

# Involvement of a Bifunctional, Paired-like DNA-binding Domain and a Transpositional Enhancer in *Sleeping Beauty* Transposition\*

Received for publication, April 24, 2002, and in revised form, June 10, 2002  
Published, JBC Papers in Press, June 24, 2002, DOI 10.1074/jbc.M204001200

Zsuzsanna Izsvák<sup>‡§</sup>, Dheeraj Khare<sup>‡¶</sup>, Joachim Behlke<sup>‡</sup>, Udo Heinemann<sup>‡||</sup>, Ronald H. Plasterk<sup>\*\*</sup>, and Zoltán Ivics<sup>‡ ‡‡</sup>

From the <sup>‡</sup>Max Delbrück Center for Molecular Medicine, D-13092 Berlin, Germany, the <sup>§</sup>Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, 6726 Szeged, Hungary, the <sup>||</sup>Institute of Chemistry/Crystallography, Free University of Berlin, 14195 Berlin, Germany, and the <sup>\*\*</sup>Hubrecht Laboratory of Developmental Biology, 3584 CT Utrecht, The Netherlands

***Sleeping Beauty* (SB) is the most active Tc1/mariner-like transposon in vertebrate species. Each of the terminal inverted repeats (IRs) of SB contains two transposase-binding sites (DRs). This feature, termed the IR/DR structure, is conserved in a group of Tc1-like transposons. The DNA-binding region of SB transposase, similar to the paired domain of Pax proteins, consists of two helix-turn-helix subdomains (PAI + RED = PAIRED). The N-terminal PAI subdomain was found to play a dominant role in contacting the DRs. Transposase was able to bind to mutant sites retaining the 3' part of the DRs; thus, primary DNA binding is not sufficient to determine the specificity of the transposition reaction. The PAI subdomain was also found to bind to a transpositional enhancer-like sequence within the left IR of SB, and to mediate protein-protein interactions between transposase subunits. A tetrameric form of the transposase was detected in solution, consistent with an interaction between the IR/DR structure and a transposase tetramer. We propose a model in which the transpositional enhancer and the PAI subdomain stabilize complexes formed by a transposase tetramer bound at the IR/DR. These interactions may result in enhanced stability of synaptic complexes, which might explain the efficient transposition of *Sleeping Beauty* in vertebrate cells.**

Mobility of DNA-based transposable elements is usually regulated by both host-encoded and element-encoded factors, which operate by imposing constraints on transposition. One important form of transpositional control is represented by regulatory “checkpoints,” at which certain molecular requirements have to be fulfilled for the transpositional reaction to proceed. These requirements can operate at different stages of transposition, including the very early steps of binding of the transposase to the transposon DNA and synaptic complex as-

sembly, the process by which the two ends of the elements are paired and held together by transposase subunits. Transposition of one of the best studied transposons, bacteriophage Mu, is tightly regulated at the level of synaptic complex assembly, which does not proceed unless certain conditions are met (1). Transpositional enhancers in diverse recombination systems might serve as catalysts of transposon/transposase complex assembly. For example, the very N-terminal domain of Mu transposase plays an important role in transposition by binding to a transpositional enhancer, the internal activating sequence (2). Similarly, the P element transposon of *Drosophila* contains enhancer sequences (3), which are believed to transiently interact with P transposase to promote transposition (4). In the Hin recombinase system, an enhancer is bound by the bacterial accessory protein FIS, which simultaneously interacts with the recombinase bound to the recombination sites to promote assembly of the synaptic complex (5).

Eukaryotic transposons often possess protein domains that are not found in prokaryotes and *vice versa*. For example, members of the Tc1/mariner superfamily contain a paired box-like DNA-binding domain. The paired domain was originally described in important developmental regulatory proteins in eukaryotes, such as the Pax family of transcription factors (6–8). The paired domain has a bipartite structure containing two helix-turn-helix (HTH)<sup>1</sup> motifs (PAI + RED = PAIRED) that are organized as modular DNA binding domains, and have evolved versatility in binding to a range of different DNA sequences through various combinations of these domains (6, 7). Both subdomains might have the ability to bind DNA, but usually the N-terminal PAI subdomain has more specific DNA recognition (6). The function of the RED subdomain is not always clear; it can enforce the specificity of the PAI subdomain (7, 9) or can be involved in protein targeting (10). The nucleotide sequences recognized by the composite paired domain are degenerate; the DNA-binding specificity is relaxed (6, 7, 9). The origin of the paired domain is not clear, but phylogenetic analyses indicate that it might have been derived from an ancestral transposase (11).

The vast majority of naturally occurring Tc1/mariner-like transposons are nonfunctional because of inactivating frameshift mutations, small deletions, and internal translational termination codons. In vertebrates, not a single active element has been found. Based on a comparative phylogenetic ap-

\* This work was supported in part by a research fellowship from EMBO (to Z. Izsvák), by Deutsche Forschungsgemeinschaft Grant GRK 8013 (to D. K.), and by research fellowships from Human Frontier Science Programme Organization and Organisation for Economic Co-operation and Development (to Z. Ivics). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>¶</sup> Supported by the graduate college “Model Studies of Structure, Properties and Recognition of Biological Macromolecules.”

<sup>‡‡</sup> To whom correspondence should be addressed: Max Delbrück Center for Molecular Medicine, Robert Rössle Str. 10, D-13092 Berlin, Germany. Tel.: 49-30-9406-2546; Fax: 49-30-9406-2547; E-mail: zivics@mdc-berlin.de.

<sup>1</sup> The abbreviations used are: HTH, helix-turn-helix; IR, inverted repeat; DR, transposase-binding site; HDR, transposase-binding half-site; SB, *Sleeping Beauty*; BS<sup>3</sup>, bis(sulfosuccinimidyl) suberate; oligo, oligonucleotide; EMSA, electrophoretic mobility shift assay; NLS, nuclear localization signal.

proach, we have reconstructed an active Tc1-like transposon from bits and pieces of inactive elements found in the genomes of teleost fish, and named this transposon *Sleeping Beauty* (SB) (12).

SB is flanked by terminal inverted repeats (IRs), which contain binding sites for the enzymatic factor of transposition, the transposase. The transposase-binding sites of SB elements are repeated twice per IR in a direct orientation (DRs). This special organization of inverted repeat, termed IR/DR, is shared by a group of Tc1-like transposons, but not by Tc1 itself (13). Specific binding to the DRs is mediated by an N-terminal, paired-like DNA-binding domain of the transposase (12), which overlaps with a nuclear localization signal (NLS) (14). A GRPR-like sequence (GRRR in SB) between the two HTH motifs is conserved in Tc1/*mariner* transposases (13, 15). The GRPR motif is characteristic to homeodomain proteins (8) and mediates interactions with DNA in the Hin invertase of *Salmonella* (16) and in the RAG1 recombinase that mediates V(D)J recombination of immunoglobulin genes in vertebrates (17).

SB mediates transposition in a variety of vertebrate species (18–20), and is more active than other members of the Tc1/*mariner* family (21). Because there is substantial interest in developing transposon technology for gene therapy and gene discovery, it is of importance to dissect the molecular mechanisms involved in transposition and its regulation. In this work we investigated questions related to substrate recognition, specificity and complex formation in SB transposition. We carried out a functional analysis of the DNA-binding domain of the transposase and DNA sequence motifs within the inverted repeats of SB transposons. We found that the two HTH subdomains and the GRPR-like motif are all contributing to DNA binding, with the N-terminal PAI subdomain and a 10-bp binding site core sequence having a dominant role in this process. Domain swapping experiments indicate that primary DNA binding is not sufficient to determine specificity of the transposition reaction. The PAI subdomain also binds to a transpositional enhancer-like sequence within the left inverted repeat of SB and mediates protein-protein interactions with other transposase subunits. Our findings suggest that the transpositional enhancer and the PAI subdomain of the transposase are stabilizing complexes formed by a transposase tetramer bound at the IR/DR. Enhanced stability of synaptic complexes is one possible explanation for the efficient transposition of *Sleeping Beauty* in vertebrate cells.

#### EXPERIMENTAL PROCEDURES

**Plasmids**—The pT/neo donor construct (see 1 (construct 1) in Fig. 2A) has been described previously (12). Constructs 2 and 3 were made by subcloning the left or the right inverted repeat sequences, respectively, into derivatives of pT/neo lacking the right or left IRs. Construct 4 was made by PCR amplification of pT/neo with primers 5'-gactatcgattatttctacttataattcactgtatc-3' and 5'-gactatcgattttggaaaatgactgtgtcatg-3', *ClaI* digestion (underlined in the primer sequences) and recircularization. Construct 5 was made with primers 5'-gtttagatctagattatttctacttataattcactgtatc-3' and 5'-atctagatctaaacaattgttggaaaatgactgtgtc-3', *BglII* digestion and recircularization. Construct 7 was made with primers 5'-aatttacttataattcactgtatcac-3' and 5'-aatgtgatgtaaacattgttggaaaatgac-3' and recircularization. Construct 8 was made with primers 5'-gtaaacaaattttctgacatttcacattctta-3' and 5'-agacagattattcagctttatttcttcac-3' and recircularization. Construct 9 was made by PCR amplification of the left IR of pT/neo with primers 5'-gtcattactagtcatt-aaaactgtttttcaaacctccac-3' and 5'-cattactagtcacttagtataaactctgaccac-3', digestion with *SpeI*, and cloning into a derivative of pT/neo containing a *SpeI* site downstream of the internal transposase-binding site in the left IR. Construct 10 was made by amplifying the transposon sequences from construct 2 with primer 5'-tacagttgaagtggaaattacacacttaggttaattgtttacagacagattatttcac-3', and cloning the PCR product into the *SmaI* site of pUC19. The plasmid containing the HDR sequence was made by cloning a double-stranded oligo into the *SmaI* site of pUC19.

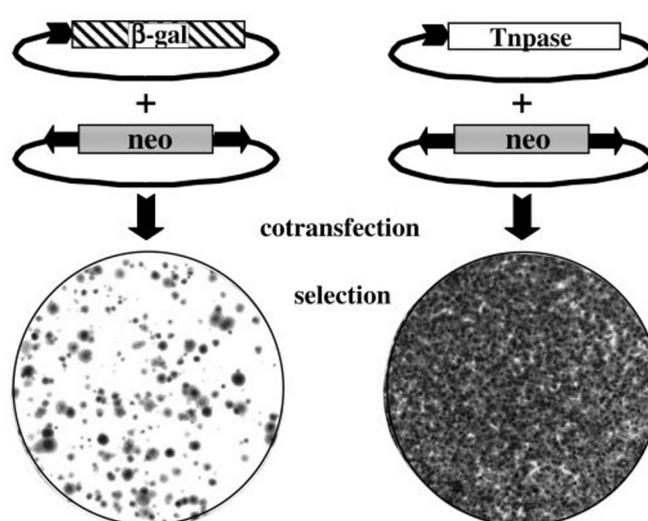
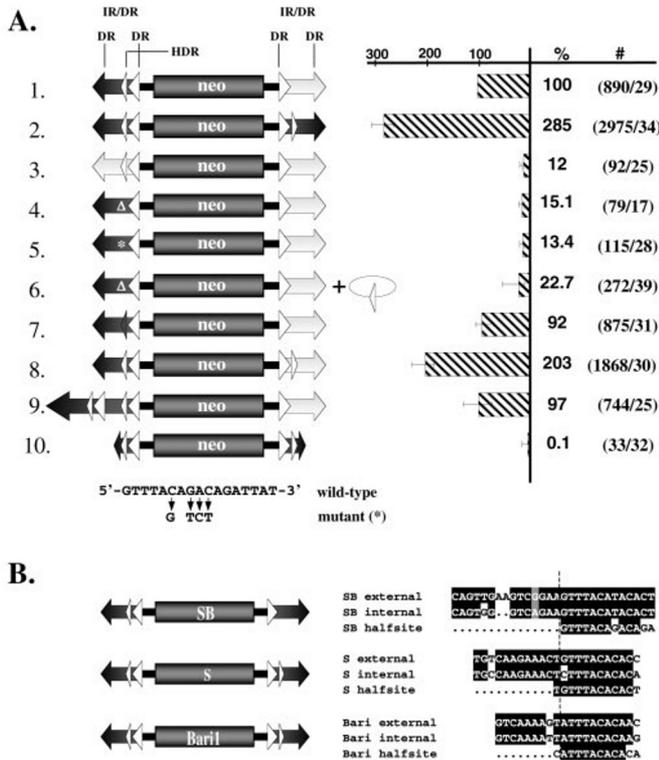


FIG. 1. **In vivo transposition assay.** Cultured cells are cotransfected with a transposon donor plasmid carrying a selectable *neo* gene together with a transposase-expressing helper plasmid. In control transfections, a plasmid expressing  $\beta$ -galactosidase is cotransfected. Cells are placed under G-418 selection, and resistant colonies are counted. Chromosomal integration of the selectable marker gene in the absence of transposase is the background of the assay. When transposase is present, the number of resistant colonies increases because of transposition. The ratio of colony numbers in the presence *versus* in the absence of transposase is a measure of the efficiency of transposition.

Eukaryotic expression plasmids are derivatives of the pCMV/SB construct described previously (12). pCMV/N57 was made by PCR-amplifying pCMV/SB with primers SV40 (5'-gctccagacatgataagatacatgtgatg-3') and N57-Stop (5'-tctagaattctagcggatgacggctgctggtc-3') and recircularization. pCMV/SB-L25P was made by PCR-amplifying pCMV/SB with primers L25P/F (5'-tcatcccggagcaatttcaaacgctg-3') and L25P/R (5'-ttgctcccgggatgaaccagactgtgagg-3'), followed by *SmaI* digestion (underlined in the primer sequences) and recircularization. pCMV/SB-PD was made by PCR amplifying pCMV/SB with primers FTC-3 (12) and L25P/R and a fragment of the Tc1 transposase gene with primers 5'-tattgcccgcgatcatgtaaaaatctgttgggtgtaaaaatc-3' and 5'-aggccaccgctctctcggcggagatttcaaacg-3', and ligating the PCR products after digestion with *EagI* and *MluI*. pCMV/SB-HD was made by PCR amplifying pCMV/SB with primers 5'-aggccaccgctgacacccatagatg-gatcg-3' and 5'-tcttctgagcgtcctgtatgctgctgctg-3', and ligating the PCR products after digestion with *BspI* and *MluI*.

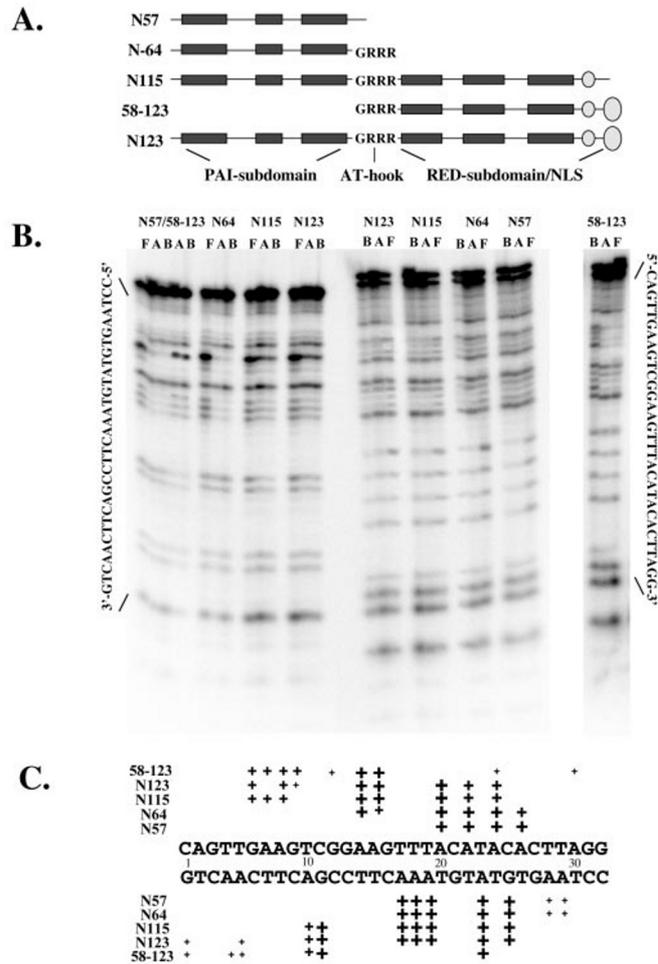
Bacterial expression vectors pET21a/N57, pET21a/N64, pET21a/N115, pET21a/N123, pET21a/N123(L25P), and pET21a/58–123 were made by PCR amplification of the corresponding fragments of the transposase gene, cutting the PCR products with *NdeI* and *XhoI* (underlined in the primer sequences), and cloning them into the respective sites in pET21a (Novagen). Primers were the following: SB-Start/*NdeI* (5'-tatacatatgggaaaatacaaacgaaatcagcc-3'); N57/*XhoI* (5'-tctactcaggcggatgacggctgctggtc-3'); N64/*XhoI* (5'-gctatactcagacagcgtctctctcagc-3'); N115/*XhoI* (5'-gctatactcaggcggccttcaggttatgctg-3'); N123/*XhoI* (5'-ggtctcagcagcagctgctctctctc-3'); 58–123-Start/*NdeI* (5'-gctatacatatg-caggaaggagacgctgctg-3'). pET21a-HD was made by PCR amplification from pCMV/SB-HD with primers SB-Start/*NdeI* and N123/*XhoI* and cloning into the respective sites of vector pET21a after digestion with *NdeI* and *XhoI*. pET21a-PD was made in a similar way after amplification from pCMV/SB-PD with primer 5'-tatacatatggtaaaaatcgt-tgggtgtaaaaatc-3' and N123/*XhoI*.

**Protein Expression, Electrophoretic Mobility Shift Assay (EMSA), and Methylation Interference Analysis**—Induction of His-tagged, recombinant protein expression was in *Escherichia coli* strain BL21(DE3) (Novagen) by the addition of 0.4 mM isopropyl-1-thio- $\beta$ -D-galactopyranoside at 0.5 optical density at 600 nm and continued for 2.5 h at 30 °C. Cells were sonicated in H-buffer (25 mM HEPES (pH 7.6), 15% glycerol, 0.25% Tween 20) containing 2 mM  $\beta$ -mercaptoethanol, 1 M NaCl, 1 $\times$  COMPLETE (Roche Molecular Biochemicals), and 20 mM imidazole (pH 8.0) was added to the soluble fraction before it was mixed with Ni<sup>2+</sup>-nitrilotriacetic acid resin (Qiagen) according to the recommendations of the manufacturer. The resin was washed with sonication buffer con-



**FIG. 2. Conservation of transposase-binding "half-sites" in different Tc1-like elements of the IR/DR group, and their effect on *Sleeping Beauty* transposition.** *A*, on the left, a schematic of various marker elements is shown. On the right, the respective transpositional activities are compared with the activity of T/neo, a neo-marked transposon that has wild-type inverted repeats (construct 1), and whose activity was chosen as reference and set as 100%. Numbers of G-418-resistant colonies in the presence versus in the absence of transposase per  $10^5$  transfected cells are given. Transposase-binding sites are represented by triangles (composite binding site, large triangles; HDR, narrow triangle). A  $\Delta$  sign in constructs 4 and 6 indicates a deletion that removes the HDR. Base pair substitutions within the HDR in construct 5 are marked with asterisks, and a comparison of such mutant sites and the wild-type sequence is shown below. *B*, on the left, a schematic of SB, Bari1, and S elements is shown with a central transposase gene (gray boxes) and flanking terminal inverted repeats (black arrows). Wide triangles represent transposase-binding sites; narrow, white triangles represent HDRs. On the right, the HDR sequences are aligned to the 3'-halves of the composite transposase-binding sites of the respective transposons.

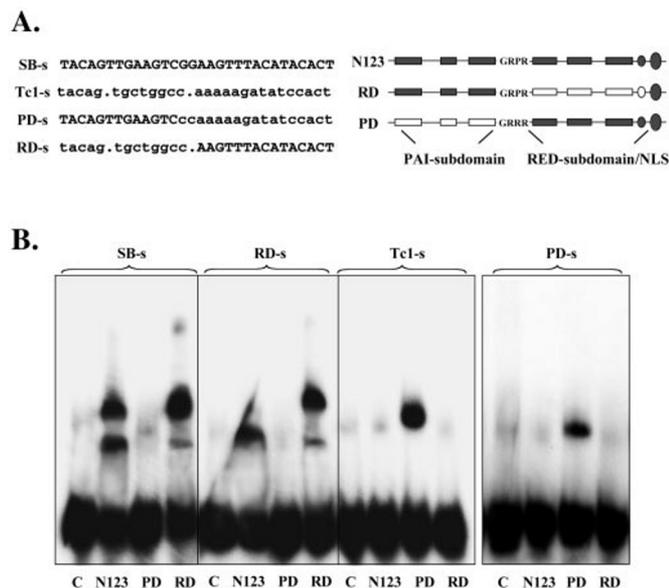
taining 30% glycerol and 50 mM imidazole; bound proteins were eluted with sonication buffer containing 300 mM imidazole and dialyzed overnight at 4 °C against a buffer containing 100 mM NaCl and 25 mM HEPES (pH 7.5). The second step in purification of N123 for molecular mass determination and sedimentation equilibrium analysis involved a Mono S column (Amersham Biosciences); a linear gradient was run from 50 to 1000 mM NaCl over 20 ml. Protein was eluted at ~500 mM NaCl. Overnight dialysis was carried out against 25 mM HEPES (pH 7.6) and 50 mM NaCl in a membrane with molecular mass cutoff of 3.5 kDa. N-terminal sequencing showed that the first terminal methionine was missing. Double-stranded oligonucleotides were made by mixing in equimolar ratios two single-stranded oligos, boiling, and annealing by slow cooling. The double-stranded oligo for the sedimentation equilibrium experiments was purified on a Mono Q column (Amersham Biosciences) in 10 mM Tris (pH 7.6) with a NaCl gradient. The peak fractions were pooled and rebuffered in 5 mM HEPES (pH 7.6) and 50 mM NaCl. For EMSA, either double-stranded oligos or an ~350-bp *EcoRI* fragment comprising the left inverted repeat of the element (Fig. 6C) were end-labeled using [ $\alpha$ - $^{32}$ P]dATP and Klenow. Nucleoprotein complexes were formed in 20 mM HEPES (pH 7.5), 0.1 mM EDTA, 0.1 mg/ml bovine serum albumin, 150 mM NaCl, 1 mM dithiothreitol in a total volume of 10  $\mu$ l. Reactions contained 0.2 pmol of labeled oligo or 100 pg of labeled fragment, 1  $\mu$ g of poly(dI)(dC), and 0.05 pmol of peptide. After a 15-min incubation on ice, 5  $\mu$ l of loading dye containing 50% glycerol and bromphenol blue was added, and the samples loaded



**FIG. 3. Mapping of base-specific contacts between the DNA-binding domain of the transposase and its binding site.** *A*, schematic representation of histidine-tagged SB transposase derivatives. Gray boxes represent predicted  $\alpha$ -helices making up potential helix-turn-helix structures. *B*, denaturing polyacrylamide gel showing interference of methylation at purine bases with binding of the peptides shown in A. On top, F means unbound DNA, and A and B represent DNA bound by 0.5 and 0.05 pmol of peptide, respectively. *C*, summary of the methylation interference results. The two strands of a transposase-binding site are shown; position 1 is the first base pair of the transposable element. Interference with binding at specific bases is indicated above and below the sequences for the upper and for the lower strand, respectively. The relative strengths of interference are indicated with + signs of different size.

onto a 5% polyacrylamide gel. Methylation interference was done as described (22).

**Molecular Mass Determination and Complex Formation between Oligonucleotide and N123 Protein**—Molecular mass determination was performed in a XL-A type analytical ultracentrifuge (Beckman) equipped with UV absorbance optics, using a 36-bp-long double-stranded oligonucleotide containing the SB binding site (DNA 36, 5'-caactaagtgtatgtaaacctccgacttcaactgt-3') and N123 protein. Sedimentation equilibrium experiments were done using externally loaded six-channel cells with 12-mm optical path length and the capacity to handle three solvent-solution pairs of ~70  $\mu$ l of liquid. Sedimentation equilibrium was reached after 2 h of overspeed at 24,000 rpm followed by an equilibrium speed of 20,000 rpm for ~30 h at 10 °C. Depending on the loading concentration, the radial absorbance in each compartment was recorded at three different wavelengths between 240 and 290 nm using the molar absorbance coefficients. Molecular mass determinations employed the global fit of three radial distributions described by Equation 1, using the program POLYMOLE (23). In these equations  $\rho$  is the solvent density,  $\bar{v}$  is the partial specific volume,  $\omega$  is the angular velocity,  $R$  is the gas constant, and  $T$  is the absolute temperature.  $A_r$  means the radial absorbance, and  $A_{r,m}$  represents the corresponding value at meniscus position.



**FIG. 4. EMSA analysis of the DNA-binding properties of hybrids between the PAI and RED subdomains of the Sleeping Beauty and Tc1 transposases.** A, DNA sequences of the SB and Tc1 transposase-binding sites, and those of two hybrid sites are given on the left. On the right, schematics of the DNA-binding domain of the SB transposase (N123) and those of two hybrid transposases are shown. Boxes represent predicted  $\alpha$ -helices making up potential helix-turn-helix structures separated by a GRPR-like sequence motif, and followed by a bipartite NLS (two ovals). B, mobility shift analysis. On top of each panel, the oligonucleotide substrate is indicated; below the panels, C represents control reaction with no protein added, followed by reactions with the N123, PD, and RD peptides.

$$A_r = A_{rm}e^{MF} \text{ with } F = [(1 - \rho\bar{v})\omega^2(r^2 - r_m^2)]/2RT \quad (\text{Eq. 1})$$

The association constants of interacting macromolecules were calculated by fitting the sum of exponential functions (23) to the experimentally obtained radial absorbance distributions. In each experiment, three absorbance profiles obtained at three different wavelengths were analyzed to determine the equilibrium constant, the stoichiometry of the reaction, as well as the partial concentrations ( $c_i$ ) of reactants and complexes using POLYMOLE. Based on the partial concentrations and the molecular masses ( $M_i$ ) of free reactants and complexes, the weight-average molecular masses ( $M_w$ ) for the different mixtures were calculated by Equation 2.

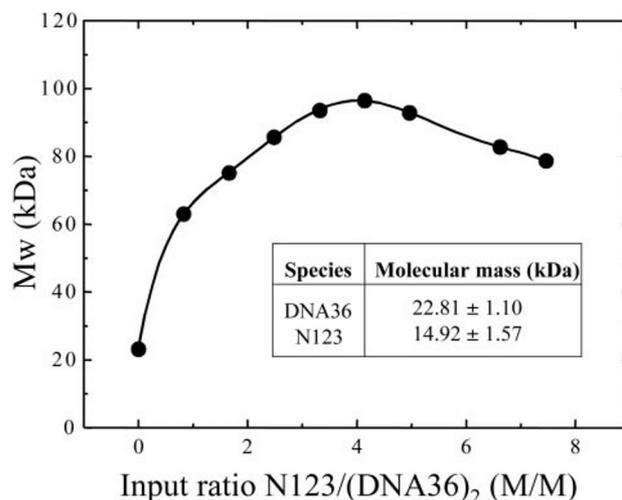
$$M_w = \sum c_i \cdot M_i / \sum c_i \quad (\text{Eq. 2})$$

**Chemical Cross-linking**—Reactions were performed in 25 mM HEPES (pH 7.5), 5 mM MgCl<sub>2</sub>, 250  $\mu$ g of bovine serum albumin, 2 mM dithiothreitol, 100 mM NaCl, and 6  $\mu$ M SB derivatives in a final volume of 20  $\mu$ l for 15 min on ice without the cross-linker. Then bis(sulfosuccinimidyl) suberate (BS<sup>3</sup>, Pierce) was added in concentrations noted in the figures, and reactions were incubated for additional 15 min at room temperature. The reaction was stopped by adding 0.5 mM glycine, and products were separated by SDS-PAGE and Western-blotted as described in Ref. 18.

**In Vivo Transposition Assay**—HeLa cells were cultured in DMEM supplemented with 10% fetal bovine serum, seeded onto six-well plates 1 day prior to transfection. Plasmid DNAs were purified through Qiagen columns. Cells were transfected with 150 ng of DNA/plasmid, using FuGENE transfection reagent (Roche Molecular Biochemicals). Selection and staining were done as previously described (12).

## RESULTS

**In Addition to the IR/DR Structure, an Extra Binding Site for the Paired-like Domain Is Needed for Transposition**—The left and right IRs of SB are not identical in sequence, raising the possibility that they are not functionally equivalent. To test this, we generated SB versions that have two left IRs or two right IRs flanking a neomycin resistance gene (*neo*). We assessed the relative transpositional efficiencies of these trans-

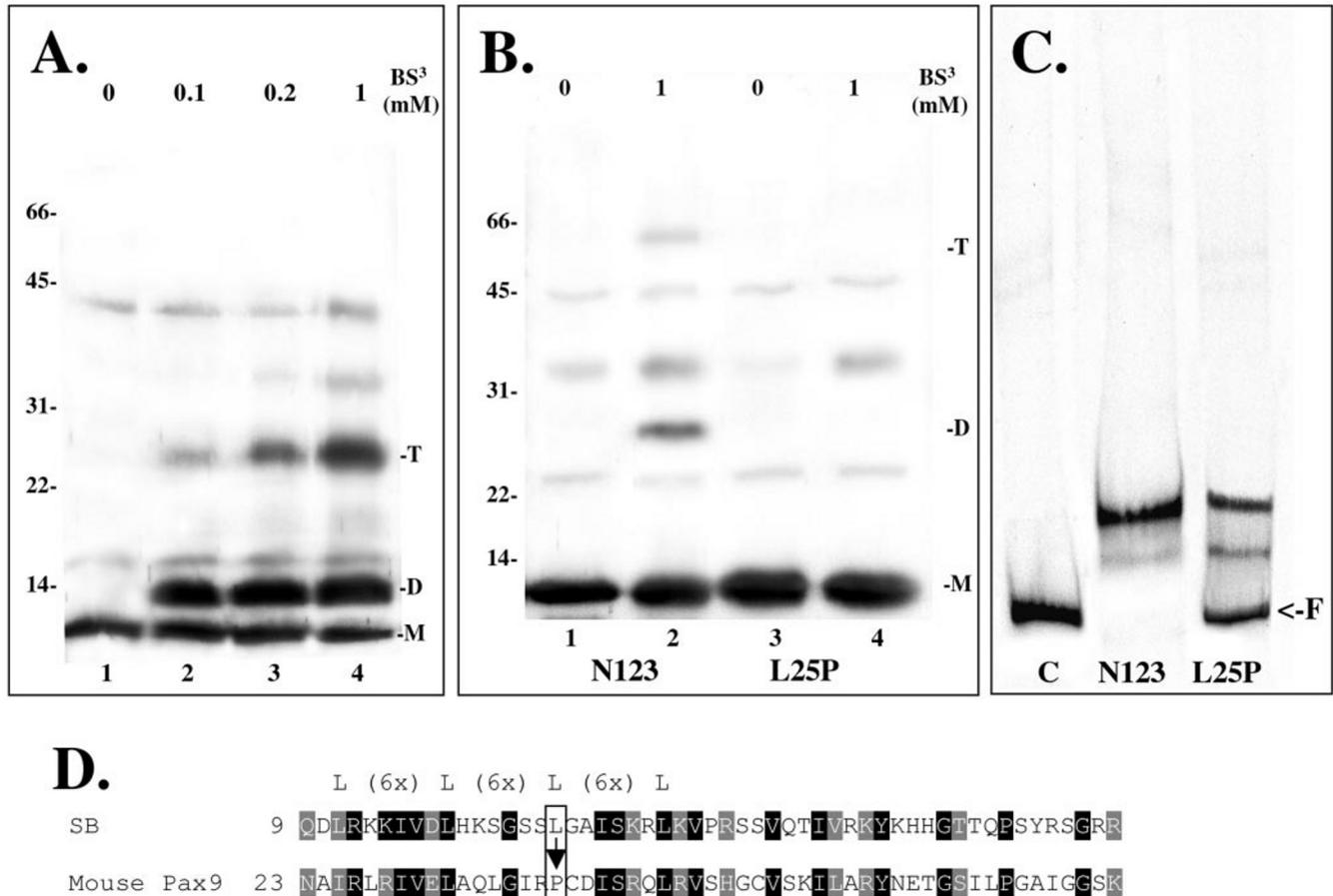


**FIG. 5. Sleeping Beauty transposase forms a tetrameric complex with DNA in solution.** Figure shows influence of N123 to oligonucleotide ratio on complex formation demonstrated by the calculated weight-average molecular masses. The loading concentration of oligonucleotide was 1.4  $\mu$ M.

posons in an *in vivo* transposition assay, which is based on cotransfection of a transposon donor plasmid and a transposase-expressing helper plasmid into cultured cells (Fig. 1) (12). Cells are then placed under G-418 selection, and numbers of resistant colonies counted. The ratio between numbers obtained in the presence and absence of transposase (Fig. 1) is the readout of the assay, and is a measure of the efficiency of transposition. Transposons flanked by two left IRs (2 (construct 2) in Fig. 2A) showed an almost 3-fold increase in transpositional efficiency compared with the wild-type element with one left and one right IR (construct 1). In contrast, transposons flanked by two right IRs (construct 3) showed severely reduced transposition: ~12% of wild-type rates. These data indicate that the two IRs of SB are functionally not equivalent, and that the left IR contains sequences that are required for high efficiency transposition.

Inspection of the left IR of the transposon revealed the presence of a 11-bp sequence that resembles the 3'-half of a transposase-binding site (DR) (Fig. 2B). Sequence motifs analogous to this "half-site" (HDR for half DR) can also be detected in two other Tc1-like transposons, the S and Bari1 elements in *Drosophila* (Fig. 2B), but not in other Tc1-like transposons. Whereas the HDR sequence is present only in the left IR of SB, these fly transposons have HDRs in both of their IRs. Because of its evolutionary conservation among transposons in phylogenetically distant species, we predicted that the HDR has some function in the transposition of these elements.

To address whether the HDR has a role in transposition, it was deleted from the IR, which decreased transposition frequency to ~15% (4 (construct 4) in Fig. 2A). Introducing point mutations into the HDR (construct 5) had a comparable effect, indicating that it is not just the spacing between the transposase-binding sites, but the sequence of the HDR that is important for transposition. Introducing a HDR on a separate molecule by cotransfecting a HDR-containing plasmid together with HDR-mutant transposons (construct 6) resulted in only a small increase of transposition frequency, indicating that the HDR sequence acts preferentially *in cis*. The HDR in SB elements does not exactly match the 3'-half of the DR (Fig. 2B). Replacing the HDR with the exact sequence of the 3'-half of the DR (construct 7) did not reveal the significance of this difference, because this engineered element transposes at approximately wild-type level. Introducing a HDR into the right IR



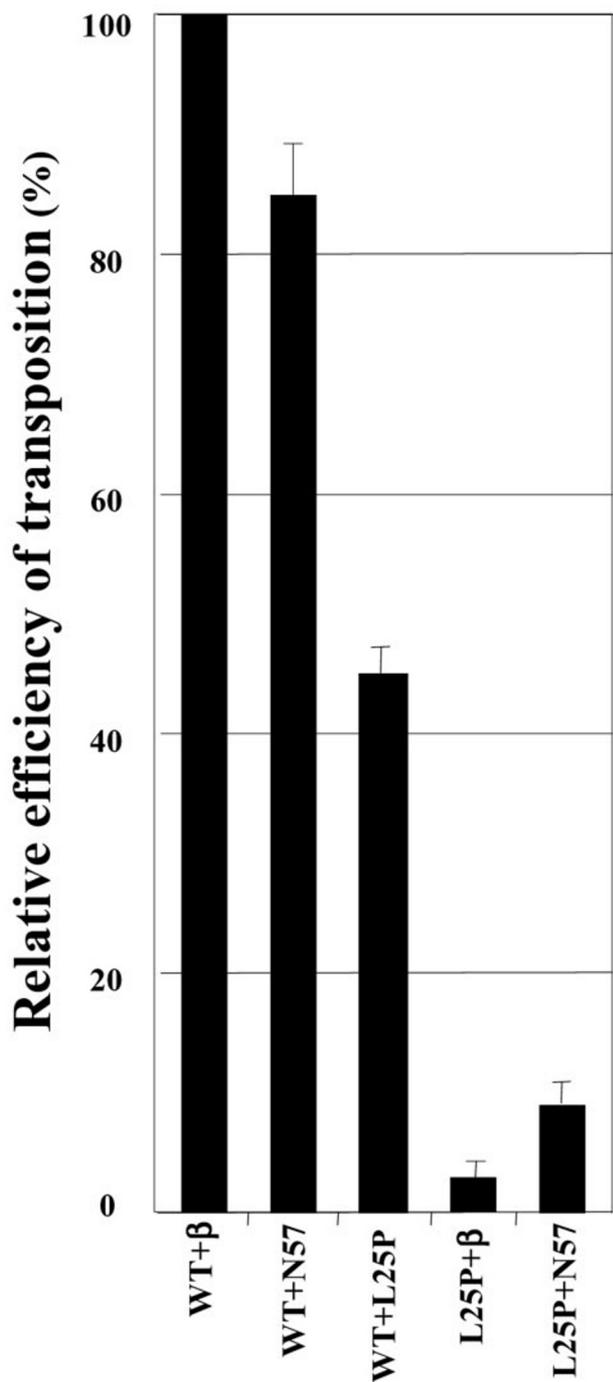
**FIG. 6. Sleeping Beauty transposase has N-terminal protein-protein interaction domains.** *A*, multimerization of N57 in the presence of 0.1, 0.2, and 1 mM BS<sup>3</sup>, 15% SDS-PAGE. *B*, multimerization of N123 in the presence of 1 mM BS<sup>3</sup>; the L25P mutation interferes with homomultimerization, 15% SDS-PAGE. *C*, EMSA analysis of the DNA-binding abilities of N123 and N123(L25P). Probe is a DNA fragment containing the left IR of SB with two transposase-binding sites. The two shifted bands represent complexes in which either one or both sites are bound. *D*, amino acid sequence alignment of the paired-like subdomain of SB transposase and the canonical paired domain. Conserved amino acid residues are *highlighted*; the leucine residue in position 25 in SB and its proline counterpart in Pax9 are *framed*. Expected molecular masses of the proteins (histidine tags inclusive) are as follows: N57: 7.45 kDa-M (monomer), 15 kDa-D (dimer); 30 kDa-T (tetramer); N123 and N123(L25P): 14.9 kDa-M, 29.8 kDa-D, and 60 kDa-T.

(construct 8), thereby doubling the signal and mimicking the structure of S and Bari1 elements, increased transposition efficiency by ~2-fold. However, doubling the complete left IR (including one DR and one HDR in construct 9) did not affect transposition rates, indicating that a mere increase in the number of transposase-binding sites and HDRs flanking the element does not influence transposition. Furthermore, moving the outer transposase-binding site into proximity (50 bp distance) of the internal site (construct 10) abolished transposition. This result indicates that the presence of four DRs and two HDRs is not sufficient for transposition, and that these sequences need to be in a proper context within the inverted repeats to allow transposition to proceed. Taken together, our results indicate that the HDR within the left IR of SB elements has an enhancer-like function, which is necessary for efficient transposition.

*The HDR Serves as a Binding Site for the PAI Subdomain of the Transposase*—Although the HDR has sequence similarity only to the 3'-region of transposase-binding sites, we hypothesized that it nevertheless can be contacted by the transposase. With the goal of mapping subdomain boundaries within the DNA-binding region, we expressed histidine-tagged, truncated derivatives of the transposase; N57 contains only the PAI subdomain (the N-terminal HTH of the paired domain), N64 contains the PAI subdomain plus the GRRR motif, "58–123" contains the RED subdomain (the C-terminal HTH of the paired

domain) plus the GRRR motif and the NLS, N115 contains the two predicted HTH motifs of the complete bipartite DNA-binding domain, and N123 contains the complete DNA-binding domain and the NLS (Fig. 3A).

Specific base pair contacts with the transposase DNA-binding domain were examined using methylation interference analysis of DNA-protein interactions (Fig. 3B). Either of the two strands of oligonucleotides containing the full transposase-binding site (Fig. 3, B and C) were radioactively labeled, treated with dimethyl sulfate, and incubated with the transposase polypeptides at two different concentrations. Free and protein-bound DNA fractions were isolated, cleaved at methylated positions by alkali boiling, and run on denaturing polyacrylamide gels. Methylation of purine bases on both strands of the DNA interfered with binding of N57 to the 3'-portion of the transposase-binding site, between base positions 17–26 (Fig. 3C). Therefore, the PAI subdomain has the potential to also bind to the HDR motif. Interference with binding of N64, which contains the GRRR motif in addition to the PAI subdomain, extended to the two adenine residues 5' to the N57 binding region. Binding at these sequences is compatible with the GRRR motif being an AT-hook, contacting the DNA in the minor groove of A-T base pairs (15). As expected, N115 and N123, both containing the complete bipartite DNA-binding domain of the transposase, showed interference covering a larger region of the probe, between base positions 6 and 25, confirm-



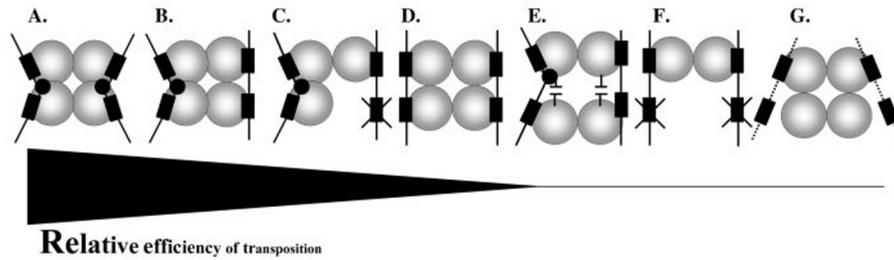
**FIG. 7. Protein-protein interactions mediated by the N-terminal region of the transposase are required for *Sleeping Beauty* transposition.** Effect of the L25P mutation on SB transposition and partial complementation of this deficiency by N57 are shown using an *in vivo* transposition assay. HeLa cells were cotransfected with pT/neo plus the expression constructs indicated at the bottom. Transposition frequency is estimated by counting G-418-resistant colonies. 100% activity was rendered to the control experiment where the transposon substrate DNA was cotransfected with a helper plasmid expressing the wild-type SB transposase (WT) plus a plasmid expressing a non-transposase related protein,  $\beta$ -galactosidase ( $\beta$ ). Each bar represents results from at least three repetition of experiments. Numbers of colonies in the presence versus in the absence of transposase per  $10^5$  transfected cells are the following: WT+ $\beta$ , 1280/40; WT+N57, 1129/42; WT+L25P, 470/35; L25P+ $\beta$ , 46/45; L25P+N57, 84/31.

ing that the RED subdomain is responsible for contacting the 5'-portion of the transposase-binding site. "58-123" alone is capable of binding to DNA, albeit with reduced specificity,

because we detected interference between positions 6 and 15 and also in the 3'-region of the binding site, at position 23. Interference at the 5'-region of the probe was generally weaker than at the 3'-region, suggesting dominant contacts between the PAI subdomain and a core binding region 5'-TTTACATACA-3', and a role of the RED subdomain in modulating sequence specificity of overall DNA binding at the ends of the transposon. In summary, the PAI subdomain has two types of binding site within the transposon IRs: the DRs, to which it binds together with the RED subdomain, and the HDR, to which it has the potential to bind alone.

*Domain Swapping Indicates That Primary DNA Binding Is Not Sufficient to Enforce Specificity of the Transposition Reaction*—The above experiments suggested that the role of the RED subdomain is to ensure specificity of transposase binding to its binding sites within the transposon. To test this prediction, we generated hybrid transposase DNA-binding domains by exchanging either the PAI or the RED subdomains of the SB and Tc1 transposases, and fused these domains at the AT-hook motifs present in both of these proteins (Fig. 4A). Protein PD contains the Tc1 PAI subdomain fused to the SB RED subdomain, whereas protein RD contains the reciprocal fusion. Along with the protein hybrids, we designed hybrid oligonucleotide substrates containing the binding regions of the respective domains of the two transposases. Oligo PD has its 5' part from the SB binding site, and its 3' part from the Tc1 binding site, whereas oligo RD represents the reciprocal combination (Fig. 4A). Interactions between the hybrid proteins and the hybrid DNA probes were analyzed using an electrophoretic mobility shift assay (Fig. 4B). Protein PD did not bind to the SB probe, but it did bind to oligo PD. Thus, the hybrid protein is indeed capable of interaction with its predicted hybrid binding site. However, protein PD also bound to the Tc1 probe, indicating that the presence of the core binding site in the DNA substrate is sufficient to allow binding of the hybrid protein. Similarly, protein RD did not bind to the Tc1 probe, but did bind to both the RD and the SB probes. Finally, in addition to its own SB substrate, N123 also bound to the RD probe, but did not bind the PD and the Tc1 probes. Apparently, the PAI subdomain of the transposase can alone determine transposase binding to a given sequence as long as it contains the binding site core, and the RED subdomain is unable to override binding if the core is in an inappropriate sequence context. Based on these observations, one could expect that an SB-derived transposase that has a PD domain combination could mobilize Tc1 transposons, and a similar transposase with an RD combination could mobilize SB transposons. We tested this possibility using the *in vivo* transposition assay described above, and found no indication of cross-mobilization by the hybrid transposases (data not shown). Therefore, specificity of the catalytic steps of transposition is probably enforced at a step subsequent to initial transposase binding.

*SB Transposase Forms a Tetrameric Complex with DNA*—We have shown previously that SB transposition is absolutely dependent on the presence of multiple transposase-binding sites within the IRs of the element (18). The requirement for the IR/DR structure indicates a tight regulatory point in transposition, possibly at the level of formation of a higher order DNA-protein complex, which is associated with the IR/DR structure. Based on the above observations, we hypothesized that multimerization of the transposase occurs at some point during the transposition reaction. Sedimentation equilibrium experiments were carried out to analyze the stoichiometry of complex formation between an oligonucleotide containing the transposase-binding site and N123. A prerequisite for such analysis is knowledge of the molecular masses of the reactants.



**FIG. 8. A proposed model of the early molecular events in *Sleeping Beauty* transposition.** Schematic representation of possible DNA-protein interactions at the inverted repeats of SB elements. Spheres represent transposase subunits, lines represent the inverted repeats of the transposon, black boxes indicate the transposase-binding sites, and the HDR is shown as a black dot. *A*, four subunits bound at the four transposase-binding sites (DRs) within the inverted repeats form a complex, which is stabilized by two HDR motifs. The extra HDR increases transposition almost 3-fold (2 (construct 2) in Fig. 2A). *B*, wild-type inverted repeats with only one HDR (1 (construct 1) in Fig. 2A). *C*, one internal DR is missing, and transpositional efficiency drops to 26% (18). *D*, the HDR is missing, and transpositional efficiency drops to 15% (4 (construct 4) in Fig. 2A); *E*, transposase is deficient in N-terminal protein-protein interactions, and transpositional efficiency drops to nondetectable levels (supported by L25P in Fig. 7). *F*, there are only two DRs in the inverted repeats, and transpositional efficiency drops to nondetectable levels (18). *G*, the DRs are not in an adequate DNA context, and transpositional efficiency drops to nondetectable levels (10 (construct 10) in Fig. 2A and results in Fig. 5).

The values obtained for the oligonucleotide and N123 protein (Fig. 5) indicate that both the oligo and the protein are in a monomeric state in solution. Mixtures consisting of 1.4  $\mu\text{M}$  oligonucleotide and variable amounts of N123 protein were centrifuged until reaching the sedimentation equilibrium, and analyzed with respect to complex formation using the program POLYMOLE (23). Although the oligonucleotide was monomeric in solution, it appeared to dimerize in the presence of a small amount of N123. The best fit of radial absorbance curves is reached assuming an oligonucleotide dimer binding four molecules of N123 (Fig. 5). This is also supported by the  $M_r$  values, which have a maximum at a 4:2 ratio of N123 to oligonucleotide, and drop at higher ratios because of the excess of free N123. Taken together, the transposase can take up a tetrameric form in solution in the presence of DNA, and the N-terminal DNA-binding region is sufficient to mediate tetramerization.

*The PAI Subdomain of SB Transposase Is Involved in Protein Multimerization, and thus Has a Dual Function*—To further map the domain(s) responsible for protein-protein interactions, N-terminal, truncated versions of SB transposase were tested using chemical cross-linking. Multimeric forms were observed using the N-terminal versions N57 (Fig. 6A) and N123 (Fig. 6B), indicating that a protein-protein interaction domain is encoded in the N-terminal 57 amino acids of SB transposase. Similarly, the Tc3 N terminus was shown to crystallize as a dimer (24). Fig. 6D shows the significant sequence similarity between the paired-like subdomain of the transposase and the corresponding region of Pax proteins (14). In contrast to the transposase, the canonical paired domain does not dimerize through physical contacts between protein molecules (25). Previously, we noted a hydrophobic heptad motif within the PAI subdomain of the transposase (14). Three of the four hydrophobic residues are also present in Pax proteins (Fig. 6D). To assess the potential role of Leu-25 in protein-protein interactions, it was changed to proline, the corresponding amino acid in this position in Pax proteins (Fig. 6D). Although a leucine to proline change can have drastic effects on protein structure, the L25P mutant retained significant DNA binding ability (Fig. 6C). However, this mutation severely inhibited the ability of the peptide to multimerize (compare lanes 2 and 4 in Fig. 6B). Taken together, the PAI subdomain of the SB transposase appears to combine the two activities of specific DNA binding and protein-protein interaction.

*Transposase Dimerization Mediated by the Paired-like Domain Is Required for Transposition*—To test the effect of the lack of N-terminal dimerization on SB transposition, we used the *in vivo* transposition assay described above. All constructs encoding mutant versions of the transposase were checked for proper expression by Western hybridizations, using an anti-SB

polyclonal antibody. The L25P mutant, which can bind to DNA (Fig. 6C) but is defective in N-terminal multimerization (Fig. 6B), was inactive in transposition (Fig. 7), consistent with a functional importance of transposase multimerization mediated by the N-terminal HTH domain. An independent confirmation of the requirement for N-terminal dimerization was sought by coexpression of the L25P mutant together with the wild-type transposase in cells, because the presence of an inactive transposase mutant can have a dominant negative effect on transposition (26). In the reference control,  $\beta$ -galactosidase was coexpressed together with the transposase (Fig. 7) to ensure that changes observed in relative transposition frequencies are not the result of transposase dilution. L25P displayed a significant negative effect on transposition ( $\sim 55\%$  repression, Fig. 7), which might indicate the disturbance of a cooperation between transposase subunits.

A plasmid expressing N57 was constructed and cotransfected into cells together with wild-type and dimerization-deficient transposases. Coexpression of N57 slightly decreased transposition by the wild-type transposase, but, more importantly, increased transposition by the L25P mutant  $\sim 2.5$ -fold (Fig. 7). Thus, an isolated functional domain can partially complement an inactive transposase, and further demonstrates the requirement for N-terminal protein-protein interactions for transposition.

## DISCUSSION

A uniform requirement among transposition reactions is the formation of a nucleoprotein complex, the synaptic complex, before the catalytic steps can take place. The molecular requirements for synaptic complex formation have been extensively studied and elucidated in prokaryotic transposition systems such as Mu (1), Tn5 (27), Tn10 (28), and IS911 (29). However, much less is known about these processes in vertebrate transposons, mainly because of the general inactivity of these elements in higher eukaryotes. The reconstructed SB transposon is a faithful representation of an ancient transposable element, which was active in fish genomes millions of years ago. Thus, this element can now be used as an experimental system to probe the molecular mechanisms and regulation of DNA transposition in vertebrates. This is especially important in light of recent reports on the development of SB as a genetic vector for insertional mutagenesis and transgenesis in vertebrate species (20, 21).

We have mapped functional subdomains within the paired-like DNA-binding domain of SB transposase. Specificity of DNA binding is predominantly determined by base-specific interactions mediated by the PAI subdomain (Fig. 3). In cooperation with the main DNA-binding domain, an AT-hook contributes to the specificity of substrate recognition (Fig. 3). The

function of the RED subdomain is not clear; although it also binds DNA (Fig. 3), it is unable to enforce specificity of DNA binding on mutant binding sites (Fig. 4). Whereas the more N-terminal motif of the bipartite NLS is included in the RED subdomain, the second part of the NLS does not appear to contribute to DNA recognition (Fig. 3).

DNA requirements for SB transposition differ in two essential points from those of the Tc1 element. First, there is an absolute demand for the IR/DR structure, *i.e.* for the presence of two transposase-binding sites within each IR, in SB transposition (18). Second, an enhancer-like sequence motif (HDR) within the left IR of SB, reminiscent of the binding site of the paired-like DNA-binding domain, is involved in transposition (Fig. 2). Both the IR/DR structure and the HDR were found in two other Tc1-like elements, suggesting structural, as well as mechanistic similarities between them.

These structural differences indicate that transposition of SB is a more complex process than that of Tc1, having at least two control points. The first involves the requirement of four transposase-binding sites (18). This suggests the binding of four molecules of the transposase to the inverted repeats, and thus the involvement of a transposase tetramer in the reaction. We mapped a protein-protein interaction function in the paired-like DNA-binding domain of the transposase (Fig. 6), and determined that this domain forms tetramers in complex with transposase-binding sites (Fig. 5). However, these tetrameric complexes contain only two DNA molecules. It is possible that under these conditions the physiologically relevant DNA-protein complexes that participate in transposition cannot be fully reconstituted. The transposase-binding sites may not have been in the proper DNA context in the form of oligonucleotides. Indeed, a transposon construct in which the outer binding sites were moved from their original positions (10 (construct 10) in Fig. 2A) is inactive in transposition. Alternatively, N123 does not contain all of the transposase domains necessary for complex assembly. This possibility is strengthened by the finding that the C terminus of the Tn5 transposase is required for synaptic complex formation (30).

The following observations suggest that the components of a second potential regulatory mechanism include the HDR motif and the N-terminal HTH domain of the transposase. 1) The HDR is important but not essential in transposition, because removal or mutation of this sequence dramatically decreases, but does not abolish transposition (Fig. 2); 2) the PAI subdomain of the transposase has the potential to specifically interact with the HDR (Figs. 3 and 4). The PAI subdomain also binds to the recombination sites (Fig. 3). Consequently, it appears to be involved in DNA-binding with two distinct specificities. The composite DRs are recognized in combination with the RED subdomain, whereas binding to the HDR is mediated by the PAI subdomain alone.

The above observations suggest that the necessary factors that are required for synaptic complex assembly of SB include the complete inverted repeats with four transposase-binding sites, the HDR motif, and tetramerization-competent transposase. We propose a model of the early molecular events that take place in SB transposition (Fig. 8). Each inverted repeat of the element is bound by two transposase molecules (one transposase per binding site), and there are both *cis* (within one IR) and *trans* (between two IRs) interactions between the transposase subunits, so that a transposase tetramer holds the two ends of the transposon in a synaptic complex (Fig. 8B). If one internal transposase-binding site is missing (Fig. 8C), transposition is reduced (18). If both internal sites are missing (Fig. 8F), or protein-protein interactions are inhibited (Fig. 8E), or transposase-binding sites are not in an adequate DNA context

(Fig. 8G), transposition is completely abolished. It seems probable that interactions between the PAI subdomain of the transposase and the HDR act via assisting the formation of a catalytic complex, or by stabilizing already formed transposase-transposon complexes through DNA-protein and protein-protein interactions (Fig. 8, A and B).

The Mu transposon can be considered a prokaryotic analogue of *Sleeping Beauty*. Mu transposition utilizes a tetramer of the transposase, and transposition is tightly controlled on the level of tetramer formation (1). One difference to SB is that complex assembly is inhibited before the tetramerization step if certain conditions, including proper orientation of the binding sites, are not fulfilled (1). In contrast, complex formation of SB can apparently proceed to tetramerization, even under inappropriate conditions (Fig. 5). Similarly to SB, the very N-terminal domain of Mu ( $\alpha$ ) is engaged in binding an enhancer-like element that is needed for synaptic complex assembly (2, 31). In contrast to Mu transposase, where the two specificities of binding to the enhancer and to the recombination sites are encoded in two distinct domains (2), the paired-like region of SB transposase combines these two functions in a single protein domain.

The IR/DR structure and the enhancer are characteristic only to a subgroup within the Tc1 transposon family. Protein-DNA interactions at these sites might contribute to the formation of a tight and stable synaptic complex, resulting in more efficient transposition. This can be one explanation why SB is more active in transposition than other members of the same family in vertebrate cells (21).

*Acknowledgments*—We thank D. Fiedler and E. Stüwe for technical assistance and members of the Transposition Group at the Max Delbrück Center for helpful comments on the manuscript. We acknowledge the help of A. Otto and E. Müller in N-terminal protein sequencing and mass spectrometry.

#### REFERENCES

- Mizuuchi, M., Baker, T. A., and Mizuuchi, K. (1992) *Cell* **70**, 303–311
- Leung, P. C., Teplow, D. B., and Harshay, R. M. (1989) *Nature* **338**, 656–658
- Mullins, M. C., Rio, D. C., and Rubin, G. M. (1989) *Genes Dev.* **3**, 729–738
- Lee, C. C., Beall, E. L., and Rio, D. C. (1998) *EMBO J.* **17**, 4166–4174
- Heichman, K. A., and Johnson, R. C. (1990) *Science* **249**, 511–517
- Czerny, T., Schaffner, G., and Busslinger, M. (1993) *Genes Dev.* **7**, 2048–2061
- Jun, S., and Desplan, C. (1996) *Development* **122**, 2639–2650
- Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wuthrich, K. (1994) *Cell* **78**, 211–223
- Pellizzari, L., Tell, G., and Damante, G. (1999) *Biochem. J.* **337**, 253–262
- Poleev, A., Okladnova, O., Musti, A. M., Schneider, S., Royer-Pokora, B., and Plachov, D. (1997) *Eur. J. Biochem.* **247**, 860–869
- Breitling, R., and Gerber, J.-K. (2000) *Dev. Genes Evol.* **210**, 644–650
- Ivics, Z., Hackett, P. B., Plasterk, R. H., and Izsvák, Z. (1997) *Cell* **91**, 501–510
- Plasterk, R. H., Izsvák, Z., and Ivics, Z. (1999) *Trends Genet.* **15**, 326–332
- Ivics, Z., Izsvák, Z., Minter, A., and Hackett, P. B. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 5008–5013
- Aravind, L., and Landsman, D. (1998) *Nucleic Acids Res.* **26**, 4413–4421
- Feng, J. A., Johnson, R. C., and Dickerson, R. E. (1994) *Science* **263**, 348–355
- Spanopoulou, E., Zaitseva, F., Wang, F. H., Santagata, S., Baltimore, D., and Panayotou, G. (1996) *Cell* **87**, 263–276
- Izsvák, Z., Ivics, Z., and Plasterk, R. H. (2000) *J. Mol. Biol.* **302**, 93–102
- Luo, G., Ivics, Z., Izsvák, Z., and Bradley, A. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 10769–10773
- Yant, S. R., Meuse, L., Chiu, W., Ivics, Z., Izsvák, Z., and Kay, M. A. (2000) *Nat. Genet.* **25**, 35–41
- Fischer, S. E. J., Wienholds, E., and Plasterk, R. H. A. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6759–6764
- Colloms, S. D., van Luenen, H. G., and Plasterk, R. H. (1994) *Nucleic Acids Res.* **22**, 5548–5554
- Behlke, J., Ristau, O., and Schonfeld, H. J. (1997) *Biochemistry* **36**, 5149–5156
- van Pouderooyen, G., Ketting, R. F., Perrakis, A., Plasterk, R. H., and Sixma, T. K. (1997) *EMBO J.* **16**, 6044–6054
- Xu, W., Rould, M. A., Jun, S., Desplan, C., and Pabo, C. O. (1995) *Cell* **80**, 639–650
- Lohe, A. R., and Hartl, D. L. (1996) *Mol. Biol. Evol.* **13**, 549–555
- Bhasin, A., Goryshin, I. Y., Steiniger-White, M., York, D., and Reznikoff, W. S. (2000) *J. Mol. Biol.* **302**, 49–63
- Sakai, J., Chalmers, R. M., and Kleckner, N. (1995) *EMBO J.* **14**, 4374–4383
- Normand, C., Duval-Valentin, G., Haren, L., and Chandler, M. (2001) *J. Mol. Biol.* **308**, 853–871
- Steiniger-White, M., and Reznikoff, W. S. (2000) *J. Biol. Chem.* **275**, 23127–23133
- Mizuuchi, M., and Mizuuchi, K. (1989) *Cell* **58**, 399–408