Transposition of a reconstructed *Harbinger* element in human cells and functional homology with two transposon-derived cellular genes

Ludivine Sinzelle*, Vladimir V. Kapitonov†, Dawid P. Grzela*, Tobias Jursch*, Jerzy Jurka†, Zsuzsanna Izsvák*‡, and Zoltán Ivics*§

*Max Delbrück Center for Molecular Medicine, 13092 Berlin, Germany; †Genetic Information Research Institute, Mountain View, CA 94043; and ‡Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, 6726 Szeged, Hungary

Edited by Susan R. Wessler, University of Georgia, Athens, GA, and approved December 29, 2007 (received for review August 17, 2007)

Ancient, inactive copies of transposable elements of the *PIF/Harbinger* superfamily have been described in vertebrates. We reconstructed components of the *Harbinger3_DR* transposon in zebrafish, including a transposase and a second, transposon-encoded protein that has a Myb-like trihelix domain. The reconstructed *Harbinger* transposon shows efficient cut-and-paste transposition in human cells and preferentially inserts into a 15-bp consensus target sequence. The Myb-like protein is required for transposition and physically interacts with the N-terminal region of the transposase via its C-terminal domain. The Myb-like protein enables transposition in part by promoting nuclear import of the transposase, by directly binding to subterminal regions of the transposon, and by recruiting the transposase to the transposon ends. We investigated the functions of two transposon-derived human proteins: HARBI1, a domesticated transposase-derived protein, and NAIF1, which contains a trihelix motif similar to that described in the Myb-like protein. Physical interaction, subcellular localization, and DNA-binding activities of HARBI1 and NAIF1 suggest strong functional homologies between the *Harbinger3_DR* system and their related, host-encoded counterparts. The *Harbinger* transposon will serve as a useful experimental system for transposon biology and for investigating the enzymatic functions of domesticated, transposon-derived cellular genes.

molecular domestication | myb domain | nuclear import | transposase | DNA binding

PIF/Harbinger is a superfamily of eukaryotic DNA transposons found in diverse genomes including plants and animals (1–6). Few *PIF/Harbinger* elements have been reported to be active. The *P instability factor* (*PIF*) and its associated miniature inverted-repeat transposable element called *mPIF* were found to actively transpose in maize (3). In rice, the *mPing* element can be mobilized upon *trans*-activation by its autonomous partner *Pong* (7, 8).

Harbinger3_DR is one of the three families of *PIF/Harbinger* transposons described in the zebrafish genome (9). The family contains five full-length elements predicted to be inactive because of mutations and ≈1,000 copies of a shorter element called *Harbinger3N_DR* (Fig. 1*A*). *Harbinger3N_DR* does not have coding capacity, but it shares most of its sequences including the terminal-inverted repeats (TIRs) with *Harbinger3_DR* (Fig. 1*A*); therefore, these elements likely used the transpositional machinery of autonomous elements for propagation. *Harbinger3_DR* contains two genes flanked by short, 12-bp TIRs and 3-bp target site duplications (TSDs) (Fig. 1*A*). The first gene encodes a transposase, whereas the second gene encodes a protein of unknown function that contains a SANT/Myb/trihelix domain, and hence is referred to as the Myb-like protein (4, 6, 9) (Fig. 1*A*). This motif is characterized by three α helices and the conservation of three bulky aromatic residues (Phe, Trp, and Trp in Fig. 1*C*) and might be involved either in a DNA-binding function similar to that observed in Myb-related transcriptional regulators (Myb-like domain) or in protein–protein interactions as described for chromatin remodeling factors (SANT-like domain) (10). Both genes encoded by *Ping* and *Pong* elements were recently found to be required for *mPing* transposition (11).

Transposons can contribute to the emergence of new genes with functions beneficial to the host via an evolutionary process referred to as "molecular domestication" (reviewed in ref. 12). More than 100 human genes have been recognized as probably derived from transposons (13, 14). The best studied example is the RAG1 gene that evolved from the *Transib* superfamily of DNA transposons (15) and that, together with RAG2, carries out V(D)J recombination, a site-directed DNA rearrangement of Ig gene segments in vertebrates (16). The primate-specific *SETMAR* gene that arose by fusion of a *mariner* transposase gene and a SET chromatin modifier domain has conserved some activities of the transposase, including binding and cleaving transposon ends (17–19).

PIF/Harbinger transposons also contributed to the evolution of cellular genes. In *Drosophila*, the *DPLG1-7* genes were recruited from at least three distinct *PIF*-like transposase sources (6). In vertebrates, the *HARBI1* gene constitutes the only known example of domesticated genes derived from a *PIF/Harbinger* transposase (Fig. 1*A*) (9). HARBI1 is conserved in all studied jawed vertebrates and is most similar to the *Harbinger3_DR* transposase with a 30–40% sequence identity. Because the putative catalytic motifs of *PIF/Harbinger* transposases (4, 9) are preserved (Fig. 1*B*), HARBI1 is expected to retain transposase-related activities.

Here, we resurrect the functional components of *Harbinger3_DR*, characterize the role(s) of the myb-like protein in the transposition process, and demonstrate functional homologies between the transposon-encoded proteins and two domesticated, transposon-derived host proteins: HARBI1 and NAIF1.

Results

The Host-Encoded HARBI1 and NAIF1 Proteins Are Closely Related to the Transposon-Encoded Transposase and Myb-Like Proteins. tBLASTn searches identified NAIF1 (nuclear apoptosis-inducing factor 1), also referred to as C9ORF90, as a protein closely related to the Myb-like protein. NAIF1 was previously characterized as a single-copy gene conserved across vertebrates (20). An alignment between the Myb-like transposon proteins and the fish, frog, bird, and mammalian orthologs of NAIF1 revealed high homology between the N-terminal region of NAIF1 (spanning residues 1–92)

Author contributions: L.S. and Z. Ivics designed research; L.S., V.V.K., D.P.G., T.J., and Z. Ivics performed research; V.V.K., D.P.G., T.J., and J.J. contributed new reagents/analytic tools; L.S. and Z. Izsvák analyzed data; and L.S. and Z. Ivics wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

§To whom correspondence should be addressed at: Max Delbrück Center for Molecular Medicine, Robert-Rössle Strasse 10, 13092 Berlin, Germany. E-mail: zivics@mdc-berlin.de.

This article contains supporting information online at www.pnas.org/cgi/content/full/0707746105/DC1.

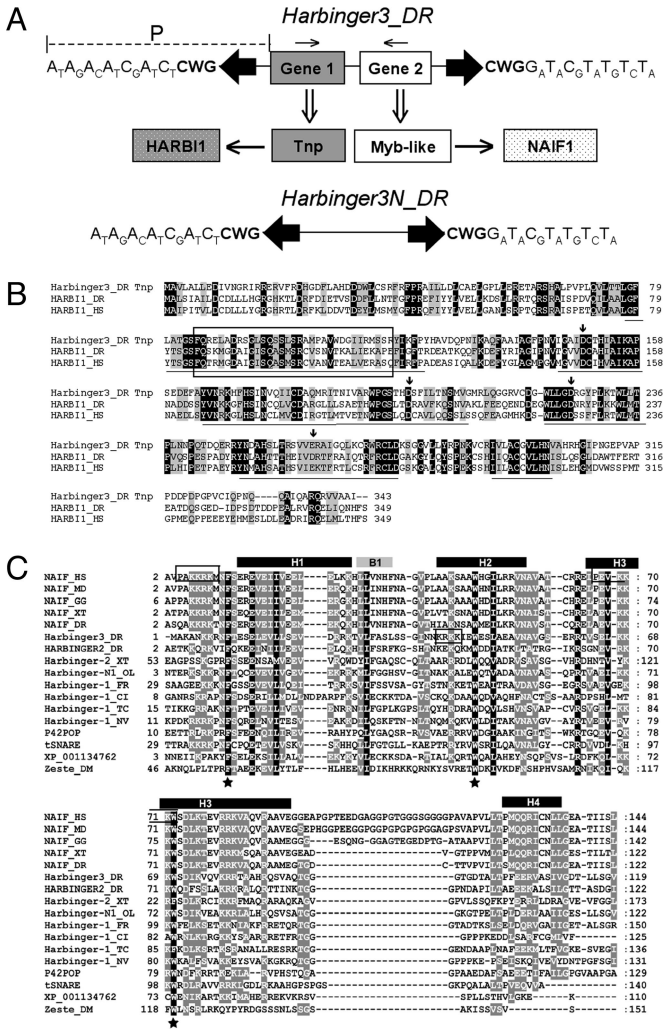

Fig. 1. Schematic representation of *Harbinger3_DR* and similarities of transposon-encoded proteins to cellular factors. (*A*) Structures of autonomous *Harbinger3_DR* and nonautonomous *Harbinger3N_DR* elements. TIRs are indicated by black arrows. The 17-bp palindromic target sequence with the alternative, most frequent nucleotides as subscript letters is indicated. Gray arrows indicate directions of transcription. ''P'' indicates the probe used in EMSA experiments. The transposase and the Myb-like protein gave rise to the domesticated vertebrate genes HARBI1 and NAIF1, respectively. (*B*) Alignment of the *Harbinger3_DR* transposase with human (HARBI1_HS) and zebrafish (HARBI1_DR) HARBI1 proteins. The six domains in *Harbinger* transposases preserved in HARBI1 proteins are underlined. The DDE triad is shown by vertical arrows, and the predicted HTH motif is boxed. (*C*) Alignment of the SANT/Myb/trihelix motif of Myb proteins and Myb-like proteins encoded by *Harbinger* transposons with the trihelix domain of NAIF1 proteins. The predicted NLSs are boxed. The three bulky residues (Phe, Trp,Trp) are indicated by stars. Predicted α helices (H) and β strands (B) are indicated.

and the N-terminal region of the Myb-like protein (spanning residues 1–90) with 36–38% of sequence identity (Fig. 1*C*). The position of the putative trihelix motif and the three bulky aromatic residues are conserved in NAIF1 (Fig. 1*C*), suggesting potential functional homology with the Myb-like protein. NAIF1 and HARBI1 are not detectable in the genomes of the jawless vertebrates *Pertomyzon marinus* (lamprey), *Ciona intestinalis* and *Ciona savignyi* (sea squirts), and *Strongylocentrotus purpuratus* (sea urchin). Therefore, it appears that both proteins emerged in a common ancestor of jawed vertebrates after its separation from jawless vertebrates some 500 million years ago. Phylogenetic analysis of NAIF1 and HARBI1 suggests that they have evolved in a similar mode, maybe because of their involvement in the same molecular pathway [supporting information (SI) Fig. 7]. Overexpression of human NAIF1 induced apoptosis and its N-terminal region was critical for its apoptosis-inducing function (20). However, the physiological role of NAIF1 remains unknown.

The Resurrected *Harbinger3_DR* Transposon Is Active in Human Cells and Transposes by a Cut-and-Paste Mechanism. Based on the consensus sequences established previously (9), transposon components projected to be sufficient for *Harbinger* transposon mobility, namely, a nonautonomous *Harbinger3N_DR* element and the coding sequences for both the transposase and the Myb-like protein were synthesized. The transposon components were used to set up a cell-based transposition assay similar to that established for *Sleeping Beauty* (SB) (21). The system consisted of a transposon donor plasmid carrying an SV40 promoter-driven neomycin-resistance gene (*neo*) inserted into the consensus *Harbinger3N_DR* element [pHarb(SV40-neo) in Fig. 2*A*] and two helper plasmids expressing the transposase and the Myb-like protein [pFV4a(Tnp) and pFV4a(Myb-like) in Fig. 2*A*]. The pHarb(SV40-neo) plasmid was transfected together with either pFV4a(Tnp) or pFV4a(Myb-like) or both in HeLa cells. Transposition, and its efficiency, was assessed from the numbers of G418-resistant colonies. Cotransfection of either the transposase- or the Myb-like protein-expressing plasmid together with the transposon-donor construct did not increase colony numbers (Fig. 2*A*). However, coexpression of both proteins produced neomycin-resistant colonies at a 2.7-fold higher rate than transfection with the donor plasmid alone, indicative of chromosomal transposition events. Because HARBI1 was found to be most closely related to the *Harbinger3_DR* transposase (9), the zebrafish ortholog of HARBI1 was also tested and was found to be deficient in transposition (Fig. 2*A*). Inactive transposase mutants might act as regulators of transposition; however, coexpression of HARBI1 together with the transposon components did not have any appreciable effect on *Harbinger* transposition (Fig. 2*A*).

We investigated the excision step of *Harbinger* transposition by a PCR-based assay (19) that generated a product consistent with transposon excision and subsequent repair of the donor plasmid from DNA samples extracted from cells transfected with the transposon, the transposase, and the Myb-like protein (Fig. 2*B*, lane 2). Sequencing of four PCR products revealed complete loss of the *Harbinger* transposon and the presence of a single CAG that restores the target site (Fig. 2*B*). This finding is consistent with excision of *mPing* in *Arabidopsis*, where the vast majority of excision sites were found to contain only the TTA target site (11).

To test integration, genomic DNA flanking the integrated transposons was isolated by inverse PCR. Transposon sequences were flanked by CAG or CTG trinucleotides, typical of transposition-mediated integration (Fig. 2*C*). The corresponding "empty" chromosomal regions showed that these integrations occurred in CWG target sites that were subsequently duplicated upon integration, consistent with the established DNA targets in zebrafish (9).

In sum, we reconstructed an active vertebrate *Harbinger* transposon. Transposition occurred in human cells by a cut-and-paste mechanism. The data strongly suggest that the Myb-like protein is an essential component for transposition of *Harbinger3_DR*.

The Reconstructed *Harbinger* Transposon Inserts Preferentially into a 15-bp Palindromic Consensus Sequence in Human Cells. *In silico* analysis of a large number of *Harbinger3_DR* and *Harbinger3N_DR* integration sites in the zebrafish genome revealed a 17-bp palindromic target site centered on the CWG triplet (9) (Fig. 1*A*). To investigate target site preferences of the reconstructed *Harbinger* element in human cells, we analyzed a total of 46 transposition events isolated from three independent transposition assays (SI Table 1). Ninety-five percent of the insertions (44 of 46) were flanked by either CAG or CTG trinucleotides (SI Fig. 8). Sequence-logo analysis revealed a 15-bp consensus sequence including the

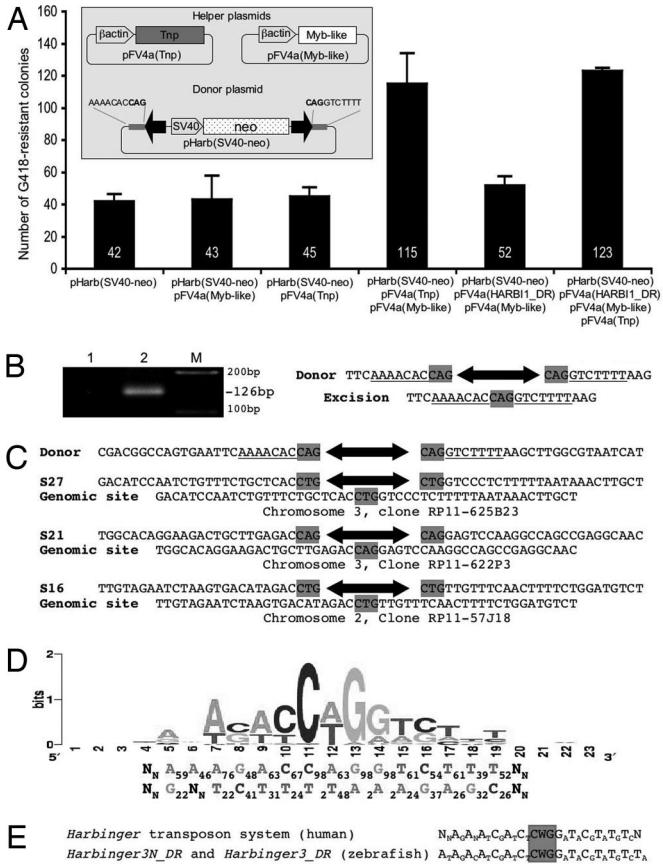

Fig. 2. *Harbinger* transposition in HeLa cells. (*A*) The numbers represent the mean values of the colony numbers in three independent assays. The error bars indicate SEM. (*Inset*) Components of the transposon system. (*B*) Agarose gel of PCR products and sequence of the donor site after transposon excision in cells cotransfected with (*i*) pFV4a and (*ii*) pFV4a(Tnp) and pFV4a(Myb-like). M, size marker. (*C*) Transposon integration sites. The TSDs are boxed in gray. (*D*) WebLogo analysis of 23-bp insertion sequences. The most frequent nucleotides at each position and the alternative, frequently appearing nucleotides are indicated with their frequencies. (*E*) Alignment of consensus target sequences derived from *de novo* integration events of the reconstructed *Harbinger* system in human cells and from *in silico* studies in zebrafish (9).

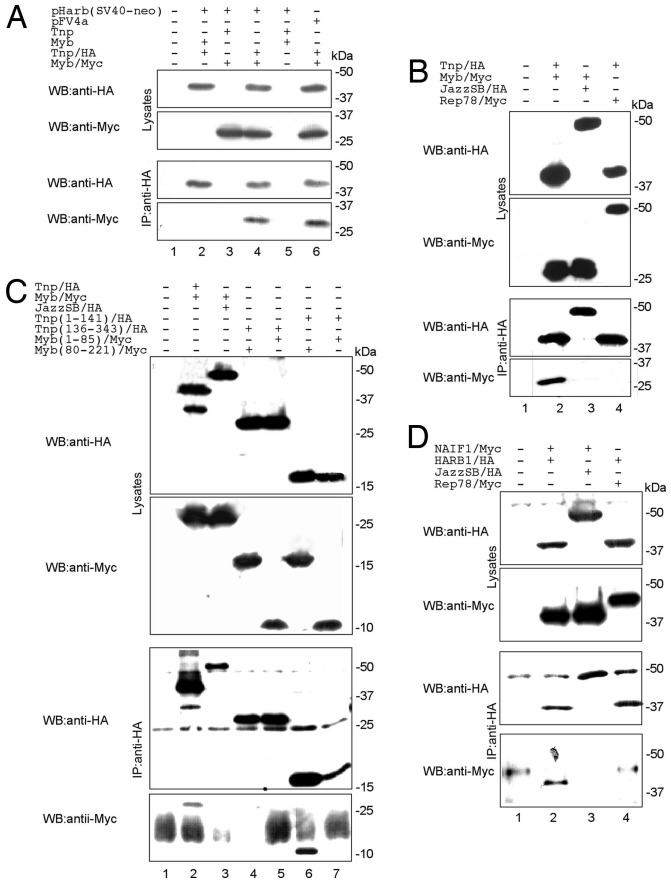

Fig. 3. Physical interactions between the transposase and the Myb-like protein and between HARBI1 and NAIF1 in human cells. (*A*) Interaction of the transposase (Tnp) with the Myb-like protein (Myb-like). Lysates and immunoprecipitates (IPs) were analyzed by Western blotting (WB) with anti-HA and anti-Myc antibodies. (*B*) Specificity of the interaction between Tnp and Myb-like protein. (*C*) Mapping of interaction domains for Tnp and Myb-like protein. (*D*) Physical interaction of HARBI1 with NAIF1.

CWG target site (Fig. 2*D*), which matches the zebrafish consensus in 15 of the 17 base pairs (positions 1 and 17 being not conserved) (Fig. 2*E*). Taking into account the alternative nucleotides at each position in the zebrafish consensus, each of the 46 integration sites retains at least 12 of the 17 base pairs (SI Fig. 9). Thus, our data demonstrate that the *Harbinger* transposon retains its target-site specificity independent of the host genome. The effect of flanking sequence composition on *Harbinger* transposition was investigated by mutating the consensus flanking region and by mutating the target site to TAA (SI Fig. 10*A*). Transposition out of these sequences was as efficient as from the consensus flanking regions, and donor sequences apparently had no effect on target-site selection (SI Fig. 10 *B–D*).

The Myb-Like Protein and NAIF1 Physically Interact with the Transposase and HARBI1. Because both the Myb-like protein and the transposase were required for the transposition process, their possible physical interaction was examined by coimmunoprecipitation. Myc-tagged Myb-like protein (Myb-like/Myc) and hemagglutinin-tagged transposase (Tnp/HA) were coexpressed in HeLa cells. The transposase was precipitated with an anti-HA antibody, and the immunoprecipitated proteins were analyzed for the presence of the Myb-like protein by immunoblotting with an anti-Myc antibody. As shown in Fig. 3*A*, the anti-HA antibody coprecipitated Myb-like/Myc (lane 4), indicating that the transposase and the Myb-like protein form a complex in cells. This interaction did not require transposon DNA (compare lanes 4 and 6 in Fig. 3*A*) and was specific, because the anti-HA antibody failed to coprecipitate either Myb-like/Myc when coexpressed with HA-tagged Jazz-SB transposase (22) or Myc-tagged Rep78 of adeno-associated virus when coexpressed with Tnp/HA (Fig. 3*B*, lanes 3 and 4). To map the regions of both the Myb-like protein and the transposase that are essential for interaction, two deletion mutants were tested for each protein by coimmunoprecipitation. Myb-like(1–85) expresses the N-terminal region, and Myb-like(80–221) lacks the N-terminal region of the Myb-like protein, whereas Tnp(1–141) contains the N-terminal 141 residues, and Tnp(136–343) is restricted to the C-terminal 209 residues of the transposase. The anti-HA antibody coimmunoprecipitated Tnp(1–141) only when coexpressed with Myb-like(80–221) (Fig. 3*C*, lane 6). These data indicate that the interaction requires domains located in the N-terminal region of the transposase and the C-terminal region of the Myb-like protein. The N-terminal region of Harbinger transposase was predicted to contain a helix–turn–helix (HTH) motif (6) (Fig. 1*B*) possibly involved in protein–protein interactions. However, a mutant transposase carrying helix-breaker proline substitutions predicted to disrupt the HTH (SI Fig. 11*A*) was fully proficient in interaction with the Myb-like protein *in vivo* (SI Fig. 11*B*), arguing against the

CELL BIOLOGY

involvement of the HTH in mediating interactions between Tnp and Myb-like.

As a first step toward a functional analysis of HARBI1, we used coimmunoprecipitation to assess its possible interaction with NAIF1 (Fig. 3*D*). Analysis of immunoprecipitates revealed efficient coprecipitation of Myc-tagged NAIF1 with HA-tagged HARBI1 (Fig. 3*D*, lane 2). No immunoprecipitation was detected for cells coexpressing either NAIF1/Myc and HA-tagged Jazz-SB (lane 3) or Myc-tagged Rep78 and HARBI1/HA (lane 4), showing specificity of the interaction.

These data provide evidence for a transposase/Myb-like protein interaction and suggest that such interaction plays a role in transposition of *Harbinger3_DR*. Similarly, HARBI1 interacts with NAIF1, suggesting functional parallels to the transposon components.

The Myb-Like Protein and NAIF1 Promote Nuclear Import of the Transposase and HARBI1. Having found that the Myb-like protein interacts with the transposase, we examined the subcellular localization of both proteins. Red fluorescent protein-tagged Myb-like protein localized to the nuclei of transiently transfected HeLa cells [confirming the prediction based on the presence of a putative nuclear localization signal (NLS) (Fig. 1*C*)], whereas enhanced green fluorescent protein-tagged transposase was found to have cytoplasmic and nuclear distribution (SI Fig. 12).

Coimmunofluorescence was next applied to investigate potential effects of the transposase/Myb-like protein interaction on subcellular localization of both proteins. When Tnp/HA was expressed alone, it predominantly localized in the cytoplasm (Fig. 4*A Top*). In contrast, when Tnp/HA and Myb-like/Myc were coexpressed, the transposase was enriched in the nucleus, where the Myb-like protein localized (Fig. 4*A Middle*, and SI Fig. 13). Cotransfection of the Tnp/HA and Myc-tagged Rep78 (that do not interact, as shown in Fig. 3*B*) showed Rep78 localization in the nucleus and intranuclear centers as expected for Rep (23) and a predominant transposase localization in the cytoplasm, similar to that observed in cells expressing transposase alone (Fig. 4*A*, compare *Bottom* and *Top*).

We next sought evidence that the interaction between the transposase and the Myb-like protein (Fig. 3*C*) is required for protein colocalization. Myb-like(1–85)/Myc localized in the nucleus (Fig. 4*B*, row 1), whereas Myb-like(80–221)/Myc mainly showed cytoplasmic expression (Fig. 4*B*, row 2). Thus, the N terminus of the Myb-like protein is critical for its nuclear localization. Tnp/HA showed mainly cytoplasmic expression when coexpressed with either Myb-like(80–221)/Myc or Myb-like(1–85)/Myc (Fig. 4*B*, rows 3 and 4), indicating that the N-terminal region of the Myb-like protein responsible for nuclear localization is not sufficient to promote nuclear import of the transposase. However, induced nuclear import of the transposase was observed when full-length Myb-like/Myc was coexpressed with Tnp(1–141)/HA but not when coexpressed with Tnp(136–343)/HA (Fig. 4*B*, rows 5 and 6). Furthermore, the HTH mutant of the transposase efficiently localized to the nucleus when coexpressed with Myb-like/Myc (SI Fig. 11*C*), consistent with its ability to interact with Myb-like. These data demonstrate that nuclear import of the transposase requires the N-terminal region (amino acids 1–85) of the Myb-like protein responsible for nuclear import (Fig. 4*B*, compare rows 5 and 7) as well as the C-terminal region (amino acids 80–221) responsible for interaction with the transposase (Fig. 4*B*, compare rows 3 and 5). Thus, the results establish a functional link between protein colocalization and interaction between the N-terminal region of the transposase and the C-terminal region of the Myb-like protein.

Subcellular localization of HARBI1 and NAIF1 was investigated by using the same experimental approach as described above. NAIF1/Myc was found exclusively in the nucleus, consistent with previously reported data (20), with a ring-like distribution (Fig. 5*A Upper*). Similar to the *Harbinger* transposase, HARBI1/HA showed localization in both the cytoplasm and the nucleus (Fig. 5*A Lower*).

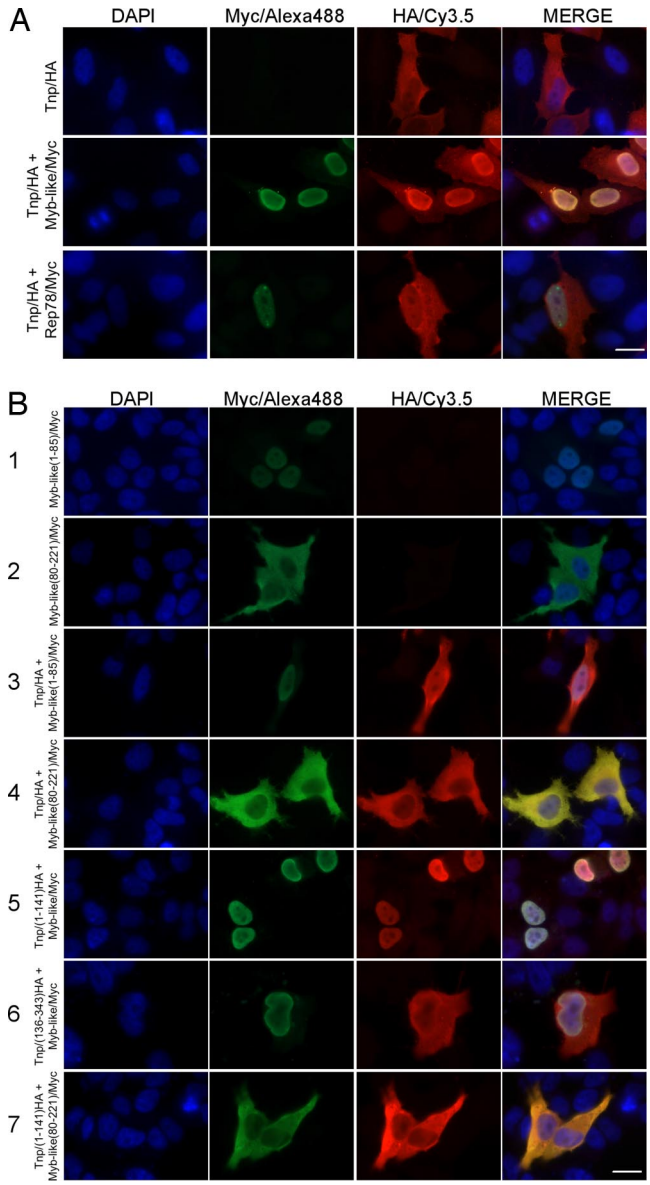

Fig. 4. Subcellular localization of the transposase and the Myb-like protein. (*A*) Colocalization assays of the full-length Tnp/HA and Myb-like/Myc proteins. (Scale bars, 20 μm.) (*B*) Representative fluorescent images of HeLa cells expressing various combinations of deletion mutants of the transposase and the Myb-like protein, as indicated on the left. (Scale bars, 20 μm.)

Both a physical interaction between HARBI1 and NAIF1 (Fig. 3*D*) and their similarities to the transposon-encoded transposase and the Myb-like protein suggest that the two proteins may colocalize in cells. Indeed, cells coexpressing HARBI1/HA and NAIF1/Myc showed a dramatic relocalization of HARBI1 to produce a nuclear pattern characteristic of NAIF1 (Fig. 5*B*). In contrast, coexpression of Myc-tagged Rep78 with HARBI1/HA did not alter the subcellular localization pattern of HARBI1 (compare Fig. 5 *A Lower* and *B Lower*). These results support the conclusion that NAIF1 promotes nuclear localization of HARBI1. In sum, both the Myb-like protein and NAIF1 are nuclear proteins that aid nuclear import of the transposase and HARBI1, respectively, an important step in biochemical reactions that involve DNA, including transposition.

The Myb-Like Protein Binds Subterminal Repeats in Transposon DNA and Recruits the Transposase to Transposon Ends. Interaction of transposase molecules with the terminal regions of the transposon

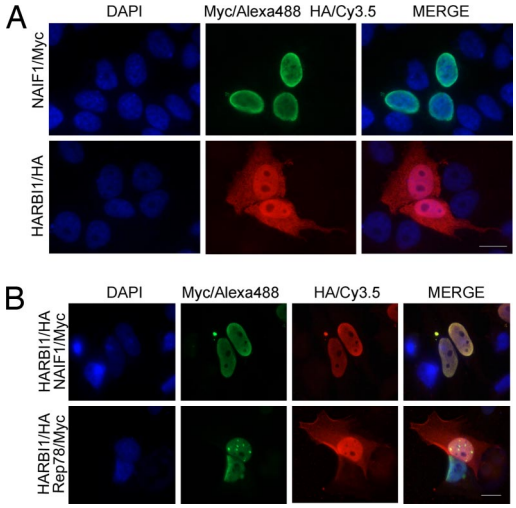

Fig. 5. Subcellular localization of HARBI1 and NAIF1. (*A*) Subcellular localization of HARBI1/HA and NAIF1/Myc. NAIF1/Myc was detected with anti-Myc and Alexa488-conjugated antibodies (green channel). HARBI1/HA was detected with anti-HA and cyanine 3.5- conjugated antibodies (red channel). (Scale bars, 20 μm.) (*B*) Colocalization assays of HARBI1/HA and NAIF1/Myc.

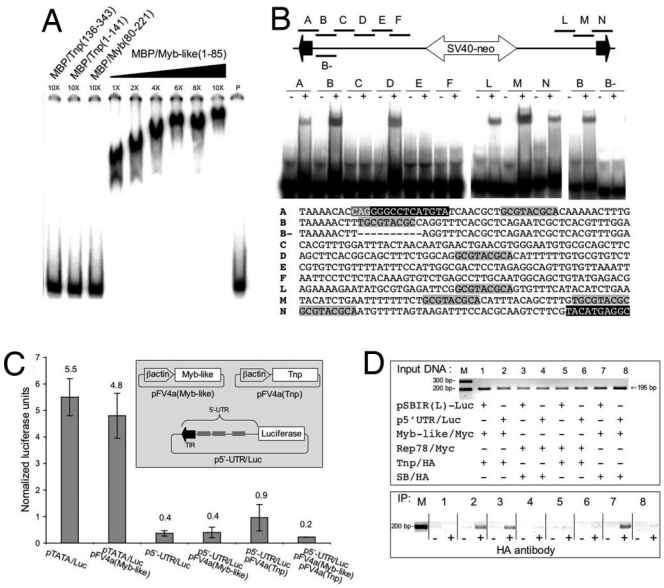

Fig. 6. DNA-binding activities of the transposase and the Myb-like protein. (*A*) EMSA of MBP/Tnp(136–343) (1× = 261 nM), MBP/Tnp(1–141) (1 × = 261 nM), MBP/Myb-like(80–221) (1 × = 291 nM), and increasing concentrations of MBP/Myb-like (1–85) (1 × = 320 nM) mixed with a 486-bp *Harbinger3_DR* transposon probe (depicted in Fig. 1*A*). (*B*) Mapping of the Myb-like protein binding sites. A schematic of the Harb(SV40-neo) element is shown with the relative positions of selected oligonucleotides used as probes in EMSA. Each reaction was performed with (+) and without (−) MBP/Myb-like (1–85) (600 nM). The sequences of the oligonucleotides are indicated with the TIRs highlighted in black (in probes A and N) and the 9-bp binding sites of the Myb-like protein are highlighted in gray. (*C*) Luciferase reporter assay. The diagram represents reporter gene expressions (indicated on the *y* axis) in HeLa cells from the plasmids indicated below, in the absence or presence of pFV4a(Myb-like) and/or pFV4a(Tnp). (*Inset*) Components of the assay. (*D*) ChIP assay. Transposase-complexed DNAs were precipitated by using anti-HA antibody. PCR was performed with total DNA (input DNA) and immunoprecipitated DNA (IP) by using primers for the luciferase coding region generating a 195-bp product. M, size marker.

is a requirement for cut-and-paste transposition. The *PIF/Harbinger* transposases and the HARBI1 proteins have been predicted to contain a single HTH motif compatible with DNA-binding capacities (Fig. 1*B*) (6). Based on the presence of a putative Myb-like trihelix domain with a highly electropositive predicted surface charge (theoretical pI = 10), the Myb-like protein is expected to have a DNA-binding activity (10).

To test the capacity of the transposase and the Myb-like protein to bind transposon DNA, EMSA was used by incubating maltose-binding protein (MBP)-tagged, purified proteins with a probe corresponding to the 5′-UTR of the *Harbinger3_DR* transposon including the left TIR and flanking consensus target sequence (Fig. 1*A*). MBP/Myb-like(1–85) produced retarded bands, whereas MBP/Myb-like(80–221) did not (Fig. 6*A*), demonstrating that the trihelix motif is necessary and sufficient to bind DNA. No shift was observed for either MBP/Tnp(1–141) or MBP/Tnp(136–343), indicating that only the Myb-like protein has the capacity to bind transposon DNA (Fig. 6*A*). Increasing concentrations of MBP/Myb-like(1–85) in the binding reaction produced more slowly migrating complexes, indicating either the presence of multiple binding sites in the probe that became saturated or multimerization of the protein upon DNA binding (Fig. 6*A*). To map the binding sites of the Myb-like protein in the transposon, an overlapping series of double-stranded oligonucleotides covering the full consensus sequence of *Harbinger3N_DR* was tested for binding. Three binding sites were identified in both ends of the transposon sharing the 9-bp palindromic sequence motif 5′-GCGTACGCA (Fig. 6*B*). This sequence motif indeed constitutes the binding site of the Myb-like protein, because an oligonucleotide lacking the site was not shifted (compare probes B and B− in Fig. 6*B*). We conclude that the Myb-like protein binds six sites in the transposon ends via its trihelix motif. Because NAIF1 is predicted to have a trihelix motif similar to that described for the Myb-like protein (Fig. 1*C*), we tested its ability to bind DNA. By using the same probe as above, NAIF1 was found to bind to DNA, but no shift was observed for HARBI1 (SI Fig. 14).

The SANT/myb/trihelix motif was found to function as a DNA-binding domain for a large number of transcription factors (10); thus, the myb-like protein may play a role in transcriptional regulation of the transposase (9). Transcriptional activation of the 5′-UTR of the transposase gene fused to a luciferase reporter was measured in an *in vivo* one-hybrid DNA-binding assay (Fig. 6*C*). The p5′-UTR/Luc and the control pTATA/Luc reporter plasmids were transfected into HeLa cells with or without pFV4a(Myb-like) and pFV4a(Tnp). The Myb-like protein apparently did not affect reporter expression (Fig. 6*C*), arguing against a role in transcriptional regulation of the transposase gene.

To investigate potential recruitment of the transposase into a complex formed by the Myb-like protein and transposon DNA, *in vivo* chromatin immunoprecipitation (ChIP) was used after co-transfection of cells with plasmid DNA containing the 5′-UTR of *Harbinger3_DR* and Tnp/HA with or without Myb-like/Myc (Fig. 6*D*). The 5′-UTR of SB together with HA-tagged SB transposase served as positive control for the assay. After cross-linking, transposase-complexed DNAs were precipitated by using anti-HA antibody coupled to agarose beads and amplified by using a diagnostic PCR. As expected, SB transposon DNA was precipitated in an SB transposase-dependent manner irrespective of coexpressed proteins (Fig. 6*D*, lanes 1, 3, 5, and 7). Similarly, precipitation of *Harbinger* transposon DNA required expression of the *Harbinger* transposase (compare lanes 2 and 8 in Fig. 6*D*) but was only seen when the transposase was coexpressed with Myb-like/Myc (Fig. 6*D*, lane 2) but not when coexpressed with Rep78/Myc (Fig. 6*D*, lane 6). PCR products were only recovered in antibody-treated samples (Fig. 6*D*). Taken together, the results suggest that the Myb-like protein contributes to *Harbinger* transposition by binding to the transposon DNA and by recruiting the transposase to the transposon ends.

CELL BIOLOGY

Discussion

We report the molecular reconstruction of functional components of the first active vertebrate *PIF*/*Harbinger* transposon and present a substantial functional analysis of its transposition. A feature displayed by the reconstructed *Harbinger* element that is unique even within the *PIF*/*Harbinger* transposon family is its highly selective target choice (Fig. 2*D*). Thus, the *Harbinger* transposon system may serve as a useful experimental tool for investigating determinants of target-site selection of mobile genetic elements, as well as for establishing technologies for site-specific transgene integration.

Consistent with studies on *mPing* mobilization (11), we found that both the transposase and the Myb-like protein are essential for transposition (Fig. 2). We provide evidence for cooperativity between the two proteins that occurs via physical interaction between domains located in the N terminus of the transposase and the C terminus of the Myb-like protein (Fig. 3) and establish that one essential role of the Myb-like protein is to promote nuclear import of the transposase (Fig. 4). Once in the nucleus, the transposase has to interact with the transposon DNA to execute recombination at the transposon ends. However, in sharp contrast to what one would expect from a transposase protein, the *Harbinger* transposase does not directly bind transposon DNA (Fig. 6). Based on the presence of a SANT/Myb/trihelix domain (Fig. 1*C*) (4, 7), the Myb-like protein has been proposed to bind to the transposon DNA, where it either acts as a transcription factor of the transposase gene (9) and/or serves as a platform for transposase binding (11). Indeed, in contrast to the transposase, the Myb-like protein binds to six sites within the *Harbinger3N_DR* transposon through its N-terminal trihelix domain (Fig. 6 *A* and *B*). However, it evidently does not act as a transcriptional regulator (Fig. 6*C*); rather, it recruits the transposase to the transposon ends (Fig. 6*D*).

Our data are compatible with a transpositional model in which the two, transposon-encoded proteins contribute distinct functions to provide a transpositionally active complex (SI Fig. 15). The Myb-like protein promotes nuclear import of the transposase and likely participates in forming a synaptic complex by directly binding to subterminal regions of the transposon and by recruiting the transposase to the transposon ends. Although quite unique among eukaryotic transposons, the requirement for multiple transposition factors is not without precedent. For example, transposition of the *En*/*Spm* element in maize was found to require two proteins, TnpA and TnpD, encoded by alternatively spliced transcripts derived from a single transcription unit (24). The differential expression of the *Harbinger* element transposase and the Myb-like protein may contribute to the regulation of transposition: an intriguing concept subject of future investigations.

Transposable element-derived genes have been identified in diverse eukaryotic kingdoms including animals, plants, and fungi (12). Conservation of these genes implies that they have been under selection for important cellular functions. The RAG proteins that catalyze V(D)J Ig gene rearrangements in jawed vertebrates provide the best studied examples for the evolution of useful functions from transposons. However, the cellular functions of the vast majority of domesticated, transposon-derived genes remain largely enigmatic. We made steps toward functional characterization of the vertebrate HARBI1 and NAIF1 genes and established functional homologies with the transposon-encoded proteins. Namely, similar to the interactions between the transposase and the Myb-like protein, NAIF1 interacts with HARBI1 (Fig. 3), promotes nuclear import of HARBI1 (Fig. 5), and acts as a DNA-binding protein (SI Fig. 8). Thus, HARBI1 is expected to function in a DNA-recombinational reaction together with NAIF1 as a cofactor. Future investigations into the mechanism of *Harbinger* transposition and its regulation should facilitate novel discoveries regarding the cellular functions of NAIF1 and HARBI1.

Materials and Methods

Cell Culture, Excision, and Transposition Assay. Transposition assays were carried out as described in ref. 21. Briefly, cells were transfected with 600 ng of transposon donor and 60 ng of each expression plasmid by using JET-PEI/-RGD (Qbiogene). Excision PCR was done as described in ref. 19, with pUC2 and pUC5 as outer-primer pair, and 19-3F and 19-3R as inner-primer pair.

Integration-Site Analysis by Inverse PCR. Genomic DNA was digested with BglII and BamHI followed by ligation with T4 DNA ligase under diluted conditions. Nested PCRs amplifying the left and right flanks of the transposons were performed by using primers ITRL1 and ITRR1 followed by ITRL2 and ITRR2. PCR products were purified with QIAquick gel extraction kit (QIAgen) and directly sequenced.

Electrophoretic Mobility Shift Assay (EMSA). A 486-bp EcoRI-SpeI fragment of pHarb(SV40-neo) containing the target sequence and the left TIR of *Harbinger3N_DR* transposons as well as double-stranded oligonucleotides end-labeled with α-^{32}P-dATP and α-^{32}P-dCTP by Klenow fill-in were used as probes. EMSA reactions were carried out as described in ref. 21.

Additional Methods. Methods for cloning all constructs used in this study, for purification of MBP fusion proteins, and for immunofluorescence, immunoprecipitation, Western blotting assays, bioinformatics as well as primer sequences are provided in *SI Text* and SI Table 2.

ACKNOWLEDGMENTS. Components of the resurrected *Harbinger* transposon system were synthesized by GENEART (Regensburg, Germany). This work was supported by a grant from the Deutsche Forschungsgemeinschaft SPP1230 ''Mechanisms of gene vector entry and persistence.''

1. Kapitonov VV, Jurka J (1999) *Genetica* 107:27–37.
2. Jurka J, Kapitonov VV (2001) *Proc Natl Acad Sci USA* 98:12315–12316.
3. Zhang X, Feschotte C, Zhang Q, Jiang N, Eggleston WB, Wessler SR (2001) *Proc Natl Acad Sci USA* 98:12572–12577.
4. Zhang X, Jiang N, Feschotte C, Wessler SR (2004) *Genetics* 166:971–986.
5. Grzebelus D, Yau YY, Simon PW (2006) *Mol Genet Genomics* 275:450–459.
6. Casola C, Lawing AM, Betran E, Feschotte C (2007) *Mol Biol Evol* 24:1872–1888.
7. Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, Wessler SR (2003) *Nature* 421:163–167.
8. Nakazaki T, Okumoto Y, Horibata A, Yamahira S, Teraishi M, Nishida H, Inoue H, Tanisaka T (2003) *Nature* 421:170–172.
9. Kapitonov VV, Jurka J (2004) *DNA Cell Biol* 23:311–324.
10. Boyer LA, Latek RR, Peterson CL (2004) *Nat Rev Mol Cell Biol* 5:158–163.
11. Yang G, Zhang F, Hancock CN, Wessler SR (2007) *Proc Natl Acad Sci USA* 104:10962–10967.
12. Volff JN (2006) *Bioessays* 28:913–922.
13. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle, M., FitzHugh W, *et al.* (2001) *Nature* 409:860–921.
14. Campillos M, Doerks T, Shah PK, Bork P (2006) *Trends Genet* 22:585–589.
15. Kapitonov VV, Jurka J (2005) *PLoS Biol* 3:e181.
16. Jones JM, Gellert M (2004) *Immunol Rev* 200:233–248.
17. Cordaux R, Udit S, Batzer MA, Feschotte C (2006) *Proc Natl Acad Sci USA* 103:8101–8106.
18. Liu D, Bischerour J, Siddique A, Buisine N, Bigot Y, Chalmers R (2007) *Mol Cell Biol* 27:1125–1132.
19. Miskey C, Papp B, Mates L, Sinzelle L, Keller H, Izsvak Z, Ivics Z (2007) *Mol Cell Biol* 27:4589–4600.
20. Lv B, Shi T, Wang X, Song Q, Zhang Y, Shen Y, Ma D, Lou Y (2006) *Int J Biochem Cell Biol* 38:671–683.
21. Ivics Z, Hackett PB, Plasterk RH, Izsvak Z (1997) *Cell* 91:501–510.
22. Ivics Z, Katzer A, Stuwe EE, Fiedler D, Knespel S, Izsvak Z (2007) *Mol Ther* 15:1137–1144.
23. Wistuba A, Kern A, Weger S, Grimm D, Kleinschmidt JA (1997) *J Virol* 71:1341–1352.
24. Frey M, Reinecke J, Grant S, Saedler H, Gierl A (1990) *EMBO J* 9:4037–4044.