Localization of the Tn5 transposase promoter using the cycling reaction of RNA polymerase

Reid C. Johnson and William S. Reznikoff

Department of Biochemistry, College of Agricultural and Life Sciences, University of Wisconsin,
Madison, WI 53706, USA

Received 14 January 1981

ABSTRACT

RNA polymerase is known to undergo a cycling process during initiation of transcription in vitro in which large amounts of small oligonucleotides are released. We have used this cycling reaction to determine the 5' end of the RNA synthesized from the outer ends of the Tn5 inverted repeats. By analyzing the size of the radiolabeled oligonucleotides synthesized using different labeled nucleoside triphosphates and in reactions deficient for a particular nucleoside triphosphate, the partial 5' sequence was obtained. This sequence was correlated with the DNA sequence of the region and an unambiguous origin for the mRNA was determined. The start site for the RNA, which is located at 97 base pairs from the outer ends of the inverted repeats, was confirmed by digesting γ -³²P-ATP labeled full length (run off) transcripts with ribonuclease T₁. The resulting γ -labeled T₁ generated oligonucleotide corresponded to the predicted size determined using the cycling reaction. With the knowledge of the RNA start site, the probable translation start sites for the protein(s) known to be required for Tn5 transposition can be predicted. Possible implications of the DNA sequence with respect to the regulation of the transposase are also discussed.

INTRODUCTION

Tn5 is a 5700 base pair transposon with 1534 base pair inverted repeats flanking a unique central region encoding resistance to certain aminoglycoside antibiotics, including neomycin and kanamycin (1,2,3). Previous studies have defined five proteins synthesized by the Tn5 element (4,5). The smallest of these proteins is the 25K Dalton neomycin phosphotransferase II enzyme which is responsible for conferring antibiotic resistance. While this protein is thought to be coded entirely within the central region of the transposon (3,5), the promoter responsible for its synthesis is located within the left inverted repeat (5,6). Two proteins differing in approximately 4000 Daltons at their NH₂ termini, have been found to be synthesized from each repeat (4,5). The proteins from the left repeat are 26 amino acids smaller than those of the right repeat due to an ochre codon located 1442 base pairs from

the outer end of the left repeat not present in the right repeat (3,5,6). One or both of the proteins from the right repeat, which migrate at approximately 58K and 54K MW on SDS - polyacrylamide gels, are crucial for transposition. Insertions or deletions within the coding sequence for these proteins eliminate the ability of the element to catalyze its own transposition (4,5). Evidence for a promoter located at each end of the transposon which could be involved in the synthesis of the repeat proteins was found by Rothstein *et al* (4). We decided to further characterize this promoter reasoning that it would give some information as to the synthesis and regulation of the transposase protein(s).

It was found that *in vitro* transcription of the Tn5 repeat promoter generated a large number of small oligonucleotides (this paper and 7). These oligonucleotides are generated by the cycling of the RNA polymerase which results in the release of small RNA transcripts followed by re-initiation without leaving the DNA template (7,8,9). This cycling reaction has been studied in some detail for the *lac*L8UV5 promoter (7,9) where it has been found that at least ten oligonucleotides are generated per full length product, leading to a rather long half-time for full length transcription (7). The Tn5 promoter has been shown to cycle at a four-fold higher rate than *lac*, resulting in a significantly longer half-time for full length transcription (7). Other promoters, including those in phage λ and T7, have also been shown to undergo this cycling reaction when one or more nucleoside triphosphates are absent from the elongation reaction. This forced cycling by RNA polymerase has been termed abortive initiation (8,10).

MATERIALS AND METHODS

Preparation of template and in vitro transcription.

The 250 base pair EcoRI-Hpa I restriction fragment containing the left end of Tn5 was used as the DNA template in most reactions. pRZ102 (ColE1-Tn5) or pRZ112 (pRZ102 deleted for the right half of Tn5) (11) was digested with EcoRI (gift of J. Gardner) and Hpa I (Bethesda Research Labs, Inc.) and electrophoresed on a 5% polyacrylamide 10% glycerol tube gel. The resulting fragment was visualized by ethidium bromide staining, excized, and eluted as described by Maxam and Gilbert (12). The DNA was further purified by phenol followed by ether extraction and finally ethanol precipitated and resuspended in a small volume of H₂O.

0.1 μ g (6 picomoles) of fragment was incubated with 0.5 μ g (10 picomoles) of purified RNA polymerase (gift of L. Maquat and S. Rothstein) in a 20 μ l

reaction volume in the buffer described by Majors (13). After 15 minutes at 37°, heparin (Sigma) was added to a final concentration of 100 µg/ml, and the incubation was continued for 1 minute to allow for dissociation of any loosely bound complexes. Transcription was initiated by the addition of nucleoside triphosphates and was terminated after 15 minutes at 37° by the addition of an equal volume of transcription stop buffer (.1M EDTA, 0.1% SDS, 7 M Urea, 89 mM Tris - borate, pH 8.3, xylene cyanol, and bromophenol blue). In standard reactions, the nucleoside triphosphate concentrations were 300 µM for ATP and GTP and 30 µM for the unlabeled pyrimidine nucleoside triphosphate and 15 µM for the labeled pyrimidine nucleoside triphosphate. 25 µCi of either α -³²P-CTP or α -³²P-UTP (Amersham, > 350 Ci/mM) was used as the labeled nucleotide. For reactions using γ -³²P-ATP as the labeled nucleotide, the unlabeled ATP concentration was lowered to 150 µM and 100 µCi of γ -³²P-ATP (Amersham, >2000 Ci/mM) was added. Unlabeled nucleoside triphosphates were purchased from PL Biochemicals Inc. and were not further purified.

The samples were boiled for 1 minute and electrophoresed in a 15 cm 8% polyacrylamide - 7 M urea gel (14) when full length transcripts were to be detected or in a 40 cm 20% polyacrylamide - 7 M urea gel (12) to resolve oligonucleotides. The labeled bands were visualized by autoradiography. Ribonuclease T1 digestion. RNAs were eluted from gels by soaking in gel elution buffer (12) plus 10 µg tRNA overnight at 37° followed by filtration through siliconized glass wool and ethanol precipitation. The isolated RNA was digested with 20 units of ribonuclease T1 (Worthington Biochemicals) for 30 minutes at 37° in 89 mM Tris - borate (pH8.3), 3 mM EDTA, 7 M urea.

RESULTS

Filter binding and in vitro transcription studies by Rothstein et al (4) had previously located the approximate position of a promoter at the ends of Tn5 which transcribed in the direction of the central region of the transposon. The EcoRI - Hpa I restriction fragment which contains 185 base pairs of the left end of Tn5 (as well as the analogous Pst I - Hpa I restriction fragment from the right end of Tn5) was filter bound in the presence of RNA polymerase and was found to program the synthesis of RNA. Polyacrylamide gel electrophoresis resolved two RNAs differing in length by about 10 bases which were estimated to originate approximately 100 base pairs from the Hpa I site. The amount of the two RNAs varied with respect to each other depending on the reaction conditions (e. g. template used, Mg⁺² concentration, and time of elongation) (4, Johnson, Rothstein, and Reznikoff, unpublished). We now

believe that the shorter RNA is due to pausing or premature termination at the 3' end of the RNA as has been observed for the lac promoter (13) and not the result of two unique start sites. In order to further define the promoter at the ends of Tn5, we determined the exact start site of the mRNA from this region utilizing the abortive initiation reaction of RNA polymerase.

The initiating nucleoside triphosphate was determined by varying the levels of nucleotides in the transcription reaction since the initiating nucleotide is required in high concentration for many promoters (15,16). Less than 5% of the standard amount of full length product was obtained when the ATP concentration was 10 μ M as compared with the amount of full length product obtained in the standard reaction using 300 μ M ATP, indicating that transcription is initiated with ATP. No effect in the amount of full length product was seen when the GTP concentration was lowered to 10 μ M. Initiation with ATP was subsequently confirmed by isolating full length product labelled with γ -³²P-ATP.

Figure 1 is an autoradiogram of the products obtained when the RNA is synthesized in the presence of α -³²P-UTP or α -³²P-CTP and is electrophoresed on a 20% polyacrylamide-7 M urea gel. Large quantities of oligonucleotides, up to 8 nucleotides in length, are found which are generated by the cycling reaction of RNA polymerase at the promoter. The sizes of the products generated from the Tn5 promoter were determined by electrophoresing in parallel the oligonucleotides synthesized from the lac L8UV5 promoter whose sizes have been confirmed by homochromatography (7). (Since the migration of short RNAs is sequence dependent in these gels, the migration of equivalent size bands is not identical).

The length of the smallest labeled oligonucleotide corresponds to that position in which the radioactive nucleoside triphosphate is first incorporated. Thus, when α -³²P-CTP is used as the labeled nucleotide, the first band present is a trimer (Fig. 1d), and when α -³²P-UTP is the labeled nucleotide, the smallest radioactive oligonucleotide is a tetramer (Fig. 1a). These results imply that the third nucleotide in the RNA sequence is a C and the fourth nucleotide is a U.

Similar experiments, where particular nucleotides are left out of the reaction, can be performed to further define the RNA sequence. In Fig. 1b and e, transcription was performed in the absence of GTP. In these cases neither octamer nor full length product was synthesized. Therefore, GTP must normally be incorporated at the eighth position in the RNA. The fact that a G was not present in the RNA sequence before the eighth position was confirmed by

Figure 1 Autoradiogram of a 20% polyacrylamide -7 M urea gel. In lanes a-c, RNA was synthesized in the presence of α -³²P-UTP in reactions containing a) ATP, GTP, TPT, CTP b) ATP, UTP, CTP c) ATP, GTP, UTP and in lanes d-f, RNA was synthesized in the presence of α -³²P-CTP in reactions containing d) ATP, GTP, UTP, CTP e) ATP, UTP, CTP f) ATP, GTP, CTP. Lane g is the oligonucleotides produced from the Hae III 203 base pair lac L8UV5 promoter containing restriction fragment in the presence of 300 μ M ATP, 30 μ M GTP, 15 μ M UTP and 25 μ Ci α -³²P-UTP. The sizes of the RNAs are identified for Tn5 and lac.

isolating the aborted octamer and digesting with ribonuclease T_1 (Fig 2b). No significant digestion by the G specific enzyme was observed.

UTP was absent from the α - 32 P-CTP labeled reaction shown in Fig. 1f. This resulted in the synthesis of only the trimer, indicating that the absence of UTP is preventing elongation past the third position. This result confirms α - 32 P-UTP reactions which indicated a U at the fourth position (see above). Likewise, since CTP was found to be incorporated before UTP (see above), no incorporation of α - 32 P-UTP would be expected if CTP was not present in the reaction. As shown in Fig. 1c, only a small amount of products migrating at the fourth and fifth position are synthesized in transcription reactions lacking CTP, possibly due to trace amounts of contaminating CTP in the labeled or unlabeled nucleotides which allows read-through past the first C. Alternatively, the 4-mer and possibly the 5-mer may have arisen from a secondary site since the migration of the 4-mer is not identical with the 4-mer in the other Tn5 lanes.

These experiments predict a partial RNA sequence of pppApXpCpUpXpXpXpG, where X is unknown. Position 2 can be inferred to be an A since the first sites of incorporation of C, U, and G were found to be at 3,4, and 8, respectively. When the DNA sequence of the region as determined by Auerswald and Schaller (3) was scanned, the sequence beginning at 97 base pairs from the left end of Tn5 matched the partial RNA sequence (see Fig. 3). No other Tn5 sequence in the DNA template used (positions 1-185) was compatible with this RNA sequence.

An additional experiment was performed to confirm that the RNA sequence as determined by analyzing the products generated by the cycling reaction of RNA polymerase is indicative of full length transcription and not an artifact of the cycling mechanism. γ - 32 P-ATP full length (runoff) RNA was isolated from an 8% polyacrylamide-7 M urea gel after using the EcoRI-Hpa I restriction fragment as template and digested with ribonuclease T_1 . As shown in Fig. 2c the resulting product migrated slightly faster than the aborted octamer but was larger than the aborted heptamer. The faster migration is presumably due to the 3' PO_4 present on the T_1 generated oligonucleotide as opposed to a 3' OH on the aborted octamer. Thus, the full length RNA contains a G at its eighth position which corresponds to the predicted sequence.

DISCUSSION

In Fig. 3 we present the DNA sequence around the RNA initiation site for

Figure 2 Autoradiogram of a 20% polyacrylamide -7 M urea gel. Lanes a and b contain isolated octamer before (a) and after (b) ribonuclease T_1 digestion. Lane c is isolated γ - 32 P-ATP labelled full length product synthesized from the EcoRI - Hpa I fragment of Tn5 followed by digestion with ribonuclease T_1 . Lane d is the RNA synthesized from the Tn5 promoter in the presence of all four nucleoside triphosphates and α - 32 P-UTP.

Figure 3 The Tn5 IR promoter sequence. The antisense strand of the DNA sequence corresponding to bases 61 to 108 from the outer ends of the Tn5 inverted repeats (3) is presented. The sequence is numbered with respect to the initiating nucleoside triphosphate (+1). Similarities to the model promoter sequence of Rosenberg and Court (17), which is shown on the top line, are underlined. On the bottom line is the RNA sequence from +1 to +12 with the predicted ribonucleotides from the analysis of the oligonucleotides generated by the cycling reaction of RNA polymerase circled. Also, on the bottom line is the first eight bases from the outer and inner ends of the Tn5 inverted repeat sequences showing the homology with the -36 to -41 region of the promoter.

the ends of Tn5. Since the sequence for the left and right ends of Tn5 are identical in this region (3) and the promoters at either end are indistinguishable with respect to RNA polymerase binding and the size of transcripts synthesized (6), we presume that this sequence promotes transcription from either repeat. While the Tn5 repeat promoter does not fit the model promoter sequence of Rosenberg and Court (17) particularly well, the highly conserved base pairs at -8, -12, and -13 in the Pribnow Box are present as well as two of the three highly conserved base pairs in the -35 region located at -35 and -36 with respect to the initiating nucleoside triphosphate.

One feature of the Tn5 repeat promoters is the striking region of homology with the outer ends of the element located in the -35 region (see Fig. 3). The six base pair sequence extending from -36 to -41 matches exactly the outer six base pairs of the Tn5 element. An inverted homologous sequence, containing an eight out of nine base pair match with the outer ends of the element is also present at the inner ends of the inverted repeat segments (Fig. 3) (3, D. Berg, personal communication). Thus, CTG^A_TCTCTT is the sequence of both ends of the inverted repeat (IS50 element) which has been shown to be capable of independent transposition (18). One could speculate that the six base pair homologous sequence in the promoter region might be responsible for mediating transcriptional control of the transposase promoter. The transposase protein synthesized from this promoter is presumably involved in the recognition of the ends of Tn5 or IS50 and thus may also bind in the -35 region of its promoter.

Figure 4 The genetic organization of Tn5. The start and stop sites for the RNAs and proteins synthesized from the inverted repeats are given where appropriate. Also shown is the single base pair difference between the inverted repeats located at position 1442, which generates a promoter and a translation stop codon in the left repeat.

Such a mechanism could serve in autoregulating the synthesis of the transposase by preventing RNA polymerase from binding to the promoter in the presence of the protein. This model may partially explain the zygotic induction of Tn5 transposition observed by Biek and Roth (19) as well as the low level of synthesis of the repeat proteins as assayed in minicells (4,5) even though the promoter is relatively active in vitro. Experiments are currently underway to test this hypothesis.

The sequence of the Tn5 inverted repeats indicates a single open reading frame from 71 base pairs from the end of the repeat to position 1441 for the left repeat and position 1519 for the right repeat (3). The first in-frame initiating codon present in the RNA occurs at position 137 (+41 on the RNA). This GUG is preceded by a possible ribosome binding site and would generate a protein whose size is consistent with that observed in minicells (4,5). In addition, a 4000 Dalton smaller protein is also synthesized from each repeat. The start site of this protein must occur after the Hpa I site (position 185) since fusions made at the Hpa I site leave this protein structurally unaltered (4,5). An in-frame AUG preceded by a possible ribosome binding site can be found at position 257 (+161 on the RNA) generating a protein which is 4833 Daltons smaller than the larger protein, consistent with the observed migration on SDS - polyacrylamide gels. The larger of these proteins, possibly together with the smaller protein, as synthesized from the right repeat has been shown to be absolutely required for transposition, and thus presumably the right repeat protein(s) encodes the Tn5 transposase activity (4,5). Fig. 4 summarizes our current understanding of the genetic organization of Tn5.

The use of the cycling reaction of RNA polymerase to define the start site of the RNA should may be applicable to other systems where the promoter has been approximately located and the DNA sequence is known. This procedure could be especially useful for weak promoters, where sequencing by more traditional methods is difficult, since the aborted RNAs are synthesized in extremely high amounts, although care must be taken to insure that the oligonucleotides synthesized are promoter specific. Munson and Reznikoff (7) have estimated that RNA polymerase generates at least 10 oligonucleotides per full length transcript at the lac L8UV5 promoter and at least 40 oligonucleotides per full length transcript at the Tn5 promoter. In the case of Tn5, oligonucleotides at 2 and 3 were not quantitated in this study so this number is clearly an underestimation of the number of abortive events. Cech et al (10) have also used the abortive initiation reaction to identify and quantitate the early promoters of phage T7.

ACKNOWLEDGEMENTS

We would like to thank Lianna Munson, Steven Rothstein, and Jerry Yin for helpful suggestions during the course of this work.

We are also grateful to Dr. Heinz Schaller for providing us with the DNA sequence prior to publication. This work was supported by grant GM19670 from the National Institutes of Health and grant PCM7910686 from the National Science Foundation. R.C.J. was supported by the National Institutes of Health training grant GM07215.

References

1. Berg, D. E., Davies, J., Allet, B. and Rochaix, J. D. (1975) Proc. Nat. Acad. Sci. U.S.A. 72, 3628-3632.
2. Berg, D. E. (1977) In DNA Insertion Elements, Plasmids and Episomes. Edited by Bukari, A. I., Shapiro, J. A. and Adhya, S. L. Cold Spring Harbor Press, Cold Spring Harbor, pp. 205-212.
3. Auerswald, E. A. and Schaller, H. (1981) Cold Spring Harbor Symp. Quant. Biol. 45 (in press)
4. Rothstein, S. J., Jorgensen R. A., Postle, K. and Reznikoff, W. S. (1980) Cell 19 795-805.
5. Rothstein, S. J., Jorgensen, R. A., Yin, J. C.P., Zang, Yongdi, Johnson, R. C. and Reznikoff, W. S. (1981) Cold Spring Harbor Symp. Quant. Biol. 45 (in press).
6. Rothstein S. J. and Reznikoff, W. S. (1980) Cell 23, 191-199.
7. Munson, L. M. and Reznikoff, W. S. (1981) Biochemistry (in press).
8. Johnston, D. E. and McClure, W. R. (1976) In RNA Polymerase. Edited by Losick, R. and Chamberlin, M. Cold Spring Harbor Press,
9. Carpousis, A. J. and Gralla, J. D. (1980) Biochemistry 19, 3245-3253.
10. Cech, C. L., Lichy, J., and McClure, W. R. (1980) J. Biol. Chem. 255, 1763-1766.
11. Jorgensen, R. A., Rothstein, S. J. and Reznikoff, W. S. (1979) Mol. Gen. Genet. 177, 65-72.
12. Maxam, A.M. and Gilbert, W. (1977) Proc. Nat. Acad. Sci. U.S.A. 74, 560-564.
13. Majors, J. (1975) Proc. Nat. Acad. Sci. U.S.A. 72, 4394-4398.
14. Gilbert, W. and Maxam, A. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 3581-3584.
15. Maizels, N. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 3585-3598.
16. Minkley, E. G. and Prihnow, D. (1973) J. Mol. Biol. 77, 255-277.
17. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genetics 13, 319-353.
18. Berg, D. E., Egner, C., Hirschel, B. J., Howard, J., Johnsrud, L., Jorgenson, R. A. and Tlsty, T. D. (1981) Cold Spring Harbor Symp. Quant. Biol. 45, (in press).
19. Biek, D., and Roth, J. R. (1980) Proc. Nat. Acad. Sci. U.S.A. 77, 6047-6051.