JOURNAL OF BACTERIOLOGY, Jan. 1982, p. 407–410
0021-9193/82/010407-04$02.00/0

Vol. 149, No. 1

Amino-Terminal Sequence of the Tn*3* Transposase Protein

MARY D. DITTO,[1] JOANY CHOU,[2] MICHAEL W. HUNKAPILLER,[4] MICHAEL A. FENNEWALD,[3] STEPHEN P. GERRARD,[3] LEROY E. HOOD,[4] STANLEY N. COHEN,[5] AND MALCOLM J. CASADABAN[2,3]*

Department of Microbiology,[1] Department of Biophysics and Theoretical Biology,[2] and Department of Biochemistry,[3] University of Chicago, Chicago, Illinois 60637; Department of Biology, California Institute of Technology, Pasadena, California 91125[4]; and Departments of Medicine and Genetics, Stanford University School of Medicine, Stanford, California 94305[5]

Received 20 July 1981/Accepted 16 September 1981

The amino-terminal sequence of the Tn*3* transposase protein was determined to be Pro-Val-Asp-Phe-Leu-Thr-Thr-Glu-Gln-Val-Glu-Ser. . . . This was determined both from an active transposase protein purified from a transposase-overproducing mutant strain and from a hybrid transposase–β-galactosidase fusion protein. The amino acid sequence corresponded to the DNA sequence of the transposase gene beginning at an ATG initiation codon, as previously predicted from the analysis of transposase–β-galactosidase gene fusions.

Tn*3* is one of a group of genetic elements capable of transposing to new DNA sites with little or no homology (2, 15). It has been found to contain a gene (*tnpA*) that encodes a transposase protein which is essential for transposition (6, 9, 10, 11). Tn*3* also contains two other genes: the *bla* gene for a β-lactamase that confers resistance to penicillin antibiotics, and the *tnpR* gene for a protein with two functions, repression of its own gene (5) and the adjacent divergently transcribed *tnpA* gene (6, 9) and specialized "site-specific" recombination between loci of two Tn*3* elements present on the same DNA molecule (1, 8, 18; J. Chou, unpublished data).

To further characterize the process of DNA transposition, the purification and characterization of the transposase protein have been undertaken (7). As the transposase protein is present in wild-type Tn*3*-containing cells in very low concentrations, mutations which overproduce this protein and increase the rate of transposition were isolated and characterized (4; M. Casadaban, J. Chou, and S. N. Cohen, Cell, in press). However, an ambiguity arose as to the site of initiation of the wild-type and overproduced proteins (6, 10). In the present study, we resolved this question by determining the amino-terminal sequence for the wild-type protein and an overproduced transposase protein.

DNA sequence analysis of this region shows two possible initiation sites for translation of the transposase protein: an ATG codon and, 33 nucleotides upstream, a GTG codon (6, 10) (Fig. 1). Mutations which result in overproduction of the transposase protein and an increase rate of DNA transposition have been isolated, and five of these have been mapped by DNA sequence analysis to be single base changes in the initiation region of the transposase protein (4; Casadaban et al., in press). Four of these are located between the GTG and the ATG triplets, and the fifth is located in the GTG triplet, changing it to ATG (Fig. 1). Because these mutations are located in the region for translation initiation for the transposase protein, it is possible that they create a new translation initiation site.

The transposase protein from the overproducing mutation *q1199* was purified to homogeneity as described previously (7). (This protein has been shown to have a single-strand DNA-binding activity [7].) Transposase coded for by the wild-type *tnpA* gene was not purified because it is synthesized at a very low rate (6) and there is no convenient assay. To show that the overproducing mutation did not alter the transposase protein sequence, a hybrid transposase–β-galactosidase protein was also purified to determine the original transposase amino-terminal sequence. The hybrid transposase–β-galactosidase could be readily purified without an overproducing mutation because of the very sensitive enzymatic assay for β-galactosidase which can be used to monitor its purification (17).

The transposase–β-galactosidase hybrid protein from the *tnpA-lacZ* gene fusion on plasmid pMC995 (6) (Fig. 1) was used. This fusion has the first 45 amino acid codons of the *tnpA* transposase gene joined in frame to the eighth and following codon of the *lacZ* gene and results in a hybrid protein that is enzymatically active for β-galactosidase (but not for transposase). Transcription and translation for this fused gene are initiated from the *tnpA* gene control sequence (3).

The procedure used for the isolation of the

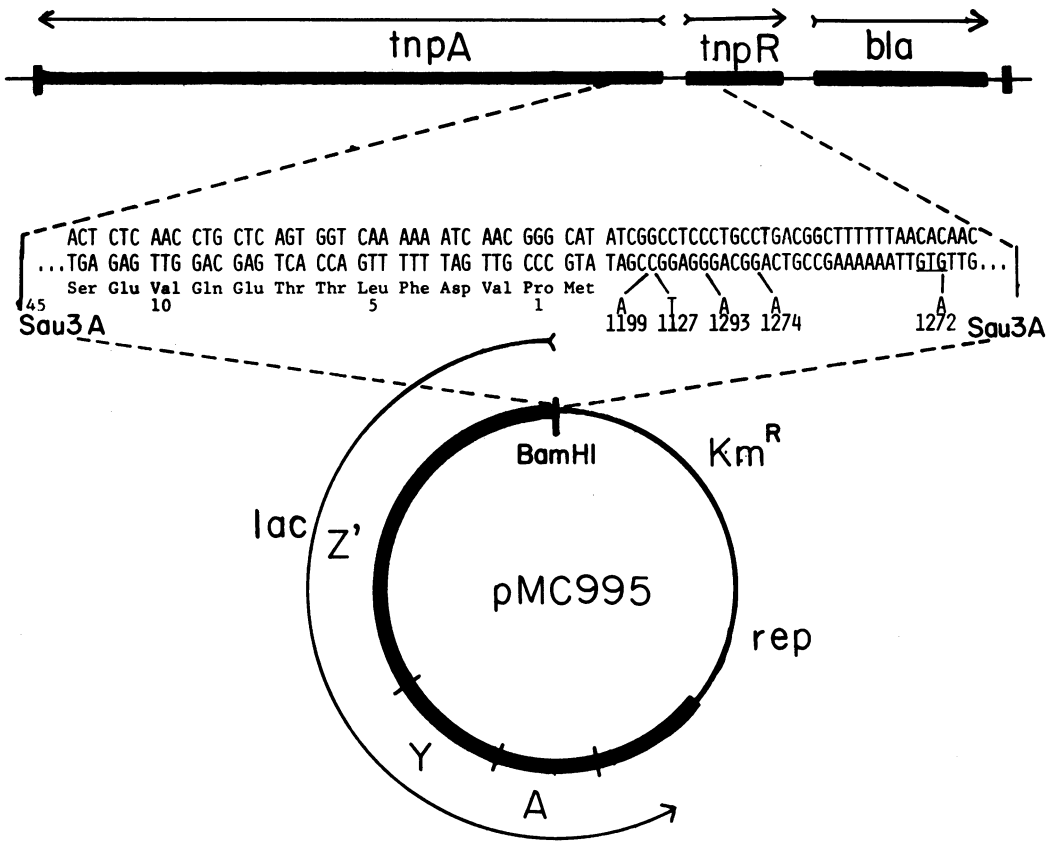

FIG. 1. Tn*3* *tnpA* translation initiation region and the pMC995 *tnpA*-transposase fusion with the *lacZ* β-galactosidase gene. The top line shows the 4,957-nucleotide Tn*3* element with three identified genes. The arrows above depict the directions of transcription. The middle line depicts the 356-nucleotide *Sau*III A fragments of Tn*3*, which contains the promoters, translation initiation sites, and control regions for both *tnpA* and *tnpR* (6). The first 12 amino-terminal amino acid codons for the transposase are shown. The alternatively proposed GTG initiation codon is underlined. The base changes for five mutations which result in transposase overproduction are shown. The lower part shows the pMC874 *lac* gene fusion plasmid vector into which the 356-nucleotide *Sau*III A fragment was inserted, by DNA cloning, to form pMC995. This creates a *tnpA-lacZ* fused gene encoding a hybrid transposase–β-galactosidase protein which retains β-galactosidase enzymatic activity but has 45 amino-terminal amino acids from the transposase, replacing the first 8 nonessential amino acids of β-galactosidase (6).

fusion protein was a modification of that described for the wild-type β-galactosidase (17). Eight liters of MC1050 (*lac* deletion strain) cells carrying pMC995 were grown to the early stationary phase and harvested. Lysis was carried out by the procedure of Ikeda et al. (13). The cell extract was adsorbed to a DEAE-52 column. β-Galactosidase activity was found to be included in 0.25 to 0.40 M NaCl fractions. The peak fractions were pooled and precipitated by 33% saturation with ammonium sulfate. The pellet was suspended, dialyzed, and loaded on a three-step sucrose gradient in an SW41 rotor. Centrifugation was run for 10 h at 40 krpm. The fractions were collected and assayed for β-galactosidase activity. The active fractions were pooled to yield 50 μg of a 122,000-molecular-weight protein which was at least 98% pure as judged on a 7.5% SDS-polyacrylamide gel (16). The buffer used throughout the isolation was 50 mM Tris (pH 7.5)– 1 mM EDTA– 5 mM mercaptoethanol (13).

The amino-terminal amino acid sequences for the proteins were analyzed by the microsequencing technique described previously (12). The first 12 amino acids sequenced for both proteins were identical and corresponded to the DNA sequence for the start at the ATG codon (6) (Table 1). From this we conclude that translation of the wild-type transposase is initiated at the ATG site. We cannot, however, rule out the possibility that translation is also initiated at the GTG codon at a very low frequency (<5%). Indeed, the most upstream of the overproducing

TABLE 1. Amino acid sequencing yields[a]

Cycle	Amino acid	Yield (pmol)	
		Overproduced transposase protein	Hybrid transposase–β-galactosidase protein
1[b]	Proline	30	25
2	Valine	28	50
3	Aspartic acid	19	30
4	Phenylalanine	25	45
5	Leucine	29	45
6	Threonine	5	8
7	Threonine	5	8
8	Glutamic acid	19	35
9	Glutamine	6	19
10	Valine	15	27
11	Glutamic acid	14	22
12	Serine	4	5

[a] Approximately 4.0 μg of the overproduced transposase protein and 8 to 12 μg of the pMC955 hybrid transposase–β-galactosidase protein were analyzed by automated Edman degradation as described previously (12). The phenylthiohydantoin amino acid-derived fractions for the first 12 cycles were identified by high-pressure liquid chromatography. The yields were corrected for background levels, which varied from 2 to 5 pmo per amino acid. The variation in individual amino acid yields reflects the relative instabilities of the particular phenylthiohydantoin amino acid. No secondary sequences, including one for a start at the GTG site or one for retention of the formylmethionine, were present at the limit of detectability, which was 5% of the main sequence.

[b] This first cycle also shows appreciable quantities of serine, glycine, aspartic acid, and glutamic acid, presumably arising from contamination of the sample with free amino acids. No other amino acid residues were detected in each of the later steps other than the one shown.

mutations we isolated (*q1272*; Fig. 1) may result in efficient initiation of translation from the GTG site which is changed by the mutation to ATG. Alternatively, this G/C-to-A/T base change mutation may create or improve a promoter site for initiation of transcription since it is located in a region resembling a promoter. The other overproducing mutations, including the *q1199* mutation used for our purification, appear from their position just before the ATG site to result in increased translation initiation (4). This conclusion that the ATG site is used for translation initiation is consistent with the DNA sequence of the related transposon γδ, which has an equivalent ATG site but no equivalent GTG site (18). On the other hand, we should point out that the transposase gene of the unrelated transposon Tn*5* has been shown to have two translation initiation sites which are used at approximately equal frequency (14).

Our location of the start site for *tnpA* translation further clarifies the genetic structure of the complex region of Tn*3* between the *tnpA* and *tnpR* genes. These genes are divergently transcribed from adjacent or overlapping gene control regions. In the space between these genes the *tnpR* gene product acts to repress transcription of both genes (5, 6, 9) and also to site-specifically recombine between two Tn*3* elements (1, 8, 18; J. Chou et al., unpublished data).

We thank Nicholas Cozzarelli for advice.

This work was supported by Public Health Service grants GM 29067 and GM 06965 from the National Institute of General Medical Sciences to M.C. and L.H., respectively, and by National Science Foundation grant PCM 80-05599 to L.H. M.F. was supported by a fellowship from the Damon Runyon-Walter Winchell Cancer Fund (DRG-260-FT).

LITERATURE CITED

1. **Arthur, A., and D. Sherratt.** 1979. Dissection of the transposition process: a transposon-encoded site-specific recombination system. Mol. Gen. Genet. **175**:267–274.
2. **Campbell, A., D. Berg, D. Botstein, E. Lederberg, R. Novick, P. Starlinger, and W. Szybalski.** 1977. Nomenclature of transposable elements in prokaryotes, p. 1522. *In* A. I. Bukhari, J. A. Shapiro, and S. L. Adhya (eds.), DNA insertion elements, plasmids, and episomes. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
3. **Casadaban, M. J., J. Chou, and S. N. Cohen.** 1980. In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for detection and cloning of translational initiation signals. J. Bacteriol. **143**:971–980.
4. **Casadaban, M. J., J. Chou, P. Lemaux, C.-P. D. Tu, and S. N. Cohen.** 1981. Tn*3*: transposition and control. Cold Spring Harbor Symp. Quant. Biol. **45**:269–273.
5. **Chou, J., M. J. Casadaban, P. G. Lemaux, and S. N. Cohen.** 1979. Identification and characterization of a self-regulated repressor of translocation of the Tn*3* element. Proc. Natl. Acad. Sci. U.S.A. **76**:4020–4024.
6. **Chou, J., P. G. Lemaux, M. J. Casadaban, and S. N. Cohen.** 1979. Transposition protein of Tn*3*: identification and characterization of an essential repressor-controlled gene product. Nature (London) **282**:801–806.
7. **Fennewald, M., S. Gerrard, J. Chou, M. Casadaban, and N. Cozzarelli.** 1981. Purification of the Tn*3* transposase and analysis of its binding to DNA. J. Biol. Chem. **256**:4687–4690.
8. **Gill, R., F. Heffron, G. Dougan, and S. Falkow.** 1978. Analysis of sequences transposed by complementation of two classes of transposition-deficient mutants of Tn*3*. J. Bacteriol. **136**:742–756.
9. **Gill, R., F. Heffron, and S. Falkow.** 1979. Identification of the transposase and transposase-specific repressor encoded by Tn*3*. Nature (London) **282**:797–801.
10. **Heffron, F., B. J. McCarthy, H. Ohtsubo, and E. Ohtsubo.** 1979. DNA sequence analysis of the transposon Tn*3*: three genes and three sites involved in transposition of Tn*3*. Cell **18**:1153–1163.
11. **Heffron, F., M. So, and B. J. McCarthy.** 1978. *In vitro* mutagenesis using synthetic restriction sites. Proc. Natl. Acad. Sci. U.S.A. **75**:6012–6016.
12. **Hunkapiller, M. W., and L. E. Hood.** 1980. New protein sequenator with increased sensitivity. Science **207**:523–525.
13. **Ikeda, J. E., A. Yudelevich, and J. Hurwitz.** 1976. Isolation and characterization of the protein coded by gene *A* of bacteriophage φX174 DNA. Proc. Natl. Acad. Sci. U.S.A. **75**:2669–2673.

14. **Johnson, R., and W. Reznikoff.** 1981. Localization of the Tn*5* transposase promoter using the cycling reaction of RNA polymerase. Nucleic Acids Res. **9**:1873–1883.
15. **Kopecko, D. J., and S. N. Cohen.** 1975. Site-specific recA-independent recombination between bacterial plasmids: involvement of palindromes at the recombinational loci. Proc. Natl. Acad. Sci. U.S.A. **75**:1373–1377.
16. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
17. **Miller, J. H.** 1972. Experiments in molecular genetics, p. 352–355. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Reed, R. R.** 1981. Resolution of cointegrates between transposons γδ and Tn*3* defines the recombination site. Proc. Natl. Acad. Sci. U.S.A. **78**:3428–3432.