

## Tn917 Transposase. Sequence Correction Reveals a Single Open Reading Frame Corresponding to the *tnpA* Determinant of Tn3-Family Elements

FLORENCE Y. AN\* AND DON B. CLEWELL\*†

\*Department of Biologic and Materials Sciences, School of Dentistry, and †Department of Microbiology and Immunology, School of Medicine, The University of Michigan, Ann Arbor, Michigan 48109

Received October 9, 1990; revised January 22, 1991

A nucleotide sequence correction on the *Enterococcus faecalis* transposon Tn917 indicates that what was formerly thought to be two open reading frames (ORF5 and ORF6) is actually one reading frame (ORF5). The latter exhibits homology with the Tn3-family transposase determinants known as *tnpA*. © 1991 Academic Press, Inc.

Tn917 was originally identified on the plasmid pAD2 in the multiply antibiotic resistant *Enterococcus faecalis* strain DS16 (Tomich *et al.*, 1979, 1980). It encodes an inducible MLS (macrolide, lincosamide, and streptogramin B) resistance and is a member of the Tn3-family of transposons (Heffron, 1983; Sherratt, 1989). There is a close general resemblance to the staphylococcal transposon Tn551 (Perkins and Youngman, 1984; Khan and Novick, 1980), except that the latter expresses MLS resistance constitutively.

The complete nucleotide sequence of Tn917 has been previously reported (Shaw and Clewell, 1985). Three open reading frames were found to be associated with transposition functions. One of these, referred to as ORF4, was found to exhibit homology with Tn3-family resolvase determinant (*tnpR*). The other two open reading frames, ORF5 and ORF6, appeared to be associated with transposase activity (Perkins and Youngman, 1984; Shaw and Clewell, 1985). Although homology of the latter two reading frames with Tn3-family transposase (*tnpA*) was not noted, subsequent computer comparisons indicated significant homology indeed existed (Murphy, 1989). This prompted a re-

examination of this region of the transposon, the results of which are reported here. Correction of sequencing errors (mainly frameshift-related) has revealed that ORF5 and ORF6 are actually combined and represent a single ORF (now ORF5 or *tnpA*).

Figure 1 shows the corrected sequence of *tnpA*. The protein product is inferred to consist of 973 amino acid residues and to have a molecular weight of 112,205. This compares favorably with the *tnpA* products of Tn3 family elements (Sherratt, 1989). The 3' end of the reading frame is 17 bp from an inverted repeat segment (positions 3098 through 3140) which would form a stem consisting of 16 bp with an 11-bp intervening loop ( $\Delta G$  is  $-25.4$  kcal [Tinoco *et al.*, 1973]). The latter might correspond to a transcription terminator; although subsequent multiple T's, generally associated with  $\rho$ -independent transcription terminators, are not present.

The corrections and their positions were as follows: G converted to C at position 687; insert A at 1803; insert G at 1804; delete G between 1807 and 1808; delete T between 1905 and 1906; insert G at 2186; insert G at 2809; insert T at 3079; insert G at 3082.

Figure 2 presents a homology matrix relating to the *tnpA* products of Tn501 (Brown *et al.*, 1985), Tn21 (Ward and Grinstead, 1987), and Tn3 (Heffron *et al.*, 1979); significant homology is clearly evident. The degrees of homology (identical residues) are 32, 32, and

The sequence data from this article has been deposited with the EMBL/GenBank Data Libraries under Accession No. M36722.

20 40 60 80 100 120  
 GAGGAAAAAAAAAGGGGCGTCACAGTAAAGGTAAAGTATCAATGTATAGCTTTGGAAATGTATGACAGCAAGGATLCCATTCGTCAAAATCTTGATGCCCTCAATTAATAAACAA  
 -35 -10  
 140 160 180 200 220 240  
 ACCTTTTACCGTTACTCAATAAAGCTATCGCTTAAGATATGGCTATGAAAAGAAATTTAACTACTTCACAGCGTGAACAACCTCTTCTGTAGACACTATCAGAAAGAGATTAAAA  
*tnpR* \* S.D. M A M K R I L T T S Q R E Q L L S V D H L S E E D F K >  
 260 280 300 320 340 360  
 CGGTATTTAGTATGATGATGGAGTTTAAATCAACACCGTGGAAAGCTCAATAAACTAGGATTGCGATACAACTTTGTTGGCCCGTACTCTGGGTGTCTTAACT  
 A Y F S F S D Y D L E V I N Q H R G K V N K L G F A I Q L C L A R Y P G C S L S >  
 380 400 420 440 460 480  
 AATTGGCCGATTAATCAACGAGCTAACTTCTTGTAGTGGAGCTGGACCTGCTATGATGCAATGATTAATGATGATCATAGATAACACCTGCAAACTCACTCAACGAGATC  
 N N P I K S T R L I T S Y V S R O L H L D A I D L N S Y D H R N T R A N H F N E I >  
 500 520 540 560 580 600  
 TTGAAGTATCAACTATCGATTCCGCTAGTCTAATACACAAAACAGTAAATAGATAATTTAAATGAACTAGCTTTAGAAAATGATGACTCTATCTAATGAAAAACAAT  
 L E V F N Y H R F P G S A N T Q K Q L I E Y L I E L A L E N D D S I Y L M K R T I >  
 620 640 660 680 700 720  
 GATTTCTTAACTGAAAGAATTTTTCCTATCTATGCTACACTTGAAGACATTTAAGCCGCTGCGAGATAAAGCAAACTTATTTCAATTACTCTGTCTTAACTA  
 D F L T R K R I I F P S I A T L E D I I S R C R D R K A E N N L F S I L L C S L T >  
 740 760 780 800 820 840  
 GATAACAAATGAAAGCTAGAGATTTGTTCAAAITATGAAGACGAAAATAACTAACTCCCTTGGCTAAAAGACATTCCAGGTAAAGCAAATCCAGAAAAGCTTTATGAGTAT  
 D I Q I E K L E S L F Q I Y E E T K I T K L A W L K D I P G K A N C P E S F M S I >  
 860 880 900 920 940 960  
 TGTAAAAAGTGAAGTATGCTTCCATGGAGCTGGGCAATTAATGCTCCCATTAATCGGAACAGTTCCTCAGCTAGCTAGACTAGGGAAATATGATGCATATGATGCT  
 C K K V E V I A S M G L C T I N V S H I N R N R F L Q L A R L C E N Y D A Y D F >  
 980 1000 1020 1040 1060 1080  
 TCCGTTTGGAGCTGAAAAGGATGACTTCTTACTTATGCTTTTTCAGTCACTCATCAATATCTGATGATCACTGATGAGATTAAGACCGATTTAGCAAGTATTAACGC  
 S R F L E K R Y S L L I A F L V N H Q Y L I D Q L I E I N D R I L A S I K R >  
 1100 1120 1140 1160 1180 1200  
 AAAGGACAGCTGATCAACAGAACAGTAAAGGAAAAGAAAATGGCTACTAAAATAATGGAACTATTGCTTCTTAAATGATGCTCTTCACTTGGCAAGACATAATGATAGTAA  
 K G T R D S Q E Q L K E K G K L A T R K K L E H Y A S L I D A L H P A K D N D S N >  
 1220 1240 1260 1280 1300 1320  
 CCTTTTGCAGAAATTAAGCAATCATGCTTGGGAAATTTAGTACAAGATGGAAGAAAGCTAAAAAGCTATTACAGGTAAATAAAATCATGGCTATTAGAAAATGGTTCGAAATAAA  
 P F F D E I E R I M P W E D L V Q D G E E A K K A I T G N K N H G Y L E M V R N K >  
 1340 1360 1380 1400 1420 1440  
 GCTAATFACCTCCGAGATACACCGCAATGTTTGAAGGACCTTCGTTCAAAAGCACTCCGGCAGCAATCCAGTCTCATGGCCCTAACTCAACTAATGATTACACAAATGTTG  
 A N Y L R R Y T P M L R T L S F K A T P A A N P V L M A L L D L H N S G >  
 1460 1480 1500 1520 1540 1560  
 AAAAGAAAATACCGCAGTACTTCTACTGTTTGTGAGTAAAAATGAAAAGCGTTGTTCGGCCGAGAAGGGGAAAATAGATCGGTCTTACATGATGATGATGATGATGATGATGAT  
 K R K I P A D T S T D F V S K K W K S L V R P E E G K I D R S Y Y E L V A F T E >  
 1580 1600 1620 1640 1660 1680  
 CTAAGACAATATTCGATCGGAGATATTCAGTTGGAAGAGATGATGATCGCAATATTGATGATFACCTAGTTGATTTATCTGCTGATGATGATGATGATGATGATGATGATGAT  
 L K N I R S G D I S V E G S M I H R N I D D Y L V D L S A C I D S E T I P D T >  
 1700 1720 1740 1760 1780 1800  
 TTTGAGGACTTTAAAGATCGGAAATAATTTAGATTTACAGCTTCAATTTATTCAGCTGATGATGAGAAATTCAGGACCAAACTTAAAGATTTGAAAAAGTTACACCTAGC  
 F E D Y L K D R E I I L D L Q L Q F Y S T V D K R I S R A N L K K L E K V T P S >  
 1820 1840 1860 1880 1900 1920  
 GAAGCAGAAATATATGAAAAAACTTTATTCAATAATCCATAAGCTTAGTGATCTTTTAAATGAGGTGGACAGTGGCAAACTTTCAAGAAATTTAGTCAATGATCTACACA  
 E A E I Y R K K L Y S I I P K I R L S D L L I E V D S W T N F S Q E F S H D S T >  
 1940 1960 1980 2000 2020 2040  
 GGGAAACCGCGAGTGAACAAGAAAATAATTTTCTGCTTTGCTGGGTATGGGATGAAATTTGGTTTGAATAAAATGGCCCAATCACTCTCGGAATTTCTTATCTCAGTGA  
 A N Y L R R Y T P M L R T L S F K A T P A A N P V L M A L L D L H N S G >  
 2060 2080 2100 2120 2140 2160  
 GCCAATGCCAAATGACCGCTTTTAAAGAACTGTCGACTGCTCACTGTTTGGTAAATATCAGTTAAAGCTTCTGTTGAGCAGTCTTTGGGGTGAAGAAAACCACTGCT  
 A N A K Q W R F Y K E A L T R A Q S V L V N Y Q L K L P V A D F W G E G K T T A >  
 2180 2200 2240 2260 2280  
 TCAGACGAAATTCGCGCTCCGATCGCGCTCTCAGCTTAAAKTCCGATGTAATCCACATACAAAAGTATGGAATAAGAGGCTACAATGATTCGATCAATAAGATAGCCATAGCCT  
 S D G M R V P V G V S A L K S D V N P H Y K S M E K G A T M I R S I N D R H T >  
 2300 2320 2340 2360 2380 2400  
 CATCATATCGAGTTGCTCACTAATACAAAGCAAGCTACTTATCCCTGATCGCTACTTTATCATGAAACAGATCTAGATATTTAGGAACTTTTACTGATCAAAATCGGTTATCT  
 H H I E V A S T N D E A T H T L G L L Y H E F D L D I E H F T D T N G Y S >  
 2420 2440 2460 2480 2500 2520  
 GATCAGTGTGTTGAAATGACCGCTACTAGGCTTTGATTTGAAGCTCGCATCAGAAAATAAAAAAATCACAATTTATTTCTAATAACACCTTCTACTACCCTAACTTATCAGAA  
 D Q V F G M T A L L G F D F E P R I R N I K K S Q L F S I K S P S Y Y P N L S E >  
 2540 2560 2580 2600 2620 2640  
 GATATAAGCGGAAAAATCAATGTAATAATTTGAAGAAAATGATGATAATTAAGCAATCGCCTTTCGATTCAAACAGGAAAATGATCTAGTTCTTTACTATTAGGAAAAGCTAGGC  
 D I S G L R I N V K I I E E N Y D E I K R I A Y S I Q T G K V S S L L L G K L >  
 2660 2680 2700 2720 2740 2760  
 TCATACGACAGTAAAGATAGAGTACTTGCACCTGAGAGAAGCTAGTGCATTGAAAAGACATTTTATGATGATATATTTACAGATAGTGAAGTACGGGCAAGGATCACTCATGGA  
 S Y A R K N R V A L A L R E L G R I E K S I F M I D Y I T D S E L R R R I T H >  
 2780 2800 2820 2840 2860 2880  
 CTAATAAGCAGAACGATTAATGCTTTACCTAGAGACTTTTTTGGCGACCGGAAAATTTAGGAGCGGATATTCGCGCAACTTCAAAGTCTAGTGGCTTAAATGCTTTA  
 L N K T E A I N A L R R E L F F G R K F M E R D I R R G Q L S A C I D S E T I P D T >  
 2900 2920 2940 2960 2980 3000  
 ATAAGTCAATGATGAAAGCGCTACTTCAAGCAGCTTATAATTTATCTCGTCAAAATAGATCCGAACTAACTAAGTATGAAAGTATGATCTCTTATTAAGTGGGAGCAT  
 I N A T S I W N A V L Q A A Y N L V K I D P E V T K Y H K H V S P I N W E H >  
 3020 3040 3060 3080 3100 3120  
 ATCCTTTTCTGGAGAGTAAATTTGACTTGTATCTATCTTAAACACTTAAAGAAATGAAATAAAAAAATAATAGGCTTGAACATTTGGTTTATGGGAAATTTGTACCCCTTA  
 I T F L G E Y K F D L L S I P K H L R E L N I K N K \*  
 -----  
 3140 3160 3180 3200 3220 3240  
 TCGATACAAATCCGACTAAGCGCTGGGACCCCTTTTTAGGATATATTTGTTTTAATGTTTAACTATCTTATTCTAGACAATAAGCTCTTTCTAATCTCTTAAATAGCTTTT  
 -----  
 3260 3280 3300 3320 3340 3360  
 TTAAGCTTATAAATTCGCAACAATAAAGATTTGTAGATAAAGAAATTAATGGACAAACCAACTCCAGACAATTTACAGCAAGCAATGCAAGGAAATGAAACCAATTAATCTCAA  
 -----  
 3380 3400  
 TCGAGAAAAGAGTTTTTGTAAAATGGAAAAATGTTTACGCTCAA

FIG. 1. The nucleotide sequence of *tnpA* (ORF5) of Tn917. The underlined regions correspond to the potential promoter, Shine Dalgarno (S. D.) site, and the inverted repeats that may represent a transcription terminator. The inferred amino acids of TnpA are shown in single-letter designations and are encoded from basepairs 160 to 3078. The designation "tnpR\*" indicates the 3' terminus of *tnpR*, which is 16 bp

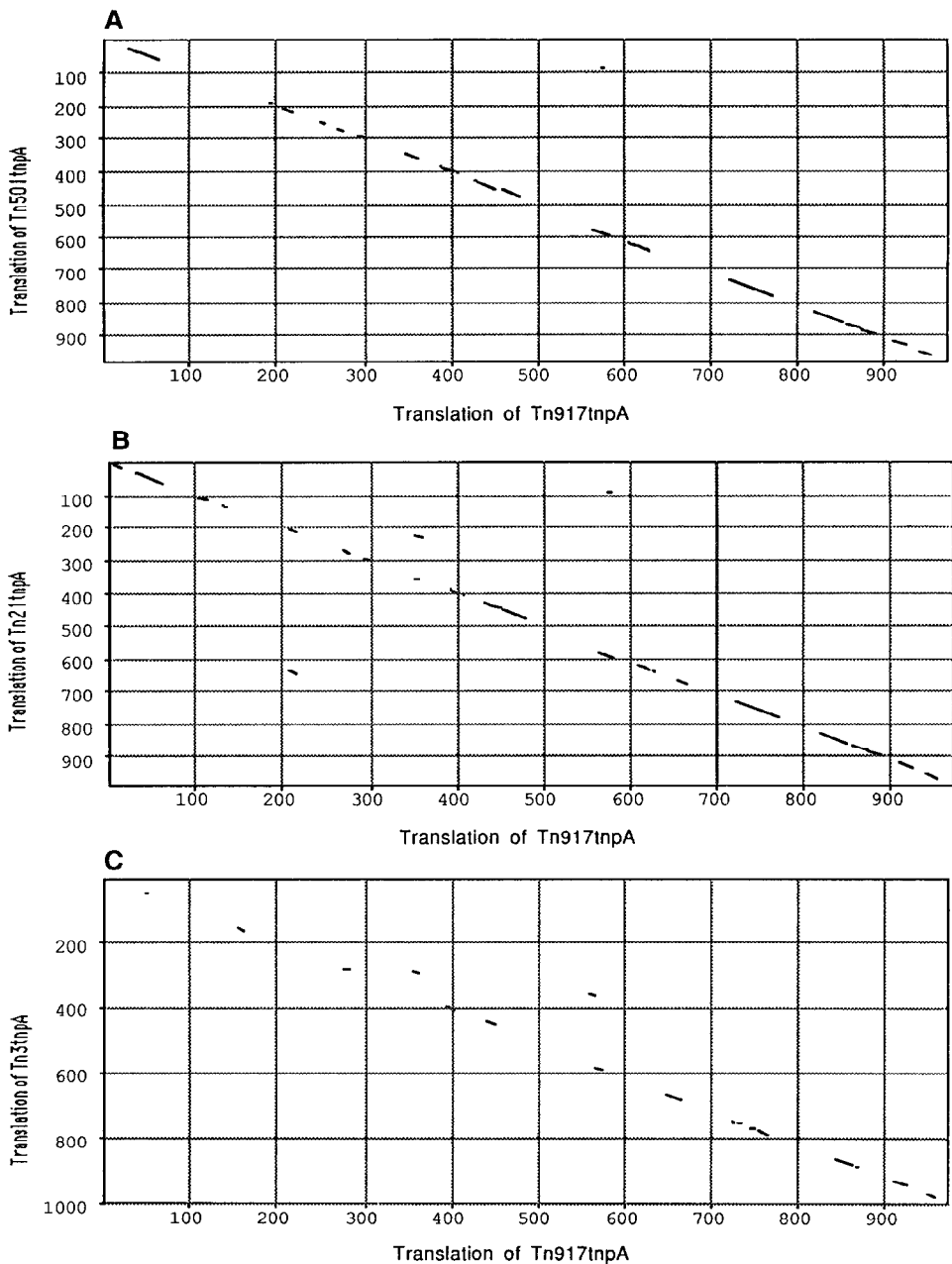


FIG. 2. Homology matrix plots comparing of *tnpA* of *Tn917* with *Tn501* (A), *Tn21* (B), and *Tn3* (C). A MacVector (IBI) program was used to generate the comparisons. Each point reflects at least 50% homology within a 14-amino acid alignment with the corresponding protein.

upstream of the start site of *tnpA*. The dashed line designates the right-end 38-bp inverted repeat. Sequencing was performed utilizing a pBluescript (Stratagene, Inc.) clone (pAM3917R) of a segment from the pAD1::*Tn917* derivative pAM938 (Ehrenfeld and Clewell, 1987). The clone contained the 3' half of the transposon and adjacent pAD1 DNA. Overlapping nested deletions [ *ExoIII*/mung bean nuclease protocol (Stratagene, Inc.)] generated from both orientations were sequenced using the double-strand DNA procedure as described by Ausubel *et al.* (1987). A Sequenase kit from U.S. Biochemical Corp. was utilized. In some cases, specific synthetic primers made at the University of Michigan Biomedical Research Core Facility were used.

17%, respectively. Homologies based on similarity are 53, 50, and 30%, respectively.

The similarity of Tn917 with other Tn3 family transposons, with respect to both *tnpR* and *tnpA*, is consistent with a close evolutionary relationship. In this regard it is noteworthy that Tn917, which originated in the gram-positive *E. faecalis*, has been shown to transpose in a gram-negative (*Escherichia coli*) background (Kuramitsu and Casadaban, 1986).

### ACKNOWLEDGMENTS

We thank N. Brown, J. Mahillon, and E. Murphy for pointing out the homologies between the ORF5 and ORF6 reading frames and *tnpA* of other Tn3-family elements, and N. Brown and D. Rouch for information on the likely position of certain frameshifts. This work was supported by USPHS Grants GM33956, AI10318, and DE02731.

### REFERENCES

- AUSUBEL, F. M., BRENT, R., KINGSTON, R. E., MOORE, D. D., SEIDMAN, J. G., SMITH, J. A., AND STRUHL, K. (1987). "Current Protocols in Molecular Biology." Wiley, New York.
- BROWN, N. L., WINNIE, J. N., FRITZINGER, D., AND PRIDMORE, R. D. (1985). The nucleotide sequence of the *tnpA* gene completes the sequence of the *Pseudomonas* transposon Tn501. *Nucleic Acids Res.* **13**, 5657-5669.
- EHRENFELD, E. E., AND CLEWELL, D. B. (1987). Transfer functions of the *Streptococcus faecalis* plasmid pAD1: Organization of plasmid DNA encoding response to sex pheromone. *J. Bacteriol.* **169**, 3473-3481.
- HEFFRON, F. (1983). Tn3 and its relatives. In "Mobile Genetic Elements" (J. A. Shapiro, Ed.), pp. 223-260. Academic Press, New York.
- HEFFRON, F., MCCARTHY, B. J., OHTSUBO, H., AND OHTSUBO, E. (1979). DNA sequence analysis of the transposon Tn3: Three genes and three sites involved in transposition of Tn3. *Cell* **18**, 1153-1164.
- KHAN, S. A., AND NOVICK, R. P. (1980). Terminal nucleotide sequence of Tn551, a transposon specifying resistance in *Staphylococcus aureus*: Homology with Tn3. *Plasmid* **4**, 148-154.
- KURAMITSU, H., AND CASADABAN, M. (1986). Transposition of the gram-positive transposon Tn917 in *Escherichia coli*. *J. Bacteriol.* **167**, 711-712.
- MURPHY, E. (1989). Transposable elements in gram-positive bacteria. In "Mobile DNA" (D. E. Berg and M. M. Howe, Eds.), pp. 269-288. Am. Soc. Microbiol., Washington, DC.
- PERKINS, J., AND YOUNGMAN, P. J. (1984). A physical and functional analysis of Tn917, a *Streptococcus* transposon in the Tn3 family that functions in *Bacillus*. *Plasmid* **12**, 119-138.
- SHAW, J., AND CLEWELL, D. B. (1985). Complete nucleotide sequence of macrolide-lincosamide-streptogramin B-resistance transposon Tn917 in *Streptococcus faecalis*. *J. Bacteriol.* **164**, 782-796.
- SHERRATT, D. (1989). Tn3 and related transposable elements: Site-specific recombination and transposition. In "Mobile DNA" (D. Berg and M. M. Howe, Eds.), pp. 163-184. Am. Soc. Microbiol., Washington, DC.
- TINOCO, I., BORER, P., DENGLER, B., LEVINE, M., UHLENBECK, O., CROTHERS, D., AND GRALLA, J. (1973). Improved estimation of secondary structure in ribonucleic acids. *Nature (London)*. *New Biol.* **246**, 40-41.
- TOMICH, P., AN, F., AND CLEWELL, D. B. (1979). A transposon (Tn917) in *Streptococcus faecalis* that exhibits enhanced transposition during induction of drug resistance. *Cold Spring Harbor Symp. Quant. Biol.* **43**, 1217-1221.
- TOMICH, P., AN, F., AND CLEWELL, D. B. (1980). Properties of erythromycin-inducible Tn917 in *Streptococcus faecalis*. *J. Bacteriol.* **141**, 1366-1374.
- WARD, E., AND GRINSTED, J. (1987). The nucleotide sequence of the *tnpA* gene of Tn21. *Nucleic Acids Res.* **15**, 1799-1806.

Communicated by Francis L. Macrina