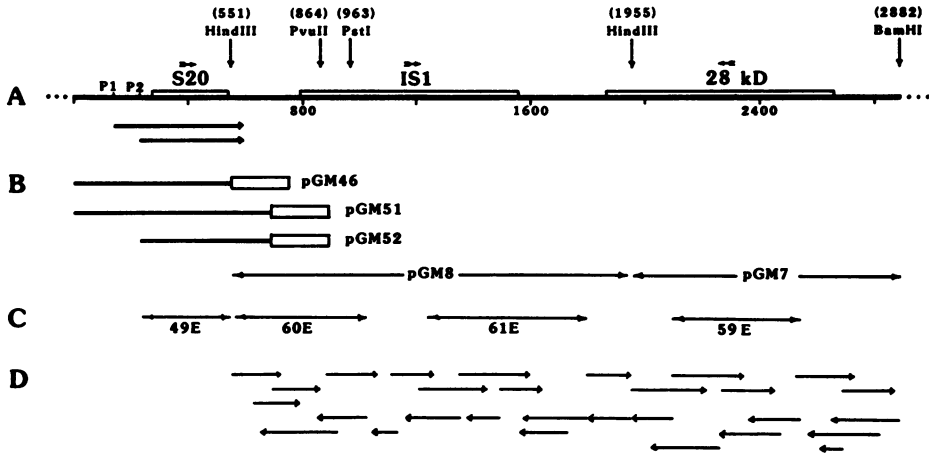independently but which overlap partially both contain a 1.4 kb

Fig. 1a. Map of the region comprised of the gene for ribosomal protein S20 (rpsT) and sequences distal to it. The open boxes denote coding sequences (S20 and 28 kD polypeptide) or other features (IS1) recognizable in the DNA sequence of Fig. 2. Small arrows denote the direction of translation in S20, IS1, and the 28 kD open reading frame. P1, P2, and the thick arrows below S20 indicate the known S20 promoters (7, 18) and the extent of the transcripts derived from them, respectively. Selected restriction enzyme cleavage sites and the corresponding coordinates are indicated by vertical arrows and numbers in parentheses. Numbers below the line represent distances in b.p. from the HincII site (position 1) as in the published sequence of S20 (7).
b. Plasmids spanning the S20 region. Solid lines indicate the extent of fragments which have been subcloned into plasmids (see also ref.4). The open boxes in plasmids pGM46, 51, and 52 represent galactokinase sequences of pKO1 (12) to which the upstream fragments have been ligated (see also Table II).
c. Single-stranded RNA probes for the S20 region. The double arrows denote the extent of single stranded RNA probes synthesized by transcription of derivatives of pSP64 (13) with SP6 polymerase and the appropriate substrates (refer to Materials and Methods). The probes are complementary to the upper strand in the sequence of Fig. 2.
d. Sequencing strategy. The arrows indicate the direction and extent to which the sequence in Fig. 2 was determined from individual phage (see Materials and Methods and Results). Each arrow represents from 1-3 experiments, each involving 1-3 gels. The strategy for sequencing residues 1-690 has been reported previously (7).

HindIII fragment corresponding to residues 552-1955 in Fig. 2. A second argument can be made from the data in Table I which summarizes the sizes of restriction fragments of lambda dapB2 (4) predicted to span this copy of IS1 compared to the sizes of IS1-

```
           10        20        30        40        50        60        70        80
AACCGGCGATTGAGTACCGAAATCAGAAAGAAGTGGCTTTTCATTGCCATGGCGCAAATCACGGGAAGAAACTGACCGCC

                                                                     P1
           90       100       110       120       130       140       150       160
TGCTGCAATTTTTATCGCGGAAAAGCTGTATTCACACCCCGCAAGCTGGTAGAATCCTGCGCAATCACTACGTAACGAGT

                                                                          P2
          170       180       190       200       210       220       230       240
GCCGGCACATTAACGGCGCTTATTTGCACAAATCCATTGACAAAAGAAGGCTAAAAGGGCATATTCCTCGGCCTTTGAAT

                                        S20
          250       260       270       280       290       300       310       320
TGTCCATATAGAACACATTTGGGAGTTGGACCTTGGCTAATATCAAATCAGCTAAGAAGCGCGCCATTCAGTCTGAAAAG

          330       340       350       360       370       380       390       400
GCTCGTAAGCACAACGCAAGCCGTCGCTCTATGATGCGTACTTTCATCAAGAAAGTATACGCAGCTATCGAAGCTGGCGA

          410       420       430       440       450       460       470       480
CAAAGCTGCTGCACAGAAAGCATTTAACGAAATGCAACCGATCGTGGACCGTCAGGCTGCTAAAGGTCTGATCCACAAAA

                                          S20
          490       500       510       520       530       540       550       560
ACAAAGCTGCACGTCATAAGGCTAACCTGACTGCACAGATCAACAAACTGGCTTAATCGCCAATTTGCTGAAGCTTTGTG

          570       580       590       600       610       620       630       640
AAAAAGCCCGCGCAAGCGGGTTTTTTTATGCCTGCTGCTTTTGTGTGCGTGAGAGAACACTACTGCCGGTTGCCAGATAG

          650       660       670       680       690       700       710       720
TCAAAAGAGACATGCCGTTGTACTATCTGAACTCTTCGTTGGCTTAAGAATCGTAGAACCATCGGCTGATACATTCTGAA

                                                                     IS1
          730       740       750       760       770       780       790       800
CAAGAATATCTACTCAATCTCCAGACTCCCCCACAAAGAATATGGATATTGTGATACACATTGAGGTAGGTGATGCTGCC

          810       820       830       840       850       860       870       880
AACTTACTGATTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAGCTGTCCCTCCTGTTCAG

          890       900       910       920       930       940       950       960
CTACTGACGGGGTGGTGCGTAACGGCAAAAGCACCGCCGGACATCAGCGCTATCTCTGCTCTCACTGCCGTAAAACATGG

          970       980       990      1000      1010      1020      1030      1040
CAACTGCAGTTCACTTACACCGCTTCTCAACCCGGTACGCACCAGAAAATCATTGATATGGCCATGAATGGCGTTGGATG

         1050      1060      1070      1080      1090      1100      1110      1120
CCGGGCAACAGCCCGCATTATGGGCGTTGGCCTCAACACGATTTTACGTCACTTAAAAAAACTCAGGCCGCAGTCGGTAAC

         1130      1140      1150      1160      1170      1180      1190      1200
CTCGCGCATACAGCCGGGCAGTGACGTCATCGTCTGCGCGGAAATGGACGAACAGTGGGGCTATGTCGGGGCTAAATCGC

         1210      1220      1230      1240      1250      1260      1270      1280
GCCAGCGCTGGCTGTTTTACGCGTATGACAGTCTCCGGAAGACGGTTGTTGCGCACGTATTCGGTGAACGCACTATGGCG

         1290      1300      1310      1320      1330      1340      1350      1360
ACGCTGGGGCGTCTTATGAGCCTGCTGTCACCCTTTGACGTGGTGATATGGATGACGGATGGCTGGCCGCTGTATGAATC

         1370      1380      1390      1400      1410      1420      1430      1440
CCGCCTGAAGGGAAAGCTGCACGTAATCAGCAAGCGATATACGCAGCGAATTGAGCGGCATAACCTGAATCTGAGGCAGC

         1450      1460      1470      1480      1490      1500      1510      1520
ACCTGGCACGGCTGGGACGGAAGTCGCTGTCGTTCTCAAAATCGGTGGAGCTGCATGACAAAGTCATCGGGCATTATCTG

                                       IS1
         1530      1540      1550      1560      1570      1580      1590      1600
AACATAAAACACTATCAATAAGTTGGAGTCATTACCAGCAATTACTACTCTTTTCAGCATCCCCATGAAGGAAGTAATCA

         1610      1620      1630      1640      1650      1660      1670      1680
GATTCCTCTATTTATTCGCCCGCGTACATATGGAAACAATGTCTCAAGCGGGCAAATAATGAGCCGAGTCAAAATCGTGG

         1690      1700      1710      1720      1730      1740      1750      1760
TAATGTACAACTAATCTATCTCCCCCACCTTAACTTTGGGGGAGCTGCCGATTAACGCACCGCTGGACTAAAAAGCGCAG

         1770      1780      1790      1800      1810      1820      1830      1840
AATAATCCGTATTGCAGATTCGCTGTACCGCCGGGTGCTGAATCATCCGCTCAGCAAAAATAGCATGGTACTCTTCCATC
```

```
      1850        1860        1870        1880        1890        1900        1910        1920
ACATTCTCGACGCGACCAATTTCTACGACAGTTTTATCGGCATAAAAGTCATATGCATAAAGCGTTGGGGCAACGAAGAT
                             ↑ 28 Kd
      1930        1940        1950        1960        1970        1980        1990        2000
TGCATTTGCATCGCCACCAAAAGGACCTGCAGCCAAGCTTTCATCAAAGCGGCATCATCAAACTCGCCGAGGATTTCTAC

      2010        2020        2030        2040        2050        2060        2070        2080
GTTTAATCCCTGGGAGTTAAACCAGTTAAGCAATTTGCGCCCTAACATTGAACGTCGCCCAGGAATCAAAAGTCGCCGTT

      2090        2100        2110        2120        2130        2140        2150        2160
CTTCCAGACAAGCCGGGAACGGTTTTTCTGGTGGTGGATTTGTACACCAGAAACTCACGCCACATTCGCCAATTCTCACG

      2170        2180        2190        2200        2210        2220        2230        2240
GAGAACAGGCCTTCCTGCTGCGTAGAGTCTATCGGACAGTCAGAAATGATCATCTCCAGTTTATGCTGACTTAATTGCTC

      2250        2260        2270        2280        2290        2300        2310        2320
CAGCAGCATTTCGTGGGTGGATTCGAAGCAGCGAAGATGAATGGGCTCGCCTTCTACCACTGCGGCGTTAAGTACGCTAC

      2330        2340        2350        2360        2370        2380        2390        2400
TGACCAGGCGTTTGGAAAGTGCATCAGCCACGCCAACGTCAAACAATAAATTGGATTCTTTGCGATAGTTCACAATATCC

      2410        2420        2430        2440        2450        2460        2470        2480
AGCATTTCCTGGCTTAAGGTGAACATTTTATCGGCATAGCGATAGACCAGTTCTCCAGCTCGCTGGGTTCGAGACCACGT

      2490        2500        2510        2520        2530        2540        2550        2560
CCCCTTGCGTTTAAATAATTTCGCTTGCAGGGCGTCTTCCAGCGCTCGAATCTGTCCGGTAATGGTTTGTGGAGTTAAAT

      2570        2580        2590        2600        2610        2620        2630        2640
AAAGCGCCTCCGCTGCGCCAACCACGGAACCTTCTTTATAGACATGCCAGAAGTAATACAAGTGGTTGTAATTGATATGA

      2650        2660        2670        2680        2690        2700        2710        2720
GACATGCTCATTTCTCTCCCTGATAACAATGAAAAGGGAGCCGTTTATGGCTCCCCGGTAAACCGTCCTGTCAAACTGAT
28 Kd ◄◄
      2730        2740        2750        2760        2770        2780        2790        2800
GGACGCAAACGAACGCGTAACCAGCTGTATCCAATTACCGCCGAAGAGATAGAACCGACCAGGATACCGAGTTTCGCCCA

      2810        2820        2830        2840        2850        2860        2870        2880
GTTAATCAGTTCTGGATCTACGCTACCAAAGGCCAGGCTGGCAATAAAGATAGACATAGTAAAACCGATACCGCACAGGA

      2890        2900        2910        2920        2930        2940        2950        2960
TC
```

Fig. 2. DNA sequence distal to (and including) ribosomal protein
S20 (rpsT). The numbering begins at a HincII site 272 residues
to the 5' side of the initiation codon for S20 as in ref. 7. A
clerical error in the reported sequence at residue 647 has been
corrected. Horizontal arrows above the sequence indicate
starting points (5' ends) for open reading frames and other
features of the sequence discussed in the text; vertical arrows
denote stopping points (3' ends). The sequence of residues 1-551
has also been reported by others (24).

containing fragments of E. coli DNA identified by Southern

blotting (33, 34, 35). It is clear that a fragment potentially

identical to one containing the S20-linked IS1 has been found in

E. coli in every case for which there is information.

    Two other copies of IS1 normally resident in the chromosome

of E. coli have been mapped and flank the argF locus, forming a

transposon-like structure (36, 37). Insufficient information is

available to determine whether the S20-linked copy of IS1 also

Table I. Sizes of restriction fragments containing endogenous
copies of IS1 in E. coli K12[a].

Enzyme	Sizes of fragments (kb)	
	S20-linked[b]	Endogenous[c]
BamHI	5.2	>30, >25, >20, >18, 7.2, 7.1, 5.1
BglII	5.3	12.9, 9.0, 7.6, 5.6, 3.8, 1.7
EcoRI + HindIII	1.4	19, 10.4, 9.7, 7.9, 3.5, 1.6, 1.2
HindIII	1.4	>30, >25, >18, >11, >10, 5.2, 1.5
PstI	4.0 + 3.1[d]	>18, >14, 7.6, 7.1, 6.6, 4.7, 3.9, 2.0

a.   Data for EcoRI fragments are not reported since there are no
sites for this enzyme in the 15 kb insert of lambda dapB2 (4).
b.   Sizes of fragments containing the S20-linked copy of IS1 are
taken from ref. 4, corrected in light of the sequence in Fig. 2.
c.   Sizes of the fragments containing endogenous copies of IS1
are taken from the following sources: BamHI, HindIII, and PstI
from ref 34; BglII, from ref 35; EcoRI + HindIII, from ref. 33.
The underlined numbers indicate the fragments which best
correspond to those of the S20-linked copy of IS1.
d.   The probe used by Nyman et al (34) originated in the 3'
portion of IS1 (to the "right" of the PstI site) and would not be
homologous to the 3.1 kb fragment of lambda dapB2.

forms one terminus of a transposon-like structure.

Efficiency of the putative S20 terminator in vitro

A G-C rich dyad symmetry followed by seven consecutive T
residues (residues 566-587) lies immediately distal to the S20
coding sequences in Fig. 2 . This structure resembles that of
well-characterized rho-independent terminators (reviewed in ref.
38). To facilitate the study of this potential terminator, the
HinfI site at residues 688-692 was repaired and a synthetic
linker containing a BamHI site attached. The potential
terminator could then be handled conveniently as a HindIII-BamHI
fragment of 145 residues, corresponding to residues 552-691 of
the DNA sequence in Fig. 2.

A template containing both S20 promoters, part of the
structural gene, and the "portable" terminator described above
was constructed from pGM32 (18) to yield pGM35 (see Fig. 3).
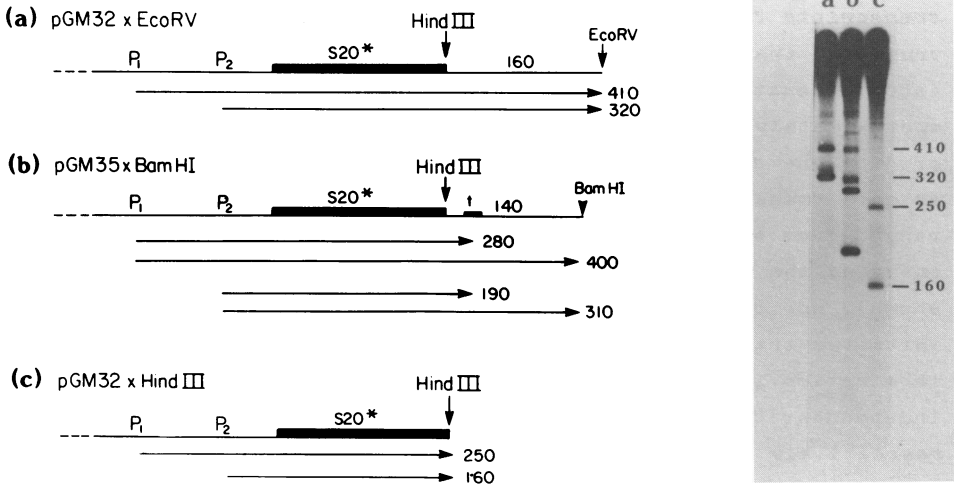Transcription of pGM35 linearized with BamHI should produce

(a) pGM32 x EcoRV

(b) pGM35 x Bam HI

(c) pGM32 x Hind III

Fig. 3. Efficiency of the putative S20 terminator in vitro. The left hand portion of the Figure illustrates the structure of three linearized templates, each containing both S20 promoters (P1 and P2), part of the structural gene to residue 390 (S20*; see also the legends to Tables II and III), with (b) or without (a, c) the putative transcriptional terminator. The arrows indicate the sizes, in nucleotide residues, of the expected transcripts. The right hand portion of the Figure illustrates the separation of in vitro transcripts labelled with [a^{32}P]-UTP and resolved on a 6% polyacrylamide gel containing 8 M urea (see Materials and Methods and ref. 18). The numbers in the far right hand margin give the sizes of the marker transcripts in lanes a and c. Other markers from an end-labelled HinfI digest of pBR322 were used to confirm these sizes (not shown). Lane (a), products transcribed from EcoRV-digested pGM32; lane (b), products transcribed from BamHI-digested pGM35; lane (c), products transcribed from HindIII-cleaved pGM32 (see ref. 18).

runoff transcripts of approximately 400 and 310 residues. If termination occurs near residue 587 (refer to Fig. 2) then these transcripts should be truncated to approximately 280 and 190 residues, respectively. For comparison, pGM32 whose transcription in vitro has been characterized previously (18) was cleaved with either HindIII or EcoRV to generate markers (refer to Fig. 3).

The products obtained after transcription of these templates in vitro were resolved by electrophoresis as shown in Fig. 3 (right side). Several conclusions have emerged from this experiment. First, the presence of the putative terminator in

pGM35 results in the synthesis of both truncated and read-through
transcripts from both S20 promoters (Fig. 3b). The sizes of the
truncated transcripts relative to the markers in lanes a and c
(and to denatured DNA fragments not shown in Fig. 3) are
approximately 300 and 190 residues, consistent with the
predictions discussed above. Quantification of the molar yields
of each transcript has shown that the efficiency of termination
ranged from 65-75%. These efficiencies have not been altered by
changing the ratio of RNA polymerase to template (data not
shown), nor do they depend on which of the S20 promoters
initiates transcription. For comparision, efficiencies of
termination in vitro at the trp operon attenuator and rho-
independent terminator (trp t) have been reported as 95% and 25%
respectively (39). The efficiency of termination in vitro from a
linearized plasmid template containing the entire structural gene
for S20, including residues 163-691 of the sequence in Fig. 2, is
comparable (70-75%) to that in pGM35 (data not shown). Thus
sequences in the 3' portion of the S20 structural gene which are
deleted in pGM35 do not appear to influence appreciably the
efficiency of termination in vitro. Taken together, the results
show that the sequence spanning residues 566-587 in Fig. 2
functions as a rho-independent terminator for transcription of
S20 in vitro.

Transcription of S20 and the adjacent IS1 in vivo

       The presence of an IS1 element whose polar properties are
well-known (40) so close to the gene for S20 raises the
possibility that the S20-linked copy of IS1 influences the
transcription of S20. The IS1 element could for example serve as
an auxiliary transcriptional terminator or as a modulator of
transcript stability. Conversely, S20 could influence expression
of IS1 (41). I have examined the predictions of these
possibilities in two ways. In one approach, I used methods
described by McKenney et al., (12) to assess the extent of
termination of transcription of the S20 gene in vivo. Various
fragments were inserted into the vector pKO1 which contains the
structural gene for galactokinase, but no promoter (12). Strains
harbouring plasmids pKO1 or pGM52 serve as controls as they lack
either S20 promoter. Comparison of the galactokinase/beta-

Table II.   Efficiency of the putative S20 terminator in vivo

Plasmid	S20 sequences[a]	Activity[b]		Galactokinase
		galK	B-lac	B-lactamase
pKO1	–	7.1	0.40	18
pGM52	232-691	16	0.55	29
pGM46	1-551	418	0.16	2580
pGM51	1-691	12	0.16	75
pGM48	1-390	936	0.16	5890
pGM45	1-390^552-691[c]	118	0.54	220
pGM54	1-272	1540	0.31	4940

a.   The numbering of the S20 sequences is taken from Fig. 2.
Refer also to Fig. 1b.   The sequences in pGM46, pGM48, and pGM54
were inserted between the EcoRI and HindIII sites of pKO1 whereas
those in pGM45, pGM51, and pGM52 were inserted between the EcoRI
and SmaI sites of the vector.   In all cases the S20 promoters
and/or coding sequences are oriented in the same sense as the
galactokinase coding sequences in pKO1.   All plasmids were
maintained in the host C600 galK (12).
b.   Toluenized extracts were prepared as described (12) and
assayed for galactokinase (galK) (12) or beta-lactamase (B-lac)
(23) under the conditions noted in Materials and Methods.
Activities are reported as nmoles product formed/min/A650 of
culture.
c.   In pGM45, a HindIII linker at a TaqI site in the S20 coding
sequence was used to join residues 1-390 to the natural HindIII
site between residues 551-552 (see Fig. 2).   The carat denotes
the deletion of residues 391-551.

lactamase ratios in the plasmid pairs pGM46-pGM51 (Table II,
lines 3 and 4) and pGM48-pGM45 (Table II, lines 5 and 6) shows
that read-through past the terminator (in pGM45 and pGM51)
amounts to 2-3% of the total transcription generated from the S20
promoters.   The absence of residues 391-551 in the pair pGM45-
pGM48 does not significantly enhance this readthrough.   Thus
potential RNA secondary structure which could form in this region
(not shown) does not appear to contribute significantly to
termination at or near residue 587.   The higher overall
activities in the pair pGM45-pGM48 relative to pGM46-pGM51 may
reflect a reduced polarity on galactokinase expression.   The
construction of pGM45 and pGM48 deletes the UAA stop codon of the
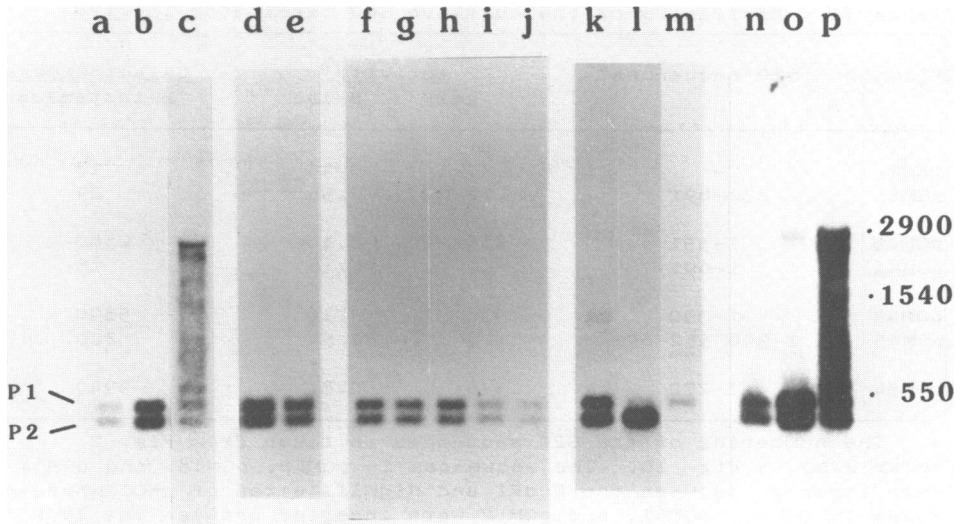S20 coding sequence and creates a read-through polypeptide which

Fig. 4. Analysis of S20 mRNAs in various strains. Total RNA
(4.0 ug; 2.0 ug in lanes i and j) from the strains enumerated
below was denatured, resolved by electrophoresis, blotted to
nylon, and probed with [^{32}P]-labelled 49E probe (see Materials
and Methods and Fig. 1c). Exceptions are noted below. The
sources of RNA and their relevant genotypes and phenotypes are:
(a) C600 galK; (b) C600 galK (pGM51) (refer to Fig. 1a and Table
II); (c) C600 galK (pGM46) (refer to Fig. 1a and Table II); (d)
N2076 (42), wild type for ribonuclease III; (e) N2077 (42),
partially deficient for ribonuclease III under the conditions
used (not shown); (f) N99 (43), wild type for nusA; (g) C205
(43), nusA1 derivative of N99; (h) SA1030 (44), wild type for rho
factor; (i) AD1600, rho15 derivative (44) of SA1030 grown at 30°
where the phenotype is partially expressed (data not shown); (j)
AD1600 grown at 30° then shifted to 37° for 30 min prior to
harvest; (k) C600 galK (pGM51) as in (b); (l) C600 galK (pGM44)
(this plasmid carries residues 164-691 cloned between the EcoR1
and SmaI sites of pKO1 and should overexpress transcripts
initiated at P2); (m) C600 galK (pGM51) probed with a nick-
translated fragment spanning residues 1-232 of the S20 sequence
which should detect only P1-initiated transcripts); (n), (o),
(p), as in lanes (a), (b), and (c) respectively, except that the
exposure was increased.

would ultimately terminate in the galactokinase leader of the

vector close to the start codon. Complete deletion of S20 coding

sequences (pGM54 in Table II) does not lead to any further

increase in the ratio of galactokinase to beta-lactamase. I

conclude that up to 98% of the transcription originating at the

S20 promoters terminates prior to residue 691 in the absence of

the IS1 element.

In a second approach, RNA extracted from various E. coli strains (see the legend to Fig. 4) was resolved by electrophoresis and analyzed by Northern blotting. The autoradiogram in Fig. 4 (lanes a and n) shows that the 49E probe specific for S20 (see Materials and Methods) detects two discrete mRNAs, and no others, in the haploid parental strain C600 galK. These transcripts are approximately 450 and 350 residues in length. These sizes are consistent with initiation of transcription at residues 141 and 232, respectively, with termination at or near residue 587, as is the case in vitro. The intensities of both transcripts, and thus their relative abundances, are roughly equal under the conditions examined (Figure 4, lanes a, d-j, and n). There is no indication of longer (read-through) or shorter (degraded) S20 transcripts in the parental haploid strain, or in any of the mutant strains tested, even with longer exposures (Fig. 4, lane n). Examination of strains partially deficient in ribonuclease III (Fig. 4, compare lanes e to d), altered in the nusA protein (Fig. 4, compare lane g to f), or temperature-sensitive for the termination factor rho (Fig. 4, compare lanes i and j to lane h) failed to uncover any alteration in the basic pattern or relative intensities of the two S20 transcripts. In other experiments (not shown) I have found that it is feasible to detect as little as 1% of the steady-state haploid level of S20 mRNA by Northern blotting. This sensitivity would be adequate to detect read-through RNAs at the level suggested by the fusion experiments above. Thus if there is read-through of the rho-independent termination signal distal to S20, none of the factors or enzymes examined (ribonuclease III, nusA, or rho) is rate-limiting in the steady-state.

RNA from the multicopy strain C600 galK (pGM51) (see Table II) displays a pattern identical to that from strain C600 galK with an approximate 5 to 6 fold increase in intensities of both S20 transcripts (Fig. 4, lanes b, k, and o). In contrast, RNA from strain C600 galK (pGM46) which lacks the "portable" terminator on the plasmid (refer to Table II) forms a smear ranging from about 350 to about 3000 residues (Fig. 4, lanes c and p). This smear, presumably, represents S20-galactokinase

fusion mRNA in various stages of synthesis and decay. RNA from strain C600 galK (pGM44) which includes residues 164-690 inserted between the EcoRI and SmaI sites of the vector pKO-1 (i.e., P2, S20 leader and coding sequences, and the terminator; refer to Fig. 1 and Fig. 2) displays a substantial intensification of the lower (350 residue) S20 transcript. Conversely, a probe generated from residues 1-232 of the S20 gene (refer to Fig. 2) detects only the longer of the two S20 mRNA species (Fig. 4 lane m). These observations strengthen the interpretation made above that sites of initiation and termination of transcription deduced from in vitro experiments reflect the situation in vivo. In addition, since the plasmid pGM51 lacks the distal IS1 sequence, the results obtained in Fig. 4 lanes b and k indicate that this IS1 is entirely dispensible for the generation of steady-state RNA species indistinguishable from those encoded by the chromosomal gene. Likewise, these results also suggest that if S20 mRNA is processed to yield the observed steady-state S20 transcripts, then the sites of termination and/or processing must lie between residues 587 and 691.

I have also used Northern blotting to search RNA from strain C600 galK doubling in 40 or 90 min for potential IS1 transcripts homologous to the probes diagrammed in Fig. 1c. None of the distal probes tested (59E, 60E, or 61E) detected discrete RNAs in this haploid strain under several conditions of stringency (data not shown). A positive control (not shown) demonstrated that the 61E probe could detect an artificially generated IS1 transcript in vivo. Thus if there is an IS1 transcript specified by the copy of IS1 distal to the gene for S20, it is either insufficiently abundant, too heterogeneous in size, or too unstable to be detected by blotting. Since IS1-mediated transposition is a rare event (40), very low levels of IS1 mRNA, beyond the sensitivity of these assays, may be entirely adequate to account for its observed level of activity.

Taken together, these experiments argue strongly that virtually all the S20 mRNA in E. coli is monocistronic and does not include detectable quantities of the distal IS1 sequence. If minor quanitities of a polycistronic mRNA containing both S20 and IS1 do exist, they are unlikely to be significant factors in the

expression of S20. The rho-independent terminator 5' to the gene
for S20 must, therefore, be highly efficient in vivo. The
absence of proteins such as nusA in the experiment of Fig. 3 may
be sufficient to explain the differences between the efficiencies
of this terminator measured in vitro and in vivo. The trpt
terminator serves as a precedent in this regard (45). The
primary and potential secondary structures conferred on S20 mRNA
by its terminator strongly resemble those at the 3' termini of
rRNA transcripts (reviewed in ref. 46). The quasi-symmetrical
structure of the S20 terminator suggests that like rRNA
terminators it should also be capable of functioning in both
orientations (cf. ref. 47). This has not been tested directly,
however.

Open Reading Frames Distal to S20

In addition to the three known open reading frames (rpsT
(S20), insA, and insB (40)) on the upper strand of the sequence
in Fig. 2, there is a fourth long open reading frame extending
from residues 2651-1863 on the opposite strand. The latter could
encode a protein of molecular weight 28,000 and is preceeded by a
potential ribosome binding site, AGGGAG, albeit in close
proximity to the initiating codon. We have previously shown that
a plasmid, pGM7, encompassing residues 1955-2882 of the sequence
in Fig. 2 can encode a 24 Kd protein in a system for coupled
transcription and translation in vitro (4). Inspection of the
sequence in Fig. 2 suggests that this product represents a fusion
of most of the coding sequences in the fourth open reading frame
to the HindIII site of pBR322. A stop codon to the 5'side of
this site in the vector would terminate translation and account
for the observed 24 Kd product of pGM7. In addition, one of the
M13 clones used in sequencing (8042: residues 2815-2208) yielded
blue plaques whereas its complement, 8051 produced white, as
expected. This observation is consistent with an open reading
frame on the lower strand of the sequence in Fig. 2 spanning
residue 2208. Thus the fourth open reading frame is clearly
translatable. A comparison of its predicted amino acid sequence
with those in the NBRF data bank using the searching procedures
of Lipman and Pearson (25) failed to detect any significant
homologies to known proteins. The function of the potential

product of this open reading frame therefore remains obscure.
Two regions of hyphenated dyad symmetry separate the IS1 copy
from the 3'-end of the fourth long open reading frame. The
significance of these symmetry elements, if any, remains to be
determined.

REFERENCES

1.   Lindahl, L. and Zengel, J.M. (1982) Adv. Genet. 21, 53-121.
2.   Nomura, M., Gourse, R., and Baughman, G. (1984) Ann. Rev.
     Biochem. 53, 75-117.
3.   Bjork, G.R. (1985) Cell 42, 7-8.
4.   Mackie, G.A. and Parsons, G.D. (1982) Can. J. Biochem. 60,
     338-346.
5.   Webster, T., Tsai, H., Kula, M., Mackie, G.A., and Schimmel,
     P. (1984) Science 226, 1315-1317.
6.   Regue, M., Remenick, J., Tokunaga, M., Mackie, G.A. and Wu,
     H.C. (1984) J. Bacteriol. 158, 632-635.
7.   Mackie, G.A. (1981) J. Biol. Chem. 256, 8177-8182.
8.   Burton, Z., Burgess, R.R., Lion, J., Moore, D., Holder, S.,
     and Gross, C.A. (1981) Nucl. Acids. Res. 9, 2889-2903.
9.   Smiley, B.L., Lupski, J.R., Svec, P., McMacken, R., and
     Godson, G.N. (1982) Proc. Nat. Acad. Sci.(USA) 79, 4550-
     4554.
10. Friesen, J.D., Parker, J., Watson, R.J., and Fiil, N.P.
     (1976) Mol. Gen. Genet. 144, 115-118.
11. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C.,
     Heynecker, H.L., Boyer, H.W., Crosa, J.H., and Falkow, S.
     (1977) Gene 2, 95-113.
12. McKenney, K., Shimatake, H., Court, D., Schmeissner, U.,
     Brady, C., and Rosenberg, M. (1981) in "Gene Amplification
     and Analysis" (Chirikjian, J.S. and Papas, T.S., eds.), pp
     383-415, Elsevier-North Holland, New York.
13. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T.,
     Zinn, K., and Green, M.R. (1984) Nucl. Acids, Res. 12,
     7035-7056.
14. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
15. Mackie, G.A. (1980) J. Biol. Chem. 255, 8928-8935.
16. Biggin, M.D., Gibson, T.J., and Hong, G.F. (1983) Proc.
     Nat. Acad. Sci. (USA) 80, 3963-3965.
17. Larsen, R. and Messing, J. (1982) Nucl. Acids, Res. 10, 39-
     49.

18. Mackie, G.A. and Parsons, G.D. (1983) J. Biol. Chem. <u>258</u>, 7840-7846.
19. Dennis, P.P. and Nomura, M. (1975) J. Mol. Biol. <u>97</u>, 61-76.
20. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., and Rutter, W.J. (1979) Biochemistry <u>18</u>, 5294-5299.
21. McMaster, G.K. and Carmichael, G.G. (1977). Proc. Nat. Acad. Sci.(USA) <u>74</u>, 4835-4838.
22. Denhardt, D.T. (1966) Biochem. Biophys. Res. Commun. <u>23</u>, 641-646.
23. Ross, G.W. and O'Callaghan, C.H. (1976) Meth. Enzymol. <u>43</u>, 69-85.
24. Kamio, Y., Lin, C.-K., Regue, M., and Wu, H.C. (1985) J.Biol. Chem. <u>260</u>, 5616-5620.
25. Lipman, D.J. and Pearson, W.R. (1985) Science <u>227</u>, 1435-1441.
26. Ohtsubo, H. and Ohtsubo, E. (1978) Proc. Nat. Acad. Sci. (USA) <u>75</u>, 615-619.
27. Johnsrud, L. (1979) Mol. Gen. Genet. <u>169</u>, 213-218.
28. Calos, M.P., Johnsrud, L., and Miller, J.H. (1978) Cell <u>13</u>, 411-418.
29. Grindley, N.D.F. (1978) Cell <u>13</u>, 419-426.
30. Reif, H.-J. and Saedler, H. (1975) Mol. Gen. Genet. <u>137</u>, 17-28.
31. Yamagata, H., Daishima, K., and Mizushima, S. (1983) FEBS Lett. <u>158</u>, 301-304.
32. Tokunaga, M., Loranger, J.M., and Wu, H.C. (1983) J. Biol. Chem. <u>258</u>, 12102-12105.
33. Read, H.A., Das Sarma, S., and Jaskunas, S.R. (1980) Proc. Nat. Acad. Sci. (USA) <u>77</u>, 2514-2518.
34. Nyman, K., Nakamura, K., Ohtsubo, H., and Ohtsubo, E. (1981) Nature <u>289</u>, 609-612.
35. Hu, M. and Deonier, R.C. (1981) Gene <u>16</u>, 161-170.
36. Hu, M. and Deonier, R.C. (1981) Mol. Gen. Genet. <u>181</u>, 222-229.
37. York, M.K. and Stodolsky, M. (1981) Mol. Gen. Genet. <u>181</u>, 230-240.
38. Holmes, W.M., Platt, T., and Rosenberg, M. (1983) Cell <u>32</u>, 1029-1032.
39. Christie, G.E., Farnham, P.J., and Platt, T. (1981) Proc. Nat. Acad. Sci. (USA) <u>78</u>, 4180-4184.
40. Calos, M.P. amd Miller, J.H. (1980) Cell <u>20</u>, 579-595.
41. Biel, S.W., Adelt, G., and Berg, D.E. (1984) J. Mol. Biol. <u>174</u>, 251-264.
42. Apirion, D., and Watson, N. (1975) J. Bacteriol. <u>124</u>, 317-324.
43. Friedman, D.E. (1971) in "The Bacteriophage Lambda" (Hershey, A.D., ed), pp. 733-738, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
44. Das, A., Court, D., and Adhya, S. (1976) Proc. Nat. Acad. Sci. (USA) <u>73</u>, 1959-1963.
45. Farnham, P.J., Greenblatt, J., and Platt, T. (1982) Cell <u>29</u>, 945-951.
46. Liebke, H. and Hatfull, G. (1985) Nucl. Acids Res. <u>13</u>, 5515-5525.
47. Postle, K. and Good, R.F. (1985) Cell <u>41</u>, 577-585.