Tn7 transposition: a multigene process. Identification of a regulatory gene product

Gillian M. Smith* and Peter Jones¹

ICI/University Joint Laboratory and ¹Department of Biochemistry, School of Biological Sciences, University of Leicester, Adrian Building, University Road, Leicester LE1 7RH, UK

Received 4 July 1986; Revised and Accepted 24 September 1986

ABSTRACT

Tn7 transposition is abolished by the deletion of a 2.2kb HindIII fragment from the central region of the transposon. Transposition is restored when the fragment is present *in trans*. When this fragment is present *in trans* with wild-type Tn7, transposition frequencies are stimulated 10-100-fold. The DNA sequence of this fragment has been determined and found to contain one long open reading frame coding for a protein of molecular weight 61 187. We have visualised this protein using a DNA-directed prokaryotic transcription-translation system. This gene may fill a regulatory role in the mechanism of Tn7 transposition. When present *in trans* the 2.2kb HindIII fragment alleviates a transcriptional block of promoter activity detected in Tn7.

INTRODUCTION

Prokaryotic transposable elements are discrete pieces of DNA found in, and capable of inserting into, the chromosomal and extrachromosomal DNA of bacteria, often leaving a copy of themselves in the donor molecules. The simplest transposable elements code only for genes involved in their transposition, the insertion sequences or IS elements. Larger transposable elements carry additional genes, for example determining antibiotic resistance, and are known as transposons (for review see 1,2).

Transposon 7 (Tn7) is a large transposable element (14 kb) which carries genes determining resistance to the antibiotics trimethoprim (Tp) and streptomycin/spectinomycin (Sm/Sp). It transposes into a variety of plasmids at a relatively high frequency, and with little site specificity (3,4,5). Tn7 transposition occurs at a much higher frequency (up to 100%) into the *E.coli* chromosome, at a specific site and in only one orientation (3,6). The chromosomal insertion site and the termini of Tn7 have been sequenced (7). There is evidence that there are also preferred sites or "hot spots" for the efficient insertion of Tn7 into the chromosomes of other bacterial genera (8,9,10,11). The high transposition frequency of Tn7 into bacterial plasmids and chromosomes render it a formidable vehicle for

the spread of antibiotic resistance genes through bacterial populations.

Analysis of the DNA sequence of the ends of Tn7 (7) reveals that the transposon does not belong to the class of transposable elements known as composite transposons, of which Tn5 and Tn10 are well studied examples, or the TnA family of transposons of which Tn3 is the best studied example (1).

Investigation of the effects of deletions in Tn7, constructed using restriction endonucleases, on its ability to transpose into conjugative plasmids (5,12) has revealed that Tn7 transposition is a multigene process. Deletion of the BglII fragment of Tn7 (Figure 1), which removes 800 base pairs (bp) of DNA inwards from the 42nd bp of the right hand end of the transposon, abolished transposition and could not be complemented *in trans*. This deletion disrupts an arrangement of four 22 bp direct repeats described by Lichtenstein and Brenner (7). This region must play an essential cis-acting role, possibly acting as a recognition site for the transacting transposition functions. Deletions in the right hand segment of Tn7 (rightwards from the EcoRI site, Figure 1) indicate that at least three genes might be required for its transposition (5,12). These transposition functions can be supplied in *trans* and are freely diffusible, as is true for those of Tn3 but distinct from the preferentially cis-acting proteins of Tn5, Tn10 and Tn903 (1).

We describe the identification of one of these gene products essential for the transposition of Tn7. The DNA sequence of this gene has been determined and the protein encoded by it identified *in vitro*. This protein appears to have a regulatory role in the mechanism of Tn7 transposition which it may effect by alleviating a transcriptional block of promoter activity, detected in Tn7, allowing the expression of gene(s) in the central region of the transposon essential for transposition.

MATERIALS AND METHODS

Bacterial strains and plasmids

Bacteria were grown and maintained as described (13) using Oxoid No.2 nutrient broth and minimal medium (14). Antibiotics were used at the following concentrations, trimethoprim, 10 $\mu\text{g.ml}^{-1}$; spectinomycin, 25 $\mu\text{g.ml}^{-1}$; chloramphenicol, 20 $\mu\text{g.ml}^{-1}$; kanamycin, 25 $\mu\text{g.ml}^{-1}$; ampicillin, 50 $\mu\text{g.ml}^{-1}$.

The plasmid used as a source of the promoterless β -galactosidase gene was pMC903 (15). Constructs were tested for β -galactosidase expression in the *E.coli* strain MC1000 which is deleted for the lactose operon (15).

Expression of β -galactosidase from the lac operon of *E.coli* was measured in the strain J53 (14) using isopropyl β -D-thiogalactopyranoside (IPTG), 8mM, to induce expression. The plasmid pYEJ001 was obtained from P-L Biochemicals, Inc., Milwaukee, Wis., U.S. The plasmid is a derivative of pBR327 containing a synthetic consensus *E.coli* RNA polymerase promoter followed by a HindIII chloramphenicol acetyltransferase (CAT) gene cartridge which contains its own Shine-Dalgarno sequence. Other plasmids used included ColE1 and ColE1:Tn7 (16,6) and the plasmid pACYC184 (17). The plasmid pHH7105 was constructed by replacing the HindIII fragment of pYEJ001 with fragment D of Tn7. The orientation of fragment D is such that expression of the gene product of fragment D must be from its own promoter.

Molecular cloning

Basic techniques of molecular cloning used, such as preparation of plasmid DNA, analysis of DNA using restriction endonuclease digestion and agarose gel electrophoresis, construction of plasmids and transformation of bacteria have been described (13).

Estimation of plasmid copy number

Plasmid copy number was estimated using a method based on that of Twigg and Sherratt (18).

DNA sequencing

The M13 phage vectors mp8 and mp9 (19) were used for DNA sequencing using the chain termination method (20). The host strain was *E.coli* JM103 (21). Fragment D was sequenced using clones of mp8 or mp9 containing the complete fragment or smaller sub-fragments generated by digestion of fragment D with one of the following restriction endonucleases, Sau3A, TaqI or HpaII. Both strands of the fragment were sequenced except for a stretch of 249 bp from bp 1285-1533 which was only sequenced on one strand. Each section of the sequence was obtained from between 2 and 7 independently generated clones. Fragment E was sequenced using clones of mp8 or mp9 containing the complete fragment, the 2 BamHI-HindIII fragments, or smaller sub-fragments generated by digestion of fragment E with Sau3A. Both strands of the fragment were sequenced using between 3 and 8 independently generated clones for each section.

In vitro transcription-translation of plasmid DNA

Plasmid DNA was prepared by a method of alkaline lysis and purified in caesium chloride-ethidium bromide gradients (13). The DNA (2.5 μ g) was translated using a prokaryotic DNA-directed translation kit obtained from Amersham International PLC, Bucks., U.K., with L-[³⁵S]methionine (>800

Ci.mmol⁻¹). The products were separated in a 17.5% (w/v) acrylamide gel (22), the gel dried and exposed to Kodak X-Omat RP X-ray film.

RESULTS

The DNA sequence of fragment D of Tn7

The nucleotide sequence of fragment D was determined and found to be 2230 bp long (Figure 2). The location of restriction endonuclease sites within the fragment agrees well with the mapping carried out in our laboratory of the positions of the BstEII, SstI and HpaI sites (to within 50 bp), and also with the positions of the AvaI, ClaI and Ball sites estimated by Gosti-Testu *et al.* (23). However, the sequence reveals an additional Ball site and does not contain the XbaI site reported (23). We have mapped this XbaI site at the same position as the central HindIII site of Tn7 and have located it just to the right of that site in fragment C using additional nucleotide sequence data.

There is only one long open reading frame which starts with a GTG codon and reads from right to left coding for a protein 538 amino acids in length, molecular weight 61 187. The identity of this protein as the active gene product of fragment D is supported by the fact that deletion of the BstEII fragment of Tn7 (Figure 1) does not affect the transposition frequency. From the sequence it can be seen that this deletion removes DNA to the left of the BstEII site and would not interfere with the integrity of the protein. A candidate Shine-Dalgarno sequence is highlighted (Figure

Figure 1. A simple map of Tn7 illustrating the regions of Tn7 involved in its transposition. Also shown is the approximate location of the promoter P₁.

4). A computer search for *E.coli*. consensus promoter sequences did not reveal a consensus sequence upstream of the open reading frame.

Identification of the protein encoded by fragment D of Tn7

Fragment D cloned into the vector pYEJ001, pHH7105, was used in a prokaryotic *in vitro* transcription-translation system along with pHH7106, in which the DNA between the internal ClaI sites of fragment D had been deleted (Figure 2). If a protein is expressed from the open reading frame in fragment D, then deletion of the ClaI fragment can be predicted from the sequence to result in the disappearance of that protein with the appearance of a smaller peptide of molecular weight 13 035.

The expression of a protein from fragment D in pHH7105 was detected (Figure 3, track 3) with an apparent molecular weight of approximately 70 000. Deletion of the internal ClaI fragment, pHH7106, resulted in the disappearance of that protein and the appearance of a smaller peptide with a molecular weight of approximately 12 000 (Figure 3, track 4).

Detection of promoter activity in Tn7

Genes for β -galactosidase (15) and chloramphenicol acetyltransferase (CAT), which lack promoters but contain their own Shine-Dalgarno sequences, were used to search for promoter activity in the central region of Tn7, known to encode diffusable gene products essential for the transposition of Tn7.

Promoter activity (P_1 , Figure 1) was detected (Table 1) when the BamHI fragment of the plasmid pMC903, carrying the promoterless β -galactosidase genes, was cloned in place of the 0.9 kb BamHI fragment E (Figure 1) of Tn7, in the orientation shown in Figure 4 (plasmid pHH7201). The direction of transcription is leftward towards the centre of Tn7.

The copy number of the plasmid pHH7201 was determined using a method based on that of Twigg and Sherratt (18) and estimated to be approximately 8 per chromosome equivalent, therefore the strength of the promoter is approximately 0.1 times that of the fully induced lac promoter of the *E.coli* chromosome (Table 1).

Detection of a transcriptional block in Tn7

The promoterless CAT gene from pYEJ001, carried on a HindIII fragment, was cloned in place of the 4.7 kb HindIII fragment (B) of Tn7 (pHH7202) in the same orientation, with respect to transcription, as the β -galactosidase fusion (Figure 4). Surprisingly little or no expression of CAT was detected (Table 2). This suggests the presence of a transcription stop signal within the 0.4 kb HindIII-BamHI fragment F of Tn7 (Figure 1).

10 20 30 40 50 60 70 80 90 100 110 120
 AAGCTTCGTCCCTTTGATCATGAAATGCCCTACTTTGACACGGGGCGGACTTCGATATTGGATTTCCTGACACGCTGTCAATCAAGATTAGTTTTTGACCGCAATGGATCAATCGTCC
 HindIII
 TTCGAAGCAGGAACTAGTACTTTACGGGATGAACTGTGCCCGCCCTGAAGCTATAACCTAAAGGACTGTGCGACAGTAGGTTCTAACTCAAAAACTGGCGCTTACCTAGTAGCAGG
 130 140 150 160 170 180 190 200 210 220 230 240
 AAGCACGTGCTCGCTGCACTTCAAGCAGGCTCTAATTTTGGCCAAGGCTCCAACATGTGCGCCGCGCTCAAGAAGCTATTGGATTACTACTATGGATCGTCTGTAATCCCCCTCAAA
 TTCGTGCAGCAGCGACGTGAAGTTCGTCCGAGATTAAAAACGGTTCGCGAGGTTTGACAGCGGCGGAGTTCCTAGTAACCTAATGATGATACCTAGCAGACATAGGGGGGAGTTT
 250 260 270 280 290 300 310 320 330 340 350 360
 ACCACGAGGCTATTGTTAGAAAAATCCATTAAATGAAGTCAAGGTGTTCCCTACGGACGAATCAGCGATGAAGGTGGTGTACCTGGCTATCCAGCGGCGACCAAGAATGGACGATG
 TGGTGCTCCGATAAACATCTTTTAAGTAATTATACTTGAGTCCACAAGGGATGCTGCTTAGTCGCTACTCCACCACATGGACCGATAGGTGCGCGCTGGTCTTTACCTGCTAC
 370 380 390 400 410 420 430 440 450 460 470 480
 CCGATCCGCACTGGAAGCCGGCCATGAACGTTTATGATCGAATTTGGTGACCGTTTAAATGGCCACCTTTAACTGACCGGCAAAAGCACTTACACAAAAGGGTTACAGGCTCTTTC
 GGCTAGGCGTTGACCTTCGGCCGTAAGTGGCAAAATACTAGCTTAAACCACTGGCAAAATTACCGTGGAATGACTGGCGGTTTTCTGTAATGTGTTTTCCCAATGTCCGAGAAAG
 BstEII*
 490 500 510 520 530 540 550 560 570 580 590 600
 AGTCATTCCTTAATGCGTAAATGCTCTTAACCACTCTCTCAGCCACCCCTTCCATATCTCTCTCGATCTTGACGCTCCCCAACCCCTTTGGATGATTACCCCTCTATGCCCCG
 TCAGTAAGGAAATACGCAATTAACGAGAGATTGGTGAGAGAGTCGGGTGGGAAGGTATAGGAGAGAGCTATGAACGTCGAGGGGGTGGGGAACCTACTAAGTGGGGAGATCGGGCA
 HisThrPheGlnGluArgValIleArgGluAlaTrpGlyGluMetAspGluArgSerValGlnLeuGluGlyLeuGlyLysProHisAsnValGlyArgHisGlyAsp
 610 620 630 640 650 660 670 680 690 700 710 720
 CTTGTCATATAATTGATCAACAACTATTGGCCAATTCAATGAGCTCTTCCCACTCCACGCGATCTTTTCAAAATCATTCGCGCATGTTTCGCTATCAAGCCACTCAACACTT
 GAACAGGTATATACTAGTTGTTTGTGATAACCGTAAAGTTACTCGAGAAGTGGTGAGGTGGCGCATAGAAAAGTTTGTAGTAAAGCGGTACAAGCGATAGTTGCGGTGAGTTGTGAA
 GlnGlyTyrLeuGlnAspPheLeuSerAsnProTrpAsnLeuSerSerLysValIleGlyArgArgIleLysGluPheAspAsnArgTrpThrGluSerAspValGlySerLeuValLys
 730 740 750 760 770 780 790 800 810 820 830 840
 TGGTAGAAGACGATTTTCCACCATCGATGCCACTCTAGCAATCAAAATCACTGCCATTACGTTGACCCCAACGCGTTAATCACTCTGGTGAGCCATCTTTACACAGATGTA
 ACCATCTTTCTGTAAGTGGGGTAGGCTAGTGAAGTGGAGATCGTTATGTTTAAAGTACGTAATGCAACGTCGGGGTGGCCGAATTAGTGAGAACCCTCGGTAGAAATGTGTCTACAT
 ThrSerLeuMetLysValGlyAspSerAlaAspValGluLeuLeuValPheGluSerGlyAsnArgGlnValGlyValAlaLysIleValArgProSerGlyAspLysCysLeuHisLeu
 850 860 870 880 890 900 910 920 930 940 950 960
 ATCGGCTACGCCCACTTTTGGCAAAACGAGTTTCTTCAAAAAGCACCCGACATGCAAAATTAGTTTTAGATTGAAAGTAGCTCATCAAAGCAGCAAAATCGATTAGCAAAATAG
 ClaI
 TAGCCGATCGGGTTGAAAACCGTTTGGTTCCAAAGAGTTTTCGTGGGCTGACGTTTAAATCAAAATCTTAACCTTCATCGAGTAGTTTTCTGCGTTTAGCTAATCGTTTTATC
 ArgSerArgGlyValLysProLeuValLeuThrGluGluPheLeuValArgCysAlaPheLysThrLysLeuIleSerLeuLeuGluAspPheAlaAlaPheArgAsnAlaPheIleSer
 970 980 990 1000 1010 1020 1030 1040 1050 1060 1070 1080
 AGTTGTAATGGTTGCATCTGCTTCCACCCCATCTGCCGCTGCTAGGACACCACCTAAGTGTGGCTCATCAGTACTAACCAATTACTATTTTCACTTGTGTTGCTCTCTCTTTT
 TCAACATTAAACCAAGTAGGACGAAGGGTGGGTGATAGCAGCGGACGATCCTGTGGTGGATTACACCGAGTAGTCATGATGGTTTAAATGATAAAAGTGAACAACACGAAGAAGAAAA
 AsnTyrAsnThrAlaAspGlnLysGlyGlyValAspAlaAlaAlaLeuValGlyGlyLeuHisProGluAspThrSerValLeuAsnSerAsnGluSerThrThrAlaGluGluLysGlu
 1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
 CTAATAATAGTTGGCTACTTTTCTTCCCTCGAGATTTTCAATTGCGTAGGCTACTAAAACCTATCCATGTGCGGCTCAGCCTCTATGACTCGTGCTGTCTGTCTATATTGGCTG
 GATTTTATCAAAACCGATGAAAAGACGGGACGCTCTAAAAGTTAAGCTGATCCAGATGATTTTGAATAGGTACAGCGAGTCGGAGTACTGAGCAGACAGACAGATATAACCGAC
 LeuIleThrGlnSerSerLysArgAlaArgArgSerLysGluIleArgSerProArgSerPheSerIleTrpThrProGluAlaGluIleValArgThrAspThrAspIleAsnAlaThr
 1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320
 TTTCTCATCGTCCAACCTGTAGATGCTCATCAACAACCTGACTCTGTTGAAACCGCTATGTGCTGCTACTACCATCACCTGCCCTCTTTCCTGAAAAGAGGCATGCGTAATAGCTG
 AAGGAGTAGCAGGTGACATCTACGAGTAGTATTGTTGACTGAGACAACCTTTGGGATACACGACGATGAGTGTAGTGGACGGAGGAAAAGGACTTTTCTCCGTACCGATTATCGAC
 GluGluAspAspLeuGlnLeuHisGluAspAspValIleSerGluThrSerValAlaIleHisGlnThrSerGlyAspGlyAlaGluLysGluGlnPheSerAlaHisSerIleAlaThr
 1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430 1440
 TTGTGCTAGGAAGCATAGCGTTGATTTCTAAACCTACTATTTCTCAACTAAATAATCTTATCTCGTTAGAAGAACGCTCTTTTACATGTAATTTCAACCTGCATGGCGGAGGGT

```

AACACGATCCTTCGTATCGCACTAAGATTGGATGATAAGGAGTTGATTATTAGGAATAGGAGCAATCTCTTCGAGGAAATGTACATTAAAGTTGGGACGTACCCGCTCCCA
ThrSerProLeuMetAlaAsnIleGluLeuGlyValIleGluGluValLeuTyrAspLysAspGluAsnSerSerArgGlyLysValHisLeuLysTrpGlyGlnMetProProAsp

1450 1460 1470 1480 1490 1500 1510 1520 1530 1540 1550 1560
CAAAGCTAAGCACCAGCTTTCACGCCGTTCTTAATTTCTCTATTTTGTTGATAGTGCTAAAAATACTTTTCATCAGATCCATAACATCTTGATCGAAAAACAACCAACCAAGCT
GTTTCGATTTCGTGGTGGAAAGTTGCGCAAGAATTAAGAGATAAAACAATATCACGGATTTTATGAAAGTATGCTTAGGTATTTGAACTAGGCTTTTGTGGTTGGTTTTCGA
PheSerPheCysTrpSerGluValGlyAsnLysIleGluArgAsnGlnGlnTyrHisArgPheIleSerGluTyrSerAspMetValAspGlnAspSerPheLeuTrpValLeuLeuGln

1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680
GCATACGCCGACTGCTCTAACGCCCTTTAGGAAACGATGAATGGGTAAAGATCCTTATCTCTAATGATCTCGCTCAACCTCATCTGAACCTGCAACTCTTGCTGTAAGCGGTAC
CGTGATCGCGGCTGACGAGATTGCGGGAAATCCTTGTCTACTTGACCCATTCTAGGAATAGAGATTACTAGAGCGAGTTGGAGTATGACTTGCAGCTTGAGAACGACATTTGCGCATG
ValValAlaSerGlnGluLeuAlaGlyLysProPheSerSerSerProLeuIleArgIleGluLeuHisAspArgGluValGluTyrGlnValAspPheGluGlnGlnLeuAlaThrSer

1690 1700 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
TGCTCAACAGCTTCGACAGAAATAGGAGTTAATTAATAATGACCGAGCTAGCTCTAAGTGGGATGTGATAACGAGTCTTGCTATATGGTGTTCAGCTTCGTAATAATGAAC
ACGAGTTTGTGGAAGTGTCTTATCTCTCAATTAATTTTATTACTGGCTCGATCGAGATTGACACCTACACTATTGCTCAGAACGATATACCACAAAGTGAAGCATTTATTACTTG
SerLeuLysSerArgCysPheTyrSerAsnIleLeuPheLeuSerArgAlaLeuGluLeuGlnProIleHisTyrArgThrLysSerTyrProThrGluAlaGluTyrIlePheSerSer

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920
TCTGCGCTCATCTGTTGGCTACGGAAGCAAGGACGAGAGATTAGGAAAATCGATTATTTGGCTCGTTGAAGATCCGATGCTTCAAACTCAATCACACGATCCGCTTATTTATTT
*Clal*
AGACGCGAGGTAGGAACAACGATCGCTTCTGTTCTCTCTAATCCTTTAGCTAATAAAACGAGCAACTTCTAGGCTACGAAGTTTGAGTTAGTGTGCTAGGCGGAATAATAAA
GlnAlaGlyAspLysAsnArgValSerAlaPheSerSerLeuAsnProPheAspIleIleLysAlaArgGlnLeuAspSerAlaGluPheGluIleValArgAspAlaLysAsnIleGln

1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030 2040
GCTTTGTTGACTGTTAACCACCTTAGCTGACTTAATAAGGCAAAAGAGTAATCGAGTCCATTTCCGTTCTTGCAATTGATTAACCAAAACAAAATACGCCACTCTTATGACCCG
CGAAACAACATGACAATTGTTGGAATCGAGCTGAATTTATTCGTTTACTCATTTAGCTCAGSTAAGGCAAGAACCTAACTAATTGTTTGTGTTTTTATGCGGTGAGGAATACTGGGC
LysThrThrSerAsnValValLysAlaArgSerLeuLeuProLeuHisThrPheArgThrTrpLysArgGluGlnMetProAsnPheTrpValPheIleArgTrpGluLysHisGlySer

2050 2060 2070 2080 2090 2100 2110 2120 2130 2140 2150 2160
AGTTACGGAATAAATGACCAATATGTACAACCTGCACATTGTTCATTAATGTAGTAGCCTAACCACTCACTCTCCCATAAACCATTTCCAGAATCTTTGTGCTCTCTCAGTTATCC
TCAATGCCCTTATTACTGTTTATACATGTTGGACGTGTAAACAGTAATTTACATCGATCGGATTGGTGAAGGCTATTTGGTAAAGTCTTAGAAACACGAAGGAGTCAATAGG
AsnArgPheLeuHisGlyIleHisValValGlnValAsnAspAsnPheThrAlaLeuArgValVal
      ←
2170 2180 2190 2200 2210 2220 2230
GCTCTTTAGATAACGTGCACTTCTTAATAATCGCCAGCGCCTAAGCTCAACATCTTCTGTTTAAGCTT
HindIII
CGAGAAATCTATTGCAACGTGAAGAATTATTAGCGGTGCGGATTCGAGTTGTAGAAGGACAAATTCGAA

```

Figure 2. The DNA sequence of the smaller internal HindIII fragment (D) of Tn7. The translation of the single long open reading frame is included, which codes for a protein of 538 amino acids with a molecular weight of 61 187. A putative Shine-Dalgarno sequence is outlined.

The DNA sequence of the 0.9 kb BamHI fragment of Tn7 (E) has been determined (Figure 5). The fragment was found to be 908 bp in length. The location of restriction endonuclease sites for HindIII, KpnI and SstI agreed well with our mapping data (to within 50 bp). In the right hand HindIII-BamHI fragment (F) candidate regions of dyad symmetry, which might act as rho-dependent terminators, are highlighted (24). There is also a long open reading frame, reading from right to left initiating 24 bp in from the right hand BamHI site, which does not terminate within fragment E and could be part of a longer open reading frame encoded by fragment C.

Figure 3. SDS polyacrylamide gel electrophoresis of proteins produced in a DNA-directed transcription-translation system. Track 1, no DNA; track 2, pYEJ001; track 3, pHH7105; track 4, pHH7106.

Table 1. Detection of promoter activity in Tn7

Bacterial strain	Plasmid(s)	β -galactosidase expression (nmol.min ⁻¹ .10 ⁸ cells ⁻¹)
MC1000	none	0
MC1000	pMC903	0
MC1000	pHH7201	65
MC1000	pHH7201+pHH7103	65
J53	none	0
J53 + IPTG	none	70

The values of β -galactosidase expression are averages from 2 or more experiments, assays were performed in duplicate according to Miller (26).

Figure 4. The structure of gene fusion constructs. a) pHH7201, the bold line indicates the insertion of the promoterless β -galactosidase gene between the BamHI sites of Tn7 carried in ColE1. b) pHH7202, the bold line indicates the insertion of the promoterless CAT gene in place of the HindIII fragments of Tn7 carried in ColE1.

Fragment D when present *in trans* alleviates the transcriptional block

Investigation of the effects of deletions in Tn7 on its transposition (5), pointed to some mechanistic interaction between regions D and E (Figure 1). Deletion of fragment E reduced the frequency of Tn7 transposition 1000-fold. Deletion of both fragments D and E gave the same 1000-fold decrease in transposition frequency, implying that fragment D is not involved in transposition. Paradoxically, deletion of fragment D alone abolished transposition. This function could be supplied *in trans* by fragment D, cloned into the plasmid pACYC184 (pHH7103, 5). Fragment D must therefore contain the complete coding requirements of a diffusible gene product that is essential for the transposition of Tn7 only when fragment E is present.

Table 2. Expression of CAT from pHH7202

<u>Bacterial strain</u>	<u>Plasmid(s)</u>	<u>CAT expression</u> (MIC Cm $\mu\text{g}.\text{ml}^{-1}$)
DS910	none	< 5
DS910	pHH7202	< 5
DS910	pHH7202+pHH7105	> 50
DS910	pHH7105	< 5
DS910	pYEJ001	> 50

```

      10      20      30      40      50      60      70      80      90      100     110     120
GGATCCGAGATATAGCTCCAAACCTGCCCCCTACGTGCAGACCGCAATCAGCCTCAAAATCTCTCCTGCTTTAGGGGTACCAATCAACATCACTGCTGACCCCAATATATACC
BamHI*
CCTAGGGCTTTATATCAGGGTTTGGGACGGGAGATGCAGCTCTGGCGTTTATGTCGAGGTTTTTASAGAGCACSAATCCCCATGGTTAGTTGTAGTACCATCGGGTTATTATAAGTGG
AspTrpPheIleAlaGlyPheGlyAlaGlyArgArgAlaSerArgLeuAspAlaGlyPheIleGlyArgAlaLysProThrGlyIleLeuMetValProValGlyIleIleAsnValMet

      130     140     150     160     170     180     190     200     210     220     230     240
ATCCTCACAAAAAGTTTCAGCATCTCTTGGATGTCACCCGAAACGAGAGCGGCTTAATGCTGAATTTTCATCAATAACCAACAACCCCTAAGCATGTGCAATTTGCTGTCATCAAA
TAGCAGTGTTTTTCAGTCTAGAGAACTCTAGTGGGCTTGTCTCTGCCBAATTTACGACTTAAAGTAGTTATTGGTTGTGGGATTTCTGACACGTAACCGATAACGCTGTAGTTT
ThrValPhePheAsnLeuMetGluGlnSerGlyGlySerArgSerArgSerLeuHisGlnIleGluAspIleValLeuLeuGlyLeuAlaHisAlaAsnAlaIleGlnSerMetLeuAla

      250     260     270     280     290     300     310     320     330     340     350     360
GCCAATCATGGTTTCTATACCATGACGCTTTTAAGCCATAACGACGCTCATAGTTCGAGCCCAAGGCTCGATCCACGCTCTGAAAAATTCAGCAGATTCTTTTAGCAACCATATTGC
CGGTTGTACCAAGATATGTTACTGCAAAATTCGGTATTGCTGCGAGTATCAAGCTCGGGTCCGAGCTAGGTTGCGAGACTTTTAAAGTTCGTAAGAAATTCGCTTGGTAATACG
LeuMetThrGluIleGlyHisArgLysLeuGlyTyrArgArgGlyTyrAsnSerGlyLeuAlaArgAspLeuAlaArgPhePheAsnLeuCysIleGluLysLeuSerGlyAsnHisSer

      370     380     390     400     410     420     430     440     450     460     470     480
GAGCAGTCTATTTTCAATACACCACCTGCTCTACATGTAGTTACGATGGTAAATCAGCTGAGGATACGTTGCGATGAGATGAGGAGGCTGCTCTCCCACTACCAAGAACCA
CTCCTCAGATAAAGTTTATGTGGTGGACGAGATGTAAGTCAAGTCTACCATTTAGTGGACTCTATGACGATCTTATGCTACTTCTCTCCAGCAGAGGGGTGATGGTCTTGTGGT
CysAspIleLysLeuTyrValValGluGlnValAsnLeuGluArgHisTyrIleValGlnProTyrThrAlaLeuIleArgHisLeuSerThrThrLysGlySerGlySerCysGlyIle

      490     500     510     520     530     540     550     560     570     580     590     600
ATTAAACAATAGCTTTTGTCCGTAGATCGTGCCTCTCAAGCGAAATGTCTCAAGCTCTCCCGTTTGAACGCTCATAAACATTTTGTAAATGCTTTTGTAAATCTCTGTTTATGGA
HindIII
TAATTTGTTATTCGAAACACGGCATCTAGCACGGAGAGTTTCGCTTTACAGAGGTTGAGAGGGCAAAGTTGTGCGAGTATTGGTAAACATTTACGAAACATTTAGAGACAAATCTCT
LeuLeuLeuSerGlnAlaThrSerArgAlaGluGluPheArgPheThrGluLeuGluGlyThrGlnValArgGluGlyTyrGlyAsnGlnLeuHisLysGlnLeuAspGlyThrLysProAsn

      610     620     630     640     650     660     670     680     690     700     710     720
TTTCTGCCATGACGACCTCGAATCATGACCGAAATACGCTCCTTACGATGCAATGCGTACCTAATGGCTGAAATATGTCATCTGGAATACGACAAATGGTATGAGCTGATAACA
AAGAGCGATGTCATCGTGGAGCTTAGTACTGCTTTATGCGAGTGAATCATCGTTTACGATGATGATTACGACTTTTATCAGTAGACCTTATGCTGTTTACCATACTCGAGACTATTGT
ArgGlyValTyrGlyGlyArgIleMetValSerIleArgGluSerLeuLeuLeuHisThrGlyLeuProGlnPheTyrAspAspProIleArgCysIleThrHisAlaArgIleValArg

      730     740     750     760     770     780     790     800     810     820     830     840
CGGGACTTTTGAACATCAGAGGAATAGCTGTAAAGAGGATTTACGATGATGACGACTATTCACTGACTCTTGTATGGTGGTAAAGGCTCGATAAAGGATTTACGATAAAGGCTCT
GCCCTGAACGTTTCAGTCTCTTCTTATTCGACATTTCTCTTAAGTCACTACGTCGTGATAGTGAATCAACATTACCACTATTCGGAGCTATTTCTTAATAGTGTATTCTGGAGA
SerLysGlnLeuAspSerSerThrLeuGlnLeuSerSerLysLeuSerAlaAlaSerAsnValSerGluGlnLeuProProLeuAlaGluIlePheProAsnAspArgTyrAlaGluVal

      850     860     870     880     890     900
AGCCCCGATATCAGATAAAGCTGCTGAATCCGGGTAGCACTCATGACTCATCTTTCTGGTGGATCC
BamHI*
TGGGGCATAGTGTATTTGACGAAGTTAGGCGCATCGTGAAGTACTGAGTAGGAAAGACCACTAGG
GlyThrAspArgTyrValAlaGlnIleArgThrAlaSerMet

```

Figure 5. The DNA sequence of the 0.9 kb BamHI fragment E of Tn7. Regions of dyad symmetry which might act as rho-dependent terminators are outlined. The translation of an open reading frame is included which spans almost the entire fragment reading from right to left and leading into fragment C.

The presence of fragment D *in trans* with the wild-type Tn7 results in 10-100 fold stimulation of the transposition frequency, it is possible that the gene product of this fragment has a regulatory role. To test if this gene product affected the promoter activity detected in Tn7, the level of

β -galactosidase expression from the gene fusion pHH7201 (Figure 4) was measured when fragment D, cloned into the plasmid pACYC184 (pHH7103), was present *in trans* (Table 1). The copy number of the β -galactosidase construct was unaffected by the introduction of the plasmid pHH7103. There was no effect on the level of expression of β -galactosidase in pHH7201. However, when fragment D cloned into the plasmid pYEJ001 (pHH7105) was present *in trans* with the CAT fusion plasmid pHH7202 (Figure 4), although the copy number of pHH7202 was apparently reduced by 75% with the introduction of pHH7105, the transcription block detected in fragment F was alleviated allowing the expression of high levels of CAT (Table 2). If DNA between the BamHI and BglII sites of pHH7202 (3kb) was removed, a deletion we would expect to remove the promoter activity, the presence of pHH7105 *in trans* no longer conferred resistance to chloramphenicol.

The gene product of fragment D does not affect expression from P₁ directly but must interact with a site to the right of fragment B (Figure 1) to alleviate transcription termination.

DISCUSSION

Previous data (5) have indicated that fragment D of Tn7 must encode a *trans* acting factor involved in the mechanism of Tn7 transposition. We have sequenced fragment D, which was found to contain only one long open reading frame coding for a protein of 538 amino acids (molecular weight 61 187). We have identified this protein pD in an *in vitro* transcription-translation system using the plasmid pHH7105 and shown the protein band to be replaced by one of the predicted lower molecular weight after deletion of the internal ClaI fragment of fragment D which should lead to the production of a truncated peptide.

The detection of promoter activity P₁, to the right of fragment F in Tn7, and the detection of a transcription block in fragment F suggests a possible role for the protein pD. The deletion of fragments D and E separately and together (5) indicated some mechanistic interaction between these two regions. For this reason we tested the effects of having fragment D present *in trans* on expression from P₁ and on the transcription block detected in fragment F. We found that expression from P₁ was not affected but that the transcription block in fragment F was alleviated leading to high expression of CAT from the gene fusion construct pHH7202 (Figure 4). The location of the promoter activity P₁ was confirmed as being to the right of fragment E and within Tn7. The gene product pD must act at a site

to the right of fragment B but does not interact with P₁ directly. This situation relates to the characteristics of antitermination described for phage lambda (25), where the immediate early gene product pN alleviates transcription termination at the terminators t_{L1} and t_{R1}, allowing expression of the delayed early genes of lambda. In this case the recognition sites of the antiterminator protein N are not the promoters or the terminators but the nut L and nut R sites which are located at different positions within the transcripts. The sequence of fragment E was determined and revealed candidate sequences which could act as rho-dependent terminators. Comparison of the amino acid sequence of the protein pD to those of the N and Q proteins of lambda showed no significant homology.

Preliminary *in vitro* transcription-translation data using ColE1::Tn7 and its deleted derivatives described before (5) shows the expression of pD from the appropriate plasmids (where the 2.1 kb HindIII fragment has not been deleted). There is evidence that deletion of the 0.9 kb BamHI fragment E of Tn7, removing the transcription block, results in the expression of a high molecular weight protein which could be the transposition function encoded by fragment C. The DNA sequence of fragment E revealed an open reading frame reading leftwards into fragment C. This might code for the N-terminal region of a transposition function encoded largely by fragment C. The result of deletion of fragment E (1000-fold decrease in transposition frequency, 5) might then be explained by the truncation of the putative protein product of fragments E and C. We have also detected two proteins encoded by the right hand end of Tn7 whose expression is not detected from plasmids deleted between the BamHI and BglII sites. Ouarts *et al.* (12) have also described preliminary evidence of two proteins encoded by the right hand end of Tn7 using *E.coli* maxicells.

From our data and that of Ouarts *et al.* (12) it would seem that at least three proteins could be involved in Tn7 transposition and its regulation. The apparently complex mechanism of Tn7 transposition may reflect its versatility as a transposon; its broad host range and extremely high efficiency transposition into prokaryotic chromosomes e.g. the *att* site of *E.coli* (7) where transposition into the specific "hotspot" is up to 100% efficient; and the high efficiency with which Tn7 can transpose into many sites in a wide variety of plasmids.

ACKNOWLEDGEMENTS

We would like to thank Professor Naomi Datta for the opportunity to do this work; we were supported by a grant to Professor Datta from the Medical Research Council of the U.K. We would also like to thank S. N. Cohen for the gift of E.coli MC1000 and the plasmids pMC903 and pMC871, S. Delaney of the Imperial College of Science and Technology and especially John Keyte of the University of Leicester for help with computing. Special thanks are due to Professor W. V. Shaw of the University of Leicester without whose support additional work necessary for this manuscript could not have been carried out.

*To whom correspondence should be addressed

REFERENCES

1. Kleckner, N. (1981) *Ann. Rev. Genet.* 15, 341-404.
2. Shapiro, J.A. (1983) *Mobile Genetic Elements*, Academic Press.
3. Barth, P.T., Datta, N., Hedges, R.W. and Grinter, N.J. (1976) *J. Bact.* 125, 800-810.
4. Barth, P.T. and Grinter, N.J. (1977) *J. Mol. Biol.* 113, 455-474.
5. Smith, G.M. and Jones, P. (1984) *J. Bact.* 157, 962-964.
6. Lichtenstein, C. and Brenner, S. (1981) *Mol. Gen. Genet.* 183, 380-387.
7. Lichtenstein, C. and Brenner, S. (1982) *Nature* 297, 601-603.
8. Fennewald, M.A. and Shapiro, J.A. (1979) *J. Bact.* 139, 264-269.
9. Caruso, M. and Shapiro, J.A. (1982) *Mol. Gen. Genet.* 188, 292-298.
10. Thomson, J.A., Hendson, M. and Magues, R.M. (1981) *J. Bact.* 148, 374-378.
11. Ely, B. (1982) *J. Bact.* 151, 1056-1058.
12. Ouarts, A., Borowski, D. and Brevet, J. (1985) *Mol. Gen. Genet.* 198, 221-227.
13. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, New York.
14. Clowes, R.C. and Hayes, W. (1968) *Experiments in Microbial Genetics*, Blackwell Scientific Publications, Oxford, U.K.
15. Casadaban, M.J., Chou, J. and Cohen, S.N. (1980) *J. Bact.* 143, 971-980.
16. Dougan, C., Saul, M., Warren, G. and Sherratt, D. (1978) *Mol. Gen. Genet.* 158, 325-327.
17. Chang, A.C.Y. and Cohen, S.N. (1978) *J. Bact.* 134, 1141-1156.
18. Twigg, A.J. and Sherratt, D. (1980) *Nature* 283, 216-218.
19. Messing, J. and Vieira, J. (1982) *Gene* 19, 269-276.
20. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
21. Messing, J., Crea, R. and Seeburg, P.H. (1981) *Nucl. Acids Res.* 9, 309-321.
22. Smith, G.M. and Shaw, W.V. (1981) *J. Gen. Microbiol.* 124, 365-374.
23. Gosti-Testu, F., Norris, V. and Brevet, J. (1983) *Plasmid* 10, 96-99.
24. Brutlag, D.L., Clayton, J., Friedland, P. and Kedes, L.H. (1982) *Nucl. Acids Res.* 10, 279-294.
25. Friedman, D. and Gottesman, M. (1983) in *Lambda II*, Hendrix, R.W. Ed., Cold Spring Harbor Laboratory, New York.
26. Miller, J. (1972) in *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, New York.