Transcriptional and Translational Initiation Sites of IS50
Control of Transposase and Inhibitor Expression

Mark P. Krebs and William S. Reznikoff

Department of Biochemistry
College of Agricultural and Life Sciences
University of Wisconsin-Madison, Madison, WI 53706, U.S.A.

(Received 18 March 1986, and in revised form 7 July 1986)

We have examined transcriptional start sites responsible for expression of the transposase and transposition inhibitor proteins encoded by IS50, and determined the likely translational start site of transposase. Amino-terminal analysis of a transposase-β-galactosidase fusion protein gave the sequence Met-Ile-Thr-Ser-Ala, which corresponds to the predicted amino acid sequence starting at position 93 of IS50. S1 nuclease mapping of IS50 RNA produced in vivo indicated that three transcripts, T1, T2 and T3, start near this position. Only T1 starts upstream from the transposase amino terminus. T2 corresponds to an in-vitro transcript described previously. Analysis of the transcripts and proteins produced from deletion derivatives of an IS50-lacZ construct suggested that the three transcripts initiate at independent but overlapping promoters clustered near the end of IS50. This analysis confirmed that only T1 can encode transposase, and that T2 is largely responsible for expression of the inhibitor protein. The coding capacity of T3 was not determined. Finally, transcripts that originate outside of IS50 are prevented from expressing transposase because of a secondary structure that is present in these transcripts only.

1. Introduction

Tn5 is a transposable element that can function in numerous Gram negative bacteria (Berg & Berg, 1983). In Escherichia coli, Tn5 transposes at a frequency of about 10^-5 events per generation. This low frequency is typical of transposable elements (Kleckner, 1981), and may reflect the need to limit the occurrence of detrimental rearrangements promoted by transposons. In some transposons, such as Tn3 and Tn10, the transposition frequency is limited predominantly by the amount of the transposase protein (Chou et al., 1979b; Gill et al., 1979; Morisato et al., 1983). The mechanisms used to regulate transposase expression vary among transposons. Synthesis of the Tn3 transposase is transcriptionally regulated by the Tn3 repressor-resolvase protein (Chou et al., 1979a; Heffron et al., 1979). Tn10 has several means of limiting transposase levels. The Tn10 transposase is produced at very low levels as a result of infrequent transcription and translation of the transposase gene (Simons et al., 1983; Raleigh & Kleckner, 1986). Translation of transposase can be reduced further by an antisense mRNA provided by multiple copies of IS10 (Simons & Kleckner, 1983; Simons et al., 1983). In addition, the transcription rate of the transposase gene is decreased by methylation of adenine residues within its promoter by the product of the dam gene (Roberts et al., 1985). If the level of transposase limits the Tn5 transposition frequency, Tn5 may use some of these strategies to regulate transposase expression.

The structure of Tn5 includes two insertion sequences, IS50L and IS50R, which flank a region encoding several antibiotic resistance genes (Berg et al., 1975; Auerswald et al., 1980; Mazodier et al., 1985). IS50R produces at least two proteins, p1 and p2, which are encoded from the same reading frame (Rothstein & Reznikoff, 1981; Rothstein et al., 1980). Truncated versions of these proteins, p3 and p4, are produced by IS50L (Rothstein et al., 1980), which differs from IS50R at a single position near the 3' terminus of the reading frame (Auerswald et al., 1980; Rothstein & Reznikoff, 1981). Only p1 and p2 appear to have functional significance. The transposase protein, p1, is required for transposition (Rothstein et al., 1980, Isberg & Syvanen, 1981). The inhibitor protein, p2, lowers the transposition frequency by an unknown mechanism.
that does not involve regulation of IS50 \( R \) protein synthesis (Johnson et al., 1982; Isberg et al., 1982). p2 inhibits transposition when it is produced in high amounts, as when it is expressed from the lacUV5 promoter (Johnson et al., 1982; Johnson & Reznikoff, 1984) or from a multiplicity plasmid containing IS50 \( R \) (Johnson et al., 1982; Isberg et al., 1982). Lower levels of p2, as produced from intact Tn3 in *Escherichia coli*, do not appear to inhibit transposition significantly (Johnson & Reznikoff, 1984). It is not known, however, whether further reductions in the level of p2 would result in an increase in the transposition frequency. Thus, the level of p1, p2 or of both proteins may contribute to the low transposition frequency of Tn3. Alternatively, the levels of host-encoded proteins that are involved in the transposition process may be limiting, as has been suggested (Johnson & Reznikoff, 1984).

As a step toward understanding the control of transposase expression, we have identified transcriptional and translational initiation sites that direct IS50 protein synthesis. Our results suggest that p1 is not expressed from a previously identified transcript, which starts at position 98 of IS50 (Johnson & Reznikoff, 1981), but is expressed from a weak transcript that initiates at position 66. The transcript that starts at position 98 is required for high-level expression of p2. Several implications of these results are discussed.

2. Materials and Methods

(a) Bacterial and phage strains

*E. coli* strains RZ201 \( (\Delta (l a c - p r o) a r a s t h i) \) and RZ211 \( (\Delta (l a c - p r o) a r a r s t e x a 5 6 1 s e l t h i) \) (Johnson et al., 1982) were used for nearly all experiments. JM101 \( (F^{+} traD36 lacT7Zm15 proA3(lac-pro) supE4 thi) \) was used as a host for bacteriophage M13mp10 (Messing, 1972). Relative error did not exceed ±5%.

(b) Media

 Cultures were grown at 37°C in LB, 2YT or M9 media prepared as described (Miller, 1979). M9 was supplemented with 4 mg glucose/ml, 2 mg Casamino acids/ml, 4 μg thiamine/ml, 40 μg proline/ml and 1 mm MgSO4. Ampicillin was used at 100 μg/ml when appropriate. X-gal was used at 40 μg/ml.

(c) Enzymes

Restriction enzymes were obtained from New England Biolabs or Bethesda Research Labs. Bal31 exonuclease was obtained from New England Biolabs. S1 nuclease, calf intestinal alkaline phosphatase, and 3'-2', macro-globulin were purchased from Boehringer-Mannheim, and bacteriophage T4 polynucleotide kinase was obtained from PL Biochemicals. Rabbit anti-β-galactosidase sera globulin were purchased from Boehringer-Mannheim, and combined and used to prepare IgG as described (Johnson et al., 1982). Phage T4 DNA ligase and the large fragment of DNA polymerase I (Klenow fragment) were gifts from Dr M. Cox (University of Wisconsin).

(d) Plasmid constructions

DNA manipulations were performed essentially as described (Maniatis et al., 1982). pRZ903 was constructed by digesting pRZ102 (Jorgensen et al., 1979) with EcoRI and HpaI, and combining this with pMC1403 (Casadaban et al., 1980) that had been cut with EcoRI and SmaI. The sample was treated with T4 DNA ligase and used to transform cells. Ampicillin-resistant colonies that were blue on plates containing X-gal were characterized further pRZ918, pRZ919 and pRZ920 were constructed from pRZ311, a derivative of pBR201 (Rothblat & Reznikoff, 1981) in which the BglII-SmaI fragment was replaced with a BglII-SmaI fragment containing the carboxy-terminal portion of the transposase gene fused to lacZ (R. Johnson, unpublished results). pRZ311 was linearized with XhoI, treated with Bal31 exonuclease, and then digested with BamHI. The sample was treated with Klenow fragment in the presence of the 4 dNTPs, then ligated with T4 DNA ligase and used to transform cells. Transformants that were blue on X-gal plates containing X-gal and ampicillin were characterized further by restriction analysis and plasmid sequencing (Wallace et al., 1981). pRZ651 through pRZ661 were constructed from pRZ309, a fusion of lacZ at the carboxy terminus of the transposase gene (R. Johnson, unpublished results). pRZ309 was linearized with EcoRI, treated with Bal31 exonuclease, and then with Klenow fragment in the presence of the 4 dNTPs. EcoRI linkers (8-mer New England Biolabs) were added using T4 DNA ligase. After the sample was digested with EcoRI and SalI, the fragment containing IS50 was purified and combined with pBR322, which also had been digested with these enzymes. After ligation and transformation, a number of colonies having a range of blue intensity on plates containing X-gal and ampicillin were characterized further by restriction analysis. Candidates were digested with EcoRI and XhoI, and ligated with M13mp19 that had been cut with EcoRI and SalI. Dideoxy sequencing was performed as described (Sanger et al., 1977).

(e) β-Galactosidase assays

Cells (RZ211) carrying plasmids were grown in supplemented minimal media as described above to an \( A_{550} \) of 0.4 to 0.6. β-Galactosidase was assayed as described (Miller, 1979), in triplicate using independent transformants of the plasmids that were being tested. Relative error did not exceed ±5%.

(f) Fusion protein purification and sequencing

Three litres of cells (RZ201) carrying pRZ903 were grown to late log phase (\( A_{550} \approx 3.0 \)) in 2YT medium containing 100 μg ampicillin/ml. Cells were harvested by centrifugation for 30 min at 3000 g and washed in 30 ml of 10 mM-Tris-HCl (pH 8.0). After centrifuging for 10 min at 10,000 g, the pellets were frozen and stored at -20°C. A modification of the method of Struck et al. (1985) was used to purify the fusion protein. The cells were thawed, resuspended in 30 ml of 10 mM-Tris-HCl (pH 8.0), 30 μg 2'-3'- macro-globulin/ml, 30 μg phenylmethylsulfonyl fluoride/ml and passed once through a French Pressure cell at 110,400 kPa. The sample was centrifuged for 10 min at

† Abbreviations used: X-gal, 5-bromo-4-chloro-3-indolyl-β-D-galactoside; Ig, immunoglobulin; PTH, phenylthiohydantoin; bp, base-pair(s).
Cellular RNA was prepared as described (Peterson & Reznikoff, 1985). Cells (RZ201) carrying pRZ204 (Rothenstein & Reznikoff, 1981) or pRZ964 were grown in supplemented minimal medium to an A₆₀₀ of 0.5 to 0.6. Purified RNA was precipitated with ethanol, resuspended in water to a final concn of about 1 mg/ml, and stored at −20°C. To prepare probes for S₁ nuclease analysis, restriction fragments from the region of interest were treated with calf intestinal alkaline phosphatase, and then with T4 polynucleotide kinase in the presence of [γ-3²P]ATP as described (Maniatis et al., 1982). After labeling, the fragment was cut with a second restriction enzyme and the appropriate end-labeled segment was purified. In initial experiments, a 185 bp HindII–HincII fragment from IS50 labeled at the HincII site was used as a probe. A 233 bp EcoRI–HinPI fragment from pRZ102 was labeled at the HinPI site and used as a probe in later experiments. A modification of the method of Taeva et al. (1981) was used for S₁ nuclease mapping; 20,000 cts/min (Cerenkov) of probe, which amounted to 0.4 pmol of the HindII–HincII fragment and 0.06 pmol of the EcoRI–HinPI fragment, was combined with 25 or 5 µg of purified RNA and precipitated with ethanol. The dried samples were resuspended in 35 µl of 40 mM-PIPES (pH 6.4), 0.4 M-NaCl, 1 mM-EDTA, 80% (v/v) formamide, and heated for 15 min at 75°C followed by a slow cooling to 37°C over 3 h. Then 5000 units of S₁ nuclease were added in 30 µl of 30 mM-sodium acetate (pH 4.5), 9.25 M-NaCl, 1 mM-ZnSO₄, 5% (v/v) glycerol and the samples were incubated for 1 h at 30°C. After precipitation with ethanol, the products were separated on an 8% (v/v) polyacrylamide/7 M-urea sequencing gel and autoradiographed. For size markers, the hydroxylapatite and electrophoresis probes were sequenced as described (Maxam & Gilbert, 1980). The intensity of S₁-protected bands was measured by scanning the autoradiograph with a Zeineh soft laser densitometer.

Cells (RZ201) carrying pRZ964 were grown to an A₆₀₀ of 0.5 to 0.6 in M9 medium supplemented as above. Equivalent A₆₀₀ units of each sample (approx. 0.9) were centrifuged for 5 min in an Eppendorf centrifuge at 4°C. The pellets were resuspended in 100 µl of sample buffer (Laemml, 1970), heated at 100°C for 5 min, and electrophoresed on a 0.1% (w/v) sodium dodecyl sulfate/5% polyacrylamide gel. Immunoblotting was performed using rabbit anti-β-galactosidase IgG and a goat-anti rabbit IgG horse radish peroxidase conjugate (Bio-Rad Laboratories) according to the manufacturer's instructions. The electrophoresed samples were transferred to nitrocellulose by electrophoresis. Bovine serum albumin (Sigma) was used at 3% (w/v) as a blocking agent and in washes, or at 1% (w/v) in solutions containing antibody. Antibody complexes were detected with 4-chloro-1-napthol. The stained bands were quantified by scanning with a Zeineh soft laser densitometer.

3. Results

(a) Location of the transposase translational start site

The large open reading frame of IS50R contains three triplets that could encode the translational start site of pl: the ATG at positions 81 to 83, the ATG at positions 93 to 95, and the GTG at positions 138 to 140 (Figs 1 and 3, see numbering below). pl was previously predicted to start at the GUG corresponding to positions 138 to 140 (Johnson & Reznikoff, 1981). We constructed a series of IS50-lacZ translational fusion plasmids to determine if the pl start site was encoded upstream from this site. The lacZ gene in these plasmids was fused to IS50 upstream from the region that encodes p2, so that β-galactosidase activity would reflect translation beginning at the pl translational start site. The location of the fusion endpoint in these plasmids is shown in Figure 1, and the β-galactosidase activity of cells carrying the plasmids is shown in Table 1.

The lacZ gene in pRZ918 was fused to IS50 after the T at position 131, so that the putative initiation codon corresponding to the GTG at positions 138 to 140 was absent. In pRZ19 and pRZ20, the fusion endpoint was located further downstream from the GTG, after the C at position 143 and the G at position 149, respectively. The β-galactosidase activity of cells harboring these plasmids differed by less than 2-fold (Table 1). If the pl translational start site had been deleted in pRZ19, we would have expected a substantially lower value for β-galactosidase activity, which was not observed. Likewise, cells carrying an IS50-lacZ fusion plasmid, in which the GTG had been deleted to 130

pRZ918

<table>
<thead>
<tr>
<th>130</th>
<th>150</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAC TGG GCA ATT AAC TCT GTC TGC CGG GCG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pRZ919</td>
</tr>
<tr>
<td></td>
<td>pRZ920</td>
</tr>
</tbody>
</table>

Figure 1. Location of the fusion endpoints of IS50-lacZ constructs. The region surrounding the GTG at positions 138 to 140 is shown as triplets that align with the large open reading frame of IS50 (see the numbering in Fig. 3). The broken lines indicate the point at which the translational fusion to lacZ begins in each construct.
Table 1

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Fusion endpoint</th>
<th>β-Galactosidase levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRZ918</td>
<td>T (131)†</td>
<td>40</td>
</tr>
<tr>
<td>pRZ919</td>
<td>C (143)</td>
<td>76</td>
</tr>
<tr>
<td>pRZ920</td>
<td>G (149)</td>
<td>48</td>
</tr>
<tr>
<td>pMC1403</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

β-Galactosidase assays were conducted on cells (RZ211) carrying the indicated plasmids using the method of Miller (1972).

† The base represents the 3' boundary (relative to the direction of transcription) of IS50 sequence remaining in the fusion constructs. The number in parentheses refers to the position of this base using the numbering of Fig. 3.

GTA, expressed p1-β-galactosidase at the same level as cells carrying the wild-type fusion construct (R. Johnson, unpublished results). These results suggested that the initiation codon was either encoded upstream from the GTG at positions 138 to 140 (at the ATG at positions 81 to 83 or the ATG at positions 93 to 95) or that more than one start site was used.

To distinguish among these possibilities, we sequenced the amino terminus of a p1-β-galactosidase fusion protein by automated Edman degradation. The protein was purified from cells carrying an IS50-lacZ construct (pRZ903) by immunoaffinity chromatography and preparative gel electrophoresis. The PTH-amino acid derivatives of the first five cycles (Table 2) corresponded to the predicted amino acid sequence starting at the ATG at positions 93 to 95.

The sequence data also indicated that different amino termini were not present, since the yield of PTH-amino acids at cycles 2 and 5 (Table 2) was close to the yield of the first cycle. If other sequenceable amino termini were present, the yield of each cycle would be distributed among several PTH-amino acid derivatives. A consistent sequence was obtained using a different method of preparing the fusion protein (data not shown).

Table 2

<table>
<thead>
<tr>
<th>Cycle</th>
<th>PTH-amino acid</th>
<th>Yield (pmol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Met</td>
<td>48</td>
</tr>
<tr>
<td>2</td>
<td>Ile</td>
<td>47</td>
</tr>
<tr>
<td>3</td>
<td>Thr</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>Ser</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>Ala</td>
<td>40</td>
</tr>
</tbody>
</table>

The yields shown were determined by comparing the area of each PTH-amino acid peak with the area of a peak corresponding to a known quantity of PTH-amino acid. The yield of the 1st cycle represented 21% of the expected yield, based on the amount of fusion protein loaded on the protein sequencer.

Figure 2. S1 nuclease mapping of transcripts from intact IS50R. RNA from cells (RZ211) carrying pRZ204 was hybridized to a HindIII-HincII fragment from IS50 labeled at the HincII site, treated with S1 nuclease and electrophoresed on an 8% polyacrylamide/7 M-urea gel. (a) Products of the G and (b) the A > C Maxam & Gilbert (1980) sequencing reactions using the labeled fragment. (c) The products of S1 nuclease mapping. The diagram shows the outer end of IS50 with inward transcripts (wavy lines) and the hybridization probe with the labeled end marked with an asterisk. Numbers refer to the positions of the 5' ends of the transcripts.

(b) Location of the 5' termini of IS50 RNA

The amino terminus of the p1-β-galactosidase fusion protein was found to be located upstream from the start site of the only IS50 RNA transcript observed in vitro (Johnson & Reznikoff, 1981). We searched for additional transcripts that could encode this protein by mapping the 5' ends of RNA produced in vivo, using S1 nuclease. Our analysis was limited to "inward" transcripts starting near the outer end of IS50. Total RNA prepared from cells carrying IS50R on a high copy plasmid (pRZ204) was hybridized to a 5' end-labeled probe covering the region shown in Figure 2. The hybridized samples were treated with S1 nuclease and electrophoresed along with the products of sequencing reactions of the hybridization probe. An autoradiograph of the electrophoresed products is shown in Figure 2.
Several groups of protected fragments were observed. The major group centered around a fragment that comigrated with a 90-base product of the Maxam & Gilbert (1980) sequencing reaction (Fig. 2, lane c). Since the hybridization probe was end labeled at position 188, this fragment corresponded to an RNA species having a 5' end at position 98. A transcript starting at this site was observed in vitro (Johnson & Reznikoff, 1981). A second, weak group of protected fragments that was mapped by similar analysis corresponded to RNA with a 5' end around position 66. This group of fragments represented approximately 1% of the total protected fragments, as determined by densitometric scanning. A mRNA transcript starting at position 66 would be capable of encoding pl. Other protected fragments may be artifacts of the S1 nuclease treatment or may represent additional RNA species produced in vitro (see below).

(c) There are three "inward" promoters in the outer end of IS50

We wanted to define the regions of IS50 responsible for the production of RNA, and to determine if the RNA species found by S1 mapping represented discrete transcripts. We were especially concerned that the RNA species starting at position 66 might not correspond to a transcriptional start site, because it was present in such small amounts. We constructed a series of deletions of the outer end of IS50 (Fig. 3). The plasmid used to make the deletions contained a translational fusion of lacZ to IS50 for monitoring protein expression (see below). As determined by sequencing, the deletions spanned a region starting outside IS50 and ending just upstream from the pl translational start site (Fig. 3). An autoradiograph of the dideoxy sequencing analysis of one deletion is presented (Fig. 4) to show an additional nucleotide, T (position 60), which was not present in a previously published IS50 sequence (Auerswald et al., 1980). We found this difference in both IS50L and IS50R (data not shown).

We analyzed the mRNA transcripts produced in strains harboring these deletion constructs using S1 nuclease mapping. The analysis was identical with that used above, except that the hybridization probe was 5' end-labeled at position 155 and covered the region shown in Figure 5. The pattern of protected fragments was similar to that of experiments performed using intact IS50.
indicating that the presence of the lacZ gene did not interfere with the production of the transcripts.

As shown in Figure 5, the deletions defined three regions that were correlated with the appearance of protected fragments in S1 mapping. Deletions that did not remove IS50 sequence past position 28 (pRZ954, pRZ955, pRZ957 and pRZ958) yielded the same pattern of protected fragments as intact IS50 (Fig. 5, lanes a to d). Deletions that extended past position 28 (pRZ962, pRZ963 and pRZ964) resulted in the loss of protected fragments corresponding to an RNA transcript (T1) starting at position 66 (Fig. 5, lanes e to i). The deletions that passed position 56 (pRZ963 and pRZ964) resulted in the loss of protected fragments corresponding to an RNA transcript (T2) starting at position 98 (Fig. 5, lanes h and i). Finally, a deletion that extended past position 73 (pRZ964) resulted in the loss of protected fragments corresponding to an RNA transcript (T3) starting at position 109 (Fig. 5, lane i). The effect of deleting these regions is consistent with the inactivation of a transcriptional promoter corresponding to each of these transcripts. The consensus E. coli promoter sequence can include sequences located as far as 38 bp 5' from the transcriptional start site (Hawley & McClure, 1983). The maximal extent of the promoter region was thus predicted to be positions 28, 60 and 71 for the RNA transcripts T1, T2 and T3, respectively. This prediction agreed well with the endpoints of the deletions that first abolished the RNA transcripts.

In addition to the previously identified transcripts, an RNA species that originated in the vector and continued through the IS50 sequence was apparent in each deletion. The size of the fragment protected by this read-through transcript decreased as homology between the hybridization probe and the RNA decreased in subsequent deletions. The amount of read-through transcript appeared to be at least as great as the amount of T1 transcript. The read-through transcript could have
been initiated at the P4 promoter of pBR322 (Stuber & Bujard, 1981), which was present in all of the deletion constructs. In one case, pRZ963, the protected fragments arising from read-through transcription partially obscured the region corresponding to T1, so it was not possible to determine if T1 was expressed (Fig. 5, lane b). We believe that T1 is not produced from pRZ963, since deletions that flank this construct (pRZ962 and pRZ964) lack protected fragments corresponding to T1 (Fig. 5, lanes g and i).

(d) Expression of IS50 proteins from deletion plasmids

As a means of confirming the S1 analysis and determining the coding capacity of IS50 transcripts, we decided to monitor the production of fusion proteins by strains containing the IS50 deletions. As mentioned above, the plasmid used to construct the deletions contained a translational fusion of lacZ to IS50, so the expression of p1- and p2-β-galactosidase fusion proteins could be analyzed. Total cell protein from cells harboring the IS50 deletion plasmids was electrophoresed, transferred to nitrocellulose filter paper, and probed with antibody against the β-galactosidase portion of the fusion proteins. The results are shown in Figure 6.

The presence of the p1- and p2-β-galactosidase fusion proteins was correlated with the presence of the RNA species analyzed by S1 mapping. Cells harboring plasmids that express the T1, T2 and T3 transcripts (pRZ954, pRZ955, pRZ957 or pRZ958) accumulate both the p1- and p2-β-galactosidase fusion proteins (Fig. 6, lanes a to d). Cells carrying deletion constructs that did not express T1 generally did not accumulate the p1-β-galactosidase fusion protein. This was the case for cells carrying plasmids pRZ958, pRZ960 or pRZ962 (Fig. 6, lanes e to g). It is interesting to note that, although these plasmids encoded a vector transcript that continued through the outer end of IS50 and was present in amounts similar to T1, the p1-β-galactosidase fusion protein was not expressed. We believe that transcripts that originate outside of IS50 cannot encode the fusion protein because of RNA secondary structure present in such transcripts, but absent from T1 (see Discussion). This putative secondary structure would encompass the translation initiation region of p1 in transcripts starting outside IS50, but would not mask this region in T1 (Fig. 7). The deletions that did not encode T1 left the region encoding this secondary structure intact (pRZ969 and pRZ960) or removed only a small portion of the hairpin stem (pRZ962). Cells carrying these constructs did not express p1-β-galactosidase.

An important exception was plasmid pRZ963, which did not encode T1 (see above), but did express the fusion protein, although at lower levels than in intact IS50 (Fig. 6, lane h). In this deletion construct, read-through transcripts would lack the proposed secondary structure, since the deletion resulted in replacement of half of the region encoding this structure with sequence from pBR322 (Fig. 7). These results suggested that the proposed secondary structure inhibited expression of p1 from transcripts that started outside IS50.

The final deletion we considered, pRZ964, encoded very little p1-β-galactosidase (Fig. 6, lane i). Although cells carrying this plasmid did not express T1, they did produce a read-through transcript that would have lacked the secondary structure discussed above, and should have allowed expression of p1-β-galactosidase. This deletion probably lacked adequate translational initiation signals for p1-β-galactosidase expression, since the region 4 bp upstream from the initiation triplet had been replaced with pBR322 sequence.

Cells containing plasmids that expressed large amounts of T2 (pRZ954, pRZ955, pRZ957, pRZ958, pRZ959, pRZ960 and pRZ962) expressed p2 β-galactosidase at high levels (Fig. 6, lane a to g). When the expression of T2 was substantially less, as in cells harboring pRZ963, the level of
p2-β-galactosidase decreased (Fig. 6, lane h), suggesting that T2 was required in part for expression of this fusion protein. The decrease in the amounts of p2-β-galactosidase in cells carrying pRZ963 was not of the same degree as the decrease in transcript levels, however. This indicated that transcripts other than T2, either T3 or the read-through transcript, were contributing to residual expression of p2-β-galactosidase. In cells carrying pRZ964, which encoded very little of T2 or T3, the expression of p2-β-galactosidase was further reduced but still detectable (Fig. 6, lane i). This suggested that the read-through transcript was capable of encoding a low level of p2-β-galactosidase.

Some of the additional proteins that reacted with anti-β-galactosidase antibodies were probably proteolytic fragments of the IS50- lacZ fusion proteins. The most obvious example of this was a polypeptide that comigrated with β-galactosidase and was present at high levels only in the strains that accumulated the pl-β-galactosidase fusion protein. Densitometric scanning of lane a of the blot in Figure 6 indicated that the ratio of p2- to pl-β-galactosidase was 1.7:1 if this band was included in the total for the pl fusion protein, but 4:1 if it was omitted. The origin of other detectable proteins was less clear; some, such as the band marked by an asterisk in Figure 6, may correspond to previously observed fusion products (Johnson et al., 1982; Johnson & Reznikoff, 1984). Whether these proteins represented independent translational starts or were degradation products of the fusion proteins was not determined.

4. Discussion

We have examined regions in the outer end of IS50 that are responsible for the expression of IS50 proteins. Our understanding of the genetic organization of the outer end of IS50 is presented in Figure 3. Although we have limited our discussion to pl, p2 and IS50R, the conclusions can be extended to p3, p4 and IS50L, since the two insertion sequences are identical over much of their length (Auerswald et al., 1980; Rothstein & Reznikoff, 1981).

Our lacZ fusion and amino-terminal sequencing data suggest that the translational start site of the Tn5 transposase is encoded by the ATG at positions 93 to 95. The amino-terminal methionine residue determined by sequencing could represent either the first amino acid of a protein starting at AUG (positions 93 to 95) or the fifth amino acid of a protein starting at AUG (positions 81 to 83), which is then processed to yield p1-β-galactosidase. We consider it unlikely that processing of a protein starting at the upstream initiation codon would result in a protein with an amino terminus that coincided with the downstream start site. This assignment of the p1 start site agrees with an earlier prediction, which was based on the nucleotide sequence (Auerswald et al., 1980).

By S1 nuclease mapping of the transcripts encoded by IS50 deletion derivatives, we were able to identify three "inward" transcripts, T1, T2 and T3, located near the outer end of IS50. We believe that these RNA species represent discrete transcripts, because deletions that did not express T1 could express T2 and T3, and deletions that lacked T1 and T2 could still express T3. So, the RNA species do not represent processing of a larger precursor transcript. In addition, the appearance of each transcript required a region located at least 21 nucleotides but no more than 42 nucleotides upstream from its 5' terminus. This is consistent with the existence of an independent promoter for each transcript.

The relative amounts of the transcripts differ greatly. T2 is present at much higher levels than T1 and T3, which are present in nearly equal amounts. This may indicate a difference in the activity of the three promoters, although we cannot exclude the possibility that the observed levels reflect transcript stability and not transcriptional initiation. However, the sequences of the three promoters are consistent with a difference in their transcriptional activity. The region upstream from the 5' end of each transcript can be aligned with the consensus E. coli promoter sequence (Hawley & McClure, 1983), as shown in Figure 3. The criteria used to obtain these matches were the same as those used by Hawley & McClure (1983). While the −10 regions of the proposed promoter sequences generally match the consensus at highly conserved positions, the −35 regions have few significant homologies. The −35 regions must be critical for the activity of these promoters, however, because transcript accumulation drops significantly when they are deleted. The −10 region of the T1 promoter has C rather than A at a highly conserved position, which may result in lower promoter activity. In addition, this region contains two recognition sites for the dam methylase of E. coli, and a recognition site for the dam methylase is located just upstream from the −10 region.

Methylation of adenine residues by the dam gene product within the −10 region of the Tn10 pIN promoter lowers its transcriptional activity in vitro (Roberts et al., 1985). Similarly, the Tn5 T1 transcript accumulates to higher levels in a dam strain than in a wild-type strain (J. Yin, unpublished results). Methylation of adenine residues within the T1 promoter may partially account for the small amounts of T1. The T2 and T3 promoters match the consensus more closely in both the −10 and −35 regions, and might be expected to have a greater transcriptional activity. However, the T3 promoter has a suboptimal spacing of 15 bp between the conserved −10 and −35 hexamers, whereas the T1 and T2 promoters have a more commonly observed spacing of 17 and 18 bp, respectively (Hawley & McClure, 1983). The spacing of the −10 and −35 regions of the T3 promoter may contribute to its low level of activity. Finally, since the three promoters overlap, the exclusion of
RNA polymerase from both the T1 and the T3 promoters by polymerase complexes at the T9 promoter might further reduce the activity of the T1 and T3 promoters.

The transcripts have different coding capacities. T1 is the only IS50 transcript that can encode transposase. Its 5' end is located upstream of the translational start site of p1, and deletions that result in the loss of T1 do not express p1-β-galactosidase. This observation confirms that the p1 translational start site does not correspond to the GTG at positions 138 to 140. The contribution of T1 to p2 levels is minimal, as cells carrying deletions that produce T2, but not T1, express p2-β-galactosidase at the same levels as intact IS50-lacZ. Both T2 and T3 may contribute to expression of p2, although the contribution of T2 appears greatest, as deletions that decrease the amount of T2 have the most significant effect on p2 levels.

The finding of multiple "inward" promoters in IS50 agrees, in part, with previous studies. An XbaI linker inserted at position 52, which resulted in the loss of p1 and decreased levels of p2, was thought to inactivate one of two IS50 promoters (Isberg et al., 1982). Although our results suggest that this mutation should inactivate the T1 promoter and thus reduce p1 expression, they cannot account for the observed decrease in p2 levels. This study also confirms the existence of a transcriptional start site at position 98 (T2) found in vitro (Johnson & Reznikoff, 1981). The T1 transcript was not found in vitro, perhaps as a result of the low level of activity of the T1 promoter.

Our results show a striking disparity in the relative amounts of T1 and T2 and the amounts of protein these transcripts encode. If we assume that the approximately 100-fold difference in the amounts of S1-protected fragments corresponding to T1 and T2 is the same as the difference in levels of functional transcript, then the mRNA capable of encoding p1 is about 100-fold less abundant than the mRNA that encodes p2. The relative amounts of the β-galactosidase derivatives of these two proteins differ by less than the amounts of mRNA that encode them, however (compare Figs 5 and 6). Previous estimates of the p2-p1-β-galactosidase ratio range from 2:1 (Isberg et al., 1982) to 4:1 (Johnson et al., 1982). Densitometric scanning of the immunoblot shown in Figure 5 (lane a) indicates that the p2 to p1 fusion protein ratio is about 2:1. A possible explanation for the discrepancy between protein and transcript levels is that the translational initiation frequency of p1 is 25 to 50-fold higher than that of p2. However, there are few distinguishing features at the level of primary structure that might explain the difference between transcript and protein levels. The translational initiation regions are 5'-U-A-A-C-U-G-G-U-A-A-C-G-U-U-C-A-U-G-G-3' for p1 and 5'-A-A-C-G-G-U-A-G-G-G-A-G-A-C-C-A-U-G-G-3' for p2, where the underlined bases are complementary to the 5' end of 16S RNA (Shine & Dalgarno, 1974) and the overlined bases are the initiation codon (Auerswald et al., 1980; this work; J. Yin, unpublished results).

The initiation regions differ only slightly in spacing between the initiation codon and the Shine & Dalgarno (1974) complementarity. The prevalence of U in the p2 region is rare among translational initiation regions (Gold et al., 1981) and may result in a lower frequency of initiation. An alternative explanation for the disparity in the amounts of transcript and protein is that the transcripts and the proteins they encode might have different relative stabilities.

The transposase protein is not expressed from transcripts that originate outside of IS50. The lack of expression of p1 from read-through transcripts may be a result of secondary structure that is present in these transcripts but not in T1 (Fig. 7). The proposed secondary structure may reduce translational initiation, as has been shown for structures that involve the Shine & Dalgarno (1974) sequence, as in phage MS2 (Kastelein et al., 1983), the initiation codon, as in lacZ derivatives (Munson et al., 1984) or both, as in Tn10 (Davis et al., 1985).

Secondary structures that reduce translational initiation of read-through transcripts have been proposed for lacZ (Reznikoff et al., 1974; Yu et al., 1984), galE (Merril et al., 1981), and Tn10 (Davis et al., 1985). Alternatively, the region defined by deletion constructs pRZ962 and pRZ963 may serve

---

Figure 7. Possible secondary structure of read-through transcripts. The region transcribed from positions 40 to 100 of IS50 is shown as a hairpin structure that has a region that has a ΔG of -12.2 kcal (calculated using the energies given by Cech et al., 1983). A similar structure, which does not include the 1st 3 paired bases has a ΔG value of -9.4 kcal. The endpoints of 3 deletion constructs are indicated by broken lines. The start of transcript T1 (wavy line) and the p1 initiation codon (continuous outline) are shown. 1 kcal = 4.184 kJ.

---

Control Sites of IS50 Expression 789
as a signal that results in the termination of read-through transcripts (termination would occur downstream from position 187, the endpoint of the probe used in our S1 analysis). Only transcripts that lack this region would be capable of encoding IS50 proteins. Some termination of read-through transcripts has been observed for Tn10 (Davis et al., 1985). In addition, the ability of Tn5 to exert polar effects upon genes it interrupts is well-documented (Berg et al., 1980), so the existence of a read-through-specific terminator would not be surprising.

The control of transposase expression in IS50 may be similar in part to that of other transposable elements and IS10 in particular. The promoter that drives transposase expression in IS10 contains a dam methylation site in its -10 region and its activity in vitro and in vivo is reduced when methylated (Roberts et al., 1985). The low level of expression of transposase from transcripts originating outside IS10 is believed to be predominantly a result of a secondary structure that masks the translational initiation region, although premature termination of read-through transcripts also may play a role (Halling et al., 1982; Davis et al., 1985).

The presence of a secondary structure that includes the translational initiation region of the transposase as well as the transcriptional start site for the transposase promoter appears to be a feature of other transposable elements, including IS3 (Timmerman & Tu, 1985), IS5 (Kroger & Hobom, 1982), and IS30 (Dalrymple et al., 1984).

In the case of IS10, where small amounts of transposase limit the transposition rate (Morisato et al., 1983, Raleigh & Kleckner, 1986), the control of transposase expression is important in determining the level of transposition. Thus, transposition of IS10 is increased in dam- cells because of increased transposase expression (Roberts et al., 1985). The frequency of Tn5 transposition increases in dam- cells (Roberts et al., 1985; J. Yin, unpublished results), perhaps as a result of increased transposase expression. In addition, Tn10 transposition is not activated by the presence of strong promoters outside the element, partially because translation of read-through transcripts is inefficient and transposase amounts do not increase (Davis et al., 1985). It is possible that such mechanisms apply also to Tn5 and IS50. We believe, however, that the control of Tn5 transposition will prove more complex because of overlapping promoters that independently control expression of the transposase and inhibitor proteins.

We thank Bob Mirendorf for help with protein purification, Ron Niecz for protein sequencing, and Frank Pugh and Brian Shutte for their gifts of Kl one and T4 DNA ligase. We thank Caroline Donnelly, Paul Lambert, and Jerry Yin for useful discussions. This work was supported by the National Institutes of Health grant GM19670 to W.S.R. M.P.K. was supported, in part, by the National Institutes of Health training grant GM07215.

References


Edited by I. Herskowitz.