DNA sequence analysis of five genes; *tnsA, B, C, D* and *E*, required for Tn7 transposition

Carlos Flores*, M.Ishtiaq Qadri[1] and Conrad Lichtenstein*
Imperial College of Science, Technology and Medicine, Centre for Biotechnology, London SW7 2AZ, UK and [1]Department of Stomatology, University of California at San Francisco, San Francisco, CA 94143, USA

Received October 31, 1989; Revised and Accepted January 12, 1990

ABSTRACT

A region of DNA sequence of the bacterial transposon Tn7, which is required for transposition, has been determined. This DNA sequence completes an 8351 base pair (bp) region containing five long open reading frames (ORF's) that correspond to the genetically defined genes, *tnsA, B, C, D* and *E*, required for Tn7 transposition. All of the ORF's are oriented in the same direction, *ie.* inward from the element's right end. The genes are in a very compact arrangement with the presumed initiation codons never more than two bases beyond the preceding termination codon. Domains with similarity to the helix-turn-helix genre of Cro-like, sequence specific DNA binding sites occur within the deduced amino acid (a.a.) sequence of the TnsA, TnsB, TnsD and TnsE proteins. Translation of the *tnsC* ORF reveals strong homology to a consensus sequence for nucleotide binding sites as well as a region of similarity to a transcriptional activator (MalT). No striking a.a. sequence similarity to other DNA recombinases is observed. The possible roles of these proteins in Tn7 transposition is discussed in light of the analysis presented.

INTRODUCTION

Transposable genetic elements are discrete DNA segments that are able to move from one position in a genome to another, or from one replicon to another within a cell. This process does not involve homologous or general recombination systems of the host, but requires one (or a few) gene product(s) encoded by the element. Specific DNA sequences at the termini that define the boundary with the host genome are also necessary in *cis*. The termini are usually composed of inverted repeat sequences of various lengths (for recent reviews on transposons see Ref. 1 and 2).

Transposon 7 (Tn7) is a large (14 kilobase pairs (Kb)), and complex transposable DNA element of bacteria that encodes resistance to trimethoprim and the aminoglycosides streptomycin and spectinomycin (3, 4).

One factor contributing to this complexity is the transpositional behavior of Tn7. While most transposons have low specificity for target site selection, Tn7 has a dual tendency; it transposes at a high frequency to a specific 'attachment' site (*attTn7*), in the chromosome of *E. coli* and at a lower frequency, (about 100×lower) to apparently random sites in plasmids or chromosomes (3, 5, 6, 7, 8, 9). Tn7 will also transpose to regions of DNA with sequence related to *attTn7*, 'pseudo-*attTn7* sites' at a similar frequency as to random (non-*attTn7*) sites (4). Transposition to *attTn7* (and pseudo-*attTn7*) sites results in the integration of Tn7 in a single orientation (4, 6,10). Surprisingly, random insertions of Tn7 in several plasmids also occur in a single orientation (11, 12, 13, 14).

Genetic analysis of Tn7-encoded transposition functions by deletion and insertional mutagenesis, in conjunction with complementation analysis has revealed five genes involved in transposition, designated: *tnsA, tnsB, tnsC, tnsD* and *tnsE* (8, 9) (see Fig. 1). This is an unprecedented number of transposition genes. The existence of two classes of target sites reflects the requirement for two alternative, overlapping sets of *tns* gene products. Transposition to *attTn7* (and the lower frequency transposition to pseudo-*attTn7* sites) requires the products of the *tnsA, tnsB, tnsC* and *tnsD* genes, whereas transposition to random sites requires the gene products from *tnsA, tnsB, tnsC,* and *tnsE* (8, 9).

The *tns* gene products act efficiently in *trans* (8, 9 and 15) unlike the transposases of some transposons (16, 17).

The structure of *attTn7* is intriguing. Deletion analysis of the *E. coli attTn7* site and a comparison to *attTn7* sequences from *Serratia marcescens* and *Klebsiella pneumoniae* indicate that the only sequences that are indispensable for *attTn7* activity are located from about 22 to 59 base pairs to one side of the insertion point (28 to 65 in the case of *Klebsiella*), (18, 19, 20). The point of insertion in *E. coli* is within a region that produces the transcriptional terminator of the glucosamine synthase (*glmS*) gene, while the sequence critical for *attTn7* activity (called the '*glmS*-box'), encodes the carboxy-terminal 12 amino acids of this enzyme (18, 19, 20, 21).

Another example of the complexity of Tn7 is that the *cis*-essential ends are structurally and functionally non-equivalent (7, 22). Although eight base pairs (bp) at the very termini of Tn7 form a perfect inverted repeat, more extensive DNA sequence including several copies of a 22 bp motif are required at each end for transposition. In the left end of Tn7 there are three copies of this 22 bp sequence separated by intervening sequences of

* To whom correspondence should be addressed

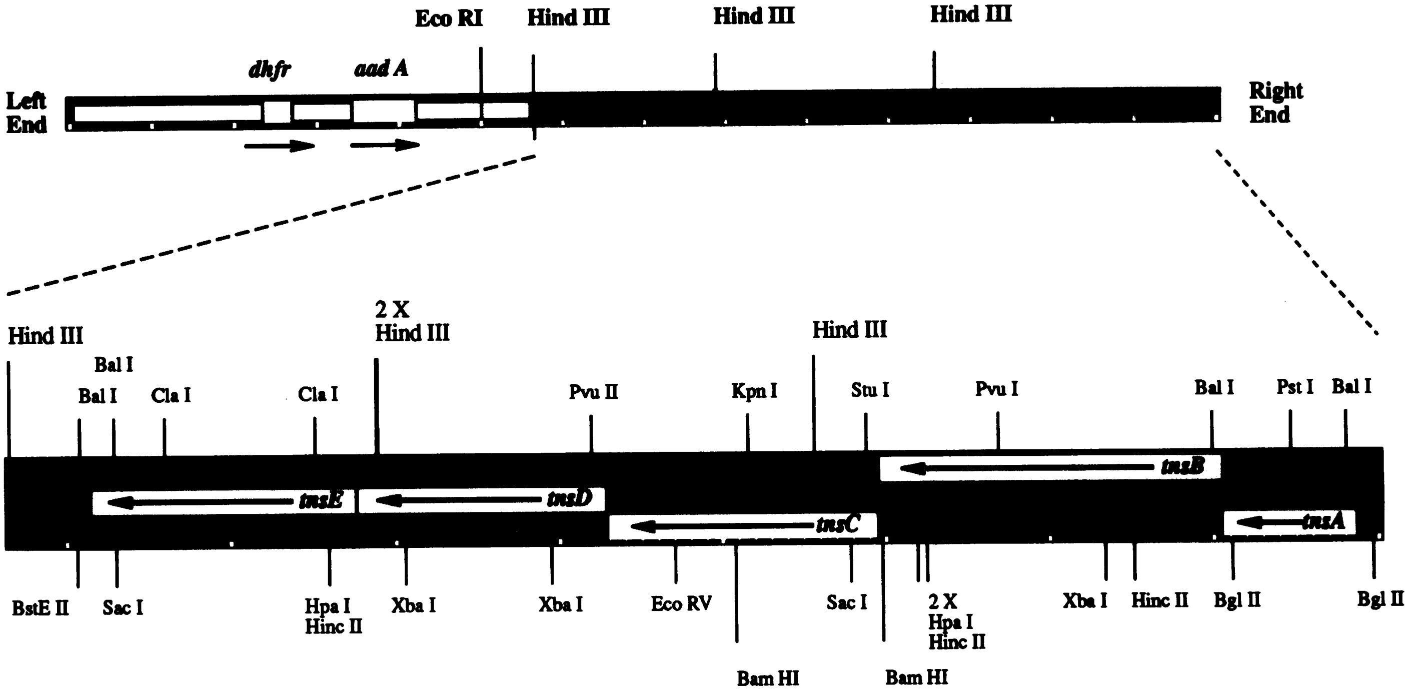


FIGURE 1. Physical and Genetic Map of Tn7. The positions of genes and some restriction enzyme cleavage sites within Tn7 are displayed. The direction of transcription is represented with arrows, and the reading frames of the *tns* genes are indicated by the vertical position of the arrows: top being the first frame. White lines in the lower edge of the rectangles demarcate the distance. Wide lines reflect increments of 1000 bases and narrow lines, increments of 100. The *dhfr* gene is responsible for resistance to trimethoprim and the *aadA* gene encodes streptomycin and spectinomycin resistance. An analysis of restriction endonuclease cleavage sites within this sequence confirms the map of Gosti-Testu *et al.* 1983 (38), with these minor corrections:(i) there are two Hind III sites, 11bp apart at positions 6111 and 6122 instead of one, (ii) there is an extra Ava I site at nt. 4263 (sites not shown here), (iii) there are two additional cleavage sites for Hpa I at position 2751 and 2811, (iv) two Hinc II cleavage sites occur close together at nt. position 2751 and 2811 instead of one, and there is an additional site at nt. number 6413. These are in addition to the corrections observed in Ref. 32.

variable length, and all three copies are required to comprise a functional left end (22). The right end contains four contiguous occurances of this motif; the three terminal copies are sufficient to allow transposition. Tn7-end derivatives that contain two left ends in the appropriate, inverted orientation do not transpose while similar derivatives with two right ends do (22).

Tn7 displays in common with some other bacterial transposons a phenomenon called transpositional immunity, which means that the presence of a copy of the transposon (or even a single transposon end) in a target replicon greatly reduces the frequency of subsequent transposition to that replicon (19, 22, 23, 24). This effect is observed even over relatively large (> 50 Kb) distances. The mechanism of immunity is best understood for the transposing bacteriophage Mu, (25) (see the discussion of possible roles of the Tns proteins).

As with almost every other transposon a duplication of target sequence accompanies Tn7 transposition. The length of target duplication is characteristic for each particular transposon; thus transposition of Tn7 generates a five base pair duplication (7).

The DNA sequence of a region that spans the five *tns* genes has been completed and compiled. The sequence contains five long open reading frames that correspond very closely to the positions determined for the five *tns* genes. This paper reports our analysis of this region.

MATERIALS AND METHODS

Bacterial strain and M13 derived clones

The *E. coli* K12 strain, TG2, was used as a host for the growth of all M13 derived clones. The genotype is: Δlac, *pro*, $suIII^+$, *thi, recA, srl*::Tn10, hsd_5, $EcoK_{r-\ m-}$, (F′ $traD_{36}$, *proAB*, $lacI_Q$ $lacZ_{\Delta M15}$), (gift of Toby Gibson).

Restriction fragments of Tn7 spanning from the Pst I site at nucleotide (nt.) 532 to the Hpa I site at nt. 6413 were purified, and cloned into M13mp10, mp11, mp18 and mp19 (26, 27). Some of the clones with larger inserts were subjected to unidirectional deletion into the Tn7 DNA using exonuclease III (28), such that long stretches of overlapping sequence data with different start-points could be collected from one original clone.

In vitro deletion using exonuclease III

The method was essentially that described by Hennikoff (28). Double stranded RF DNA was prepared and cleaved with two restriction enzymes: the one proximal to the universal priming site creates 5′ recessed ends (which are protected from exonuclease III attack), and the Tn7-insert proximal one creates 5′ overhanging or blunt ends (which are sensitive). Exo III digestions were carried out at a temperature appropriate for the rate of digestion desired, aliquots were removed at various time points, and treated with S1 nuclease to remove the single stranded ends before recircularizing with T4 DNA ligase and transforming competent *E. coli* (TG2) cells, (29).

DNA sequence determination

Single-stranded DNA templates were prepared and sequenced by the dideoxy chain termination method (30), as modified by Biggin *et al.* (31), using [^{35}S] α thio-dATP (Amersham International plc).

```
                                                                            -35                         -10   ➡ PR
                            30      Bgl II                60                            90                        |  120
TGTGGGCGGACAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAATAAAGTCTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAAAAGCAT

                           150                           180                           210                           240
ACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCTGAAGTGCAAATTGCCCGTCGTATTAAAGAGGGGCGTGGCCAAGGGCATGGTAAAGACTATATTCCATGGCTAACAGTAC
               M  A  K  A  N  S  S  F  S  E  V  Q  I  A  R  R  I  K  E  G  R  G  Q  G  H  G  K  D  Y  I  P  W  L  T  V

            ➡ tnsA
                           270                           300                           330                           360
AAGAAGTTCCTTCTTCAGGTCGTTCCCACCGTATTTATTCTCATAAGACGGGACGAGTCCATCATTTGCTATCTGACTTAGAGCTTGCTGTTTTTCTCAGTCTTGAGTGGGAGAGCAGCG
Q  E  V  P  S  S  G  R  S  H  R  I  Y  S  H  K  T  G  R  V  H  H  L  L  S  D  L  E  L  A  V  F  L  S  L  E  W  E  S  S

                           390                           420                           450                           480
TGCTAGATATACGCGAGCAGTTCCCCTTATTACCTAGTGATACCAGGCAGATTGCAATAGATAGTGGTATTAAGCATCCTGTTATTCGTGGTGTAGATCAGGTTATGTCTACTGATTTTT
V  L  D  I  R  E  Q  F  P  L  L  P  S  D |T  R  Q  I  A  I  D  S  G  I  K  H  P  V  I  R  G  V  D| Q  V  M  S  T  D  F

                           510                      Pst I540                           570                           600
TAGTGGACTGCAAAGATGGTCCTTTTGAGCAGTTTGCTATTCAAGTCAAACCTGCAGCAGCCTTACAAGACGAGCGTACCTTAGAAAAACTAGAACTAGAGCGTCGCTATTGGCAGCAAA
L  V  D  C  K  D  G  P  F  E  Q  F  A  I  Q  V  K  P  A  A  A  L  Q  D  E  R  T  L  E  K  L  E  L  E  R  R  Y  W  Q  Q

                           630                           660                           690                           720
AGCAAATTCCTTGGTTCATTTTTACTGATAAAGAAATAAATCCCGTAGTAAAAGAAAATATTGAATGGCTTTATTCAGTGAAAACAGAAGAAGTTTCTGCGGAGCTTTTAGCACAACTAT
K  Q  I  P  W  F  I  F  T  D  K  E  I  N  P  V  V  K  E  N  I  E  W  L  Y  S  V  K  T  E  E  V  S  A  E  L  L  A  Q  L

                           750                           780                           810                           840
CCCCATTGGCCCATATCCTGCAAGAAAAAGGAGATGAAAACATTATCAATGTCTGTAAGCAGGTTGATATTGCTTATGATTTGGAGTTAGGCAAAACATTGAGTGAGATACGAGCCTTAA
S  P  L  A  H  I  L  Q  E  K  G  D  E  N  I  I  N  V  C  K  Q  V  D  I  A  Y  D  L  E  L  G  K  T  L  S  E  I  R  A  L

                           870                           900 Bgl II                    930                           960
CCGCAAATGGTTTTATTAAGTTCAATATTTATAAGTCTTTCAGGGCAAATAAGTGTGCAGATCTCTGTATTAGCCAAGTAGTGAATATGGAGGAGTTGCGCTATGTGGCAAATTAATGAG
                                                                                                       M  W  Q  I  N  E
T  A  N  G  F  I  K  F  N  I  Y  K  S  F  R  A  N  K  C  A  D  L  C  I  S  Q  V  V  N  M  E  E  L  R  Y  V  A  N  *  *

                                                                                                    ➡ tnsB
                           990                          1020                          1050                          1080
GTTGTGCTATTTGATAATGATCCGTATCGCATTTTGGCTATAGAGGATGGCCAAGTTGTCTGGATGCAAATAAGCGCTGATAAAGGAGTTCCACAAGCTAGGGCTGAGTTGTTGCTAATG
 V  V  L  F  D  N  D  P  Y  R  I  L  A  I  E  D  G  Q  V  V  W  M  Q  I  S  A  D  K  G  V  P  Q  A  R  A  E  L  L  L  M

                          1110                          1140                          1170                          1200
CAGTATTTAGATGAAGGCCGCTTAGTTAGAACTGATGACCCTTATGTACATCTTGATTTAGAAGAGCCGTCTGTAGATTCTGTCAGCTTCCAGAAGCGCGAGGAGGATTATCGAAAAATT
 Q  Y  L  D  E  G  R  L  V  R  T  D  D  P  Y  V  H  L  D  L  E  E  P  S  V  D  S  V  S  F  Q  K  R  E  E  D  Y  R  K  I

                          1230                          1260                          1290                          1320
CTTCCTATTATTAATAGTAAGGATCGTTTCGACCCTAAAGTCAGAAGCGAACTCGTTGAGCATGTGGTCCAAGAACATAAGGTTACTAAGGCTACAGTTTATAAGTTGTTACGCCGTTAC
 L  P  I  I  N  S  K  D  R  F  D  P  K  V  R  S  E  L |V  E  H  V  V  Q  E  H  K  V  T  K  A  T  V  Y  K  L  L  R| R  Y

                          1350                          1380                          1410                          1440
TGGCAGCGTGGTCAAACGCCTAATGCATTAATTCCTGACTACAAAAACAGCGGTGCACCAGGGGAAAGACGTTCAGCGACAGGAACAGCAAAGATTGGCCGAGCCAGAGAATATGGTAAG
 W  Q  R  G  Q  T  P  N  A  L  I  P  D  Y  K  N  S  G  A  P  G  E  R  R  S  A  T  G  T  A  K  I  G  R  A  R  E  Y  G  K

                          1470                          1500                          1530                          1560
GGTGAAGGAACCAAGGTAACGCCCGAGATTGAACGCCTTTTTAGGTTGACCATAGAAAAGCACCTGTTAAATCAAAAAGGTACAAAGACCACCGTTGCCTATAGACGATTTGTGGACTTG
 G  E  G  T  K  V  T  P  E  I  E  R  L  F  R  L  T  I  E  K  H  L  L  N  Q  K  G  T  K  T  T  V  A  Y  R  R  F  V  D  L

                          1590                          1620                          1650                          1680
TTTGCTCAGTATTTTCCTCGCATTCCCCAAGAGGATTACCCAACACTACGTCAGTTTCGTTATTTTTATGATCGAGAATACCCTAAAGCTCAGCGCTTAAAGTCTAGAGTTAAAGCAGGG
 F  A  Q  Y  F  P  R  I  P  Q  E  D  Y  P  T  L  R  Q  F  R  Y  F  Y  D  R  E  Y  P  K  A  Q  R  L  K  S  R  V  K  A  G

                          1710                          1740                          1770                          1800
GTATATAAAAAAGACGTACGACCCTTAAGTAGTACAGCCACTTCTCAGGCGTTAGGCCCTGGGAGTCGTTATGAGATTGATGCCACGATTGCTGATATTTATTTAGTGGATCATCATGAT
 V  Y  K  K  D  V  R  P  L  S  S  T  A  T  S  Q  A  L  G  P  G  S  R  Y  E  I  D  A  T  I  A  D  I  Y  L  V  D  H  H  D

                          1830                          1860                          1890                          1920
CGCCAAAAAATCATAGGAAGACCAACGCTTTACATTGTGATTGATGTGTTTAGTCGGATGATCACGGGCTTTTATATCGGCTTTGAAAATCCGTCTTATGTGGTGGCGATGCAGGCTTTT
 R  Q  K  I  I  G  R  P  T  L  Y  I  V  I  D  V  F  S  R  M  I  T  G  F  Y  I  G  F  E  N  P  S  Y  V  V  A  M  Q  A  F

                          1950                          1980                          2010                          2040
GTAAATGCTTGCTCTGACAAAACGGCCATTTGTGCCCAGCATGATATTGAGATTAGTAGCTCAGACTGGCCGTGTGTAGGTTTGCCAGATGTGTTGCTAGCGGACCGTGGCGAATTAATG
 V  N  A  C  S  D  K  T  A  I  C  A  Q  H  D  I  E  I  S  S  S  D  W  P  C  V  G  L  P  D  V  L  L  A  D  R  G  E  L  M

                          2070                          2100                          2130                          2160
AGTCATCAGGTCGAAGCCTTAGTTTCTAGTTTTAATGTGCGAGTGGAAAGTGCTCCACCTAGACGTGGCGATGCTAAAGGCATAGTGGAAAGCACTTTTAGAACACTACAAGCCGAGTTT
 S  H  Q  V  E  A  L  V  S  S  F  N  V  R  V  E  S  A  P  P  R  R  G  D  A  K  G  I  V  E  S  T  F  R  T  L  Q  A  E  F

                          2190                          2220                          2250                          2280
AAGTCCTTTGCACCTGGCATTGTAGAGGGCAGTCGGATCAAAAGCCATGGTGAAACAGACTATAGGTTAGATGCATCTCTGTCGGTATTTGAGTTCACACAAATTATTTTGCGTACGATC
 K  S  F  A  P  G  I  V  E  G  S  R  I  K  S  H  G  E  T  D  Y  R  L  D  A  S  L  S  V  F  E  F  T  Q  I  I  L  R  T  I

                          2310                          2340                          2370                          2400
TTATTCAGAAATAACCATCTGGTGATGGATAAATACGATCGAGATGCTGATTTTCCTACAGATTTACCGTCTATTCCTGTCCAGCTATGGCAATGGGGTATGCAGCATCGTACAGGTAGT
 L  F  R  N  N  H  L  V  M  D  K  Y  D  R  D  A  D  F  P  T  D  L  P  S  I  P  V  Q  L  W  Q  W  G  M  Q  H  R  T  G  S

                          2430                          2460                          2490                          2520
TTAAGGGCTGTGGAGCAAGAGCAGTTGCGAGTAGCGTTACTGCCTCGCCGAAAGGTCTCTATTTCTTCATTTGGCGTTAATTTGTGGGGTTTGTATTACTCGGGGTCAGAGATTCTGCGT
 L  R  A  V  E  Q  E  Q  L  R  V  A  L  L  P  R  R  K  V  S  I  S  S  F  G  V  N  L  W  G  L  Y  Y  S  G  S  E  I  L  R

                          2550                          2580                          2610                          2640
GAGGGTTGGTTGCAGCGGAGCACTGATATAGCTAGACCTCAACATTTAGAAGCGGCTTATGACCCAGTGCTGGTTGATACGATTTATTTGTTTCCGCAAGTTGGCAGCCGTGTATTTTGG
 E  G  W  L  Q  R  S  T  D  I  A  R  P  Q  H  L  E  A  A  Y  D  P  V  L  V  D  T  I  Y  L  F  P  Q  V  G  S  R  V  F  W

                          2670                          2700                          2730                          2760
CGCTGTAATCTGACGGAACGTAGTCGGCAGTTTAAAGGTCTCTCATTTTGGGAGGTTTGGGATATACAAGCACAAGAAAAACACAATAAAGCCAATGCGAAGCAGGATGAGTTAACTAAA
 R  C  N  L  T  E  R  S  R  Q  F  K  G  L  S  F  W  E  V  W  D  I  Q  A  Q  E  K  H  N  K  A  N  A  K  Q  D  E  L  T  K

                          2790                          2820                          2850                          2880
CGCAGGGAGCTTGAGGCGTTTATTCAGCAAACCATTCAGAAAGCGAATAAGTTAACGCCCAGTACTACTGAGCCCAAATCAACACGCATTAAGCAGATTAAAACTAATAAAAAAGAAGCC
 R  R  E  L  E  A  F  I  Q  Q  T  I  Q  K  A  N  K  L  T  P  S  T  T  E  P  K  S  T  R  I  K  Q  I  K  T  N  K  K  E  A
```

```
                          2910                          2940                          2970                          3000
GTGACCTCGGAGCGTAAAAAACGTGCGGAGCATTTGAAGCCAAGCTCTTCAGGTGATGAGGCTAAAGTTATTCCTTTCAACGCAGTGGAAGCGGATGATCAAGAAGATTACAGCCTACCC
 V  T  S  E  R  K  K  R  A  E  H  L  K  P  S  S  S  G  D  E  A  K  V  I  P  F  N  A  V  E  A  D  D  Q  E  D  Y  S  L  P

                   Bam HI 3030                          3060                          3090                          3120
ACATACGTGCCTGAATTATTTCAGGATCCACCAGAAAAGGATGAGTCATGAGTGCTACCCGGATTCAAGCAGTTTATCGTGATACGGGGGTAGAGGCTTATCGTGATAATCCTTTTATCG
 T  Y  V  P  E  L  F  Q  D  P  P  E  K  D  E  S  *
                                                M  S  A  T  R  I  Q  A  V  Y  R  D  T  G  V  E  A  Y  R  D  N  P  F  I
                                              ➡ tnsC

                          3150                          3180                          3210                          3240
AGGCCTTACCACCATTACAAGAGTCAGTGAATAGTGCTGCATCACTGAAATCCTCTTTACAGCTTACTTCCTCTGACTTGCAAAAGTCCCGTGTTATCAGAGCTCATACCATTTGTCGTA
 E  A  L  P  P  L  Q  E  S  V  N  S  A  A  S  L  K  S  S  L  Q  L  T  S  S  D  L  Q  K  S  R  V  I  R  A  H  T  I  C  R

                          3270                          3300                          3330                          3360
TTCCAGATGACTATTTCAGCCATTAGGTACGCATTTGCTACTAAGTGAGCGTATTTCGGTCATGATTCGAGGTGGCTACGTAGGCAGAAATCCTAAAACAGGAGATTTACAAAAGCATT
 I  P  D  D  Y  F  Q  P  L  G  T  H  L  L  L  S  E  R  I  S  V  M  I  R  G  G  Y  V  G  R  N  P  K  T  G  D  L  Q  K  H

                          3390                          3420                Hind III  3450                          3480
TACAAAATGGTTATGAGCGTGTTCAAACGGGAGAGTTGGAGACATTTCGCTTTGAGGAGGCACGATCTACGGCACAAAGCTTATTGTTAATTGGTTGTTCTGGTAGTGGGAAGACGACCT
 L  Q  N  G  Y  E  R  V  Q  T  G  E  L  E  T  F  R  F  E  E  A  R  S  T  A  Q  S  L  L  L  I  G  C  S  G  S  G  K  T  T

                          3510                          3540                          3570                          3600
CTCTTCATCGTATTCTAGCCACGTATCCTCAGGTGATTTACCATCGTGAACTCAATGTAGAGCAGGTGGTGTATTTGAAAATAGACTGCTCGCATAATGGTTCGCTAAAAGAAATCTGCT
 S  L  H  R  I  L  A  T  Y  P  Q  V  I  Y  H  R  E  L  N  V  E  Q  V  V  Y  L  K  I  D  C  S  H  N  G  S  L  K  E  I  C

                          3630                          3660                          3690                          3720
TGAATTTTTTCAGAGCGTTGGATCGAGCCTTGGGCTCGAACTATGAGCGTCGTTATGGCTTAAAACGTCATGGTATAGAAACCATGTTGGCTTTGATGTCGCAAATAGCCAATGCACATG
 L  N  F  F  R  A  L  D  R  A  L  G  S  N  Y  E  R  R  Y  G  L  K  R  H  G  I  E  T  M  L  A  L  M  S  Q  I  A  N  A  H

                          3750                          3780                          3810                          3840
CTTTAGGGTTGTTGGTTATTGATGAAATTCAGCATTTAAGCCGCTCTCGTTCGGGTGGATCTCAAGAGATGCTGAACTTTTTTGTGACGATGGTGAATATTATTGGCGTACCAGTGATGT
 A  L  G  L  L  V  I  D  E  I  Q  H  L  S  R  S  R  S  G  G  S  Q  E  M  L  N  F  F  V  T  M  V  N  I  I  G  V  P  V  M

                          3870                          3900                   Bam HI 3930                          3960
TGATTGGTACCCCTAAAGCACGAGAGATTTTTGAGGCTGATTTGCGGTCTGCACGTAGAGGGGCAGGGTTTGGAGCTATATTCTGGGATCCTATACAACAAACGCAACGTGGAAAGCCCA
 L  I  G  T  P  K  A  R  E  I  F  E  A  D  L  R  S  A  R  R  G  A  G  F  G  A  I  F  W  D  P  I  Q  Q  T  Q  R  G  K  P

                          3990                          4020                          4050                          4080
ATCAAGAGTGGATCGCTTTTACGGATAATCTTTGGCAATTACAGCTTTTACAACGCAAAGATGCGCTGTTATCGGATGAGGTCCGTGATGTGTGGTATGAGCTAAGCCAAGGAGTGATGG
 N  Q  E  W  I  A  F  T  D  N  L  W  Q  L  Q  L  L  Q  R  K  D  A  L  L  S  D  E  V  R  D  V  W  Y  E  L  S  Q  G  V  M

                          4110                          4140                          4170                          4200
ACATTGTAGTAAAACTTTTTGTACTCGCTCAGCTCCGTGCGCTAGCTTTAGGCAATGAGCGTATTACCGCTGGTTTATTGCGGCAAGTGTATCAAGATGAGTTAAAGCCTGTGCACCCCA
 D  I  V  V  K  L  F  V  L  A  Q  L  R  A  L  A  L  G  N  E  R  I  T  A  G  L  L  R  Q  V  Y  Q  D  E  L  K  P  V  H  P

                          4230                          4260                          4290                          4320
TGCTAGAGGCATTACGCTCGGGTATCCCAGAACGCATTGCTCGTTATTCTGATCTAGTCGTTCCCGAGATTGATAAACGGTTAATCCAACTTCAGCTAGATATCGCAGCGATACAAGAAC
 M  L  E  A  L  R  S  G  I  P  E  R  I  A  R  Y  S  D  L  V  V  P  E  I  D  K  R  L  I  Q  L  Q  L  D  I  A  A  I  Q  E

                          4350                          4380                          4410                          4440
AAACACCAGAAGAAAAAGCCCTTCAAGAGTTAGATACCGAAGATCAGCGTCATTTATATCTGATGCTGAAAGAGGATTACGATTCAAGCCTGTTAATTCCCACTATTAAAAAAGCGTTTA
 Q  T  P  E  E  K  A  L  Q  E  L  D  T  E  D  Q  R  H  L  Y  L  M  L  K  E  D  Y  D  S  S  L  L  I  P  T  I  K  K  A  F

                          4470                          4500                          4530                          4560
GCCAGAATCCAACGATGACAAGACAAAAGTTACTGCCTCTTGTTTTGCAGTGGTTGATGGAAGGCGAAACGGTAGTGTCAGAACTAGAAAAGCCCTCCAAGAGTAAAAAGGTTTCGGCTA
 S  Q  N  P  T  M  T  R  Q  K  L  L  P  L  V  L  Q  W  L  M  E  G  E  T  V  V  S  E  L  E  K  P  S  K  S  K  K  V  S  A

                          4590                          4620                          4650                          4680
TAAAGGTAGTCAAGCCCAGCGACTGGGATAGCTTGCCTGATACGGATTTACGTTATATCTATTCACAACGCCAACCTGAAAAAACCATGCATGAACGGTTAAAAGGAAAGGGGTAATAG
 I  K  V  V  K  P  S  D  W  D  S  L  P  D  T  D  L  R  Y  I  Y  S  Q  R  Q  P  E  K  T  M  H  E  R  L  K  G  K  G  V  I

                          4710                          4740                          4770                          4800
TGGATATGGCGAGCTTATTTAAACAAGCAGGTTAGCCATGAGAAACTTTCCTGTTCCGTACTCGAATGAGCTGATTTATAGCACTATTGCACGGGCAGGCGTTTATCAAGGGATTGTTAG
                                      M  R  N  F  P  V  P  Y  S  N  E  L  I  Y  S  T  I  A  R  A  G  V  Y  Q  G  I  V  S
 V  D  M  A  S  L  F  K  Q  A  G  *    ➡ tnsD

                          4830                          4860                          4890                          4920
TCCTAAGCAGCTGTTGGATGAGGTGTATGGCAACCGCAAGGTGGTCGCTACCTTAGGTCTGCCCTCGCATTTAGGTGTGATAGCAAGACATCTACATCAAACAGGACGTTACGCTGTTCA
  P  K  Q  L  L  D  E  V  Y  G  N  R  K  V  V  A  T  L  G  L  P  S  H  L  G  V  I  A  R  H  L  H  Q  T  G  R  Y  A  V  Q

                          4950                          4980                          5010                          5040
GCAGCTTATTTATGAGCATACCTTATTCCCTTTATATGCTCCGTTTGTAGGCAAGGAGCGCCGAGACGAAGCTATTCGGTTAATGGAGTACCAAGCGCAAGGTGCGGTGCATTTAATGCT
  Q  L  I  Y  E  H  T  L  F  P  L  Y  A  P  F  V  G  K  E  R  R  D  E  A  I  R  L  M  E  Y  Q  A  Q  G  A  V  H  L  M  L

                          5070                          5100                          5130                          5160
AGGAGTCGCTGCTTCTAGAGTTAAGAGCGATAACCGCTTTAGATACTGCCCTGATTGCGTTGCTCTTCAGCTAAATAGGTATGGGGAAGCCTTTTGGCAACGAGATTGGTATTTGCCCGC
  G  V  A  A  S  R  V  K  S  D  N  R  F  R  Y  C  P  D  C  V  A  L  Q  L  N  R  Y  G  E  A  F  W  Q  R  D  W  Y  L  P  A

                          5190                          5220                          5250                          5280
TTTGCCATATTGTCCAAAACACGGTGCTTTAGTCTTCTTTGATAGAGCTGTAGATGATCACCGACATCAATTTTGGGCTTTGGGTCATACTGAGCTGCTTTCAGACTACCCCAAAGACTC
  L  P  Y  C  P  K  H  G  A  L  V  F  F  D  R  A  V  D  D  H  R  H  Q  F  W  A  L  G  H  T  E  L  L  S  D  Y  P  K  D  S

                          5310                          5340                          5370                          5400
CCTATCTCAATTAACAGCACTAGCTGCTTATATAGCCCCTCTGTTAGATGCTCCACGAGCGCAAGAGCTTTCCCCAAGCCTTGAGCAGTGGACGCTGTTTTATCAGCGCTTAGCGCAGGA
  L  S  Q  L  T  A  L  A  A  Y  I  A  P  L  L  D  A  P  R  A  Q  E  L  S  P  S  L  E  Q  W  T  L  F  Y  Q  R  L  A  Q  D

                          5430                          5460                          5490                          5520
TCTAGGGCTAACCAAAAGCAAGCACATTCGTCATGACTTGGTGGCGGAGAGAGTGAGGCAGACTTTTAGTGATGAGGCACTAGAGAAACTGGATTTAAAGTTGGCAGAGAACAAGGACAC
  L  G  L  T  K  S  K  H  I  R  H  D  L  V  A  E  R  V  R  Q  T  F  S  D  E  A  L  E  K  L  D  L  K  L  A  E  N  K  D  T

                          5550                          5580                          5610                          5640
GTGTTGGCTGAAAAGTATATTCCGTAAGCATAGAAAAGCCTTTAGTTATTTACAGCATAGTATTGTGTGGCAAGCCTTATTGCCAAAACTAACGGTTATAGAAGCGCTACAGCAGGCAAG
  C  W  L  K  S  I  F  R  K  H  R  K  A  F  S  Y  L  Q  H  S  I  V  W  Q  A  L  L  P  K  L  T  V  I  E  A  L  Q  Q  A  S

                          5670                          5700                          5730                          5760
TGCTCTTACTGAGCACTCTATAACGACAAGACCTGTTAGCCAGTCTGTGCAACCTAACTCTGAAGATTTATCTGTTAAGCATAAAGACTGGCAGCAACTAGTGCATAAATACCAAGGAAT
  A  L  T  E  H  S  I  T  T  R  P  V  S  Q  S  V  Q  P  N  S  E  D  L  S  V  K  H  K  D  W  Q  Q  L  V  H  K  Y  Q  G  I
```

```
                          5790                          5820                          5850                          5880
TAAGGCGGCAAGACAGTCTTTAGAGGGTGGGGTGCTATACGCTTGGCTTTACCGACATGACAGGGATTGGCTAGTTCACTGGAATCAACAGCATCAACAAGAGCGTCTGGCACCCGCCCC
 K  A  A  R  Q  S  L  E  G  G  V  L  Y  A  W  L  Y  R  H  D  R  D  W  L  V  H  W  N  Q  Q  H  Q  Q  E  R  L  A  P  A  P

                          5910                          5940                          5970                          6000
TAGAGTTGATTGGAACCAAAGAGATCGAATTGCTGTACGACAACTATTAAGAATCATAAAGCGTCTAGATAGTAGCCTTGATCACCCAAGAGCGACATCGAGCTGGCTGTTAAAGCAAAC
 R  V  D  W  N  Q  R  D  R  I  A  V  R  Q  L  L  R  I  I  K  R  L  D  S  S  L  D  H  P  R  A  T  S  S  W  L  L  K  Q  T

                          6030                          6060                          6090                 Hind III 6120
TCCTAACGGAACCTCTCTTGCAAAAAATCTACAGAAACTGCCTTTGGTAGCGCTTTGCTTAAAGCGTTACTCAGAGAGTGTGGAAGATTATCAAATTAGACGGATTAGCCAAGCTTTTAT
 P  N  G  T  S  L  A  K  N  L  Q  K  L  P  L  V  A  L  C  L  K  R  Y  S  E  S  V  E  D  Y  Q  I  R  R  I  S  Q  A  F  I

Hind III                  6150                          6180                          6210                          6240
TAAGCTTAAACAGGAAGATGTTGAGCTTAGGCGCTGGCGATTATTAAGAAGTGCAACGTTATCTAAAGAGCGGATAACTGAGGAAGCACAAAGATTCTTGGAAATGGTTTATGGGGAAGA
 K  L  K  Q  E  D  V  E  L  R  R  W  R  L  L  R  S  A  T  L  S  K  E  R  I  T  E  E  A  Q  R  F  L  E  M  V  Y  G  E  E

   ___                    6270                          6300                          6330                      ___ 6360
GTGAGTGGTTAGGCTAGCTACATTTAATGACAATGTGCAGGTTGTACATATTGGTCATTTATTCCGTAACTCGGGTCATAAGGAGTGGCGTATTTTTGTTTGGTTTAATCCAATGCAAGA
 *  V  V  R  L  A  T  F  N  D  N  V  Q  V  V  H  I  G  H  L  F  R  N  S  G  H  K  E  W  R  I  F  V  W  F  N  P  M  Q  E
   ➥tnsE

                          6390                          6420                          6450                          6480
ACGGAAATGGACTCGATTTACTCATTTGCCTTTATTAAGTCGAGCTAAGGTGGTTAACAGTACAACAAAGCAAATAAATAAGGCGGATCGTGTGATTGAGTTTGAAGCATCGGATCTTCA
 R  K  W  T  R  F  T  H  L  P  L  L  S  R  A  K  V  V  N  S  T  T  K  Q  I  N  K  A  D  R  V  I  E  F  E  A  S  D  L  Q

                          6510                          6540                          6570                          6600
ACGAGCCAAAATAATCGATTTTCCTAATCTCTCGTCCTTTGCTTCCGTACGCAACAAGGATGGAGCGCAGAGTTCATTTATTTACGAAGCTGAAACACCATATAGCAAGACTCGTTATCA
 R  A  K  I  I  D  F  P  N  L  S  S  F  A  S  V  R  N  K  D  G  A  Q  S  S  F  I  Y  E  A  E  T  P  Y  S  K  T  R  Y  H

                          6630                          6660                          6690                          6720
CATCCCACAGTTAGAGCTAGCTCGGTCATTATTTTTAATTAACTCCTATTTCTGTCGAAGCTGTTTGAGCAGTACCGCTTTACAGCAAGAGTTCGACGTTCAGTATGAGGTTGAGCGAGA
 I  P  Q  L  E  L  A  R  S  L  F  L  I  N  S  Y  F  C  R  S  C  L  S  S  T  A  L  Q  Q  E  F  D  V  Q  Y  E  V  E  R  D

                          6750                          6780                          6810                          6840
TCATTTAGAGATAAGGATCTTACCCAGTTCATCGTTTCCTAAAGGGGCGTTAGAGCAGTCGGCCGTAGTGCAGCTTTTGGTTTGGTTGTTTTCGGATCAAGATGTTATGGATTCGTATGA
 H  L  E  I  R  I  L  P  S  S  S  F  P  K  G  A  L  E  Q  S  A  V  V  Q  L  L  V  W  L  F  S  D  Q  D  V  M  D  S  Y  E

                          6870                          6900                          6930                          6960
AAGTATTTTTAGGCACTATCAACAAAATAGAGAAATTAAGAACGGCGTTGAAAGCTGGTGCTTTAGCTTTGACCCTCCGCCCATGCAGGGTTGGAAATTACATGTAAAAGGACGTTCTTC
 S  I  F  R  H  Y  Q  Q  N  R  E  I  K  N  G  V  E  S  W  C  F  S  F  D  P  P  P  M  Q  G  W  K  L  H  V  K  G  R  S  S

                          6990                          7020                          7050                          7080
TAACGAGGATAAGGATTATTTAGTTGAGGAAATAGTAGGTTTAGAAATCAACGCTATGCTTCCTAGCACAACAGCTATTAGCCATGCCTCTTTTCAGGAAAAGGAGGCAGGTGATGGTAG
 N  E  D  K  D  Y  L  V  E  E  I  V  G  L  E  I  N  A  M  L  P  S  T  T  A  I  S  H  A  S  F  Q  E  K  E  A  G  D  G  S

                          7110                          7140                          7170                          7200
TACGCAGCACATAGCGGTTTCAACAGAGTCAGTTGTTGATGATGAGCATCTACAGTTGGACGATGAGGAAACAGCCAATATAGACACAGACACACGAGTCATAGAGGCTGAGCCGACATG
 T  Q  H  I  A  V  S  T  E  S  V  V  D  D  E  H  L  Q  L  D  D  E  E  T  A  N  I  D  T  D  T |R  V  I  E  A  E  P  T  W|

                          7230                          7260                          7290                          7320
GATAAGTTTTAGTAGACCTAGTCGAATTGAAAAATCTCGCAGGGCAAGAAAAAGTAGCCAAACTATTTTAGAAAAAGAAGAAGCAACAACAAGTGAAAATAGTAATTTGGTTAGTACTGA
|I  S  F  S  R  P  S  R  I  E  K| S  R  R  A  R  K  S  S  Q  T  I  L  E  K  E  E  A  T  T  S  E  N  S  N  L  V  S  T  D

                          7350                          7380                          7410                          7440
TGAGCCACACTTAGGTGGTGTCCTAGCAGCGGCAGATGTGGGTGGAAGCAGGATGCAACCAATTACAACTCTATTTTTGCTAATCGATTTGCTGCTTTTGATGAGCTACTTTCAATTCT
 E  P  H  L  G  G  V  L  A  A  A  D  V  G  G  K  Q  D  A  T  N  Y  N  S  I  F  A  N  R  F  A  A  F  D  E  L  L  S  I  L

                          7470                          7500                          7530                          7560
AAAAACTAAATTTGCATGTCGGGTGCTTTTTGAAGAAACCTTGGTTTTGCCAAAAGTTGGGCGTAGCCGATTACATCTGTGTAAAGATGGCTCACCAAGAGTGATTAAAGCCGTTGGGGT
 K  T  K  F  A  C  R  V  L  F  E  E  T  L  V  L  P  K  V  G  R  S  R  L  H  L  C  K  D  G  S  P  R  V  I  K  A  V  G  V

                          7590                          7620                          7650                          7680
GCAACGTAATGGCAGTGAATTTGTATTGCTAGAGGTGGATGCATCGGATGGGGTGAAAATGCTTTCTACCAAAGTGTTGAGTGGCGTTGATAGCGAAACATGGCGGAATGATTTTGAAAA
 Q  R  N  G  S  E  F  V  L  L  E  V  D  A  S  D  G  V  K  M  L  S  T  K  V  L  S  G  V  D  S  E  T  W  R  N  D  F  E  K

                          7710                          7740                          7770                          7800
GATACGGCGTGGAGTGGTGAAGAGCTCATTGAATTGGCCAAATAGTTTGTTTGATCAATTATATGGACAAGACGGGCATAGAGGGGTGAATCATCCAAAGGGGTTGGGGGAGCTGCAAGT
 I  R  R  G  V  V  K  S  S  L  N  W  P  N  S  L  F  D  Q  L  Y  G  Q  D  G  H  R  G  V  N  H  P  K  G  L  G  E  L  Q  V

                          7830                          7860                          7890                          7920
ATCGAGAGAGGATATGGAAGGGTGGGCTGAGAGAGTGGTTAGAGAGCAATTTACGCATTAAAGGAATGACTGAAAGAGCCTGTAAACCCTTTTGTGTAAGTGCTTTTGCCGGTCAGTTAA
 S  R  E  D  M  E  G  W  A  E  R  V  V  R  E  Q  F  T  H  *

                          7950                          7980                          8010                          8040
AGGTGGCCATTTAAACGGTCACCAAATTCGATCATAAAACGGTTCATGGCCGGCTTCCAGTTGCGGATCGGCATCGTCCATTTCTTGGTCGCCGCCTGGATAGCCAGGTACACCACCTTC

                          8070                          8100                          8130                          8160
ATCGCTGATTCGTCCGTAGGGAACACCTTGAGTTCATATTAATGGAATTTTCTACAAATAGCCTCCGTGGTTTTGAGGGGGGATTACAGACGATCCATAGTAGTAATCCAATGAGTTCTT

                          8190                          8220                          8250                          8280
GAGCGCGGCGACATGTTTGGACGCCTTGGCAAAAATTAGAGCCTGCTTGAAGTGCAGGCGAGCACGTGCTTGGACGATTGATCCATTCGCGGTCAAAAACTCAATCTTGGATGACAGCGT

                          8310                          8340    Hind III
GTCAGGAAATCCAATATCGAAGTCCCGCCCCGTGTCAAAGTAGGGCATTTCATGATCAAAGGACGAAGCTT
```

FIGURE 2. Nucleotide Sequence of the Right End of Tn7. The DNA sequence of the right end of Tn7 to the leftmost Hind III cleavage site is shown. Amino acid sequences appear below the five long open reading frames that correspond to *tnsA* to *tnsE*. In each case conceptual translation begins with the first initiation codon which is overlined, as are some internal, alternative start codons. Tentative ribosome binding sites are underlined. Within the translated (a.a.) sequence of *tnsA*, *tnsB*, *tnsD* and *tnsE*, regions with similarity to the helix-turn-helix, DNA binding domains are boxed. The terminal eight bases in the right end of Tn7 are indicated by a hatched bar underneath and similarly, the four 22 base repeats are indicated by solid black bars. A promoter (from ref. 21) is labeled P_{LE}, with the presumed −35 and −10 regions double overlined. The arrow at nucleotide number 111 marks the 5′ end of transcripts they identified. The locations of recognition sites for the restriction enzymes Bam HI, Bgl II, Hind III and Pst I are labeled and overlined.

Computer programs

The DNA sequence data generated was compiled using the MERGE program of MicroGenie™ (Beckman). Various other analyses of the sequence, such as promoter searches, were done using the Staden™ programs (Amersham International plc). Homology searches against protein sequence databases were done with the very kind help of Roger Staden and John Collins on the VAX at the Laboratory of Molecular Biology, Cambridge and the Distributed Array Processor, University of Edinburgh respectively.

RESULTS AND DISCUSSION

Derivation of DNA Sequence

The newly determined DNA sequence taken together with that presented by Lichtenstein and Brenner (7), Smith and Jones (32), and Gay *et al.* (21), completes an 8351 bp segment, from the right end of Tn7 to the final leftward Hind III site (figures 1 and 2), that encompasses all of the genes required for transposition. The numbering of the nucleotide sequence referred to in this paper begins with the first base of the right end of Tn7 and continues leftward. The sequence from nt. position 1 to 537 is taken from Lichtenstein and Brenner (7), and Gay *et al.* (21), and the sequence from nt. position 3024 to 3926, and from nt. 6122 to 8351 has been presented previously by Smith and Jones (32). We used the chain termination method of Sanger *et al.* (30), to determine the sequence of restriction fragments of Tn7 cloned into the M13 vectors, (26, 27) and of deleted derivatives made *in vitro* by Exonuclease III digestion (see Materials and Methods).

Physical and Genetic Organization

Analysis of the sequence is summarized in figures 1, 2, 3 and 4 and in table 1. This A+T rich sequence (43.5% G+C vs. ≈ 51.7% for *E. coli*) contains five long ORF's, all oriented from right to left, that cover 92% of it. No other ORF's of greater than 128 codons occur in either direction.

The five long open reading frames are in a dense array with adjacent ones either abutting or overlapping, and they coincide very closely to the positions of the *tns* genes mapped genetically by Rogers *et al.*(9) and Waddell and Craig, (8) (See Fig 3 and table 1). This curious arrangement of ORF's is typical of operons where translational coupling occurs (33, 34). In cases of translational coupling, translation of one gene in an operon is dependant on the prior translation of the gene immediately upstream (35).

If translational coupling does occur in Tn7 the genes involved must be co-transcribed, but the DNA sequence reveals very little about transcription of the *tns* genes. There is evidence to imply that *tnsA* and *tnsB* are co-transcribed. Waddell and Craig (8), using cloned fragments of Tn7 to complement insertion mutants of the *tns* genes, found that some insertion mutations in the *tnsA* region could not be complemented by a fragment containing *tnsA*, (*ie.* had polar effects on *tnsB*). A deletion mutant of the *tnsA* gene was complemented by the same fragment. In addition, a fragment containing *tnsA* and *tnsB* could complement both mutations. Results of similar experiments are consistent with the view that the other three *tns* genes are independent transcriptional units (8), but proof of this will require mapping the end-points of authentic transcripts.

	DNA Sequence Analysis				Genetic/Biochemical Analysis		
	Initiation codon ATG/GTG*	Termination codon	Number of a.a.'s	Protein M_r (daltons)	Positions of Genetic Loci (Kb) Begins	 Ends	Apparent Protein M_r (kd)¥
TnsA	135 * 165 * 360 465	954	273 263 198 163	31 275 30 204 23 772 19 886	§ > 0.0	< 1.485	30
					† < 0.1	> 0.8 > 0.95	
TnsB	943 * 964 1024 1078	3049	702 695 675 657	80 825 79 924 77 580 75 657	§ > 0.899	< 3.024	83 - 85
					† > 0.9 < 0.95	> 2.85 < 3.55	
TnsC	3048 *3147 3303	4713	555 522 470	62 995 59 274 53 526	§ > 3.024	< 4.808	54 - 56 / 40 -42
					† < 3.5	> 4.65 < 4.7	
TnsD	4718 *4823 *4841 *4877	6242	508 473 467 455	59 140 55 188 54 470 53 382	§ > 4.299 < 4.808	> 6.111 < 6.494	54 / 40‡
					† < 5.0	> 6.5 < 6.2	
TnsE	*6245 6353	7859	538 502	61 180 56 864	§ > 6.122	< 8.345	85 / 70 - 75
					† < 6.2	> 7.8 < 7.85	

Table 1. Data regarding gene boundaries and protein molecular mass from the DNA sequence is related to published, empirically determined data. The correspondence of the gene locations is very good, while the calculated molecular mass (Mr) of the proteins are in some instances at odds with the apparent Mr's observed. Symbols indicate the source of data; § indicates information from (9), † from (8), ¥ from (4). ‡ indicates that the protein observed was truncated (see text).

No compelling promoters could be located by comparing the *E. coli* promoter consensus to this sequence. It is conceivable that the failure to identify promoters is due to low levels of *tns* gene expression, (weak promoters often show a poor resemblance to the consensus, (36, 37)); yet a promoter has been located in the right end of Tn7 by mapping the 5′ end-point of transcripts to ≈ nt. 111 (21). Based on this experimental evidence the presumed −10 and −35 elements of the promoter have been identified (21). Surprisingly, moderately strong expression of transcriptional and translational fusions occurs down-stream of this promoter (9). This expression is modestly repressed by the presence of the *tnsB* region in *trans*, (see below) (9).

tnsA

The sequence of the right end of Tn7 as well as the start of the *tnsA* ORF has been reported previously (7, 21, 38). The first ATG in this ORF is located at nt. position 135, and is preceded by only a poor match to the ribosome binding site (r.b.s.) consensus (39) (See Fig. 3). The ORF ends with TAA at position 954. The predicted protein is 273 amino acids long, with a molecular mass (M_r) of 31 kilodaltons (kd), and an estimated pI of 5.6.

Although alternative ATG/GTG start codons in this reading frame occur at positions 165, 360 and 465, the first ATG is presumed to be the site of translation initiation based on four lines of evidence; (i) a protein of 30 kd apparent Mr is encoded in this region (4, and references within; Rogers and Sherratt personal communication), (ii) a promoter has been mapped immediately upstream (21) (see above), (iii) although poor, there is a better match to the r.b.s. consensus before this ATG than the following start codon and (iv) a gene fusion that connects a strong promoter and appropriately spaced r.b.s. to the ATG at nt. 135 results in the over-expression of a protein with an apparent Mr of 30 kd (Flores *et al.*, unpublished observations).

There is a region within the deduced a.a. sequence, starting at a.a. number 90, which is comparable to the rather loose consensus for Cro-like, DNA binding domains (40, 41) (see Fig. 4), however the score on the weight matrix of Dodd and Egan (42) is too low to be predictive of this style of DNA binding in the absence of any other evidence that the protein does bind DNA. (The PIR protein sequence database, Release 7.0 (43), includes 108 proteins that score between 1100 to 1399; about 7% of those are judged to be Cro-like based on known properties of these proteins).

tnsB

The ORF corresponding to the *tnsB* gene is much longer than that of *tnsA* and is capable of specifying a protein of 702 amino acids, with a calculated Mr of 81 kd. The first ATG, at position 943, is preceded by an appropriately spaced sequence that matches the r.b.s. consensus well (Fig. 3). This potential start codon is within the 3′ end of the *tnsA* ORF, 11 b.p. before it terminates. The *tnsB* ORF closes with a TGA codon at position 3049. Other potential initiation codons occur at nt. positions 964 (GTG), 1024 (ATG) and 1078 (ATG). A gene fusion joining strong transcriptional and translational start signals to initiate translation at the ATG at nt. 943, leads to the production of a protein with an apparent Mr of 85 kd (Flores *et al.*, unpublished observations).

The predicted TnsB protein is rich in basic amino acids (estimated pI of 8.9), and also contains a region of similarity to the Cro-like, helix-turn-helix genre of site-specific DNA

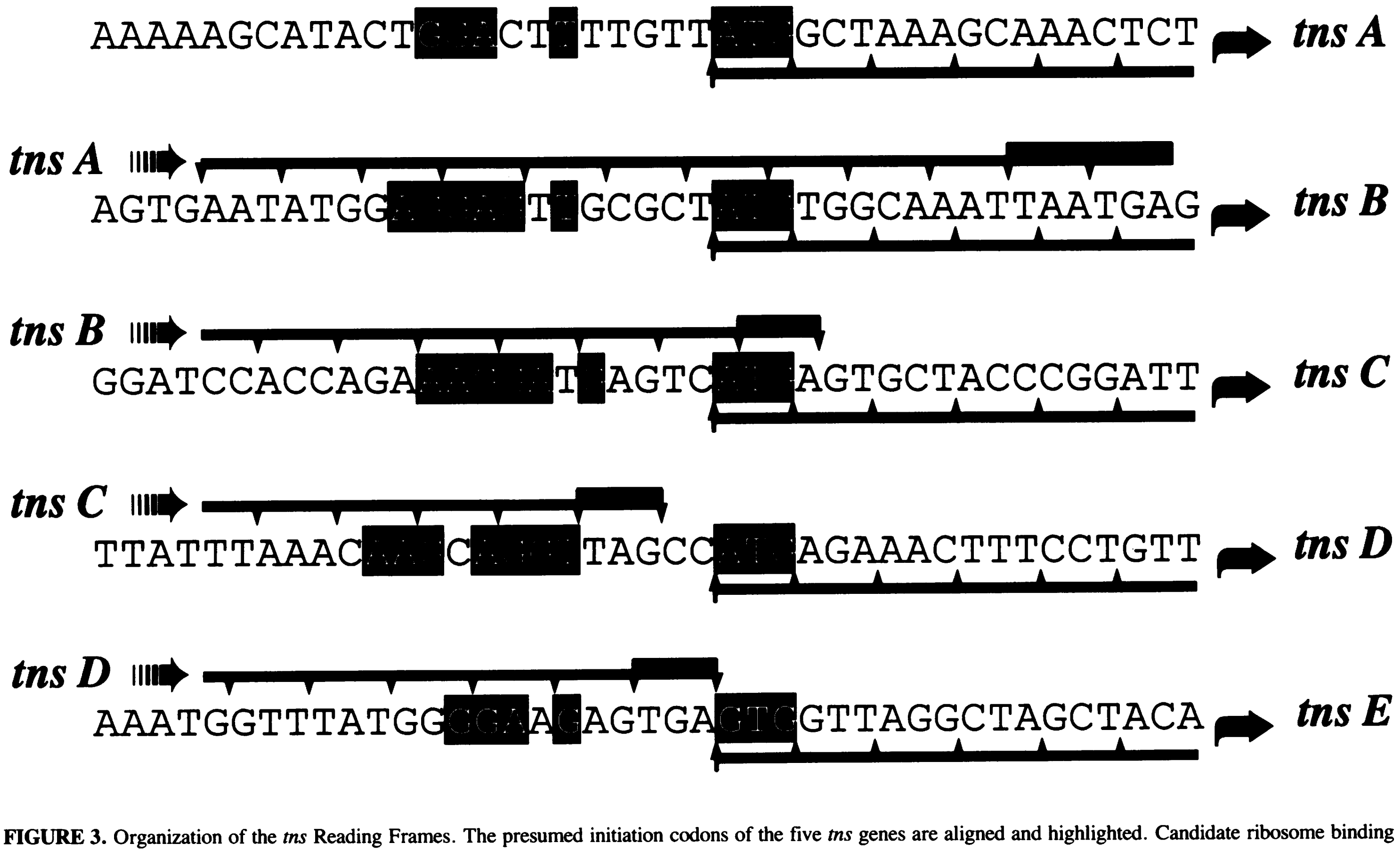
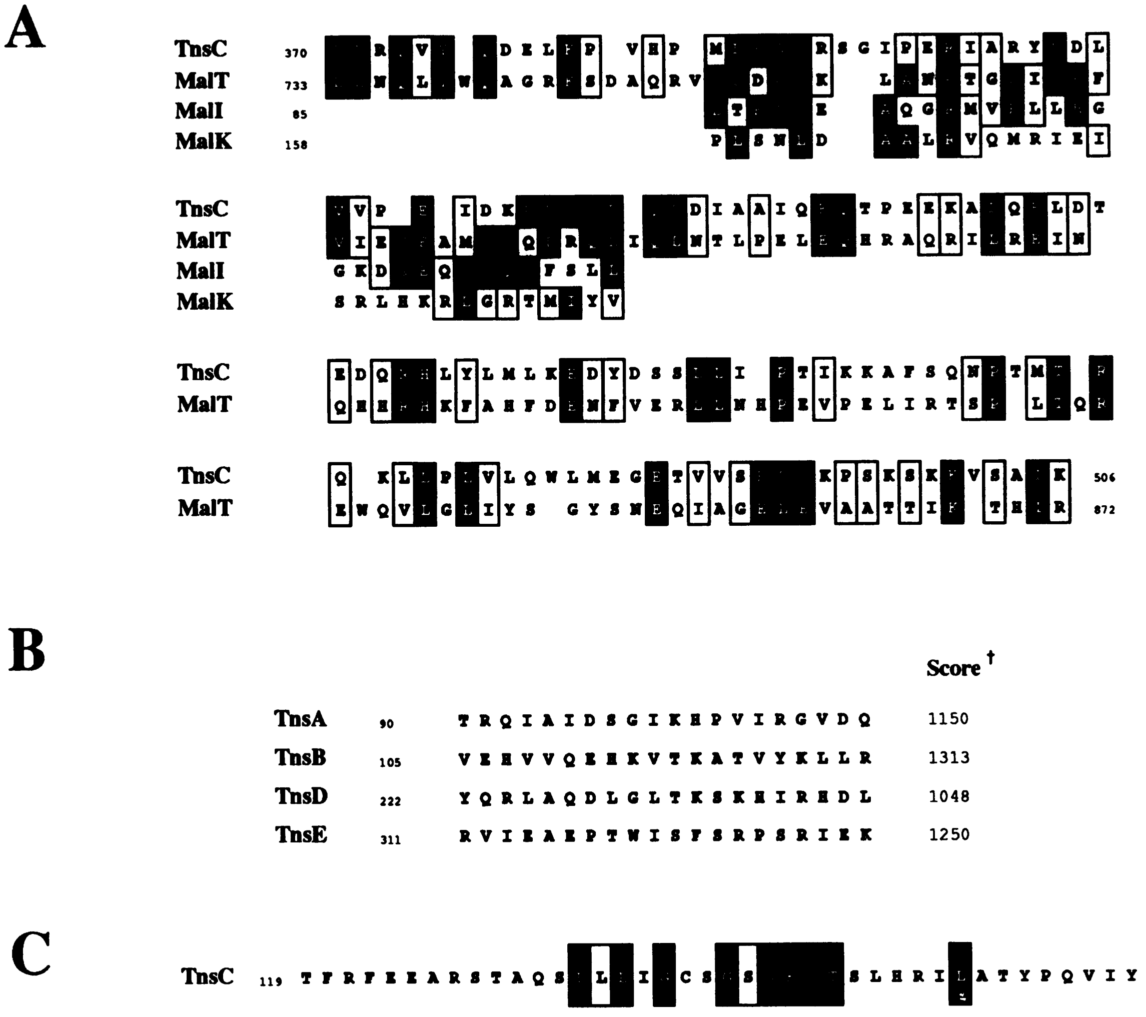

AAAAAGCATACT CT TTGTT GCTAAAGCAAACTCT
tns A
tns A
AGTGAATATGG T GCGCT TGGCAAATTAATGAG
tns B
tns B
GGATCCACCAGA T AGTC AGTGCTACCCGGATT
tns C
tns C
TTATTTAAAC C TAGCC AGAAACTTTCCTGTT
tns D
tns D
AAATGGTTTATGG A AGTGA GTTAGGCTAGCTACA
tns E

FIGURE 3. Organization of the *tns* Reading Frames. The presumed initiation codons of the five *tns* genes are aligned and highlighted. Candidate ribosome binding sites are also highlighted. Note the very compact assembly of open reading frames.



			Score†
TnsA	90	T R Q I A I D S G I K H P V I R G V D Q	1150
TnsB	105	V E H V V Q E H K V T K A T V Y K L L R	1313
TnsD	222	Y Q R L A Q D L G L T K S K H I R H D L	1048
TnsE	311	R V I E A E P T W I S F S R P S R I E K	1250

FIGURE 4. **(A)** An alignment of the amino acid sequences of TnsC, MalT, MalI and MalK is presented to reveal similarities. Identical amino acids are highlighted. Boxes indicate the occurrence of similar amino acids. The numbers before and after the alignment refer to the positions within these proteins of the initial and final amino acids displayed. **(B)** This diagram lists the amino acid sequences of regions within *tnsA, tnsB, tnsD* and *tnsE* that might interact with DNA through Cro-like, helix-turn-helix structures. The number of the first amino acid in the domain as well as the score (†) according to the weight matrix of Dodd and Egan, 1987 (42) is labeled. The significance of these scores is discussed in the text. **(C)** A region within the sequence of TnsC has homology to nucleotide binding domains found in many proteins. Matches to the most highly conserved positions in the consensus are highlighted while matches to the less well conserved positions are boxed. Again, the number to the left refers to the first amino acid listed.

binding domains (41) (See Fig. 4). This domain scores higher on the weight matrix of Dodd and Egan (42), than the similar region in TnsA. This score, combined with the fact that *tnsB* has been shown to be required for specific binding to the 22 bp repeats at the termini of Tn7 (44), implies that this region is likely to be a true Cro-like, DNA binding domain.

The fourth (final) 22 bp repeat in the right end of Tn7 overlaps with the promoter for the *tnsA* gene; and Rogers *et al.* (9) have presented evidence that TnsB represses transcription from that promoter, presumably by binding to the 22 bp repeat(s). If *tnsA* and *tnsB* are co-transcribed as Waddell and Craig believe (discussed above), the TnsB protein may autoregulate its own expression.

Complementation results of Rogers *et al.* (9), show that a cloned fragment of Tn7 from the right end to the rightmost Bam HI site, plus another fragment from this same site extending leftward, are able to provide all *trans*-acting functions necessary for Tn7 transposition. The DNA sequence reveals that this would result in the removal of 7 a.a. from the carboxy terminus of TnsB, (as they predicted from the limited DNA sequence available). This truncation has little, or no effect on transposition.

tnsC

The first initiation codon of the *tnsC* ORF starts one base before the TGA stop codon of the *tnsB* ORF, (Figs. 2 and 3). Alternative ATG/GTG start codons in this frame are found at positions 3147 (GTG) and 3303 (ATG). The translational reading frame terminates after 555 codons at nt. position 4713 with a TAG codon. The protein ostensibly encoded would be 63 kd in molecular mass and basic, with an approximate pI of 8.7. A gene fusion similar to those described for the *tnsA* and *tnsB* ORF's, to the ATG at nt. 3048 results in the appearance of a protein with an apparent M_r of 60 kd (Flores *et al.*, unpublished observations).

The derived amino acid sequence contains a very strong sequence similarity to a type-A nucleotide binding domain (45) (Fig. 4). In addition, another region of 136 a.a. situated toward the carboxy terminus of the inferred TnsC protein sequence exhibits considerable similarity (52% identical + similar amino acids in the alignment) to that of the transcriptional activator of the maltose operons of *E. coli,* MalT (46) (see Fig. 4). Within this region there is a short stretch of similarity between MalT and two other Mal proteins involved in the utilization of maltodextrins (Fig. 4). The significance of these similarities is not immediately obvious but is discussed more fully below.

tnsD

Translation of the *tnsD* ORF could start at nt. number 4718 where the first ATG is encountered and terminate with TGA at position 6242. A candidate ribosome binding sequence precedes this ATG. Possible alternative initiation codons occur at positions 4823, 4841, 4877 (GTG's) and 5003 (ATG). If the indicated ORF is translated a 508 a.a. protein of Mr 59 kd should result. This protein would be highly basic, with an estimated pI of about 9.5.

Gene fusions of the type described for *tnsA*, *B*, and *C*, to the first ATG codon of the *tnsD* ORF result in the accumulation of a protein of 55 kd apparent Mr (Flores *et al.*, unpublished observations). Other workers have ascribed proteins of 54 kd and 40 kd apparent Mr to this region (4 and references within, 47, Rogers and Sherratt personal communication;). By comparing the DNA sequence to the cloning sites used by Brevet *et al.* (47), it appears that the 40 kd protein that they observed was a truncated version of TnsD, with 42 amino acids removed from the carboxy terminus; however, the predicted Mr for this polypeptide (54 kd) does not explain the anomaly.

As stated earlier the *tnsD* gene is only required for transposition to *attTn7.* The a.a sequence of TnsD also has a region that matches the Cro-like, DNA binding consensus (Fig. 4) (42), and like TnsB has been implicated in DNA binding (9, 18). Waddell and Craig (18) have described a *tnsD*-dependant, *attTn7*-specific, DNA binding activity that may well be the TnsD protein itself. If that is the case, this region of TnsD between a.a. number 222 and 241, may be involved (Fig. 4).

tnsE

The *tnsE* ORF proceeds in the same frame (after only three base pairs that comprise the the stop-translation signal of *tnsD*), commencing with a GTG initiation codon. This ORF which potentially encodes a protein 538 a.a. in length, has been previously noted in the work of Smith and Jones (32). The predicted pI and Mr of the derived TnsE protein is 5.7 and 61 kd respectively. There is a plausible ATG start codon at position 6353, but it is likely that translation initiates at the earlier GTG sequence because: (i) it is preceded by a closer match to the r.b.s. consensus, and (ii) proteins of high apparent molecular mass have been observed that correlate with this region (*ie.* the 2224 bp Hind III fragment). Brevet *et al.* (47) attributed this area to the production of an ≈85 kd protein in maxicells, whereas Smith and Jones (32), and Craig (4 and references within), have seen 70−75 kd proteins specified from this region. We note that a translational fusion joining a strong promoter and r.b.s. as well as 13 foreign codons to the (presumed) 58th codon (at the Hpa I site, at nt. 6413), leads to the production of a protein with apparent Mr of 69 kd (Flores *et al.* unpublished observations).

Another match to the Cro-like DNA binding domain occurs in the a.a. sequence translated from *tnsE* (Fig. 4). Again however, in the absence of any evidence that TnsE does indeed bind to DNA, the significance of this similarity is uncertain. Because TnsE is essential and TnsD dispensable for transposition to plasmids, it has been speculated that TnsE may have a role equivalent to TnsD, directing transposition to random (non-*attTn7*) sites by binding to potential target sequences. However, Cro-like domains have not been implicated in this type of non-specific binding. Alternatively TnsE may not interact with target DNA directly, but through protein-protein interactions function to relieve the requirement for TnsD-*attTn7* assisted synapse formation.

Possible roles of Tns proteins

By analogy to other recombination systems, site-specific recombination (*ie.* integration/ excision/ resolution/ inversion) and transposition (semi site-specific, as half of the recombining sites are specific) (1), it is likely that one of the Tns proteins (or that several in concert), induce the breakage of phosphodiester bonds at the ends of Tn7 and at the target site, and that the same one(s) or other(s) catalyse the formation of bonds joining the two: but which ones, and how? Some clues may be gleaned from the predicted protein sequences presented here, while more direct biochemical evidence is beginning to emerge.

The TnsB and TnsD proteins must at least play a role in DNA sequence recognition since they have been shown to be essential for specific binding to the 22 bp repeats near the ends of Tn7 and to a region of *attTn7* respectively, (44, 18). Domains within TnsB and TnsD responsible for this binding have been tentatively mapped (Fig. 4). Presumably these proteins are also involved in protein-protein interactions that compose a higher order structure analogous to the 'intasome' formed during phage lambda integration (48).

It is intriguing that in both cases the sequence specific binding appears to be at sites a short distance away from the points of bond breakage rather than encompassing them. It seems that protein-DNA contacts at the points of bond breakage would be imperative and these could involve, (i) different domains of the same proteins, *ie.* TnsB and TnsD, (ii) other Tn7 encoded proteins, or (iii) host encoded proteins. In each case binding to break-point sites could be either by sequence specific binding or by non-specific interactions that are directed by the specific binding of *eg.* TnsB and TnsD to their respective sites. The facts that (i), the DNA sequence of *attTn7* at the point of insertion can be replaced with several other unrelated sequences without effecting the frequency or point of insertion, (18, 19; Sannuga unpublished observations), and that (ii), *attTn7* sites from the chromosomes of other bacteria are highly related to *E. coli attTn7* but only at regions distant from the insertion point (the '*glmS*-box') (20), suggest that contacts at the precise point of insertion are not sequence specific. Conversely the mere fact that the terminal 8 base pair sequence is perfectly conserved at each end of Tn7 may imply that sequence specific binding occurs here.

The mechanism of immunity for Tn7 is unknown, however the extent of DNA sequence in the right end of Tn7 required for immunity is roughly the same as that required for transposition (22). TnsB and TnsD may have directly analogous roles in transpositional immunity to those of Mu A and Mu B proteins, because there is an apparent similarity of binding to transposon ends and to targets respectively. Mu B binds DNA without specificity. Binding of Mu B to potential targets greatly enhances the frequency of transposition to those targets. Transposition immunity with bacteriophage Mu results from the fact that binding of Mu A (the transposase protein) to repeats at Mu's termini destabilizes the binding of Mu B to the same replicon. The instability of Mu B on molecules that contain the ends of Mu (and therefore also bound Mu A protein) causes transpositional immunity (25).

The functions of the TnsA, TnsC and TnsE proteins in Tn7 transposition are unclear. No DNA binding activity has been reported for these proteins, though it cannot be ruled out. If any of these proteins bind to DNA it could be non-specific or low affinity binding, perhaps requiring cooperativity or a higher order structure.

The implications of the similarity between the sequence of MalT and that proposed for TnsC are uncertain. MalT is known to bind maltotriose, Mg^{2+}, ATP/dATP and DNA (site-specifically), and to function as a transcriptional activator (49, 50). Unfortunately it is not known whether the region of MalT that is similar to TnsC interacts with DNA, RNA polymerase, maltotriose, Mg^{2+}, or has some other function.

Binding of MalT to a region of about 16 bp. centred on the 6 bp. 'malT box' (*ie.* the cognate DNA sequence) is thought to cause the wrapping of adjacent DNA around a core of MalT protein(s) (51). Perhaps this indicates weaker, non-specific interactions with DNA in addition to the specific ones. So far no evidence of a similar role for TnsC in DNA binding has emerged.

Transcriptional activation can be effected either by (i), increasing the apparent affinity of RNA polymerase for a promoter (*eg,* by binding DNA and transiently associating with RNA polymerase), or by (ii) simply perturbing the structure of the DNA such that open complex formation is facilitated (52, 53, 54). TnsC could be an activator or a repressor, regulating the expression of *tns* or host genes involved in Tn7 transposition.

Both TnsC and MalT have domains toward the amino terminal end that match a type-A nucleotide binding consensus, and which is not within the region of homology. This feature strengthens their similarity. ATP and/or dATP have been shown to be positive effectors of MalT binding to the 'malT-box' (49). Purified MalT protein has a low, intrinsic ATPase activity that is specifically stimulated two to three fold by maltotriose, yet ATP binding only and not hydrolysis is required for the activation of open complex formation by MalT at maltose promoters. It is not known whether hydrolysis is involved at a later step *eg.* promoter clearance, but the critical function of ATP at the early stage is presumed to be allosteric (51).

Although we anticipate that TnsC binds nucleotides (perhaps ATP) this has yet to be investigated, and if it does, the function is uncertain. It could be an allosteric effector as it appears to be for MalT, or it could serve to provide the energy required to form phosphodiester bonds ligating the ends of Tn7 to target sites; analogous to type II topoisomerases (*eg. E. coli* DNA gyrase), or as in T4 DNA ligase (55). However, there may be no need for a high energy co-factor in this reaction, because religation can be accomplished by two alternative means. (i) The energy could be conserved from bond breakages through covalent protein-DNA intermediates and transferred during religation (as in phage lambda integration/ excision (56, 57) and the DNA inversion reactions of Cin, Gin and Hin (58, 59, 60, 61), and type I topoisomerases), (62). (ii) Or religation may occur by direct transfer through nucleophilic attack of the target by the transposon ends while they are held (non-covalently) in the 'transpososome' complex, as appears to be the case for transposition of Mu (63).

Another possible role for nucleotides is demonstrated by the phage Mu system; where ATP (as well as Mu A) is required for displacement of the Mu B protein from potential targets, thus causing immunity (25, 64). The result is that intermolecular transposition is favoured over intramolecular events.

A short segment within the region of similarity between MalT and TnsC matches with two other proteins involved in maltose/maltodextrin utilization, MalK and MalI (Fig. 4). MalK appears to be a member of a family of binding proteins involved in active transport systems. It also has some role in regulation of the *mal* operons (*eg.* may be responsible for degrading an internal inducer)(65). It also contains a type-A nucleotide binding site near its amino terminus, (as do the other members of the periplasmic binding proteins). The sequence of the recently discovered MalI protein is highly homologous to three repressor proteins over its entire length (65). As well as the short stretch of similarity to MalT and TnsC, MalI contains a longer region of similarity to MalK.

The region of similarity of 31 amino acids found in the three Mal proteins has been proposed to be the binding site for some unknown inducer related to maltodextrins (65). If the corresponding region in TnsC has a similar function it is difficult to imagine what role such a molecule could have in transposition. We are left with many tantalizing clues that require further experiments to resolve.

ACKNOWLEDGMENTS

We wish to thank Dr. Smith, and Dr. Jones for allowing us to include previously published sequence of Tn7 that they determined (32), and Roger Staden for assistance and advice in analysing the sequence, and Professor John Collins for carrying out database searches. We also thank Dr. Susan Cotterill for helpful suggestions and critical reading of the manuscript, and Dr. Elaine Nimmo and Professor Dave Sherratt for sharing their unpublished sequence data with us, and Dr. Ian Anton for evaluating the nucleotide binding domain. This work was supported by the MRC grant number G8620799CB and the SERC grant number GR/C62086.

REFERENCES

1. Craig, N. L.; Kleckner N. (1987) In Neidhardt, F.; Ingraham, J.; Low, K.; Magasanik, B.Schaechter, M.; Umbarger, H. (eds.), *Escherichia coli* and *Salmonella typhimurium*: *Cellular and Molecular Biology*. American Society for Microbiology, Washington, D.C. p. 1054–1070.
2. Berg, D. E.; Howe, M. M. (eds.) (1989) *Mobile DNA*. American Society for Microbiology, Washington, D.C.
3. Barth, P.; Datta, N.; Hedges, R.; Grinter, N. (1976) J. Bacteriol., **125**, p. 800–810.
4. Craig, N. L. (1989) In Berg and Howe (eds.), *Mobile DNA* American Society for Microbiology, Washington, D.C. p. 211–226.
5. Barth, P.; Grinter, N.; Bradley, D. (1978) J. Bacteriol., **133**, p. 43–52
6. Lichtenstein, C.; Brenner, S. (1981) Mol. Gen. Genet., **183**, p. 380–387.
7. Lichtenstein, C,; Brenner, S. (1982) Nature (London), **297**, p. 601–603
8. Waddell, C. S.; Craig, N. L. (1988) Genes Dev., **2**, p. 137–149.
9. Rogers, M.; Ekaterinaki, N.; Nimmo, E.; Sherratt, D. (1986) Mol. Gen. Genet., **205**, p. 550–556.
10. McKown, R. L.; Orle, K. A.; Chen, T.; Craig, N. L. (1988) J. Bacteriol., **170**, p. 352–358.
11. Barth, P.; Grinter, N. (1977) J. Mol. Biol., **113**, p. 455–474.
12. Krishnapallai, V.; Nash, J.; Lanka, E. (1984) Plasmid, **12**,p. 170–180.
13. Moore, R. J.; Krishnapallai, V. (1982) J. Bacteriol., **149**, p. 276–283.
14. Ogawa, H.; Tolle, C.; Summers, A. (1984) Gene, **32**, p. 311–320.
15. Smith, G. M.; Jones, P. (1984) J. Bacteriol., **157**, p. 962–964.
16. Derbyshire, K. M.; Hwang, L.; Grindley, N. D. F. (1987) Proc. Natl. Acad. Sci. USA, **84**, p. 8049–8053.
17. Morisato, D.; Way, J. C.; Kim, H. J.; Kleckner, N. (1983) Cell **51**, p. 101–111.
18. Waddell, C. S.; Craig, N. L. (1989) Proc. Natl. Acad. Sci. USA, **86**, p. 3958–3962.

19. Qadri, M. I.; Flores, C. C.; Davis, A. J.; Lichtenstein, C. P. (1989) J. Mol. Biol., **207**, p. 85–98.
20. Qadri, M. I.; Flores, C. C.; Lichtenstein, C. P. manuscript in preparation.
21. Gay, N. J.; Tybulewicz, V. L. J.; Walker, J. E. (1986) Biochem. J., **234**, p. 111–117.
22. Arciszewska, L. K.; Drake, D.; Craig, N. L. (1989) J. Mol. Biol., **207**, p. 35–52.
23. Robinson, M. K.; Bennett, P. M.; Grinsted, J.; Richmond, M. H. (1977) J. Bacteriol., **129**, p. 407–414.
24. Huang, C. J.; Heffron, F.; Twu, S.; Schloemer, R. H.; Lee, C. H. (1986) Gene, **41**,p. 23–31.
25. Adzuma, K.; Mizuuchi, K. (1988) Cell, **53**, p. 257–266.
26. Messing, J.; Vieira, J. (1982) Gene, **19**, p. 269–276.
27. Yanish-Perron, C.; Vieira, J.; Messing, J, (1985) Gene, **33**, p. 103–119.
28. Hennikoff, S.; (1987) In *Methods in Enzymol.* **155**, p. 156–166.
29. Maniatis, T.; Fritsch, E. F.; Sambrook, J.; (eds.) (1982) *Molecular Cloning: a laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.
30. Sanger, F.; Nicklen, S.; Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA, **74**, p. 5463–5467.
31. Biggin, M.; Gibson, T. J.; Hong, G. F. (1983) Proc. Natl. Acad. Sci. U. S. A., **80**, p. 3963–3965.
32. Smith, G. M.; Jones, P. (1986) Nucleic Acids Res., **14**, p. 7915–7927.
33. Lindahl, L.; Zengel, J. M. (1986) Annu. Rev. Genet., **20**, p. 297–326.
34. Lindahl, L.; Archer, R. H.; McCormick, J. R.; Freedman, L. P.; Zengel, J. M. (1989) J. Bacteriol., **171**, p. 2639–2645.
35. Sor F.; Bolotin-Fukuhara, M.; Nomura, M.; (1987) J. Bacteriol., **169**, p. 3495–3507.
36. von Hippel, P. H.; Bear, D. G.; Morgan, W. D.; McSwiggen, J. A. (1984) Annu. Rev. Biochem., **53**, p. 389–446.
37. Harley, C. B.; Reynolds, R. P. (1987) Nucleic Acids Res., **15**, p. 2343–2361.
38. Gosti-Testu, F.; Brevet, J. (1982) C. R. Seances Acad. Sci., Ser. 3, **294**, p. 193–196.
39. Shine, J.; Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA, **71**, p. 1342–1346.
40. Sauer, R. T.; Yocum, R. R.; Doolittle, R. F.; Lewis, M.; Pabo, C. O. (1982) Nature (London), **298**, p. 447–451.
41. Ohlendorf, D. H.; Anderson, W. F.; Matthews, B. W.; (1983) J. Mol. Evol., **19**, p. 109–114.
42. Dodd, I. B.; Egan, B. J.; (1987) J. Mol. Biol., **194**, p. 557–564.
43. Barker, W. C., Hunt, L. T., George, D. G., Yeh, L. S., Chen, H. R., Blomquist, M. C., Seibel-Ross, E. I., Hong, M. K., Bair, J. K., Chen, S. L. Ledley, R. S. (1985) Protein Sequence Database, Release 7.0, Nov. 27 1985 of the Protein Identification Resource (PIR) of the Nat. Biomed. Res. Found., Georgetown Univ. Med. Cen.
44. McKown, R. L.; Waddell, C. S.; Arciszewska, L. K.; Craig, N. L. (1987) Proc. Natl. Acad. Sci. U. S. A., **84**, p. 7807–7811
45. Walker, J. E.; Sarast, M.; Runswick, M. J.; Gay N. J. (1982) EMBO J., **1**, p. 947–951.
46. Cole, S. T.; Raibaud, O. (1986) Gene, **42**, p. 201–208.
47. Brevet, J.; Faure, F.; Borowski, D. (1985) Mol. Gen. Genet., **201**, p. 258–264.
48. Echols, H.; (1986) Science, **233**, p. 1050–1056.
49. Richet, E.; Raibaud, O. (1989) EMBO J., **8**, p. 981–987.
50. Richet, E.; Raibaud, O. (1987) J. Biol. Chem., **262**, p. 12647–12653.
51. Raibaud, O.; Vidal-Ingigliardi, D.; Richet, E. (1989) J. Mol. Biol., **205**, p. 471–485.
52. De Crombrugghe, B.; Busby, S.; Buc, H. (1984) Science, **224**, p. 831–838.
53. Ptashne, M. A. (ed) (1986) *A Genetic Switch* Cell and Blackwell Scientific Press, Cambridge and Palo Alto
54. Liu-Johnson, H. N.; Gartenberg, M. R.; Crothers, D. M. (1986) Cell, **47**, p. 995–1005.
55. Weiss, B.; Jacquimin-Sablon, A.; Live, T.R.; Fareed, G. C.; Richardson, C. C. (1968) J. Biol. Chem., **243**, p. 4543–4555.
56. Mizuuchi, K.; Gellert, M.; Nash, H. (1978) J. Mol. Biol., **121**, p. 375–392.
57. Craig, N. L.; Nash, H. A. (1983) Cell, **35**, p. 795–803.
58. Iida, S.; Huber, H.; Hiestand-Naur, R.; Meyer, J.; Bickle, T. A.; Arber, W. (1984) Cold Spring Harbor Symp. Quant. Biol., **49**, p. 769–777.
59. Mertens, G.; Hoffman, A.; Blocker, H.; Frank, R.; Kahmann, R. (1984) EMBO J., **3**, p. 2415–2421.
60. Plasterk, R. H. A.; Simon, M. I.; Barbour, A. G. (1984) Proc. Natl. Acad. Sci. USA, **81**, p. 2689–2692.
61. Johnson, R. C.; Bruist, M. F.; Glaccam, M. B.; Simon, M. I. (1984) Cold Spring Harbor Symp. Quant. Biol., **49**, p. 751–760.
62. Been, M. D.; Champoux, J.J. (1980) Proc. Natl. Acad. Sci. U. S. A., **78**, p. 2883–2887.
63. Craigie,R.; Mizuuchi, K. (1987) Cell, **51**, p. 493–501.
64. Maxwell, A.; Craigie. R.; Mizuuchi, K. (1987) Proc. Natl. Acad. Sci. U.S.A., **79**, p. 151–155.
65. Reidl, J.; Romisch, K.; Ehrmann, M.; Boos, W.; (1989) J. Bacteriol., **171**, p. 4888–4899.