Journal of Cellular Biochemistry 89:778–790 (2003)

Nuclear Localization of the Hermes Transposase Depends on Basic Amino Acid Residues at the N-Terminus of the Protein

K. Michel and P.W. Atkinson*

Department of Entomology, University of California, Riverside, California 92521

Abstract For the *Hermes* transposable element to be mobilized in its eukaryotic host, the transposase, encoded by the element, must make contact with its DNA. After synthesis in the cytoplasm, the transposase has to be actively imported into the nucleus because its size of 70.1 kDa prevents passive diffusion through the nuclear pore. Studies in vitro using transient expression of a Hermes–EGFP fusion protein in *Drosophila melanogaster* Schneider 2 cells showed the transposase was located predominantly in the nucleus. In silico sequence analysis, however, did not reveal any nuclear localization signal (NLS). To identify the sequence(s) responsible for localization of Hermes transposase in the nucleus, truncated or mutated forms of the transposase were examined for their influence on sub-cellular localization of marker proteins fused to the transposase. Using the same expression system and a GFP–GUS fusion double marker, residues 1–110 were recognized as sufficient, and residues 1–32 as necessary, for nuclear localization. Amino acid K25 greatly facilitated nuclear localization, indicating that at least this basic amino acid plays a significant role in this process. This sequence overlaps the proposed DNA binding region of the Hermes transposase and is not necessarily conserved in all members of the *hAT* transposable element family. J. Cell. Biochem. 89: 778–790, 2003.

Key words: hermes; nuclear localization; NLS; transposase; transformation

The *Hermes* transposable element, originally isolated from the genome of *Musca domestica*, is used to genetically transform a wide variety of insects [Berghammer et al., 1999; O'Brochta et al., 2000; Pinkerton et al., 2000; Michel et al., 2001]. *Hermes* is a member of the *hAT* family of transposable elements, with members found in all eukaryotic kingdoms. The element is assumed to be mobilized by a single enzyme, the Hermes transposase, which is encoded by the element itself, and movement is achieved by a DNA intermediate. The element transposes by cut-and-paste mechanism in higher dipterans (Brachycera), creating a characteristic 8 bp target site duplication upon integration. The *Hermes* element moves by the same mechanism in the soma of lower dipterans [Nematocera, Sarkar et al., 1997b]. Transposition in the germline of Nematocera, however, occurs by a yet to be identified type of replicative mechanism, leading to integration that contains additional plasmid sequences and lacks target site duplications [Jasinskiene et al., 2000; Allen et al., 2001]. What causes this switch in the element's behavior is currently not well understood.

As an eukaryotic element, the Hermes transposase is faced with the problem of being physically separated between its place of synthesis, the cytoplasm, and its place of function, the nucleus. Due to its size of 70.1 kDa, the Hermes transposase is excluded from the nucleus [Rhee et al., 2000], and needs to be actively transported across the nuclear envelope through the nuclear pore complex (NPC). A growing number of signals in form of short amino acid sequences within the cargo have been identified, that facilitate the movement of the cargo through the NPC [Goerlich, 1997; Nigg, 1997]. Growing evidence indicates that the mechanism of transport across the NPC is conserved between yeast, plants, and animals [Smith and Raikhel, 1999]. This is not surprising given that most of the

Grant sponsor: National Institutes of Health; Grant number: AI45741.

*Correspondence to: P.W. Atkinson, Department of Entomology, University of California, Riverside, California 92521. E-mail: peter.atkinson@ucr.edu

Received 14 October 2002; Accepted 11 March 2003

DOI 10.1002/jcb.10554

fundamental cytological processes such as the cell cycle, cell signaling, and translation, rely on nucleo-cytoplasmatic shuttling of macromolecules [Gama-Carvalho and Carmo-Fonseca, 2001].

Import occurs usually via the importin/Ran pathway, where protein cargo is bound at the nuclear localization signal (NLS) to an importin α/β heterodimer in the cytoplasm, transported through the NPC, and released from the importin heterodimer by binding of Ran-GTP to importin β in the nucleus [Yoneda et al., 1999]. Classic NLSs contain basic amino acid residues that are recognized and bound by importin α. The monopartite NLS consists of short single stretches of basic amino acids with a consensus sequence of $(K/R)_{4-6}$. The bipartite NLS is characterized by two short clusters of basic amino acids separated by a spacer of variable length with a consensus sequence of $(K/R)_2\,X_{10-35}\,(K/R)_3$. The prototype of the monopartite NLS is contained in the SV40 large T antigen [Kalderon et al., 1984], while *Xenopus laevi* nucleoplasmin contains a bipartite signal [Dingwall and Laskey, 1991]. To date, the bipartite NLS is the most commonly found signal for nuclear import in eukaryotic proteins. Recently, a different type of nuclear import signal has been described for the heterogeneous nuclear ribonucleoprotein (hnRNP) A1 protein. It has a 38 amino acid long sequence, termed M9 that does not contain any clusters of basic residues. The sequence is responsible for nuclear import and export of hnRNP A1, and shuttling does occur via transportin rather than importin β. The overall mechanism of import appears to be related to the importin β/Ran pathway, because Ran functions similarly in the transportin pathway [Pollard et al., 1996]. Over the last few years, the development of functional assays that allow screening of a large number of proteins for the presence of nuclear import signals [Rhee et al., 2000] has increased our knowledge of sequences facilitating nuclear import of proteins. Most of these sequences are longer and not basic in character, and in almost all cases these sequences seem quite unique to the protein or group of proteins in which they were discovered [Christophe et al., 2000].

Hermes is mobile in a broad variety of insect species [Sarkar et al., 1997a,b]. Nuclear import of the transposase, therefore most likely relies on a common signal that can be readily identified across various species. Accordingly, the first step was to use in silico sequence analysis tools to identify possible NLSs within the Hermes transposase open reading frame. PSORT II sequence analysis [Nakai and Horton, 1999] did not reveal any classical NLS present in the Hermes transposase, not surprising given the multitude of NLSs with poorly described consensus sequences that are not included in the search algorithms of the program.

The transposase of *Activator* (Ac), a transposable element from *Zea mays*, and member of the *hAT* family, contains three NLSs that have been analyzed experimentally [Boehm et al., 1995]. While two of the NLSs are classical bipartite sequences, the third signal does not conform to the consensus sequences. Comparing the Hermes and Ac transposase sequences showed that the amino acid residues responsible for nuclear import in Ac are not conserved in Hermes, and thus did not lead to the identification of possible NLSs in the Hermes transposase.

The aim of the experiments reported here was to identify sequences within the Hermes transposase that facilitate nuclear localization in vivo. *D. melanogaster* Schneider 2 cells were chosen as the in vitro experimental system, because Hermes has been shown previously to function in the soma and germ line of this species [Atkinson et al., 1993; O'Brochta et al., 1996]. The Hermes transposase was found to contain a novel NLS within the first 110 amino acids, basic at least in some residues, not necessarily conserved in other *hAT* elements, and overlapping the proposed DNA binding motif.

MATERIALS AND METHODS

Expression Plasmid Constructs

All expression constructs are based on plasmid pMT/V5-HisC (Invitrogen, Carlsbad, CA), which placed the fusion constructs under the control of the inducible metallothionein (MT) promoter from *D. melanogaster* [Bunch et al., 1988].

pDev8–EGFP and pMTV–EGFP. The full length Hermes–GFP fusion construct was cloned in two steps. First, a Hermes ORF fragment cut from pCRS Hermes (R. Hice and P.W. Atkinson, unpublished) with *Eco*RV/*Xho*I was ligated into pMT/V5-HisC digested with the same enzymes. The resulting clone Dev8, could be used to express Hermes full-length transposase with no extra aa (R. Hice and P.W. Atkinson, unpublished).

Second, EGFP-1 (Clontech, Palo Alto, CA) was cut with *Bgl*II and *Not*I, filled in and cloned into pDEV8, cut with *Eco*47III, creating pDev8-EGFP, in which the Hermes ORF coding for aa 1–568, was located upstream of EGFP. The control plasmid pMTV–EGFP, to express EGFP alone, was constructed by cutting pEGFP-1 with *Eco*RI/*Not*I and cloned into pMT/V5-H isC, digested with the same enzymes.

pMTV–EG. The EGFP–GUS fusion construct was cloned the following way: pEGFP–GUS was created by digesting pFastBac-gus (Invitrogen) with *Bsu*36I, filled in/*Not*I and cloning the resulting fragment into pEGFP-1, digested with *Bsr*GI, filled in/*Not*I. pEGFP–GUS was cut with *Kpn*I/*Not*I and ligated into pMT/V5HisC cut with the same enzymes. In the resulting construct, pMTV–EG, EGFP was located upstream of GUS.

Hermes–EG deletion constructs. All Hermes fragments were amplified by PCR in 20 cycles using Vent polymerase (New England Biolabs, Beverly, MA) and 10 ng of pKHSHH [Sarkar et al., 1997a] as template. PCR fragments were cut with *Acc*65I/*Xma*I and cloned into pMTV-EG, cut with Acc65I/XmaI. All 5′ primers contained the 5′ tail, 5′-GGGGTACCA-CCATGG-3′, with a Kozak consensus sequence, the start codon, and *Acc*65I restriction site. The 3′ primers contained a 5′ tail, 5′-TCCCCCC-GGGC-3′, providing the *Xma*I site. For primer pairs and sequences see Table I. All clones were sequenced to verify correct cloning and PCR amplification. All constructs contained an 8 aa long linker (ARDPPVAT) between the N-terminal Hermes fragment and the EGFP–GUS fusion.

pMTV Hermes1-110-EG mutation clones. All mutations were introduced by PCR in 20 cycles using Vent polymerase and 10 ng of pKHSHH as template. Primer pairs and sequences are listed in Table II. To simplify screening for clones containing the correct mutation, the primers introduced additionally restriction sites, which did not alter the amino acid composition of Hermes (*Nae*I for K25A; *Apa*LI for K59A; *Nhe*I for R70A). The amplicons were self-ligated, and the resulting products were screened by restriction enzyme digests, and subsequently sequenced to verify the introduction of the correct mutations. The regions coding for the first 110 aa of the mutated Hermes ORF forms were amplified in 20 cycles with Vent polymerase using primer pair Hermes1-110F and Hermes1-110R. PCR fragments were cut with *Acc*65I and *Xma*I and cloned into pMTV-EG, cut with *Acc*65I and *Xma*I. Primer sequences are shown in Table I.

Cell Culture, Cell Transfection, and Induction of Protein Expression

Cell maintenance. The *Drosophila* expression system from Invitrogen was used to express the Hermes–GFP and Hermes fragment–GFP–GUS fusion proteins. *Drosophila* Schneider 2 cells [Schneider, 1972] were maintained at 28°C in DES expression medium with L-glutamine (Invitrogen), supplemented with 10%

TABLE I. PCR Primers Used for Hermes-EG Deletion Constructs

Oligo name	Sequence 5′–3′
Herm1-32R	TCCCCCCGGGCGTTCCAAATAAAACTTGTTCCTTTATG
Herm33-68F	GGGGTACCACCATGGTTTTAGCGGATATACAGAAAGAAGAC
Herm33-68R	TCCCCCCGGGCTAAGTTTGATGTCTGCCTAGTTGTG
Herm69-110F	GGGGTACCACCATGGCTTGTCGTCATAAATGCTGTGC
Herm1-110R	TCCCCCCGGGCCGAAAAAGGCCGACAATC
Herm1-110F	GGGGTACCACCATGCAGAAAATGGACAATTTGG
Herm111-214F	GGGGTACCACCATGGCCGTCTCTGGATCCG
Herm111-214R	TCCCCCCGGGCTAAGGACTTTAAACCTAAAATTAGATCTCG
Herm215-333F	GGGGTACCACCATGGATTTTGAAAGATCCACAGCAG
Herm215-333R	TCCCCCCGGGCCCAGTTGTCGAGAATAGATCGAAG
Herm334-451F	GGGGTACCACCATGGAAAGCGTGATTCAAATATTAAGTGAG
Herm334-451R	TCCCCCCGGGCTTTAATTTGTGCCACTTTCTCTTG
Herm452-517F	GGGGTACCACCATGGAATTTTGCTTATCCAAAATGGAAG
Herm452-517R	TCCCCCCGGGCTGGTGGCTCACGACTATTG
Herm518-597F	GGGGTACCACCATGGTGTGTCCAAGCGATGAATTTG
Herm518-579R	TCCCCCCGGGCCAAGCTGTCGACAGTTTGTTG

Table lists the primers used for the various Hermes fusion constructs. The numbers in primer names refer to amino acid residues that were encoded by the amplicon. All forward primers contained the 5′ tail, 5′-GGGGTACCACCATGG-3′, with a Kozak consensus sequence, the start codon, and *Acc*65I restriction site. The reverse primers contained a 5′ tail, 5′-TCCCCCCGGGC-3′, providing the *Xma*I site. F and R refer to location of the primer, forward and reverse, respectively.

TABLE II. PCR Primers Used for Hermes1-110-EG Mutation Constructs

Oligo name	Sequence 5′–3′
HermK25A	GCCGGCACAAGTTTTATTTGGAACGTTTTAGC
Herm46-72R	ATGTCGCGGAGTAATTTTATATAATCC
HermK59A	GTGCTAGCATACACAACTAGGCAGACATCAAAC
Herm182-207R	ACATAAGTTTGATGTCTGCCTAGTTG
HermR70A	GCACATAAATGCTGTGCCTCTCTAAAG
Herm152-168R	TTTTTCGCATTTTCGGC

Table lists the primers used for the three Hermes mutation constructs. The numbers in the forward primers refer to amino acid position, and the letters follow the one-letter code for amino acids. Numbers in the reverse primers (indicated by R), show location of the primer in the Hermes ORF indicated by base pair position. To simplify screening for correct mutations, the primers introduced restriction sites, which did not alter the amino acid composition of Hermes (*Nae*I for K25A, *Apa*LI for K59A, *Nhe*I for R70A).

heat-inactivated fetal bovine serum (FBS) and 50 U/ml penicillin G and 50 μg/ml streptomycin.

Transient transfections. Cells were transfected using the GeneShuttle™ system (Quantum, Montreal, Canada). Briefly, 1×10^6/ml cells were seeded in 6-well plates at a total volume of 2 ml/well and grown for 24 h. Cells were then washed three times in serum free medium with 2 mM L-glutamine, and transfected using 2 μg of DNA and 8 μl of GeneShuttle in serum-free medium with 2 mM L-glutamine. The medium was replaced after 8–12 h with serum-containing medium plus antibiotics.

Expression of recombinant proteins. Forty eight hours after transfection, gene expression of the fusion proteins was induced from the MT promoter. Without changing the medium, $CuSO_4$ was added to the medium to a final concentration of 500 μM. Cells were monitored for EGFP fluorescence the next day, and, if not indicated otherwise, were fixed and stained 36–48 h post induction.

Histological Procedures

Cell fixation. To fix the cells, cells were resuspended by pipetting the medium up and down in the wells, and were then washed 1× in phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4; pH7.4). Cells were then incubated for 15 min at room temperature in 1% paraformaldehyde (Ted Pella, Redding, CA), and subsequently washed twice in PBS.

Nuclear counter stain. After the last wash, cell nuclei were immediately stained with 10 ng/ml bis-benzimide (Hoechst 33258), in PBS for 5 min at room temperature. Cells were washed three times in PBS and resuspended in 50% glycerol, 50% PBS mounting medium, and transferred onto glass slides. Slides were sealed with nail polish, and kept at −20°C until further analysis.

GUS assay. GUS stainings were performed with modifications according to Sundaresan et al. [1995] only on fixed S2 cells after Hoechst staining (see above). After the last wash from the nuclear counter stain, cells were resuspended in PBS, and mixed with equal volume of GUS staining solution (2 mM x-Gluc [3-bromo-4-chloro-3-indolyl β-D-glucuronide] cyclohexylamine salt, Sigma, St. Louis, MO); 100 mM sodiumphosphate buffer, pH 7.0; 10 mM EDTA, 0.1% Triton X-100; 100 μg/ml chloramphenicol, 2 mM K-ferricyanide, 2 mM K-ferrocyanide). Cells were incubated at 37°C for 30 min, and then immediately analyzed by light and fluorescent microscopy.

Microscopy, Analysis of Subcellular Localization

Fluorescent microscopy. Cells were examined at 600× magnification using PlanApo 60/1.40 Oil objective and Microphot-FXA microscope (Nikon, Tokyo, Japan), with a Xenon 75 W XBO Lamp as the UV source. EGFP fluorescence was observed with a FITC filter set from Chroma (HQ480/40 × EX, Q505LP BS, HQ535/50m EM), and Hoechst staining was detected using a DAPI filter set from Chroma (D360/40x EX, 400DCLP BS, D460/50m EM). For documentation, pictures of the cells were taken with a SPOT™ camera (Diagnostic Instruments, Sterling Heights, MI), and images were processed using Adobe Photoshop software.

Localization assay. Subcellular localization was assayed by classifying GFP fluorescence into three categories: N; fluorescence predominantly in the nucleus (determined by co-localization with Hoechst stain). N/C; fluorescence in the nucleus and the cytoplasm. C;

EGFP fluorescence excluded from the nucleus. Experiments were performed in triplicate for each construct, and 150 cells were recorded per replicate.

RESULTS

Hermes Transposase is Predominantly Located in the Nucleus

In vitro expression of Hermes–GFP fusion protein in *D. melanogaster* Schneider 2 cells (S2 cells [Schneider, 1972]), revealed fluorescence to be predominantly located in the nucleus (Fig. 1, left panel). In most cases, a small region within the nucleus, corresponding to the nucleolus, could be identified as devoid of fluorescence, indicating the fusion protein was absent from this area (Fig. 1, gray arrow). The upper size limit for proteins to enter the nucleus by diffusion lies between 40 and 60 kDa [Allen et al., 2000]. The calculated size of the fusion protein was approximately 92 kDa and therefore would not be expected to passively diffuse into the nucleus. EGFP alone, with a molecular weight of 26.9 kDa, can, however, enter the nucleus by diffusion, as seen in Figure 1, right panel. EGFP is not a nuclear protein and does not contain any signals for active nuclear import [Grebenok et al., 1997]. The predominant presence of fluorescence in the nucleus, when the fusion protein was expressed, was therefore a direct consequence of the Hermes transposase sequence being present *cis* to EGFP.

GFP–GUS Fusion Protein is a Viable Marker for Nuclear Import in S2 Cells

To prevent leakage of relatively small recombinant proteins into the nucleus by means of diffusion, the constructs were linked at the C-termini to β-glucuronidase (GUS), a marker from *Escherichia coli*, commonly used to test for nuclear localization in plant systems [Kunik et al., 1998]. As shown previously in plant cells, the combination of GFP and GUS was an efficient marker for nuclear import analysis

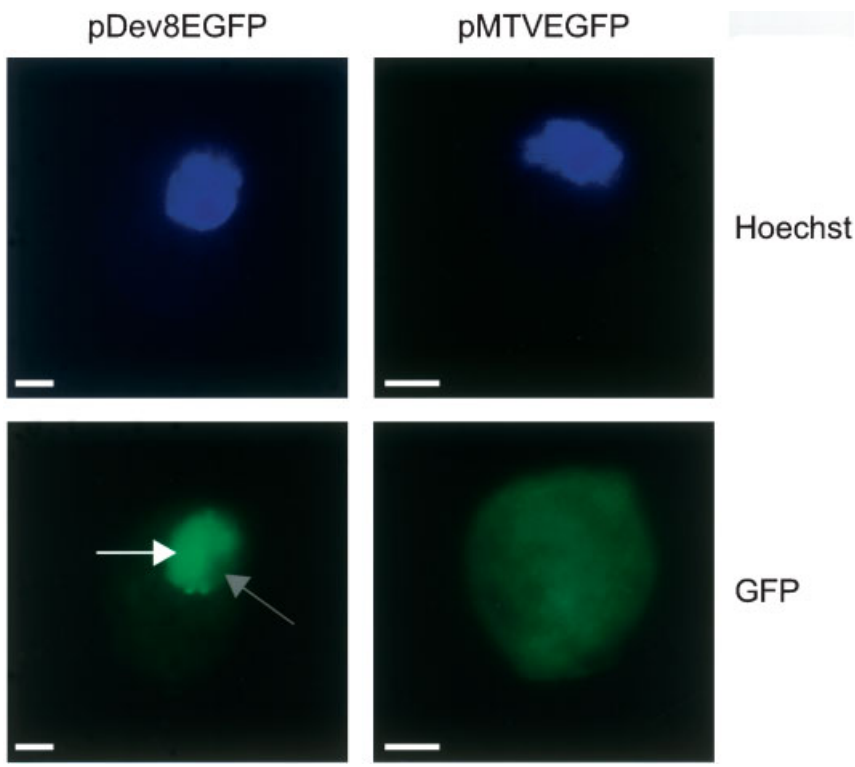


Fig. 1. Hermes transposase is located in the nucleus. Schneider 2 cells were transfected with pDEV8EGFP, or with pMTVEGFP. Expression of Hermes1-568–EGFP fusion protein (I) or EGFP (II) was induced with 500 μM $CuSO_4$ 24 h after transfection. After additional 24 h, cells were fixed with 1% paraformaldehyde, stained with Hoechst dye, and examined for fluorescence. While GFP alone was small enough to enter the nucleus by means of passive diffusion, the Hermes–GFP fusion protein (calculated 92 kDa), was imported actively into the nucleus, probably due to sequence signals present in Hermes. **Upper row**: Hoechst stain, nuclei colored in blue. **Lower row**: GFP fluorescence in green. Position of nuclei indicated in white. Position of nucleoli indicated in gray. Scale bar represents 1 μm.

[Grebenok et al., 1997]. The linkage of GFP to GUS created a fusion protein of 79.6 kDa in size, and therefore prevented the passive diffusion of GFP into the nucleus efficiently (Fig. 2, right panel). GUS was shown previously to function as a reporter gene in mammalian cells [Lorincz et al., 1996], and was shown here to be also detectable in S2 cells (Fig. 2, left and right panel). Although not determined quantitatively, it appeared that GUS was the more sensitive marker under the detection systems used, compared to GFP (see Fig. 2, left and right panel). In all experiments, GFP fluorescence was excluded from the nucleus, when fused to GUS alone, indicating expression of enzymatically active, full-length fusion protein.

The First 110 aa of the Hermes Transposase are Sufficient for Nuclear Localization

As mentioned above, active nuclear import is most commonly mediated by the importin/Ran pathway, which requires the presence of a NLS

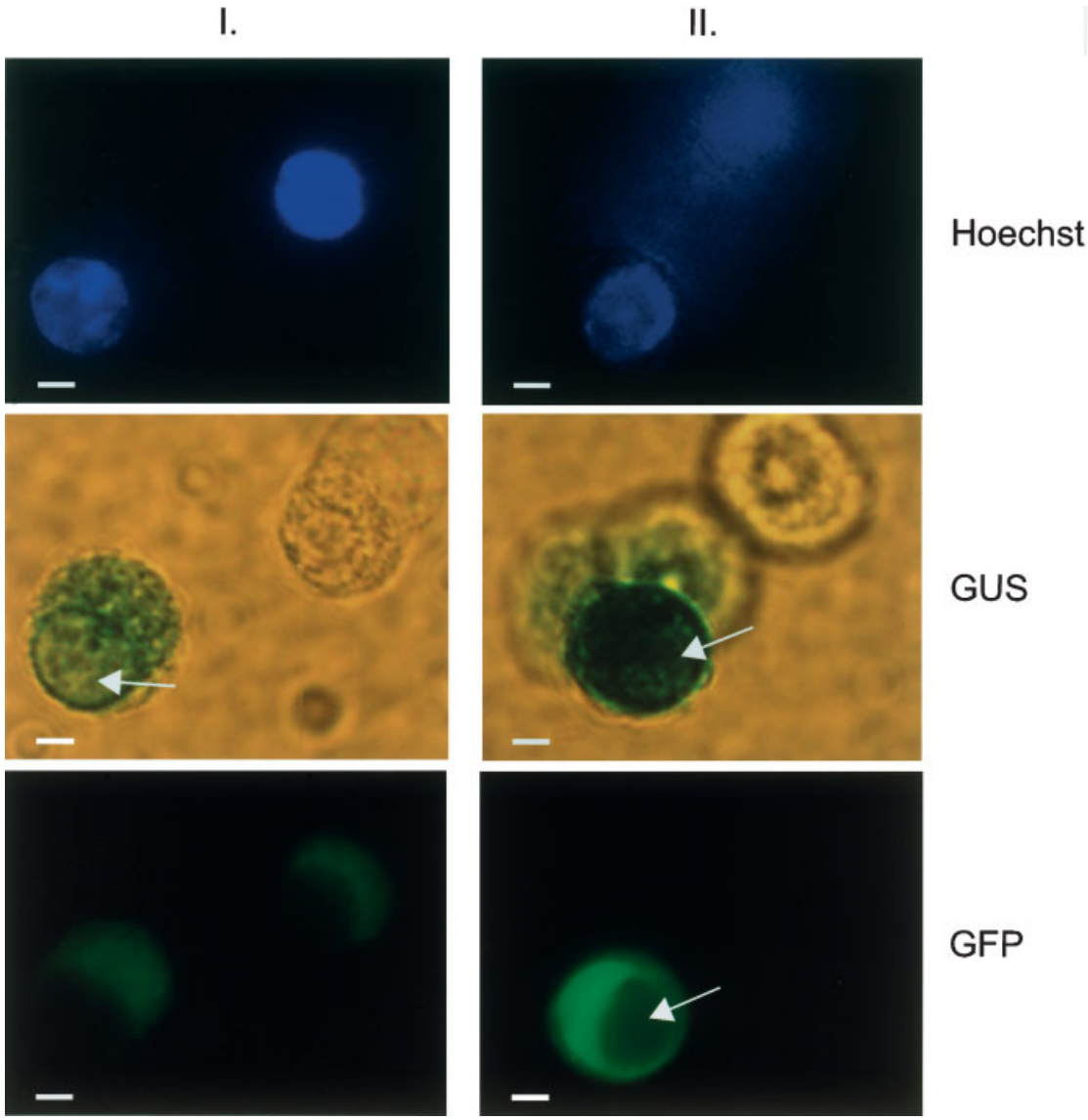


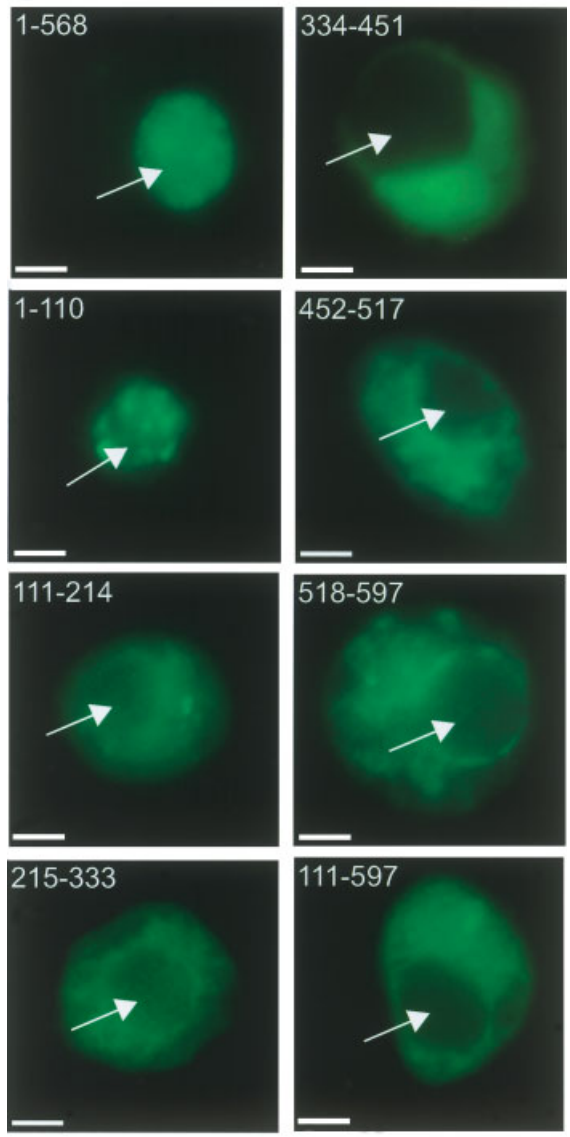
Fig. 2. EGFP–GUS fusion protein is excluded from the nucleus of S2 cells. Figure shows microscopy results of two representative S2 cells transfected transiently with pMTVEG. Due to lack of a nuclear import signal and the relatively large size of the EGFP–GUS protein (79.1 kDa), recombinant protein was excluded from the nucleus. This indicated that EGFP–GUS could be used as a marker to examine nuclear localization properties of small peptides in S2 cells. In general, GUS seemed to be the more sensitive marker, as indicated by comparison of **panels I** and **II**. While GUS activity was readily detected in panel I, GFP fluorescence was not visible. Only higher expression levels, as seen in panel II, allowed detection of GFP expression by means of fluorescence. Position of nuclei are indicated by arrows. Scale bar represents 1 μm.

in the primary structure of the protein to be imported. Sequence analysis using the PSORT II program package [Nakai and Horton, 1999] predicted localization for the Hermes transposase to be nuclear, based on Reinhardt and Hubbard [1998] neural network method for prediction of subcellular locations of proteins. The analysis did not reveal any classical NLS present in Hermes transposase, which was not too surprising, given that the sequence diversity of described NLS is increasing constantly [Christophe et al., 2000].

To identify the region responsible for nuclear localization of the Hermes transposase, small regions of about 100 aa each were expressed as GFP–GUS fusion proteins (peptides are listed in Fig. 3B). All sizes of the various Hermes fragment–EGFP–GUS fusions, exceeded the exclusion limit for passive diffusion. Figure 3 shows representative fluorescence distributions in S2 cells after transfection with, and expression of, the Hermes fusion constructs. Numbers indicated the region of amino acids of the Hermes transposase present in the fusion protein. Recombinant proteins were excluded from the nucleus, if they did not contain the first 110 aa of the Hermes transposase. To exclude the possibility that sub-cloning interrupted any NLSs, Hermes transposase lacking the first 110 aa was expressed (Hermes111-597) and was also excluded from the nucleus. Therefore, the N-terminal part of the transposase was not only necessary but also the only region in the protein facilitating nuclear localization.

To further specify the signal mediating nuclear localization of Hermes, smaller N-terminal peptides of the transposase were fused to GFP–GUS. The results of these experiments are shown in Figure 4. Because nuclear localization seemed less efficient for some of the fusion proteins and the distribution of fluorescence was not identical in all cells, sub-cellular distribution of fluorescence was quantified. The diagrams in Figure 4 summarize the results of

A



B

Hermes-peptide (aa)	Localization
1-568	N
1-110	N
111-214	C
215-333	C
334-451	C
452-517	C
518-597	C
111-597	C

Fig. 3. Subcellular localization of Hermes transposase deletions fused to EGFP–GUS. **A**: Fluorescent microscopy of S2 cells expressing various Hermes–GFP–GUS fusion constructs. The first 110 amino acids of the Hermes transposase are necessary and sufficient for translocation of recombinant protein into the nucleus. Numbers indicate amino acids of the Hermes transposase present in the fusion proteins. Arrows show position of nuclei. **B**: Summary of the subcellular localization studies shown in A. N, nuclear localization; C, fluorescence in the cytosol. Scale bar represents 1 μm.

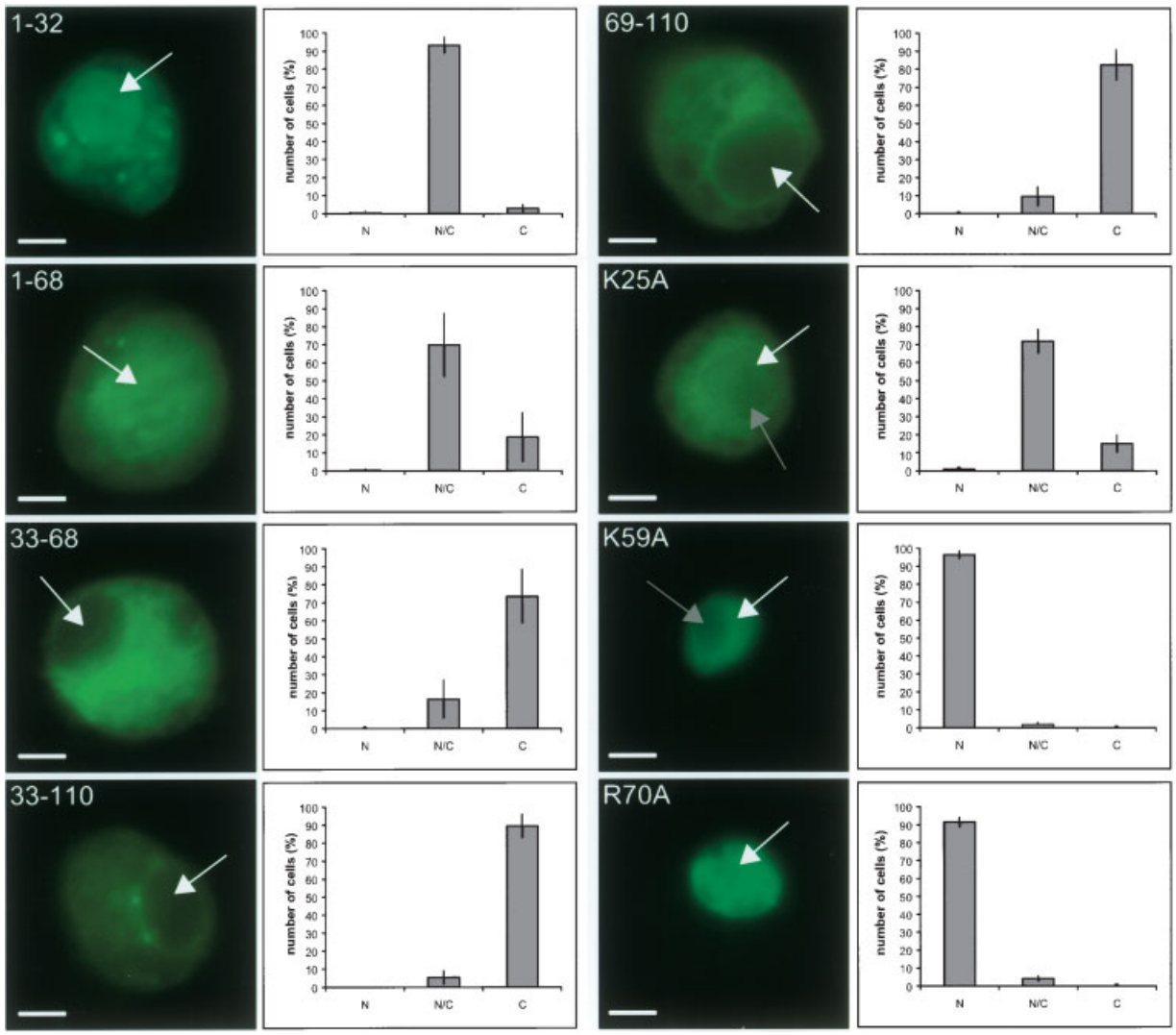


Fig. 4. Analysis of the NLS region in the Hermes transposase by deletions and amino acid substitutions. Fluorescent microscopy of S2 cells expressing partial Hermes–GFP–GUS fusion constructs. Results of the quantitative analyses are shown in the graphs to the right of each construct. Bars show the average number of cells in percent, determined in three independent experiments. Error bars indicate one standard deviation. For comparison, the distribution of phenotypes in experiments performed with fragment 1–110 were as follows: N = 98% ± 2, N/C = 2% ± 2, C = 0. N, fluorescence predominantly in the nucleus, N/C, fluorescence present in nucleus and cytoplasm, C, fluorescence excluded from nucleus. Numbers indicate amino acids of the Hermes transposase present in each of the fusion proteins; the letters follow the one-letter code for amino acids. White arrows show position of nuclei. Grey arrows point to nucleoli, where GFP fluorescence is absent. Scale bar represents 1 μm.

three independent experiments, where at least 150 cells per experiment were counted. Three categories for possible locations for fluorescence were chosen: N, fluorescence predominantly in the nucleus; N/C, fluorescence present in nucleus and cytoplasm; C, fluorescence excluded from nucleus. Of the three initial fragments tested (1–32, 33–68, and 69–110) only cells expressing fusion proteins containing the first 32 aa of the Hermes transposase showed fluorescence in the nucleus. However, the nuclear localization of the proteins seemed less dominant because a substantial amount of fluorescence was still observable in the cytosol. Nonetheless, only in about 4% of all cells examined was Hermes 1–32 GFP–GUS protein excluded from the nucleus. When expressing Hermes GFP–GUS fusion proteins containing Hermes fragments 33–68 or 69–110, 80% or more cells were devoid of nuclear fluorescence. To determine which additional sequences are necessary to enhance nuclear localization larger

fragments were fused to GFP–GUS. Expanding the Hermes peptide to region 1–68 did not enhance nuclear localization, if any, the localization in the nucleus was slightly reduced and variation was increased between experiments. The region 32–110 did, in the majority of the cells, not localize in the nucleus, just as the two smaller peptides spanning the same region were also insufficient to enable nuclear localization of the fusion proteins.

To test whether the sequence determining nuclear localization of the Hermes transposase is basic in character, three basic amino acid residues were exchanged with the neutral amino acid alanine. K59A and R70A mutations in the N-terminal 110 aa peptide did not change subcellular distribution of the fusion protein and were undistinguishable from wild-type peptide. Mutation K25A however, changed the distribution of fluorescence, and the fragment containing this mutation was less efficient to facilitate nuclear import of Hermes fragment 1–110 (see Fig. 4). While an average of 98% of cells exhibited nuclear fluorescence when expressing 1–110 peptide ($N = 98\% \pm 2$, $N/C = 2\% \pm 2$, $C = 0$; $\pm$ one standard deviation), the peptide 1–110 containing the K25A mutation was localized to about 75% in both, nucleus and cytoplasm, and was excluded from the nucleus in the rest of the S2 cells.

To determine whether the nuclear localization sequence found in Hermes is conserved in any other hAT transposase, sequence alignments of N-terminal regions of various hAT transposases were performed (Fig. 5). The NLSs previously described for the Ac transposase [Boehm et al., 1995] overlapped only partially with the region necessary for nuclear import of Hermes, and the distribution of basic residues differed between the proteins.

In summary, the minimal sequence responsible for nuclear import is region 1–110 in the Hermes transposase, with region 1–32 being required for nuclear localization. The integrity of the first 100 amino acids appears to be the determinant for efficient nuclear localization since the expression of fusion proteins containing the first 32 or 68 amino acids resulted in partial nuclear localization. The complete N-terminal sequence mutated only at position 25 (K25A) gave a similar phenotype indicating that at least this basic amino acid in this region is required for the localization of the Hermes transposase.

DISCUSSION

Green fluorescent protein from the jellyfish *Aequorea victoria* is an extremely useful marker for sub-cellular localization, because it does not require host factors for function and seems to function as part of a fusion protein extremely well. Furthermore, it can be readily detected in living cells, and also be used for time-course experiments [van Roessel and Brand, 2002]. For these reasons, GFP is considered superior compared to such markers as the *E. coli* enzymes β-glucuronidase and β-galactosidase (X-Gal). One of the disadvantages of GFP however is its small size. As seen in Figure 1, although GFP is devoid of any nuclear import signal, it readily entered the nucleus by passive diffusion. The fusion of GFP with GUS protein already proved its value for examining nuclear import properties in plants [Grebenok et al., 1997]. In the present study, an EGFP–GUS fusion also proved to be a viable marker for sub-cellular localization studies in *D. melanogaster* S2 cells. With an estimated size of 79 kDa, roughly equal to the native Hermes transposase, the marker protein was well over the size limit of nuclear entry through passive diffusion, and was only found in the cytoplasm of the insect cells (see Fig. 2). The exclusion of the fused marker proteins from the nucleus allowed examination of gradual changes between different small Hermes fragments in their ability to direct the fusion protein to the nucleus. Partial localization in the nucleus would not have been detectable with GFP as the only marker, as can be seen in the work by Otterlei et al. [1998], Becker et al. [1998]. In these studies, peptides of about 100 amino acids were fused to GFP, and detection of fluorescence in nucleus and cytosol, was interpreted as lack of any NLS in the peptide tested.

The combination of the two markers also increased flexibility in the experimental design. The use of GFP eliminated the need for staining procedures, which are usually associated with these types of experiments, and, therefore, reduced the time and materials needed for the immunohistochemistry. The higher sensitivity of GUS compared to GFP [see Fig. 2; Quaedvlieg et al., 1998] would allow the use of promoters, that could only facilitate weaker expression levels of the fusion proteins. Additionally, various fluorescent substrates have been developed for GUS, so that the detection of the fusion protein could occur at different wavelengths if required.

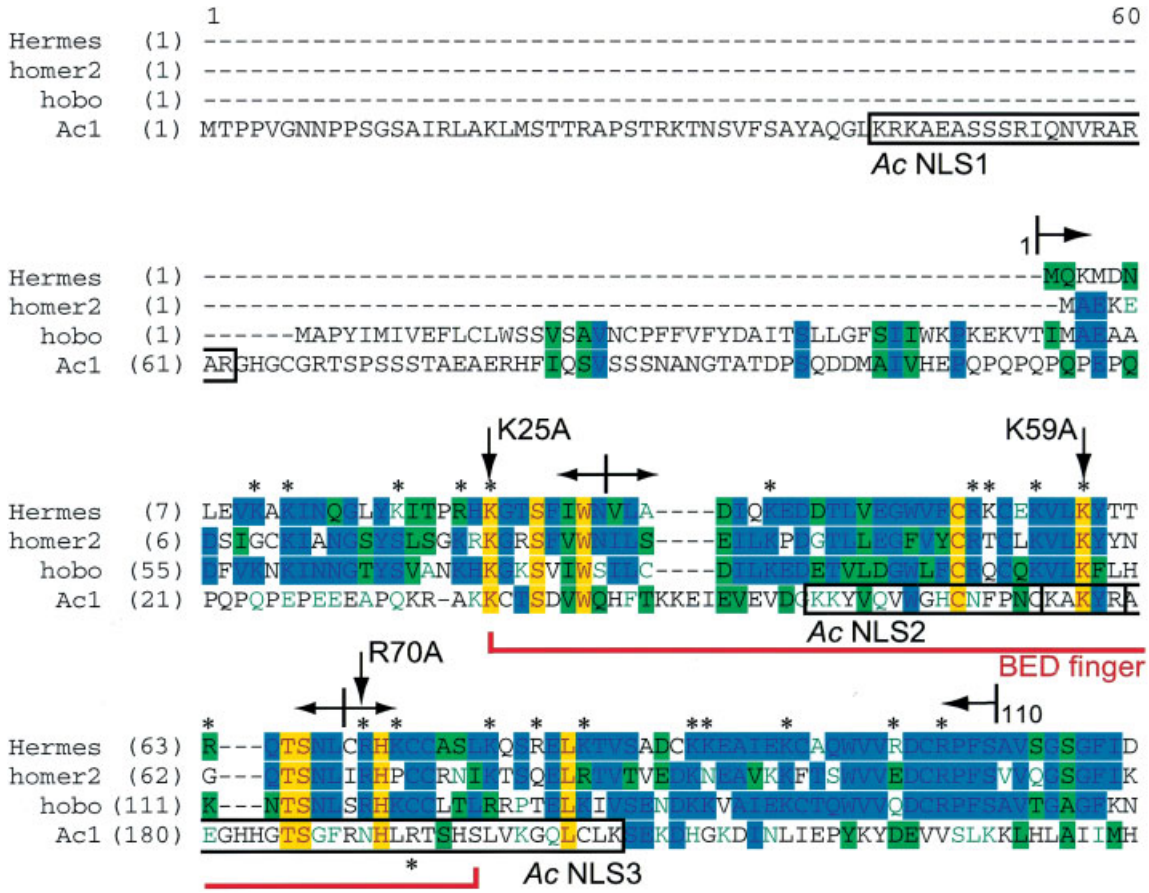


Fig. 5. Alignment of the N-terminal region of hAT transposases. The figure shows the N-terminal region of Hermes transposase responsible for nuclear import, aligned to various hAT transposase sequences. There are three NLS described for the Activator (Ac) transposase (marked by black rectangles, but homologous sequences in Hermes are either absent, as in case of NLS1 or the sequences are not conserved, as the case for NLS2 and 3. The region also contains a BED finger (red bar), the predicted DNA binding region of hAT transposases, and other proteins. Horizontal arrows indicate the various Hermes deletion constructs, and vertical arrows show positions of mutations. Asterisks mark basic amino acid residues in Hermes, of which some are conserved between hAT transposases. GenBank accession numbers are: AAB60236 (Hermes), A39652 (hobo), AAD03082 (Homer), P08770 (Ac).

Using this double marker system, we were able to show that the Hermes transposase contained a NLS within the N-terminal part of the protein. The first 110 amino acids were sufficient to direct the fusion proteins to the nucleus when expressed in S2 cells. While the region contained various clusters of basic amino acid residues, no classic NLS could be identified using in silico analysis. Several smaller fragments of the N-terminal region of the Hermes transposase (1–32, 1–68, 33–68, 33–110, and 69–110) were tested, but none of these could completely localize the protein in the nucleus. Only residues 1–32 facilitated partial nuclear localization, indicating that the Hermes transposase contains a single NLS. While the regions 33–68 and 69–110 did not lead to nuclear localization when tested alone, they are needed additionally to region 1–32 for efficient nuclear localization, either by providing further residues or by preserving an overall structure of the signal necessary for nuclear localization. The latter could also explain the slight decrease of the N/C phenotype of the Hermes 1–68 fragment compared to 1–32. It remains to be shown if the partial localization of the Hermes transposase fragments described above is a consequence of either their impaired import or impaired retention in the nucleus (for example by DNA binding).

While the first 110 amino acids of Hermes did not resemble any localization signal, such as

the M9 sequence found in hnRNP A1 and A2 proteins (17), the clustering of basic residues in this region of the transposase suggested a variation of a classic NLS. Several basic residues within the first 110 amino acids were substituted with alanine, and while K59A and R70A did not have any effect on sub-cellular distribution of Hermes 1–110, the exchange of lysine at position 25 to alanine severely impaired nuclear localization. As shown previously, partial loss of basic residues in classic NLSs lead to reduced nuclear localization [Boehm et al., 1995]. Thus, this result suggests that the NLS in Hermes, containing basic residues, may indeed be a variation of a classic NLS sufficiently varied to be cryptic to in silico analysis.

Members of the hAT family of transposable elements belong to evolutionary diverse groups of organisms. Elements were found in insects, such as *Hermes*, *hobo* from *D. melanogaster* [Blackman et al., 1989], *hermit* from *Lucilia cuprina* [Coates et al., 1996], *hopper* from *Bactrocera dorsalis* [Handler and Gomez, 1997], and *homer* from *Bactrocera tryoni* [Pinkerton et al., 1999]. Various *hAT* elements have also been described in plants [*Ac* from *Zea mays*, [Kunze et al., 1987]; and *Tam3* from *Antirrhinum majus*, [Sommer et al., 1985]] and fungi [*restless* from *Tolypocladium inflatum*, [Kempken and Kuck, 1996]; and Tfo1 from *Fusarium oxysporum* [Okuda et al., 1998]]. The transposases of these elements share sequence similarities in three distinct regions, with the C-terminal region III being the most conserved and present in all members of the group [Calvi et al., 1991]. It was suggested that these regions of sequence similarities represented distinct functional parts of the transposases, and that the associated functions were conserved among the different transposases [Warren et al., 1994].

The NLSs previously described for the Ac transposase [Boehm et al., 1995] overlapped partially with the region necessary for nuclear import of Hermes. The residues responsible for import were not conserved between the proteins, indicating convergence of the function to this region rather than common origin of the NLS. Sequence similarity was the highest between the insect hAT transposases, Hermes, hobo, and homer. Not all of basic residues in the first 100–150 amino acids are, however, conserved among these four proteins and further analyses are needed to show if the NLSs are identical between these elements. While sequence similarity was also observed to the *hermit* element, the predicted hermit transposase open reading frame contained several frame shifts, and was therefore excluded from further analysis. The exception among insect *hAT* transposable elements was *hopper*, which is more similar to *Ac1* than to any other insect *hAT* element [Handler and Gomez, 1997]. Most of the basic residues found in the N-terminal region of the other insect hAT transposases were not conserved in hopper. Alignment of the amino acid sequences of hopper and Ac1 transposases did not indicate that the functional NLSs described for Ac1 are conserved in hopper (data not shown). In summary, due to the lack of strong sequence similarity and different patterns of distributions of basic amino acid residues in the N-terminal regions of various hAT transposases, it is unlikely that the NLS are conserved at the sequence level among *hAT* elements.

While the sequences of the NLSs of Hermes and Ac were not conserved, both were located within the same region of the transposase. In addition to NLSs the N-terminus of hAT transposases harbors a DNA binding motif called BED finger [Aravind, 2000]. The BED finger, a type of zinc finger with the signature Cx_2Cx_n Hx_{3-5}[H/C], is shared by diverse classes of proteins such as chromatin-boundary-element-binding proteins and transposases. The domain was described originally by sequence alignment, and DNA binding activity was verified experimentally only in a few cases [Aravind, 2000]. Interestingly, while a bipartite DNA binding domain was found in this region for the Ac transposase, it does not contain a complete BED signature, because the second cysteine is substituted by phenylalanine.

The overlap of DNA binding domains and NLSs is commonly observed [Liu et al., 1999; Cokol et al., 2000]. Proteins with bHLH/bZIP, zinc finger, and homeobox domains are also known to contain nuclear localization domains within their DNA binding domain [LaCasse and Lefebvre, 1995]. The combination of these two functions is not too surprising taking into account that most DNA binding proteins need to be located into the nucleus. DNA binding region and NLS both need to be located on the surface of the protein to ensure interaction with DNA or factors for nuclear import, respectively. Also, classic NLSs as well as many DNA binding

domains, contain clusters of basic amino acid residues. Nonetheless, often the two functions can be separated experimentally by point mutations, and nuclear localization is observed without DNA-binding activity of the protein [Boehm et al., 1995; Li et al., 2001].

Hermes transposase is the first protein in which an overlap of the BED finger with sequences required for nuclear localization has been shown. It is currently unclear, however, if the BED finger represents the functional DNA binding domain in the Hermes transposase. The DNA binding domain is currently not available for experimental studies, because isolation of functional protein has not been possible so far. Once an in vitro system is established, it will possible to test various point mutations in the N-terminus of Hermes transposase, and dissect which residues are responsible for DNA binding and nuclear retention versus nuclear import activity.

ACKNOWLEDGMENTS

The authors thank Dr. Glenn I. Hatton for access to his fluorescence microscope, and Dr. Frederic Mercier for technical advice on fluorescent microscopy and imaging. K.M. would also like to thank Robert Hice for advice on cell culture, and Dr. Patricia S. Springer for technical advice and many helpful discussions. The work was supported by National Institutes of Health grant AI45741 to P.W.A.

REFERENCES

Allen T, Cronshaw J, Bagley S, Kiseleva E, Goldberg M. 2000. The nuclear pore complex: Mediator of translocation between nucleus and cytoplasm. J Cell Sci 113:1651–1659.

Allen ML, O'Brochta DA, Atkinson PW, Levesque CS. 2001. Stable, germ-line transformation of *Culex quinquefasciatus* (Diptera: Culicidae). J Med Entomol 38:701–710.

Aravind L. 2000. The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. Trends Biochem Sci 25:421–423.

Atkinson PW, Warren WD, O'Brochta DA. 1993. The hobo transposable element of *Drosophila* can be cross-mobilized in houseflies and excises like the Ac element of maize. Proc Natl Acad Sci USA 90:9693–9697.

Becker W, Weber Y, Wetzel K, Eirmbter K, Tejedor FJ, Joost H-G. 1998. Sequence characteristics, subcellular localization, and substrate specificity of DYRK-related kinases, a novel family of dual specificity protein kinases. J Biol Chem 273:25893–25902.

Berghammer AJ, Klingler M, Wimmer EA. 1999. A universal marker for transgenic insects. Nat (Lond) 402:370–371.

Blackman RK, Koehler MM, Grimaila R, Gelbart WM. 1989. Identification of a fully-functional hobo transposable element and its use for germ-line transformation of *Drosophila*. EMBO J 8:211–217.

Boehm U, Heinlein M, Behrens U, Kunze R. 1995. One of three nuclear localization signals of maize activator (Ac) transposase overlaps the DNA-binding domain. Plant J 7:441–451.

Bunch TA, Grinblat Y, Goldstein LS. 1988. Characterization and use of the *Drosophila* metallothionein promoter in cultured *Drosophila melanogaster* cells. Nucleic Acids Res 16:1043–1061.

Calvi BR, Hong TJ, Findley SD, Gelbart WM. 1991. Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: Hobo, activator, and Tam3. Cell 66:465–471.

Christophe D, Christophe-Hobertus C, Pichon B. 2000. Nuclear targeting of proteins: How many different signals? Cell Signal 12:337–341.

Coates CJ, Johnson KN, Perkins HD, Howells AJ, O'Brochta DA, Atkinson PW. 1996. The hermit transposable element of the Australian sheep blowfly, *Lucilia cuprina*, belongs to the hAT family of transposable elements. Genetica 97:23–31.

Cokol M, Nair R, Rost B. 2000. Finding nuclear localization signals. EMBO Rep 1:411–415.

Dingwall C, Laskey RA. 1991. Nuclear targeting sequences—a consensus? Trends Biochem Sci 16:478–481.

Gama-Carvalho M, Carmo-Fonseca M. 2001. The rules and roles of nucleocytoplasmic shuttling proteins. FEBS Lett 498:157–163.

Goerlich D. 1997. Nuclear protein import. Curr Opin Cell Biol 9:412–419.

Grebenok RJ, Pierson E, Lambert GM, Gong F-C, Afonso CL, Haldemann-Cahill R, Carrington JC, Galbraith DW. 1997. Green-fluorescent protein fusions for efficient characterization of nuclear targeting. Plant J 11:573–586.

Handler AM, Gomez SP. 1997. A new hobo, Ac, Tam3 transposable element, hopper, from *Bactrocera dorsalis* is distantly related to hobo and Ac. Gene (Amsterdam) 185: 133–135.

Jasinskiene N, Coates CJ, James AA. 2000. Structure of hermes integrations in the germline of the yellow fever mosquito, *Aedes aegypti*. Insect Mol Biol 9:11–18.

Kalderon D, Roberts BL, Richardson WD, Smith AE. 1984. A short amino acid sequence able to specify nuclear location. Cell 39:499–509.

Kempken F, Kuck U. 1996. Restless, an active Ac-like transposon from the fungus *Tolypocladium inflatum*: Structure, expression, and alternative RNA splicing. Mol Cell Biol 16:6563–6572.

Kunik T, Palanichelvam K, Czosnek H, Citovsky V, Gafni Y. 1998. Nuclear import of the capsid protein of tomato yellow leaf curl virus (TYLCV) in plant and insect cells. Plant J 13:393–399.

Kunze R, Stochaj U, Laufs J, Starlinger P. 1987. Transcription of transposable element activator (Ac) of *Zea mays* L. EMBO J 6:1555–1564.

LaCasse EC, Lefebvre YA. 1995. Nuclear localization signals overlap DNA- or RNA-binding domains in nucleic acid-binding proteins. Nucl Acids Res 23:1647–1656.

Li PP, Nakanishi A, Shum D, Sun PC-K, Salazar AM, Fernandez CF, Chan S-W, Kasamatsu H. 2001. Simian virus 40 Vp1 DNA-binding domain is functionally separable from the overlapping nuclear localization signal

and is required for effective virion formation and full viability. J Virol 75:7321–7329.
Liu L, White MJ, MacRae TH. 1999. Transcription factors and their genes in higher plants: Functional domains, evolution, and regulation. Eur J Biochem 262:247–257.
Lorincz M, Roederer M, Diwu Z, Herzenberg LA, Nolan GP. 1996. Enzyme-generated intracellular fluorescence for single-cell reporter gene analysis utilizing *Escherichia coli* beta-glucuronidase. Cytometry 24:321–329.
Michel K, Stamenova A, Pinkerton AC, Franz G, Robinson AS, Gariou-Papalexiou A, Zacharopoulou A, O'Brochta DA, Atkinson PW. 2001. Hermes-mediated germ-line transformation of the Mediterranean fruit fly *Ceratitis capitata*. Insect Mol Biol 10:155–162.
Nakai K, Horton P. 1999. PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci 24:34–36.
Nigg EA. 1997. Nucleocytoplasmic transport: Signals, mechanisms, and regulation. Nature 386:779–787.
O'Brochta DA, Warren WD, Saville KJ, Atkinson PW. 1996. Hermes, a functional non-Drosophilid insect gene vector from *Musca domestica*. Genetics 142:907–914.
O'Brochta DA, Atkinson PW, Lehane MJ. 2000. Transformation of *Stomoxys calcitrans* with a hermes gene vector. Insect Mol Biol 9:531–538.
Okuda M, Ikeda K, Namiki F, Nishi K, Tsuge T. 1998. Tfol: An Ac-like transposon from the plant pathogenic fungus *Fusarium oxysporum*. Mol Gen Genet 258:599–607.
Otterlei M, Haug T, Nagelhus TA, Slupphaug G, Lindmo T, Krokan HE. 1998. Nuclear and mitochondrial splice forms of human uracil-DNA glycosylase contain a complex nuclear localisation signal and a strong classical mitochondrial localisation signal, respectively. Nucl Acids Res 26:4611–4617.
Pinkerton AC, Whyard S, Mende HA, Coates CJ, O'Brochta DA, Atkinson PW. 1999. The Queensland fruit fly, *Bactrocera tryoni*, contains multiple members of the hAT family of transposable elements. Insect Mol Biol 8:423–434.
Pinkerton AC, Michel K, O'Brochta DA, Atkinson PW. 2000. Green fluorescent protein as a genetic marker in transgenic *Aedes aegypti*. Insect Mol Biol 9:1–10.
Pollard VW, Michael WM, Nakielny S, Siomi MC, Wang F, Dreyfuss G. 1996. A novel receptor-mediated nuclear protein import pathway. Cell 86:985–994.
Quaedvlieg NEM, Schlaman HRM, Admiraal PC, Wijting SE, Stougaard J, Spaink HP. 1998. Fusions between green fluorescent protein and beta-glucuronidase as sensitive and vital bifunctional reporters in plants. Plant Mol Biol 38:861–873.
Reinhardt A, Hubbard T. 1998. Using neural networks for prediction of the subcellular location of proteins. Nucl Acids Res 26:2230–2236.
Rhee Y, Gurel F, Gafni Y, Dingwall C, Citovsky V. 2000. A genetic system for detection of protein nuclear import and export. Nat Biotechnol 18:433–437.
Sarkar A, Coates CJ, Whyard S, Willhoeft U, Atkinson PW, O'Brochta DA. 1997a. The hermes element from *Musca domestica* can transpose in four families of cyclorrhaphan flies. Genetica 99:15–29.
Sarkar A, Yardley K, Atkinson PW, James AA, O'Brochta DA. 1997b. Transposition of the hermes element in embryos of the vector mosquito, *Aedes aegypti*. Insect Biochem Mol Biol 27:359–363.
Schneider I. 1972. Cell lines derived from late embryonic stages of *Drosophila melanogaster*. J Embryol Exp Morph 27:353–365.
Smith HMS, Raikhel NV. 1999. Protein targeting to the nuclear pore. What can we learn from plants? Plant Physiol 119:1157–1163.
Sommer H, Carpenter R, Harrison BJ, Saedler H. 1985. The transposable element Tam3 of *Antirrhinum majus* generates a novel type of sequence alterations upon excision. Mol Gen Genet 199:225–231.
Sundaresan V, Springer P, Volpe T, Haward S, Jones JDG, Dean C, Ma H, Martienssen R. 1995. Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes Dev 9:1797–1810.
van Roessel P, Brand AH. 2002. Imaging into the future: Visualizing gene expression and protein interactions with fluorescent proteins. Nat Cell Biol 4:E15–E20.
Warren WD, Atkinson PW, O'Brochta DA. 1994. The hermes transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the hobo, Ac, and Tam3 (hAT) element family. Genet Res 64:87–97.
Yoneda Y, Hieda M, Nagoshi E, Miyamoto Y. 1999. Nucleocytoplasmic protein transport and recycling of Ran. Cell Struct Funct 24:425–433.