Insertion of transposon Tn7 into the *Escherichia coli* *glmS* transcriptional terminator

Nicholas J. GAY, Victor L. J. TYBULEWICZ and John E. WALKER*

M.R.C. Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

The transposon Tn7 is unusual as it transposes at high frequencies from episomal elements to a unique site in the *Escherichia coli* chromosome. This unique site is within a region of dyad symmetry that we have demonstrated to be the transcriptional terminator of the *glmS* gene which encodes the glutamine amidotransferase, glucosamine synthetase. Transposition of Tn7 abolishes termination of *glmS* transcription at this site; the transcripts now extend into the left end of Tn7 and terminate at a new site, t_m , 127 base pairs from the left end of Tn7. This region of the transposon contains a long open reading frame which encodes a protein sequence that is significantly related to the transposase proteins of the transposable elements IS1 and Tn3. A weak transcript has been identified that emanates from a promoter on the 5' side of this reading frame. This promoter is over-run by *glmS* transcripts and so it appears that expression of the Tn7 transposase may be regulated by promoter occlusion.

INTRODUCTION

The transposon Tn7 is a large transposable element, 14 kb in size, that encodes resistance to the antibiotics trimethoprim, spectinomycin and streptomycin (Barth *et al.*, 1976). It was first isolated as part of a conjugative R-plasmid which had spread rapidly through the populations of enteric bacteria as a consequence of the heavy use of antibiotics (Barth *et al.*, 1976). Tn7 is unusual amongst transposons as it transposes from episomal elements to a single site within the chromosome of both *E. coli*, *Caulobacter crescentus*, and most probably many other enteric bacteria, at the very high frequency of about 10^{-2} /cell per generation (Barth *et al.*, 1978; Fennewald & Shapiro, 1979; Ely, 1982). All characterized transpositions of Tn7 to the *E. coli* chromosome have been found at a single site and in a unique orientation near minute 82 of the *E. coli* chromosome. Tn7 can also transpose to secondary episomal sites at a lower frequency of 10^{-6} /cell per generation which is similar to the transposition frequency of other transposons, (Lichtenstein & Brenner, 1981). DNA sequence analysis of the unique chromosomal site of transposition and of a secondary site in the plasmid ColE1 showed in both cases that the position to which Tn7 transposes lies within sequences that have the potential to form a secondary structure. Transposition generates a 5 bp direct duplication at the target site. Sequence analysis of the two termini of the transposon revealed the presence of extensive imperfect inverted repeat sequences. (Lichtenstein & Brenner, 1982). Recently, more extensive and accurate DNA sequences in the vicinity of minute 82 of the *E. coli* chromosome has located the Tn7 unique site in the non-coding region between the genes *glmS* and *phoS* (Walker *et al.*, 1984). These essential genes encode the glutamine amidotransferase, glucosamine-6-phosphate synthetase, and a phosphate transport system (Surin *et al.*, 1984) respect-

ively. Here we report that the Tn7 unique site is located within the *glmS* transcriptional terminator. We also describe experiments concerning the effects of Tn7 transposition on termination of *glmS* transcription. In addition we suggest that a gene encoding a transposase is found in the left end of Tn7 and expression of this gene is regulated by promoter occlusion.

MATERIALS AND METHODS

DNA sequence analysis

This was performed by the dideoxy chain termination method (Sanger *et al.*, 1977) as modified by Biggin *et al.* (1983).

Preparation of RNA

Cultures of *E. coli* strains Q358 or GD92::Tn7 (25 ml) were grown at 37 °C in minimal glucose medium,

Table 1. Probes used for transcriptional mapping with S_1 nuclease

Probe	Start*	Length†
1	194	267
2	194	546
3	266	307
4	266	339
5	569	495
6	507	433

* Position in nucleotide sequence in Fig. 1(a) (probes 1 and 2) or Fig. 1(b) (probes 3–6).

† All probes extend leftwards in Fig. 1. Probes 1 and 2 had an extra 25 bases, and probes 3–6 an extra 48 bases derived from the universal M13 sequencing primer and sequences complementary to the polylinker of the mp series of M13 vectors (Messing & Vieira, 1982).

Abbreviations used: bp, base pairs; kb, kilobases (or kilobase pairs).

* To whom requests for reprints should be sent.

Fig. 1. DNA sequences at the 3' end of the *E. coli glmS* gene and the left end of transposon Tn7

(a) DNA sequence including a non-coding region between *glmS* and *phoS* in wild type *E. coli* with no inserted Tn7 (Walker *et al.*, 1984). The tandem stop codons of the *glmS* gene are indicated by asterisks. The *phoS* gene starts at base 326 (Surin *et al.*, 1984). Also shown by opposing discontinuous arrows are DNA sequences at the 3' end of *E. coli glmS* forming a region of dyad symmetry followed by a run of T residues. This is the *glmS* transcriptional terminator. Transcripts of *glmS* shown by a continuous arrow in (a) terminate at base 40. Tn7 is shown inserted between nucleotides 37 and 38. (b) DNA sequence determined in the *E. coli* strain GD92::Tn7 in which Tn7 has been inserted into the site indicated above in (a). Thus, nucleotides 38 onwards are the sequence of the left end of Tn7. Now transcripts of *glmS* (continuous arrow) terminate at t_m (nucleotide 164). Tn7-URF1 is the N-terminal region of a putative protein that is related to other transposases (see Fig. 4). This gene is probably transcribed from the weak promoter P_{LE}, transcription being initiated at I, nucleotide 148 ± 5. The boxed elements -35 and -10 correspond to regions that are related to conserved elements of other *E. coli* promoters.

supplemented with proline, histidine and tryptophan (each $50 \mu\text{g} \cdot \text{ml}^{-1}$). Trimethoprim was added to $100 \mu\text{g} \cdot \text{ml}^{-1}$ where indicated. Cells were harvested at mid-exponential phase ($A_{600} = 0.5$) and resuspended in lysis buffer (2 ml) containing 0.5% sodium dodecyl sulphate, 20 mM-sodium acetate and 1 mM-EDTA, pH 5.5, and then extracted twice at 60°C with phenol (5 ml) that had been pre-equilibrated with lysis buffer. The RNA was precipitated with 0.1 vol. of 3 M-sodium acetate, pH 5.0, and 2.5 vol. of 95% ethanol and redissolved in water ($500 \mu\text{l}$).

S_1 nuclease mapping

This was based upon the procedure of Berk & Sharp (1978). Single-stranded DNA probes were made by the

'prime-cut' method from M13 templates (Farrell *et al.*, 1983). This produces probes that are radiolabelled continuously throughout their lengths. Single-stranded probes (approx. 10^5 c.p.m.) and RNA (in $7 \mu\text{l}$ of solution) were coprecipitated with ethanol and resuspended in buffer ($15 \mu\text{l}$) containing 80% formamide, 0.5 M-NaCl, 40 mM-Pipes, pH 6.4 and 1 mM-EDTA. The RNA was denatured by heating the solution to 80°C for 10 min. Then it was annealed to the probe for 15 h at 53°C . To the RNA-DNA duplexes were added ice-cold buffer ($200 \mu\text{l}$) containing 30 mM-sodium acetate, pH 4.5, 0.25 M-NaCl, 1 mM- ZnSO_4 , 5% (v/v) glycerol and S_1 nuclease (400 units). The solution was incubated at 37°C for 30 min. Then the DNA was extracted with phenol, concentrated by precipitation with ethanol and analysed by electrophoresis under denaturing conditions in a 6% polyacrylamide gel (Sanger & Coulson, 1978).

Computer analysis

The homologies between the IS1 transposase (Ohtsubo & Ohtsubo, 1978) and Tn7-URF1 and the Tn3 transposase and Tn7-URF1 were assessed with the aid of the computer program DIAGON (Staden, 1982).

RESULTS

DNA sequence at the 3' end of the *E. coli glmS* in the absence and presence of integrated Tn7

The DNA sequence at the 3' end of the *glmS* gene, quoted in Fig. 1(a) [from Walker *et al.* (1984)], shows that the tandem stop codons of *glmS* are followed closely by a region with dyad symmetry and a run of T residues. These features suggest that this structure is a *rho*-independent transcriptional termination signal (Adhya & Gottesman, 1982) for the *glmS* gene. This has been verified by experimentation as shown below. Sequence analysis of DNA from this region of an *E. coli* strain, GD92::Tn7, in which Tn7 has been inserted shows that Tn7 is inserted after nucleotide 37 and thus, that insertion has interrupted the putative transcriptional termination signal (Fig. 1b).

The DNA sequences in Figs. 1(a) and 1(b) differ from those reported earlier by Lichtenstein & Brenner (1982) as follows. In (a) an additional G has been found at position 38, an additional A at position 140 and an additional T at position 211. In (b) an additional A is present at position 279.

The unique site of insertion of Tn7 is within the *glmS* transcriptional terminator

In order to confirm that the region of dyad symmetry and associated T residues indeed constitute the signal for termination of transcription of *glmS*, we have mapped the 3' end of transcripts in this region with S_1 nuclease (Berk & Sharp, 1978) as shown in Fig. 2. It was found that RNA from *E. coli* Q358 (no Tn7, track c) protected fragments of lengths 113 and approx. 390 nucleotides respectively. This maps the wild-type termination of *glmS* transcription to nucleotide 40 in Fig. 1(a), within the run of T residues following the region of dyad symmetry. RNA from *E. coli* GD92::Tn7 (tracks d and e) protected fragments three nucleotides shorter. Thus, this is the position at which Tn7 has integrated into the genome of *E. coli* GD92. Moreover, note that no transcripts covering *glmS* but proceeding beyond this terminator could be detected (see

Fig. 2. Mapping of the site of termination of the *glmS* transcription in *E. coli* Q358 (no Tn7)

1 and 2 refer to the probes 1 and 2 (see Table 1). M, size marker, being an *MspI* digest of plasmid pBR322 DNA end-labelled with $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$ and unlabelled dGTP in the presence of the Klenow fragment of DNA polymerase; sizes are indicated in bases. a, Untreated probe; b, probe digested with S_1 nuclease in the absence of RNA; c, d and e, probes hybridized to RNAs isolated from *E. coli* Q358, *E. coli* GD92::Tn7 grown in the absence of trimethoprim, and *E. coli* GD92::Tn7 grown in the presence of trimethoprim respectively, and then digested with S_1 -nuclease.

Fig. 3. Site of termination of transcription of *glmS* in *E. coli* GD92::Tn7

3 and 4 refer to probes described in Table 1. For description of M and a-e see legend to Fig. 2.

Fig. 2). RNAs were made from *E. coli* GD92::Tn7 grown either in the presence or absence of trimethoprim and both were used in S_1 mapping experiments in case the presence of the antibiotic altered patterns of gene expression in this region of Tn7. However, no differences were ever detected between the RNA derived from cells grown in the presence or absence of trimethoprim.

Transposition of Tn7 abolishes termination of *glmS* transcription

Transposition of Tn7 to the *glmS* transcriptional terminator interrupts the region of the dyad symmetry and separates it from the run of T residues within which transcription would normally terminate. Thus, it appeared likely that this mutation would abolish termination of *glmS* transcripts. To test this suggestion S_1 nuclease mapping was performed using DNA probes derived from an *E. coli* strain GD92::Tn7 in which Tn7 DNA has been inserted into the *glmS* terminator. As shown in Fig. 3 it was found that RNA from *E. coli* GD92 (::Tn7) (tracks d and e) protects fragments 205 and 238 bases in length respectively. This maps the main termination point of

glmS transcripts in *E. coli* GD92::Tn7 to nucleotide 164 in Fig. 1(b). RNA from *E. coli* Q358 (track c) protected fragments of probe 4 of around 110 bases in length. This marks the point at which wild-type sequences diverge from Tn7. These experiments show that *glmS* transcripts are not terminated to a significant extent at the vestigial *glmS* terminator structure. Instead, *glmS* transcription proceeds into the Tn7 DNA and is terminated at a new site, t_m , 127 bp into the left end of Tn7. A small proportion of *glmS* transcripts read through t_m , but these

Fig. 4. Homology between Tn7-URF1 and the transposases of IS1 and Tn3

The homologies between the IS1 transposase and Tn7-URF1 and the transposase of Tn3 and Tn7-URF1 were assessed with the computer program DIAGON (Staden, 1982) using a window length of 25 amino acids. Each dot represents sequences, 25 amino acids long, which have double matching probabilities (McLachlan, 1971) of 0.3×10^{-8} and 0.45×10^{-8} for the IS1 and Tn3 comparisons respectively.

Tn3	148	ATSWLNQNHDIIPGATTITRLISVVEKATLRIN
Tn7-URF1	47	HTSKTISRVNELLSDLELAVFLSEVSSVLDIE
IS1	77	ETMATLRLNELLSPDVVINTDPLYESLLE
Tn3	183	KIALIPFAQRSCLEMLSPITDCSNLILNLLKKG
Tn7-URF1	82	QFLLPSPDTETIILDSGKHPVIRGVQGVETIEL
IS1	112	KLVVISKRYTQREERNLNRQLARLGRKSLSE

Fig. 5. Alignment of protein sequences of Tn7-URF1 and transposases of Tn3 and IS1

Identities and conservative substitutions are boxed. The number indicates the number of the first amino acid residue in each comparison.

terminate at a second site 61 bp beyond t_m . No read through by *glmS* transcripts can be detected beyond this point. The termination site, t_m , is not found near to sequences with dyad symmetry. Thus, it is probable that this transcriptional termination site that is dependent on the *rho* termination factor of RNA polymerase (Platt & Bear, 1983).

An unidentified reading frame (Tn7-URF1) in the left end of Tn7 has significant homology with the transposases of Tn3 and IS1

The DNA sequence of the left end of Tn7 (Fig. 1b) contains an open reading frame, Tn7-URF1. This potential gene commences with a methionine codon and is preceded by a potential ribosomal binding site (Shine & Dalgarno, 1974). It extends for 134 amino acids to the edge of the known DNA sequence of the left end of Tn7. Comparison of the amino acid sequence of the product of Tn7-URF1 with those of the transposases of Tn3 (Heffron *et al.*, 1979) and IS1 (Ohtsubo & Ohtsubo, 1978) reveals highly significant homologies (see Figs. 4 and 5). A match reaching these scores would occur by chance in only 1 in 300 000 and 1 in 200 000 sequences of similar length and composition, respectively. In contrast no significant homologies were detected between Tn7-URF1 (and IS1 and Tn3 transposases) and other DNA binding proteins (namely the *lac*, *gal* and *trp* repressors and the *E. coli* *cro* protein). These conclusions are consistent with the results obtained by Smith & Jones (1984) and by Hauer & Shapiro (1984). The latter find that sequences to the left of the *Pst*I site in Tn7 [nucleotide 569 in Fig. 1(b)] are required in *trans* for transposition. This implies that a freely diffusible product, probably a protein, is coded for by this region. Tn7-URF1 is the only open reading frame in this region and thus probably encodes this protein product. These findings suggest that Tn7-URF1 may form part of the Tn7 transposase.

Identification of a promoter (P_{LE}) in the left end of Tn7 which directs transcription of Tn7-URF1

Transcriptional mapping with S_1 nuclease of RNA from *E. coli* GD92 bearing Tn7 has revealed the presence of a relatively weak transcript emanating from a position adjacent to Tn7-URF1, as shown in Fig. 6. In this experiment it was observed that RNA from *E. coli* GD92::Tn7 (tracks d and e) protects fragments, derived from the probes 5 and 6, 420 and 360 bases in length respectively. This maps the promoter P_{LE} to nucleotide 148 in Fig. 1(b). In addition a protected fragment 152 bases in length is seen with RNA from *E. coli* GD92::Tn7 annealed to probe 6. This may represent a secondary

Fig. 6. Site of initiation of transcription from the weak promoter P_{LE}

5 and 6 refer to probes described in Table 1. For description of M and a-e see legend to Fig. 2.

termination site of *glmS* transcripts at nucleotide 226, 61 bases beyond t_m . However, a corresponding band was not detected with probe 5.

As anticipated, it was found that RNA from *E. coli* Q358 (track c) offers no protection to either probe 5 or 6, both probes being contained entirely within the sequence of Tn7.

To the 5' side of the transcriptional initiation site, P_{LE} , are found two sequences which resemble the consensus sequence for *E. coli* promoters (Pribnow, 1978). This putative promoter (see Fig. 1b) could direct transcription of Tn7-URF1.

DISCUSSION

Regulation of P_{LE} by promoter occlusion

As shown in Fig. 1 the promoter P_{LE} is over-run by the *glmS* transcript; the termination site t_m is 17 bp on the 3' side of the transcriptional initiation base of P_{LE} . That *glmS* transcripts are terminated by t_m but P_{LE} transcripts appear to be unaffected is consistent with the proposed mechanism of action of termination protein *rho*. (Platt & Bear 1983). Thus, termination protein *rho* is believed to bind at recognition sites in the transcript between 100 and 250 bp 5' to the termination site. The protein then moves in an ATP-dependent manner along the nascent transcript and effects termination by binding to the RNA polymerase holoenzyme at a transcriptional pause site. As transcripts directed by P_{LE} would not contain the 5'

Fig. 7. Structure of the *glmS* transcriptional terminator and the site of transposition of Tn7 within the structure

rho recognition sequences, these transcripts could not be terminated by t_m .

Occlusion of P_{LE} by *glmS* transcripts should cause a reduction of initiation of transcription from this promoter. This phenomenon of promoter occlusion has been observed previously. For example, the *gal* promoter is repressed by transcription from P_{LE} during induction of an adjacent λ lysogen (Adhya & Gottesman, 1982).

The *glmS* gene product is part of the biosynthetic pathway leading to *N*-acetylglucosamine (Imada *et al.*, 1977), a component of the cell wall peptidoglycan. Its repression appears to be under growth-rate-dependent transcriptional regulation, since the level of transcript detected in exponentially growing cells is very much greater than that in stationary phase cells (results not shown). This could be due either to increased mRNA degradation or to decreased transcription in stationary phase cells. If the latter is true then the transposon may use *glmS* transcription to effect growth rate-dependent regulation of Tn7-URF1 by occlusion of P_{LE} . Under conditions of rapid growth the transposon is being rapidly replicated by the host cell along with the chromosome and thus does not need to use transposition to propagate itself. In contrast, under conditions of little or no growth, the transposon has everything to gain by transposing to episomal elements such as conjugative plasmids which might allow horizontal transfer of the transposon to other bacteria. Occlusion of P_{LE} by a growth-rate-dependent transcript provides a mechanism of regulating transposase expression in just such a manner. However if the growth rate dependent mRNA level were due to mRNA degradation, this would invalidate this suggestion.

Evolutionary significance

The transposition of Tn7 with high frequency to a transcriptional terminator may be of evolutionary advantage to the transposon, as transposition to this site causes only a neutral mutation of the host, and does not affect expression of a structural gene.

Indeed, an element which transposed with such a high frequency to random sites in the *E. coli* chromosome could not be sustained as it would induce high levels of lethal mutation.

The presence of a terminator, t_m , and the occlusion of P_{LE} may also be of evolutionary significance. In the absence of t_m , transcriptional readthrough into the left end of Tn7 from *glmS* would result in extensive expression of the transposase, since the amount of *glmS* transcription during rapid growth appears to be very much greater than that directed by P_{LE} . Extensive expression of the transposase would itself cause a significant increase in the rate of transposition of Tn7 to secondary insertion sites, which in turn is likely to lead to lethal mutation, thus destroying both host and transposon.

We thank Dr. C. P. Lichtenstein for sending us *E. coli* GD92::Tn7 and useful M13 clones. N.J.G. was supported by an M.R.C. Training Fellowship and V.L.J.T. by an M.R.C. Research Studentship.

REFERENCES

- Adhya, S. & Gottesman, M. (1978) *Annu. Rev. Biochem.* **47**, 967-996
- Adhya, S. & Gottesman, M. (1982) *Cell* **29**, 939-944
- Barth, P. T., Datta, N., Hedges, R. W. & Grinter, N. S. (1976) *J. Bacteriol.* **125**, 800-810
- Barth, P. T., Grinter, N. J. & Bradley, D. E. (1978) *J. Bacteriol.* **133**, 43-52
- Berk, A. J. & Sharp, P. A. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1274-1278
- Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963-3965
- Ely, B. (1982) *J. Bacteriol.* **151**, 1056-1058
- Farrell, P. J., Deininger, P. L., Bankier, A. & Barrell, B. G. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1565-1569
- Fennewald, M. A. & Shapiro, J. A. (1979) *J. Bacteriol.* **139**, 264-269
- Hauer, B. & Shapiro, J. A. (1984) *Mol. Gen. Genet.* **194**, 149-158
- Heffron, F., McCarthy, B. J., Ohtsubo, H. & Ohtsubo, E. (1979) *Cell* **18**, 1153-1163
- Imada, A., Yukimasa, N., Kawashima, F. & Yoneda, M. (1977) *J. Gen. Microbiol.* **100**, 329-337
- Karn, J., Brenner, S., Barnett, L. & Cesareni, G. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5172-5176
- Lichtenstein, C. P. & Brenner, S. (1981) *Mol. Gen. Genet.* **183**, 380-387
- Lichtenstein, C. P. & Brenner, S. (1982) *Nature (London)* **297**, 601-603
- McLachlan, A. D. (1971) *J. Mol. Biol.* **61**, 409-424
- Messing, J. & Vieira, J. (1982) *Gene* **19**, 269-276
- Ohtsubo, H. & Ohtsubo, E. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 615-619
- Platt, T. & Bear, D. G. (1983) in *Gene Function in Prokaryotes* (Beckwith, J., ed.), pp. 123-161, Cold Spring Harbor Laboratory
- Pribnow, D. (1978) in *Biological Regulation and Development*, vol. 1 (Goldberger, R. F., ed.), pp. 433-485, Plenum Press, New York
- Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87**, 107-110

- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **79**, 5463–5467
 Shine, J. & Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U.S.A. **71**, 1342–1346
 Smith, G. M. & Jones, P. (1984) J. Bacteriol. **157**, 962–964

- Staden, R. (1982) Nucleic Acids Res. **10**, 2951–2961
 Surin, B. P., Jans, D. A., Fimmel, A. L., Shaw, D. C. & Cox, G. B. (1984) J. Bacteriol. **157**, 772–778
 Walker, J. E., Gay, N. J., Saraste, M. & Eberle, A. N. (1984) Biochem. J. **224**, 779–815

Received 12 August 1985/1 October 1985; accepted 16 October 1985