Functional Characterization of the Tn*5* Transposase by Limited Proteolysis*

Lisa A. Mahnke Braam‡ and William S. Reznikoff§

From the Department of Biochemistry and the M.D./Ph.D. Program, University of Wisconsin-Madison, Madison, Wisconsin 53706

The 476 amino acid Tn*5* transposase catalyzes DNA cutting and joining reactions that cleave the Tn*5* transposon from donor DNA and integrate it into a target site. Protein-DNA and protein-protein interactions are important for this tranposition process. A truncated transposase variant, the inhibitor, decreases transposition rates via the formation of nonproductive complexes with transposase. Here, the inhibitor and the transposase are shown to have similar secondary and tertiary folding. Using limited proteolysis, the transposase has been examined structurally and functionally. A DNA binding region was localized to the N-terminal 113 amino acids. Generally, the N terminus of transposase is sensitive to proteolysis but can be protected by DNA. Two regions are predicted to contain determinants for protein-protein interactions, encompassing residues 114–314 and 441–476. The dimerization regions appear to be distinct and may have separate functions, one involved in synaptic complex formation and one involved in nonproductive multimerization. Furthermore, predicted catalytic regions are shown to lie between major areas of proteolysis.

Transposable DNA elements range in diversity from simple bacterial insertion sequences to complex retroviruses that transpose via RNA intermediates (1). Tn*5* is a composite transposon found in Gram-negative bacteria and consists of two insertion sequences, IS*50*R and IS*50*L, that flank a central region containing antibiotic resistance genes (for a review, see Reznikoff (2)). IS*50*R encodes the 476 aa[1] transposase (Tnp) and a nearly identical protein, the transposition inhibitor (Inh). Tnp preferentially acts on Tn*5* elements located *cis* to its site of synthesis, whereas Inh is a *trans*-acting factor that decreases transposition. The Inh gene is read from the same reading frame as Tnp, but is under the control of separate transcription and translation start sites so that Inh lacks 55 aa of Tnp at the N terminus. *In vitro*, Tnp is necessary and sufficient for Tn*5* transposition in the presence of Mg^{2+} and specific DNA sequences called outside ends (OE) (3).

Transposition is a multistep process relying on both protein-DNA and protein-protein interactions. In general, the ends of a transposon are bound, synapsed, and cleaved by two or more transposase molecules acting in concert. Subsequently, the protein-transposon complex recognizes target DNA and catalyzes strand transfer, an event which is also called integration. In the case of Tn*5*, Tnp binds and synapses pairs of OE sites (4, 5). Then, Tnp cleaves the transposon free from donor DNA, leaving blunt ends with 3′-hydroxyl groups (3). Most likely, each cleavage is the result of two separate nucleophilic attacks on the phosphodiester backbone by water molecules; subsequently, the 3′-hydroxyl groups act as nucleophiles to attack the target DNA and integrate the transposon (6–8). By analogy to the closely related Tn*10* Tnp, we hypothesize that these catalytic reactions occur within the same active site with one monomer of a dimeric Tnp unit acting at each OE end (9). Inh is known to inhibit Tn*5* transposition via nonproductive multimerization with Tnp (4), and recently, Tnp itself has been shown to inhibit transposition in *trans* (10, 11). The *cis* restriction of Tnp is also thought to be attributable to nonproductive complex formation (12). Although Inh does not specifically bind DNA, it can form a three-way complex with an OE-bound Tnp monomer (4, 13). Inh heterodimerizes with Tnp in solution[2] and forms homodimers under certain conditions[2] (4).

During transposition, different domains and, probably, conformational changes in Tnp are responsible for the following functions: OE binding, synapsis involving Tnp multimerization, catalysis, and target recognition. The N terminus of Tnp is predicted to contain a DNA binding domain (14, 15). A C-terminal region has been identified as important for protein-protein interactions (15). Tn*5* Tnp belongs to the IS*4* family of transposases and shares, by sequence alignment with the IS*3* and IS*15* families and with retroelement integrases, a characteristic transposase/integrase motif probably important for catalysis (16). Here, Tnp is examined by limited proteolysis to further dissect functional aspects of the transposase.

EXPERIMENTAL PROCEDURES

Purification of Tnp and Inh—The overexpression of Tnp and Inh in *Escherichia coli* has been previously described (17), except the vectors used were pRZ7075 (17) for Tnp and pRZ4862 (constructed by M. D. Weinreich) for Inh. Cells were harvested, resuspended in 0.1 M NaCl TEG (20 mM Tris-HCl, pH 7.9, 1 mM EDTA, 10% glycerol) and 0.1 mM phenylmethylsulfonyl difluoride. For the purification of Tnp, 0.1% Triton X-100 was added. The cells were lysed with a French press at 16,000 p.s.i. Lysates were cleared by centrifugation at 38,000 × *g* and loaded onto a heparin acrylic bead (Sigma) column. The column was washed with 0.2 M NaCl TEG, and the protein was eluted with a 0.2 to 1.0 M NaCl TEG gradient. Peak fractions were diluted approximately 5-fold with 0.1 M NaCl TEG and loaded onto a reactive yellow no. 3 dye-agarose (Sigma) column. The column was washed with 0.2 M NaCl TEG, and the protein was eluted with a 0.2 to 0.6 M NaCl TEG gradient. Tnp degradation products co-purified with Tnp. These contaminants eluted

[1] The abbreviations used are: aa, amino acid; Tnp, transposase; Inh, transposase inhibitor; OE, transposon outside end; PAGE, polyacrylamide gel electrophoresis; Tricine, *N*-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine.

[2] L. A. M. Braam and W. S. Reznikoff, unpublished results.

at slightly lower salt concentrations than full-length Tnp during both chromatography steps; Inh also eluted at slightly lower salt concentrations.

Purification of Fusion-Tnp and Fusion-Inh—pRZ10100 was constructed by a polymerase chain reaction amplifying the coding region of Tnp with tailed primers containing *Eco*RI restriction sites and using *Pfu* polymerase (Stratagene). The polymerase chain reaction product was digested with *Eco*RI (Promega) and cloned into the *Eco*RI site in pET33b(+) (Novagen). The resulting plasmid overexpresses a Tnp fusion that consists of 41 aa containing a His$_6$ tag and a protein kinase recognition site at the Tnp N terminus. pRZ10200 is an overexpression vector for the analogous fusion with Inh, which was constructed in a similar fashion. The fusion proteins were overexpressed exactly as described for Tnp and Inh except 200 mg/ml kanamycin was used instead of ampicillin. Cells were harvested, resuspended in 0.4 M NaCl, 0.1% Triton X-100 TG (20 mM Tris-HCl, pH 7.9, 10% glycerol), and cleared lysates were loaded onto 1 ml of HiTrap columns (Pharmacia Biotech Inc.) chelated with Ni^{2+}. The columns were washed with 1.75 M NaCl, 30 mM imidazole, 20 mM Tris-HCl, pH 7.9, and the proteins were eluted using step gradients of 50–300 mM imidazole in 0.4 M NaCl TG.

Partial Proteolysis of Tnp, Inh, and Fusion Proteins—Protein samples were exchanged to HNKG buffer (100 mM NaCl, 100 mM potassium glutamate, 10% glycerol, 20 mM HEPES, pH 7.4) using 10 ml of Bio-Gel P-6DG desalting columns (Bio-Rad). Final protein concentrations were 0.5–1.0 mg/ml. Trypsin (Sigma) or chymotrypsin (Sigma) stocks were 10 mg/ml in 1 mM HCl. The proteases were diluted 10-fold in HNKG buffer before adding to protein samples at protein weight:protease weight ratios of approximately 200:1. The reactions were incubated at 30 °C and were stopped by boiling in the presence of SDS-PAGE loading buffer at indicated times.

Proteolytic Fragment Mapping—Proteolytic reactions were analyzed by 12% SDS-PAGE for large fragments or by Tris-Tricine SDS-PAGE for fragments less than 25 kDa. The peptides were visualized by Coomassie staining, and molecular weights were calculated based on standard relative mobilities. Fragments were blotted onto polyvinylidene difluoride membranes (Bio-Rad) and stained with Amido Black. The blots were analyzed at the Medical College of Wisconsin Protein/Nucleic Acid Shared Facility (Milwaukee, WI) for N-terminal sequence determination by Edman degradation. The locations of C termini were estimated based on fragment molecular weights and the primary aa sequence of Tnp.

CD Spectroscopy—Fusion-Tnp and fusion-Inh preparations (>98% pure based on Coomassie-stained SDS-PAGE gels) were exchanged to 0.3 M NaCl, 20 mM Tris-HCl, pH 7.9. Protein concentrations were measured by Bradford assay using a bovine serum albumin standard, and the proteins were diluted to 0.1 mg/ml. The spectra were recorded at 25 °C on a AVIV 62ADS circular dichroism spectrophotometer (AVIV Associates, Inc., Lakewood, NJ) using a 1-mm quartz cuvette over a range of 200–260 nm at 1-nm steps with an averaging time of 5 s. A blank was recorded and subtracted from each sample spectrum. Spectra were converted to molar ellipticity using 517 aa, 57,745 Da for fusion-Tnp and 462 aa, 52,009 Da for fusion-Inh.

N-terminal Protein Radiolabeling—Bovine heart muscle cAMP-dependent protein kinase-catalytic subunit (Promega) was added to reactions containing fusion proteins and [γ-^{32}P]ATP (NEN Life Science Products) in 1× kinase buffer (100 mM NaCl, 12 mM MgCl$_2$, 1 mM dithiothreitol, 10% glycerol, 20 mM Tris-HCl, pH 7.4). Reactions were incubated at room temperature for 30 min and exchanged to HNKG buffer using BioSpin P6 spin columns (Bio-Rad).

OE DNA Fragment Labeling—The following oligonucleotides were annealed: GCTCGGTACCCTGACTCTTATACACAAGT and GCTACT-TGTGTATAAGAGTCAGGGTACCGAGC. The resulting 29-base pair fragment contains the OE 19-base pair sequence with a 3-base overhang. 25 pmol of DNA and 33 pmol of [α-^{32}P]dATP were incubated with super reverse transcriptase (Molecular Genetics Resources) according to the manufacturer's instructions for 30 min at 37 °C. The reaction was desalted using a Microspin G-25 spin column (Pharmacia Biotech Inc.) to remove unincorporated nucleotides.

Southwestern Analysis—Fusion-Tnp, fusion-Tnp partially digested with trypsin, or fusion-Inh were electrophoresed using 12% SDS-PAGE and blotted onto nitrocellulose. The blot was incubated in SW hybridization buffer (75 mM potassium glutamate, 2 mM MgCl$_2$, 0.1 mM ZnCl$_2$, 1 mM dithiothreitol, 0.05% Igepal CA-630 (Sigma), 1% nonfat dried milk, 10% glycerol, 20 mM HEPES, pH 7.2) overnight. The blot was probed for 3 h using fresh buffer containing 1 pmol/ml ^{32}P-labeled OE fragment and was rinsed, dried, and analyzed by phosphorimaging.

OE Protection Experiment—^{32}P-labeled fusion-Tnp was incubated for 30 min in HNKG buffer with a 10-fold molar excess of cold OE fragment
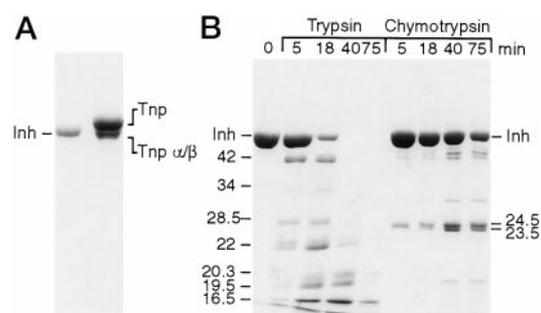

FIG. 1. **SDS-PAGE analysis of proteolytic fragments.** The gels were stained with Coomassie Blue. *A*, purified Tnp and Inh. Tnpα/β represent degradation products of Tnp. *B*, Inh was treated with limiting amounts of trypsin or chymotrypsin for various lengths of time. Mapped fragments are labeled according to kDa.

or in a buffer control. The wild type transposase used here is not known to detectably cleave OE DNA *in vitro*. Trypsin was added, and the reactions were stopped by boiling in the presence of SDS-PAGE loading buffer at various times. The reactions were electrophoresed on Tris-Tricine SDS-PAGE with prestained molecular weight markers (Bio-Rad). The gels were dried and analyzed by phosphorimaging; molecular weights were calculated based on standard relative mobilities.

Far Western Assay—Inh was partially digested with trypsin or chymotrypsin, electrophoresed on Tris-Tricine SDS-PAGE, and blotted onto nitrocellulose. FW hybridization buffer was made exactly as SW hybridization buffer except 75 mM KCl was substituted for potassium glutamate. Blots were incubated in this buffer overnight and then probed using fresh FW hybridization buffer containing 25 μg/ml ^{32}P-labeled fusion-Inh or ^{32}P-labeled fusion-Tnp for 3 h. The blots were rinsed, dried, and analyzed by autoradiography.

RESULTS

Stable Proteolytic Cleavage Fragments of Tnp and Inh—To map characteristic proteolytic degradations of Tnp that we observe in our preparations, purified Tnp and Inh (Fig. 1*A*) were studied by N-terminal sequencing. The degradation products have previously been shown to be N-terminal truncations termed Tnpα and Tnpβ (4) (Fig. 1*A*). Sequencing results are summarized in Table I. Tnpα was found to be a doublet, here termed Tnpα1 and Tnpα2 (Fig. 2*A*). A serine protease(s) probably produces Tnpα1, Tnpα2, and Tnpβ as these cleavages occur after R or K residues.

Next, we wished to fragment Tn5 Inh using partial proteolysis to define surface exposed areas and buried and/or folded regions. Treatment of Inh with trypsin results in seven stable fragments ranging from 16.5 to 42 kDa (Fig. 1*B*). Chymotrypsin proteolysis produces two major fragments of molecular masses of 24.5 and 23.5 kDa (Fig. 1*B*). These peptides were sequenced at the N terminus and mapped. Inh results are reported here in the context of the Tnp reading frame (see "Discussion"). N termini and protease cleavage sites are summarized in Table I. Mapping is summarized in Fig. 2*A*. Trypsin can potentially cleave 14% of the total residues in either Tnp or Inh in a nearly random distribution as diagrammed in Fig. 2*B*.

Proteolytic Cleavage Patterns of Tnp and Inh Fusion Proteins—To compare the proteolytic digestions of Tnp and Inh, 41 aa N-terminal fusions were constructed. When fusion-Tnp was purified under high salt conditions, Tnpα1,2 and Tnpβ were not present (Fig. 3, *lane 1*). Fusion-Inh was similarly purified, and both proteins were subjected to trypsin proteolysis (Fig. 3). We found the fusion-Tnp and fusion-Inh tryptic cleavage patterns to be similar. Tnp fragment 1a is roughly the same size as fusion-Inh and Tnp fragment 1b corresponds, approximately, to Inh fragment 1. We will demonstrate below that Tnp fragment 1a results from proteolysis within the fusion. The early fragments are stable until at least 10 min (Fig. 3, compare *lanes 2–5* with *lanes 10–13*). Tnp fragment 2 approximates Inh fragment 2 in size; Tnp and Inh fragments 3 and 4 are less prom-

TABLE I
N-terminal sequencing results of full-length proteins and proteolytic fragments

The first 10–14 aa of each protein were determined by N-terminal sequencing as described under "Experimental Procedures" and were located in the Tnp primary sequence. The N-terminal proteolytic cleavage sites are indicated.

Fragment[a]	N-terminal protease cleavage site[b]
Tnp (full-length)	Not cleaved
Inh (full-length)	Not cleaved
Tnpα_1	Arg-30*Leu-31
Tnpα_2	Lys-40*Tyr-41
Tnpβ	Arg-62*Phe-63
Chymotrypsin 264–476	Leu-263*Arg-264
Chymotrypsin 62–263	Tyr-61*Arg-62
Trypsin 63–431	Arg-62*Phe-63
Trypsin 114–412	Lys-113*Leu-114
Trypsin 63–314	Arg-62*Phe-63
Trypsin 253–440	Lys-252*Arg-253
Trypsin 253–431	Lys-252*Arg-253
Trypsin 255–418	Lys-254*Asn-255
Trypsin 114–256	Lys-113*Leu-114
Tnp fragment 1α (fusion)	Arg--21*Ala--20[c]

[a] Numbering indicates Tnp primary sequence aa.
[b] Asterisks indicate positions of the cleaved peptide bonds.
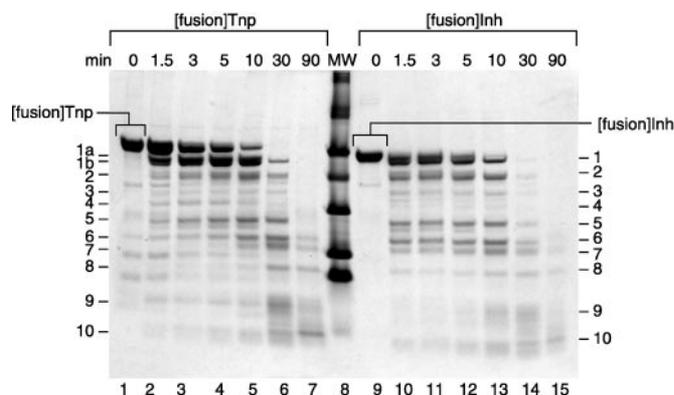[c] Minus signs indicate numbers of aa within the N-terminal fusion from Tnp residue 1.
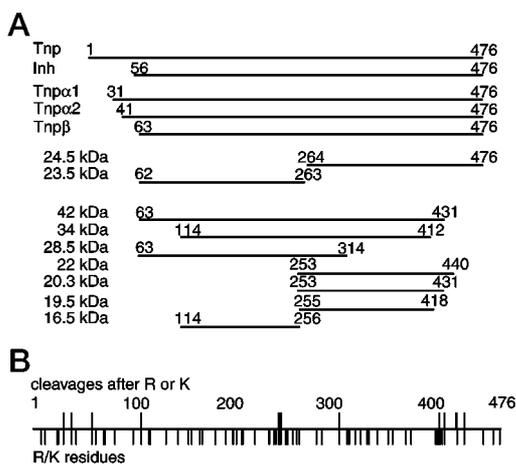
FIG. 2. **Proteolytic cleavage mapping.** *A*, Tnp, Inh, Tnpα1,α2,β, and proteolytic fragments of Inh are located relative to the primary sequence of Tnp. Inh proteolytic fragments are labeled based on molecular weight. Fragment N termini were obtained by sequencing (Table I). SDS-PAGE molecular weight calculations combined with Tnp primary sequence information were used estimate the C-terminal end points. *B*, trypsin cleavages and cleavages associated with Tnpα1,2 and Tnpβ (*upper bars*) are compared with all R and K residues (*lower bars*) in the Tnp primary sequence.

FIG. 3. **Limited trypsin digests of Tnp and Inh fusion proteins.** Trypsin digests were stopped at the indicated times and the reactions were analyzed by SDS-PAGE. The protein bands were visualized by Coomassie staining. Stable fragments are labeled *1–10*. *MW* indicates molecular weight standards.
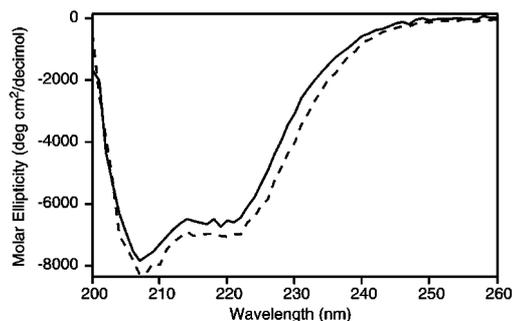
FIG. 4. **CD spectroscopy.** Spectra of fusion-Tnp (*dashed line*) and fusion-Inh (*solid line*) in 0.3 M NaCl, 20 mM Tris-HCl, pH 7.9, were recorded using a cell path length of 1 mm. The protein concentrations were 0.1 mg/ml, as determined by Bradford assay.

inent, especially Inh fragment 4, but are roughly equivalent (Fig. 3, compare *lanes 2–6* with *lanes 10–14*). At later times, Tnp fragments 5–8 are produced. These fragments are stable and correspond to Inh fragments 5, 6, 7, and 8 (Fig. 3). Finally, smaller fragments, 9 and 10, appear in the digest patterns of both Tnp and Inh (Fig. 3, compare *lanes 5–7* with *lanes 13–15*).

CD Spectroscopy—Since Tnp shares 88% identity with Inh and shows similar trypsin cleavages, we hypothesize that the two proteins are folded similarly. CD spectra of [fusion]Tnp and [fusion]Inh were collected in 0.3 M NaCl, Tris-HCl buffer without glycerol. Under these conditions, both proteins are monomeric.[2] The spectra are shown in Fig. 4. The shape and magnitude of the spectra are very similar and both have minima at about 222 and 208 nm, which is indicative of an α-helical structure. The slight difference in magnitude is within the error expected for the protein concentration determinations used.

Southwestern Analysis—Previous studies have suggested that the N terminus of Tnp is important for OE DNA binding (14, 15), and we sought to identify the extent of such a domain. Trypsin fragments were tested for OE binding activity. We used a fusion protein preparation because it was not contaminated with Tnpα1,2 and Tnpβ (see above). Only the full-length fusion and Tnp fragment 1a bound OE DNA (Fig. 5). As a negative control, fusion-Inh did not bind OE, as expected (data not shown). Tnp fragment 1a was mapped and was found to be truncated within the fusion inside the protein kinase site (Table I). Thus, Tnp fragment 1a contains an intact Tnp N terminus and, approximately, a complete C terminus.

OE Protection Assay—Because the Southwestern experiment did not reveal a DNA binding domain and because the N terminus of Tnp is apparently sensitive to proteolysis, we asked whether OE DNA could protect Tnp from trypsin attack and in doing so define the DNA binding region. Radiolabeled fusion-Tnp was subjected to trypsin proteolysis in the presence and absence of OE DNA (Fig. 6*A*). The reactions were analyzed by phosphorimaging so that only those fragments with intact N-terminal fusion kinase sites were apparent. In the absence of DNA, no stable fragments were observed except for part of the fusion itself near the bottom of the gel and a faint band migrating at approximately 10 kDa (Fig. 6*A*, *lanes 2–7*). The intact fusion is 4.4 kDa; the fusion fragment was estimated to be 1.0–2.0 kDa and, therefore, has probably been cleaved by trypsin both N-terminal and C-terminal to the site of phosphorylation. Very likely, the N-terminal cleavage occurs after the 17th residue (R) from the beginning of the fusion, which is -24 residues from the start of Tnp. In the presence of OE DNA, stable fragments of molecular masses of 14.5, 10, 7, and 3.3
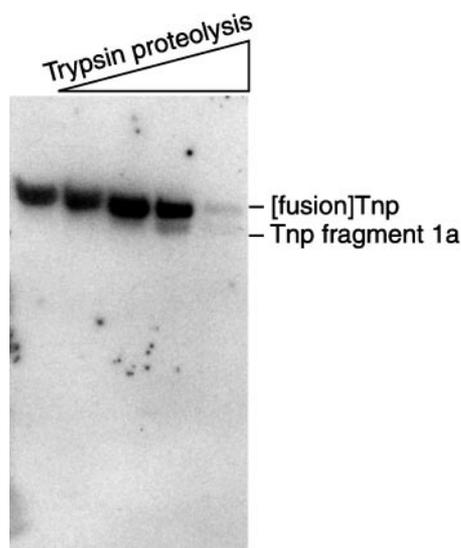
FIG. 5. **Southwestern analysis.** Fusion-Tnp was digested for 30 s and 1, 5, 10, and 30 min with limiting amounts of trypsin. The reactions were blotted onto nitrocellulose. The blots were probed with a ^{32}P-labeled OE DNA fragment and were analyzed by phosphorimaging. The *triangle* indicates increasing time of digestion.
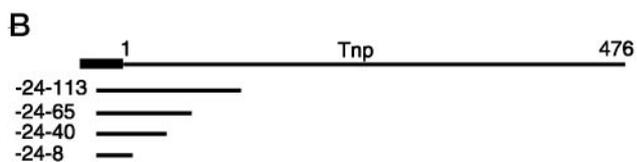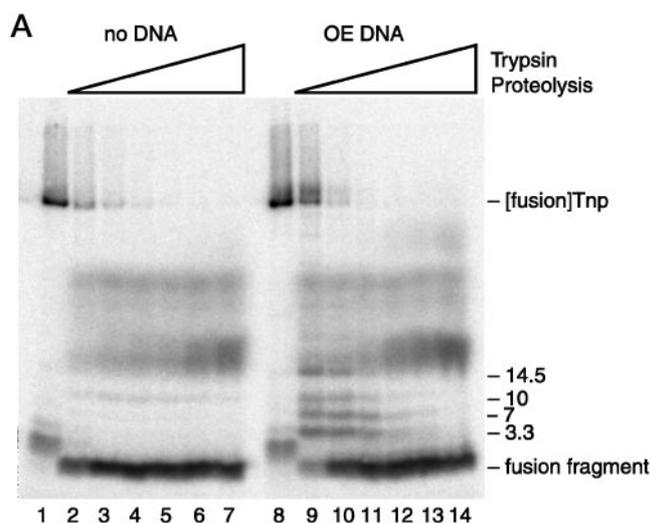

FIG. 6. **OE protection assay.** *A*, ^{32}P-labeled fusion-Tnp was treated with limiting amounts of trypsin in the absence (*no DNA*) or presence (*OE DNA*) of excess OE DNA fragment for 30 s and 1, 3, 5, 10, and 30 min. Reactions were analyzed by SDS-PAGE and phosphorimaging. Fusion-Tnp, protected fragments (labeled according to kDa) and the fusion fragment are indicated. *Triangles* indicate increasing times of trypsin digestion. *B*, the protected fragments were mapped assuming a cleavage within the fusion as described under "Results." The *black box* indicates the fusion sequence. Numbering reflects aa in the Tnp primary sequence with −24 indicating a distance of 24 aa to the N terminus of Tnp residue 1.

kDa were observed (Fig. 6*A*, *lanes 9–13*), indicating partial protection by the DNA against proteolysis in this region. The protected bands were mapped as shown in Fig. 6*B*.

Far Western Studies—The C terminus of Tnp/Inh has been implicated in protein dimerization (15). To further define the locations of dimerization regions, proteolytic fragments of Inh were probed with end-labeled Inh or Tnp. Full-length Inh binds
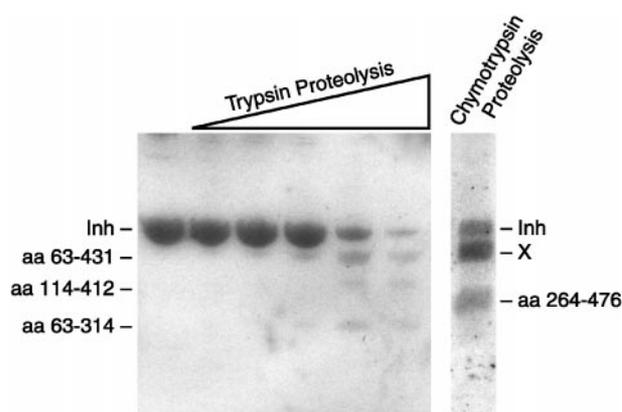

FIG. 7. **Far Western analysis.** Inh was treated with limiting amounts of trypsin for 30 s and 1, 3, 10, and 30 min, or with chymotrypsin for 10 min. The reactions were blotted onto nitrocellulose. The blots were probed with ^{32}P-labeled fusion-Inh and were visualized by autoradiography. The *triangle* indicates increasing trypsin digestion. Inh and mapped fragments are indicated. *X* indicates a fragment that had not been previously mapped.

to a ^{32}P-labeled fusion-Inh probe as do fragments of 63–43, 114–412, and 63–314 aa (Fig. 7). Fragments of 253–440, 253–431, 255–418, and 114–256 aa do not bind under these conditions. Identical results were obtained using ^{32}P-labeled fusion-Tnp as the probe (data not shown). When partial chymotryptic digests of Inh were examined in the same way, full-length Inh and fragment of 264–476 aa, but not fragment of 62–263 aa, bound to ^{32}P-labeled fusion-Inh (Fig. 7) or ^{32}P-labeled fusion-Tnp (data not shown).

DISCUSSION

Tnp and Inh Are Folded Similarly—The primary sequence of Inh is identical to the C-terminal 421 aa of the 476-aa Tnp, but Inh is produced from a unique start codon, and it is possible that nascent Inh adopts a distinct fold upon translation. Here, the cleavages in Tnp, which produce the natural N-terminal truncations Tnpα1,2 and Tnpβ, have been identified, and interestingly, the Tnpβ cleavage corresponds to a trypsin-sensitive region in Inh (aa 61/62), indicating similar folding in this region. Furthermore, the patterns of Tnp and Inh fusion protein trypsin digests were found to match closely, indicating overall folding likenesses. When Tnp and Inh were examined by CD spectroscopy, no evidence was found to indicate major secondary structure differences. Therefore, Tnp and Inh are presumed to exhibit similar secondary and tertiary folding in the long C-terminal region that they share. Thus, we have adopted the convention of labeling the primary sequence of Inh based on the Tnp sequence.

Limited Proteolysis Reveals Tnp/Inh Surface-exposed Regions—The N terminus of Tnp and the first part of Inh appear to be particularly susceptible to proteolysis with cleavages occurring after aa 30, 40, 61/62, and 113. Overall, we have classified proteolytic sites into major and minor cleavage regions: three major regions occur at 61/62, 252–263, and 412–440 aa; two minor cleavages were detected after aa 113 and 314. The cleavages indicate accessible regions which are likely to be surface exposed. The trypsin-sensitive sites and the sites associated with Tnpα1,2 and Tnpβ are specific; there is a relatively high content of R and K residues in the Tnp primary sequence that are not cleaved (Fig. 2*B*), and both trypsin and chymotrypsin showed cleavage within the 61/62 aa and the 252–263 aa major sites.

An N-terminal Tnp Fusion Can Bind OE DNA—When fusion-Tnp trypsin fragments were studied in the Southwestern experiment, the full-length fusion and a fusion cleaved within
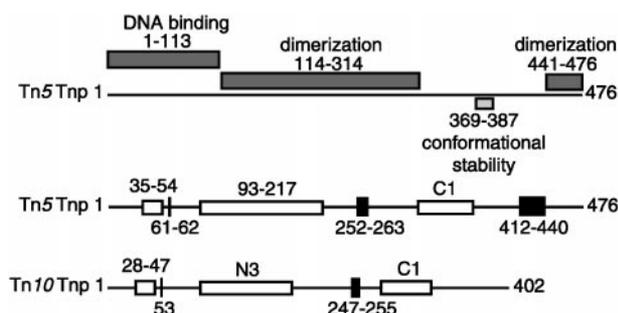
FIG. 8. **Tnp functional analysis summary.** Numbering indicates aa. Tn*5* Tnp functional regions are shown as *gray boxes*. Tn*10* Tnp was aligned with Tn*5* Tnp based on the C1 regions (16) and the N-terminal proteolytic sites aa 61/62, Tn*5*, and aa 53, Tn*10* (22). *Open boxes* indicate extended regions of homology. Tn*5* region 93–217 aa corresponds to Tn*10* region N3. Tn*10* Tnp region 28–47 aa is homologous to the Tn*5* Tnp predicted helix-turn-helix region, aa 35–54 (A. S. Silbergleit, V. A. Lanzof, and N. Benuch, personal communication). *Black rectangles* indicate major proteolytic cleavage regions as defined here and in Kwon *et al.* (22).

the leader sequence (Tnp fragment 1a) showed OE DNA binding activity. Thus, the fusion itself did not interfere with binding even though we predict protein-DNA contacts very close to the Tnp N terminus (see below). Since no Tnp primary sequence truncations demonstrated DNA binding, we hypothesize that the fragments have been cleaved within a DNA binding region somewhere in the N terminus.

The N Terminus of Tnp Can Be Protected against Proteolysis by DNA and Has a DNA Binding Domain—When fusion-Tnp labeled at the N terminus was subjected to proteolysis and was analyzed by phosphorimaging, no stable fragments were apparent; however, we know that fusion-Tnp is initially cleaved by trypsin within the fusion kinase site. Interestingly, OE DNA could protect the N terminus of Tnp, and the estimated protection extends precisely to the previously mapped trypsin-sensitive site at aa 113. Thus, we propose aa 1–113 to be important for DNA binding (Fig. 8). This hypothesis agrees with other observations as follows: 1) Inh does not bind OE DNA (4) and, thus, the first 55 aa of Tnp are associated with DNA binding; and 2) a series of point mutations at positions 41, 47, and 54 have been shown to alter OE DNA binding affinity or specificity (14). The smaller N-terminal protected fragments ending at aa 65 and 40, approximately, are also near protease accessible sites (aa 61/62 and 40). Finally, the shortest protected fragment extends to aa 8, suggesting that DNA contacts are made very near the N terminus. This observation is in agreement with deletion studies that showed Tnp truncated by 3 aa at the N terminus could bind OE DNA but Tnp truncated by 11 aa could not (15).

The 1–113 aa region encompasses a predicted helix-turn-helix structure (aa 35–54)[3] that may serve as the DNA-binding motif (Fig. 8), and this region is susceptible to proteolysis, as shown by the partial proteolytic mapping and OE DNA protection experiment and as suggested by the Southwestern probing. Upon binding, residues 1–113 either directly contact the DNA or are sterically occluded from proteases by proximal DNA. A more complex model can also be envisioned such that N-terminal folding associated with DNA binding contributes to the observed inhibition of proteolysis. Unstructured DNA binding regions are known. For example, in the lac repressor, a hinge helix undergoes a coil-to-helix transition in the presence of the lac operator DNA binding site (18). This phenomenon correlates with an "induced fit" mechanism of protein-DNA

interactions (19) and is consistent with a protein/DNA co-crystal structure of Tc*3* Tnp (20). We note that observed context-dependent effects of OE DNA mutations on Tnp binding might also be explained by an induced fit model (21).

The Shared Tnp/Inh Sequence Contains at Least Two Protein-Protein Interaction Domains which Are Distinct—In the Far Western assay, Inh fragment aa 264–476 was found to interact with both Inh and Tnp. Shorter fragments, aa 253–440, 253–431, and 255–418, could not dimerize, and we hypothesize that region aa 441–476 contains determinants for dimerization (Fig. 8). While it is possible that the nondimerizing C-terminal fragments are simply destabilized or unfolded, there is a relatively small difference in overall size between fragment aa 264–476 and 253–440, and we suspect that the latter fragment may fold significantly. Our hypothesis gains further merit when two C-terminal point mutations are considered (see below).

A separate region of dimerization was detected in fragments aa 63–431 and 63–314. Further deletion to aa 62–263 abolished the dimerization. Correspondingly, the deletion of fragment 114–412 aa to 114–256 aa resulted in loss of dimerization. Therefore, we propose a second protein-protein interaction region between aa 114 and 314 (Fig. 8). This region covers a trypsin-sensitive area spanning aa 252–263 and may be divided by a surface-exposed loop.

The two dimerization regions appear to be distinct because the absence of aa 441–476 does not negate the protein-protein interactions of aa 114–314, an observation in agreement with DNA binding studies described below. Conversely, deletion of most of the 114–314 aa region does not eliminate dimerization at the C terminus. Nevertheless, we cannot rule out the possibility that aa 264–314 interacts functionally with determinants within or near aa 441–476. Furthermore, each dimerization region may represent a suboptimal domain that is fully functional only in the context of other parts of the protein.

In any case, the locations of dimerization determinants found here were surprising in that they did not include aa 369–387. This region was identified in a Tnp C-terminal deletion study in which Tnp truncated to aa 387 could bind OE DNA and form protein-protein interactions in the dead-end three-way complex (a DNA-bound Tnp monomer interacting with another Tnp or Inh molecule) (13, 15). However, Tnp truncated to aa 369, while retaining DNA binding activity, was completely defective in forming the three-way complex (15). Thus, it has been hypothesized that aa 369–387 falls within a dimerization domain (15). We now speculate that the region itself is not involved in making protein-protein contacts directly but is positioned so that it can conformationally affect dimerization, perhaps through positioning or stabilization of globular domains (Fig. 8). In the three-way complex, the DNA binding event might cause allosteric effects which promote dimerization.

Possible Functions of the Dimerization Regions—Tn*10* Tnp is closely related to Tn*5* Tnp and has been divided into proteolysis-resistant regions by trypsin digestion (22). The 402 aa Tn*10* Tnp has a trypsin-sensitive site after aa 53 and a linker region spanning aa 247–255 (22) (Fig. 8). These cleavage regions are similar to Tn*5* Tnp major proteolytic sites aa 61/62 and 252–263 (Fig. 8). The Tn*10* Tnp-resistant regions appear to align well with N-terminal Tn*5* Tnp proteolysis-resistant regions defined here (Fig. 8). However, there appears to be an extra region in the Tn*5* Tnp C terminus, aa 441–476, that is important for dimerization (Fig. 8). We suspect that this region may be involved in nonproductive protein-protein interactions distinct from interactions used for synapsis. Our reasoning is based partly on a hypertransposing Tnp mutation, L372P,

---

[3] A. S. Silbergleit, V. A. Lanzov, and N. Benuch, personal communication.

which is *trans*-active and is defective for *trans*-inhibition (12). Based on L372P, the *cis*-restriction phenomenon has been attributed to the nonproductive multimerization seen during inhibition (12). However, the mutant protein must be able to efficiently multimerize to form synaptic transposition complexes. Therefore, it is possible that two multimerization domains in Tnp are used for different functions. Tn*10* represents a different case in which the *cis*-preference has been ascribed to low protein abundance and to regulation at the level of translation (23, 24). Moreover, Tn*10* Tnp is apparently not *trans*-inhibitory, and the element does not encode for a transposition inhibitor similar to Inh. It seems likely that Tn*10* does not utilize nonproductive multimerization to down-regulate transposition as does Tn*5*, possibly due to the fact that Tn*10* Tnp does not encode an extra Tn*5*-like domain.

The L372P mutation falls within the 369–387 aa sequence (discussed above) and may exert its effect through a conformational change in the protein via global positioning of dimerization region aa 441–476. Two other Tnp mutations that map within aa 441–476 further support its role in inhibition and *cis*-restriction: 1) a temperature-sensitive mutation, L449F, shows higher *trans* transposition relative to *cis* and has reduced *trans*-inhibitory activity (10); and 2) E451Q is a hyperactive mutant (25), which could be a defective inhibitor.

Nearer the N terminus, multimerization region aa 114–314 overlaps with predicted catalytic regions (see below and Fig. 8). We speculate that it has a role in synapsis. Protein-protein interactions between catalytic regions have also been observed for the HIV-1 and ASV integrases (26, 27).

Predictions about the Catalytic Domain and Further Tn10 Tnp Comparisons—Rezsohazy *et al.* (16) have defined conserved catalytic regions, N3 and C1, which include the characteristic DDE triad found in transposases and retroviral integrases (1). Tn*5* Tnp C1, aa 313–365 (16), lies between two major cleavage sites (Fig. 8) and is adjacent to a minor cleavage, approximately aa 314. It is likely that this region is folded into all or part of a globular domain and probably contributes to the Tnp catalytic pocket. Hypothetical N3 (aa 230–257 as predicted in Rezsohazy *et al.* (16)) is not analogous to Tn*10* Tnp N3 (16) and overlaps with the 252–263 aa-sensitive site. Moreover, the homology in this region is low, and we think it is not part of the catalytic pocket. We aligned the Tn*5* and Tn*10* Tnp primary sequences based on C1 (16) and the major proteolytic cleavage regions and could approximate the Tn*5* "N3" region at aa 93–217 relative to Tn*10* (16) (Fig. 8). Interestingly, we also detected significant homology in Tn*10* Tnp at the Tn*5* predicted helix-turn-helix region[3] adjacent to the aligned N-terminal major proteolytic sites (Fig. 8). As a final comparison, we note that the Tn*5* proteolytic sensitive site aa 252–263 nearly aligns with the Tn*10* "linker" region, aa 247–255 (22), but is shifted slightly to the N terminus based on homology identified here.

REFERENCES

1. Polard, P., and Chandler, M. (1995) *Mol. Microbiol.* **15,** 13–23
2. Reznikoff, W. S. (1993) *Annu. Rev. Microbiol.* **47,** 945–963
3. Goryshin, I. Y., and Reznikoff, W. S. (1998) *J. Biol. Chem.*, **273,** 7367–7374
4. de la Cruz, N. B., Weinreich, M. D., Wiegand, T. W., Krebs, M. P., and Reznikoff, W. S. (1993) *J. Bacteriol.* **175,** 6932–6938
5. Goryshin, I. Y., Kil, Y. V., and Reznikoff, W. S. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 10834–10838
6. Engelman, A., Mizuuchi, K., and Craigie, R. (1991) *Cell* **67,** 1211–1221
7. Mizuuchi, K., and Adzuma, K. (1991) *Cell* **66,** 129–140
8. van Gent, D. C., Mizuuchi, D., and Gellert, M. (1996) *Science* **271,** 1592–1594
9. Bolland, S., and Kleckner, N. (1996) *Cell* **84,** 223–233
10. Delong, A., and Syvanen, M. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 6072–6076
11. Wiegand, T. W., and Reznikoff, W. S. (1992) *J. Bacteriol.* **174,** 1229–1239
12. Weinreich, M. D., Gasch, A., and Reznikoff, W. S. (1994) *Genes Dev.* **8,** 2363–2374
13. York, D., and Reznikoff, W. S. (1996) *Nucleic Acids Res.* **24,** 3790–3796
14. Zhou, M., and Reznikoff, W. S. (1997) *J. Mol. Biol.* **271,** 362–373
15. Weinreich, M. D., Mahnke-Braam, L., and Reznikoff, W. S. (1994) *J. Mol. Biol.* **241,** 166–177
16. Rezsohazy, R., Hallet, B., Delcour, J., and Mahillon, J. (1993) *Mol. Microbiol.* **9,** 1283–1295
17. Wiegand, T. W., and Reznikoff, W. S. (1994) *J. Mol. Biol.* **235,** 486–495
18. Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. (1996) *Science* **271,** 1247–1254
19. Spolar, R. S., and Record, M. T. Jr. (1994) *Science* **263,** 777–784
20. Pouderoyen, G., Ketting, R. F., Perrakis, A., Plasterk, R. H. A., and Sixma, T. K. (1997) *EMBO J.* **16,** 6044–6054
21. Zhou, M., Bhasin, A., and Reznikoff, W. S. *J. Mol. Biol.*, in press
22. Kwon, D., Chalmers, R. M., and Kleckner, N. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 8234–8238
23. Kleckner, N. (1989) in *Mobile DNA* (Berg, D. E., and Howe, M. M., eds) pp. 227–268, American Society for Microbiology, Washington, D. C.
24. Jain, C., and Kleckner, N. (1993) *Mol. Microbiol.* **9,** 249–260
25. Krebs, M. P., and Reznikoff, W. S. (1988) *Gene* (*Amst.*) **63,** 277–285
26. Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A., and Skalka, A. M. (1995) *J. Mol. Biol.* **253,** 333–346
27. Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., and Davies, D. R. (1994) *Science* **266,** 1981–1985