The proteins encoded by the *pogo*-like *Lemi*\(^1\) element bind the TIRs and subterminal repeated motifs of the *Arabidopsis Emigrant* MITE: consequences for the transposition mechanism of MITEs

Céline Loot, Néstor Santiago, Alicia Sanz and Josep M. Casacuberta*

Departament de Genètica Molecular, Laboratori de Genètica Molecular Vegetal CSIC-IRTA, Jordi Girona 18, 08034 Barcelona, Spain

Received May 22, 2006; Revised August 10, 2006; Accepted September 7, 2006

ABSTRACT

MITEs (miniature inverted-repeated transposable elements) are a particular class of defective DNA transposons usually present within genomes as high copy number populations of highly homogenous elements. Although an active MITE, the *mPing* element, has recently been characterized in rice, the transposition mechanism of MITEs remains unknown. It has been proposed that transposases of related transposons could mobilize MITEs in *trans*. Moreover, it has also been proposed that the presence of conserved terminal inverted-repeated (TIR) sequences could be the only requirement of MITEs for mobilization, allowing divergent or unrelated elements to be mobilized by a particular transposase. We present here evidence for a recent mobility of the *Arabidopsis Emigrant* MITE and we report on the capacity of the proteins encoded by the related *Lemi*\(^1\) transposon, a *pogo*-related element, to specifically bind *Emigrant* elements. This suggests that *Lemi*\(^1\) could mobilize *Emigrant* elements and makes the *Lemi*\(^1*/Emigrant* couple an ideal system to study the transposition mechanism of MITEs. Our results show that *Lemi*\(^1\) proteins bind *Emigrant* TIRs but also bind cooperatively to subterminal repeated motifs. The requirement of internal sequences for the formation of proper DNA/protein structure could affect the capacity of divergent MITEs to be mobilized by distantly related transposases.

INTRODUCTION

Transposable elements (TE) can be divided into two classes according to their structure and transposition mechanism. Class 1 elements transpose by a replicative mechanism involving an RNA molecule that is reverse transcribed before integration, while class 2 elements are mobilized by a cleavage and strand-transfer mechanism usually known as ‘cut and paste’. Although the enzymes required for transposition can be encoded by the mobile element itself, non-autonomous defective elements that can be mobilized in *trans* exist for both classes of elements. Autonomous DNA elements encode transposases that are able to bind to the terminal sequences of the element and that catalyse the cleavage and strand-transfer reactions. As most non-autonomous DNA elements are mutation derivatives of their autonomous counterparts, they usually show sequence similarity to them and can often be mobilized by the related transposases.

MITEs (miniature inverted-repeated transposable elements) are usually classified as non-autonomous DNA transposons because they share structural characteristics with these elements. MITEs contain terminal inverted-repeated (TIR) sequences and do not have any coding capacity. The existence of putative transposons sharing extensive sequence similarities to some MITE families and potentially coding for transposases has led to propose that MITEs could be deletion derivatives of DNA transposons that are mobilized in *trans* by these elements (1,2).

Most MITEs have been characterized by computer-assisted searches and, until very recently, the proposal of a precise mechanism of transposition of these elements has been prevented by the lack of an actively transposing MITE. The characterization of a rice MITE named *mPing*, whose transposition is induced in anther-derived cell cultures (3) leaving excision footprints behind (4), confirmed that MITEs can transpose by a ‘cut and paste’ mechanism typical of DNA transposons. Nevertheless, although *mPing* has extensive sequence similarities to a DNA transposon potentially coding for a transposase of the *mariner* superfamiliy, the *Ping* element, the rice varieties where *mPing* was mobilized do not contain any complete *Ping* element that could account for the mobilization of these elements. For this reason it was proposed that a different DNA transposon, *Pong*, which is...
only distantly related to Ping/mPing elements, could be the source of transposase (5). Different phylogenetic analysis indicated that transposases not directly related to a particular MITE family could be responsible for the mobilization of these elements (6, 7). Moreover, although it has not yet been proved that the capacity of any DNA transposase (either of the same or of a different family) can mobilize a MITE copy, a recent report shows that rice mariner-like transposases can bind in vitro and in yeast assays to the TIRs of non-related Stowaway MITEs (8). The presence of similar TIR sequences could thus be the only requirement for interaction with transposases non-directly related to the MITE, in line with the hypothesis that MITEs could be mobilized by distantly related transposases (9).

The Emigrant MITE (10) is present in some 500 copies in the genome of Arabidopsis and other Brassicaceae. A phylogenetic analysis has shown that Arabidopsis contains different subfamilies of the Emigrant elements consistent with the amplification of one or a few 'master' copies at different times during the evolution of these species (11). The Columbia ecotype of Arabidopsis contains a single-copy poso-related element called Lemi1 that has extensive sequence homology with Emigrant elements from which the latter was probably derived by internal deletion (12).

Here we show that some of the Emigrant insertions are polymorphic among different Arabidopsis ecotypes or even among different individuals of the same ecotype, suggesting that they have transposed in a recent past and that Arabidopsis probably contains an enzymatic activity capable of mobilizing these elements. We also show that the proteins encoded by Lemi1 specifically bind to Emigrant TIRs, suggesting that Lemi1 could provide the transposase mobilizing Emigrant elements. Moreover, we show that Lemi1 proteins cooperatively bind to subterminal repeated motifs, suggesting that the internal sequences of MITEs, and not only the TIR sequences, could also be important for transposase interaction and mobilization of some of these elements.

MATERIALS AND METHODS

Amplification and cloning of Emigrant and Lemi1 sequences

For the analysis of Emigrant polymorphisms, the loci containing Emigrant insertions were amplified by PCR with primers corresponding to flanking genomic sequences. For Emi158, the primers were e158-5′ (5′-CCATATTCAACATTCT-3′) and e158-3′ (5′-CTGAAAATAGAAAAAGAG-3′).

Lemi1 sequences of different Arabidopsis ecotypes were amplified by PCR using the 372 (5′-CTCTGTCTTTGATCACTACAAAA-3′) and 1616 (5′-GCTTAAATAAATAGAAAGAG-3′) primers. PCR products were cloned into the pCRII-TOPO vector (Invitrogen) and sequenced. The sequences were aligned using the Genedoc program.

PCR products obtained by amplification of Arabidopsis Columbia-0 genomic DNA with the primers 372–1616 (5′-GGTGAACATCGTTGATTTTCTCTC-3′) and 5′-GGCTCGAGGAGTTAGTTTATAGTTAG-3′ were cloned into the pTZ57R vector (Fermentas) to generate the pLemi2 plasmid. For gel retardation assays, the Lemi1 proteins were produced as fusion proteins linked to the glutathione S-transferase (GST) using the pGEX-KG system.

The Lemi1 sequences were amplified by PCR from pLemi0, pLemi1, pLemi2 with the EcoRI-390 and 1684-SacI primers (5′-CCGAATTCCGAATTCGCTGAGGAGG-3′) and 5′-GGCTCGAGGAGTTAGTTTATAGTTAG-3′, respectively, and from pLemi5 using the EcoRI-390 and 1949-SacI primers (5′-GGCTCGAGGAGTTTATAGTTAGTTAG-3′). The PCR products were digested by EcoRI and SacI and ligated into the pGEX-KG vector, resulting in pGEX5 and pGEX5. The pGEX5 co-1 was generated by replacing an AccI–AccI fragment of pGEX5 by the fragment containing the 372–1616 Lemi1 sequence obtained by PCR from Arabidopsis Coimbra-1 genomic DNA.

Electrophoretic mobility shift assays (EMSAs)

The region of Arabidopsis Columbia-0 genomic DNA containing the Emi126 element was amplified with the primers e126-5′ (5′-CCAGCGATAACATTGATTTTCTC-3′) and e126-3′ (5′-GATGATCTCAGACGATTTCTC-3′) and cloned in the pTZ57R vector (Fermentas) to generate the pEmi1 plasmid. This plasmid was used to obtain the pEmi1 TIR1 and pEmi1 TIR2 probes, as well as a non-specific (NS) competitor DNA fragment. The pEmi1 TIR1 probe was produced by PCR amplification using the 5′-TIR1C1 and 3′-TIR1C1 primers (5′-ATATAATTGTCGAGCTAGACAAACATG-3′ and 5′-CCGAATTCCGAATTCGCTGAGGAGG-3′), respectively, and the pEmi1 TIR2 probe was obtained using the 5′-TIR2C1 and 3′-TIR2C1 primers (5′-GAAGATCTTATATATTTGTTGAGA-3′ and 5′-CGATTTAAAAGTAAGTTTGAGT-3′), respectively. The non-specific competitor DNA was generated by PCR on the same pEmi1 plasmid with primers flanking the Emi126 insertion: e126c and e126-3′ (5′-ATATAATTGTCGAGCTAGACAAACATG-3′) and 5′-CCGAATTCCGAATTCGCTGAGGAGG-3′, respectively). The PCR products were cloned into the pTZ57R vector (Fermentas) to generate, respectively, the pEmi6, pEmi3 and pEmi4 plasmids. The pEmi1 TIR1 probe (99 bp of the pEmi1 3′-terminal sequence plus 69 bp of flanking genomic sequence), the pEmi1 TIR2 probe (64 bp of pEmi1 5′-terminal sequence constructs by PCR site-directed mutagenesis using the 497W (5′-CTGATTTCACTTTTTGAAAGG-3′) and 520W (5′-CTTCAAGCTTTTTAATCGG-3′) primers. PCR products were cloned into the pTZ57R vector to produce the pLemi0 and pLemi4 plasmids. The intron of orf1-2 was removed by PCR from pLemi4 using partially complementary primers (W1′: 5′-CTGATCCAGAGGTATGGAGTACCTT-3′ and W13′: 5′-CGATCTAGTCCTTAACTTACGTG-3′) that span the donor and acceptor splicing sites. The PCR product obtained was cloned into the pTZ57R vector to generate the pLemi5 plasmid. During the PCR amplification of the Lemi1 sequences, a mutation was introduced leading to the formation of a stop codon and therefore generating a truncated protein of 264 amino acids corresponding to the entire DNA-binding domain and a truncated catalytic domain. The PCR product was cloned into the pTZ57R vector to obtain the pLemi2 plasmid.
plus 105 bp of flanking genomic sequence) and the NS probe (169 bp of genomic sequences flanking Emi126) were generated by Sacl–BamHI digestion of the pEmi6, pEmi3 and pEmi4 plasmids, respectively, and radioactively labelled, when necessary, with [α-32P]dCTP using Klenow polymerase (Roche) by standard procedures.

The Emi126-L probe was produced by EcoRI–HindIII digestion of the pEmi3 plasmid to give a fragment of 209 bp containing 64 bp of Emi126 and 146 bp of flanking sequences.

The Emi158 locus was amplified by PCR on Arabidopsis Columbia-0 genomic DNA using the e158-5′ (5′-CCATATTCAAAATTTAC-3′) and e158-3′ (5′-GCTTAATAAAATAAGAAG-3′) primers, and cloned in pTZ57R to generate the pEmi158 plasmid. This plasmid was used to produce the Emi158 TIR1 and Emi158 TIR2 probes by PCR using, respectively, the e158-5′TIR1 and e158-3′TIR1 primers (5′-TTTGAAGAATTCATTTATATAT-3′, 5′-TTTGAAGAATTCATTTATATAT-3′) and the e158-5′TIR2 and e158-3′TIR2 primers (5′-GAAGAATTCATTTATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT
by a STOP codon in position 39 of the protein and a frameshift in position 385 of the protein ([12] and Figure 2A). It has been proposed that the splicing of a putative intron could allow overcoming the frameshift (12). The sequence of the putative donor and acceptor splicing sites (data not shown and Figure 2B, respectively) perfectly fit the consensus for plant introns (14) and a consensus branch point sequence is found at the correct distance from the acceptor AG (data not shown). In order to get insight on the original structure of the Lemi1 element we have analysed the Lemi1 sequences of different Arabidopsis ecotypes. Lemi1 is present as a single-copy element in all the Arabidopsis ecotypes that we have analysed, and we have not been able to obtain evidences of mobility of Lemi1 in those genomes (data not shown). Using primers complementary to internal Lemi1 sequences we have amplified, sequenced and compared Lemi1 sequences obtained from seven different Arabidopsis ecotypes. In spite of the high similarity found (from 94 to 100% identical over 1245 bp), the Lemi1 sequences are polymorphic at particularly important positions (Figure 2B). The Lemi1 sequence does not contain the STOP codon found in Columbia in four different ecotypes, Ms-0, RLD and Dijon-G, and is probably present in two different alleles (only one of them containing the STOP) in the Tsu-0 ecotype, as we have obtained two types of sequences when amplifying this region from this ecotype. On the other hand, the Lemi1 sequence has an insertion of 56 bp in the region of the frameshift that restores the coding capacity in a single orf in Coimbra-1 and Coimbra-4 (Figure 2B). The presence of this insertion in Lemi1-related sequences of Gossypium hirsutum, Solanum demissum and Medicago truncatula (data not shown) suggests that the original Lemi1 element had a single orf of 1559 bp. Nevertheless, Coimbra-1 and Coimbra-4 also have a difference with respect to the Lemi1 sequences found in all other ecotypes at one of the two invariable nucleotides of the putative splicing acceptor site (Figure 2B). The presence of an acceptor splicing consensus site exclusively in the Lemi1 sequences where the orf is interrupted by a frameshift (Figure 2B) could also indicate the functionality of a spliced protein.

The proteins encoded by Lemi1 specifically and cooperatively bind Lemi1 and Emigrant sequences

Transposase binding to the terminal regions of the mobile element is the first step of the transposition process. In order to test the capacity of the proteins encoded by the original Lemi1 transposon to bind Lemi1 and Emigrant sequences, we reconstructed the consensus Lemi1 coding sequence by replacing the STOP codon found in the Columbia Lemi1 sequence with the tryptophan coding triplet found in Ms-0, RLD, Dijon-D and Tsu-0 ecotypes by site-directed mutagenesis. We expressed in Escherichia coli the two proteins encoded by the modified Lemi1 element as GST fusions: the protein encoded by the first orf (GST–Orf1) and the protein that would be produced by splicing the predicted Lemi1 intron and consisting in a fusion of most of Orf1 and Orf2 (GST–Orf1–2) (Figure 3). On the other hand, we generated a construct containing the insertion found in Coimbra-1 and Coimbra-4 (GST–Orf1–2+) by replacing an Accl–Accl fragment (see Figure 2A) of the Columbia sequence by that

flanking Emigrant insertions (Figure 1C). These sequence differences to the expected empty site probably represent excision footprints and suggest that although most of the polymorphisms detected are probably due to differential Emigrant insertions, a few of them were probably generated by Emigrant excisions. Our results thus suggest that in the recent past the genome of Arabidopsis contained an enzymatic activity that was able to excise and reinsert Emigrant MITEs.

Analysis of the Emigrant-related pogo-like transposon Lemi1 in different Arabidopsis ecotypes

It has been proposed that Emigrant elements originated by a severe deletion of a putative pogo-like transposon called Lemi1 that shows extensive sequence homology with Emigrant elements (12). Lemi1 is present only in one copy of the Arabidopsis Columbia ecotype, and displays an orf potentially coding for a pogo-like transposase interrupted
of the Coimbra-1. As control proteins, we expressed in *E. coli* GST fusions with truncated Lemi1 proteins: the GST construct, in which the STOP codon of the Columbia Lemi1 sequence was maintained and that encodes a GST protein fused to a short polypeptide of 38 amino acids, and the GST–BD construct, in which a STOP codon was introduced at position 264 of the protein and that encodes for a GST protein fused with a truncated Orf1 protein that contains the whole DNA-binding domain and a truncated catalytic domain. These proteins were used to perform EMSA with radioactively labelled probes corresponding to the terminal sequences of Lemi1 and Emigrant elements. We first tested the ability of Lemi1 proteins to bind its own terminal sequences. EMSA analysis showed that the GST–Orf1 protein binds Lemi1 TIR1 and TIR2 probes giving one major retarded band, B1 (TIR1), or three retarded bands, B1–B3 (TIR2) (Figure 4), suggesting that the reconstructed protein has retained the ability to specifically bind the transposon TIR sequences. In order to test if Lemi1 proteins could also specifically bind the TIRs of Emigrant elements we performed EMSA analysis with the TIRs of the Emi126 element, a polymorphic Emigrant element belonging to the young EmiA family (11). These analyses showed that Lemi1 proteins specifically bind Emigrant probes in a similar way, suggesting that the
C-terminal part of the catalytic domain, which coincides with the polypeptide encoded by orf2, does not participate in the binding of Lemi1 protein(s) to DNA. The GST–BD polypeptide binds Emigrant TIR1 and TIR2 sequences but, although its binding of TIR1 is very similar to that of GST–Orf1, GST–Orf1-2 and GST–Orf1-2+ (Figure 5A), its binding to TIR2 appears to be very different. While GST–Orf1, GST–Orf1-2 and GST–Orf1-2+ proteins gave three retarded bands with TIR2, GST–BD gave only one retarded band even at high-protein concentrations (Figure 5B, lanes 5–7). Therefore, although the DNA-binding domain of Lemi1 protein(s) is sufficient for specific binding, a region of the catalytic domain absent from the GST–BD protein is needed for multiple binding to TIR2. A suggestive possibility is that this region, which is absent from the GST–BD polypeptide and is located within the catalytic domain, could mediate protein–protein interactions allowing multiple binding to DNA. Transposase dimerization domains can be located within the DNA-binding domain, e.g. in the mariner-like Sleeping Beauty transposase (15), in the C-terminal part of the protein, in many hAT transposases such as Hermes (16), or within both the DNA-binding domain and the catalytic domain similar to the one in the case of the mariner-like Mos1 transposase (17–19).

Lemi1 proteins bind to the Emigrant TIR and to subterminal repeated motifs

A key step of transposition is the formation of a precise DNA/protein structure that requires transposase dimerization and allows DNA cleavage and strand-transfer reactions (20). In most cases this structure consists of a synaptic complex in which the transposase catalyses the reaction in trans to the transposon end at which it is bound, but transposases can also catalyse the reaction in cis, as it has been shown for the mariner Himar1 element (21). Transposase dimerization is essential to form active complexes, and transposase dimerization can take place in cis, with a single transposase unit bound to the DNA (21), or between transposases bound to the DNA side-by-side, or in trans to form paired end complexes (17). In order to determine whether the different retarded bands obtained with Emigrant TIR1 and TIR2 probes are the result of multiple protein units bound to a single DNA molecule or of complexes containing multiple DNA molecules, we have performed EMSA analysis in the presence of competitor molecules of different sizes. The competition of GST–Orf1 binding to Emi126 TIR2 probe with different unlabelled Emi126 TIR2 fragments of different sizes did not reveal any difference in the mobility of retarded bands that could be an indication of the presence of multiple DNA fragments in protein/DNA complexes (Figure 6). Thus, the retarded bands obtained in EMSA experiments are probably the result of the binding of multiple Lemi1 proteins to a single DNA molecule.

We performed DNase I footprinting analysis to determine the Lemi1-binding sites in Emi126. These experiments showed that GST–Orf1-2 binds Emigrant TIRs but also other internal sequences (Figure 7). In the case of TIR2 the DNase I protection covers a continuous region of 56 bp including the TIR and two repeated motifs that coincide with the 3′-half of the TIR sequence (Figure 7B and see Figure 5B for sequence details). The TIR2 footprint is flanked by a DNase I hypersensitive band indicating that binding of Lemi1 protein(s) induce(s) a distortion of the target DNA. Protein binding often affects DNA structure and in particular transposases often distort DNA upon binding (22). The DNase I footprinting analysis of TIR1 shows a protection that covers two regions: 23 bp of the TIR itself and a 29 bp region consisting of two repeats of a sequence coinciding with the 3′-half of the TIR in opposite orientation and separated from the TIR by 22 bp (Figure 7A and see Figure 5A for sequence details).

These experiments show that multiple binding to Emi126 TIR1 and TIR2 sequences revealed by EMSA experiments is the result of recognition of the TIR itself and the subterminal repeated sequences by Lemi1 proteins. Binding to one, two or three of these sequences could explain the different retarded bands obtained in EMSA. Nevertheless, although Lemi1 proteins gave three retarded bands in EMSA with the TIR2 probe, they gave only two with the TIR1 probe (Figure 5). This suggests that GST–Orf1, GST–Orf1-2 and GST–Orf1-2+ proteins can simultaneously bind the three binding motifs present in TIR2 while they bind, but not simultaneously, the three binding motifs found in TIR1. As there are no major differences in the binding motifs found in both TIRs, the different binding should be explained by the different arrangement of the binding motifs in both TIRs. Indeed, the TIR and the two subterminal motifs are contiguous and in the same orientation in TIR2 while in TIR1 the subterminal repeats are found in reverse orientation and separated from the TIR sequence (Figure 5). The need for an internal region of the catalytic domain of Lemi1 proteins for multiple binding to TIR2 suggests that protein–protein interactions play an important role in Lemi1 binding and different arrays of DNA-binding motifs would probably modify the protein–protein interactions that can take place.
Figure 5. DNA-binding analysis of the different Lemi1 proteins to the Emi126 TIRs. Increasing concentrations of the indicated proteins (left and middle panels) or a fixed concentration of GST-Orf1 protein (right panel) where incubated with radioactively labelled probes corresponding to Emi126 TIR1 (99 bp of the 5'-terminal sequence of Emi126 element plus 69 bp of the flanking genomic sequence) (A) or Emi126 TIR2 (64 bp of the 3'-terminal of Emi126 plus 105 bp of the flanking genomic sequence) (B), and were analysed by EMSA. Control reactions with no recombinant protein (–) were also included as controls. Competition experiments (right panel) were performed by including in the reaction mixture increasing concentrations (ratios: 1/5 and 1/50) of ice-cold non-specific competitor (NS, 169 bp of genomic sequence flanking Emi126 insertion), Emi126 TIR1(A) or TIR2 (B) fragments, Emi158 TIR1 (99 bp of the 3'-terminal of Emi158 plus 88 bp of the flanking genomic sequence) (A) or Emi158 TIR2 (64 bp of the 3'-terminal of Emi158 plus 123 bp of the flanking genomic sequence) (B), Lemi1 TIR1 (106 bp of the 3'-terminal of Lemi1 plus 93 bp of the flanking genomic sequence) (A) or Lemi1 TIR2 (64 bp of the 3'-terminal of Lemi1 plus 134 bp of the flanking genomic sequence) (B). The migrating position of the free probes (F) and the different retarded bands (B1–B3) is shown on the right. An alignment of the region comprising the TIR and subterminal sequences of the Emi126, Emi158 and Lemi1 is shown on the bottom. The position of the TIR and subterminal sequences is indicated by grey and black arrows. The position of the target site duplication is shown by a solid line. Nucleotides are numbered from the first nucleotide of the TIR.
The existence of subterminal repeats with a capacity to bind the transposase and required for proper activity has also been recently reported for some rice mariner-like transposases (8) as well as other mariner/Tc1 elements such as Sleeping Beauty (15) or pogo (23), and for transposons of other families, such as the CACTA and hAT families [reviewed in (24)]. The cooperative binding of Lemi1 protein(s) to Emigrant subterminal motifs indicates that this can also be the case for MITEs, suggesting that the TIRs could not always be the only requirement for transposase binding to MITE sequences and transmobilization.

**Lemi1 proteins bind differently to Emigrant and Emigrant sequences**

In order to test the capacity of Lemi1 proteins to bind different Emigrant-related sequences and analyse the importance of subterminal repeated motifs for binding, we performed binding and competition experiments with different sequences. In addition to Emi126, we analysed the binding of Lemi1 proteins to Emi158, another polymorphic Emigrant element (see Figure 1) that has well-conserved TIR and subterminal repeated motifs and Lemi1, which has consensus TIR sequences but presents important differences in the subterminal regions of TIR1 (Figure 5). The competition experiments with ice-cold Emi126, Emi158 and Lemi1 probes show that both Emi158 and Lemi1 bind Lemi1 proteins with an efficiency similar to that of Emi126 (Figure 5A and B, right panels). This suggests that TIR itself is the primary determinant for Lemi1 binding. Thus, Emigrant-related elements that have diverged in their subterminal regions can efficiently bind Lemi1 proteins, similarly to what has been recently found for the interaction of Stowaway MITEs with related but distinct transposon families (8). However, EMSA analysis using Lemi1 TIRs as probes showed that, although elements that have diverged in their subterminal repeats can bind Lemi1 proteins, their binding is different. Indeed, although Lemi1 binding to the well-conserved Emigrant TIR2 is similar to that of Emi126, the binding to Lemi1 TIR1 is different. Indeed, Lemi1 binding to Lemi1 TIR1 produces only one major retarded band (Figure 4) instead of the two obtained with Emi126 (Figure 5). This suggests that Lemi1 proteins only bind Lemi1 TIR motif itself and not the divergent subterminal motifs found in this sequence. These results thus show that multiple binding to Emigrant-related elements depends on the conservation of the sequence and the relative arrangement of the subterminal repeated motifs. Thus, although Lemi1 proteins could efficiently bind the TIRs of divergent Emigrant elements that have conserved the TIRs, their final DNA–protein structure will greatly depend on the number and the relative arrangement of the subterminal repeated motifs. It has been shown that the formation of a proper nucleoprotein complex, which can depend on the

---

**Figure 6.** Competition experiments with Emi126 TIR2 fragments of different sizes. GST–Orf1 protein was incubated with Emi126 TIR2 probe in the presence of increasing concentrations (ratios: 1/1, 1/10 and 1/30) of ice-cold Emi126 TIR2 (E126) or a DNA fragment containing the same Emi126 region flanked by longer genomic and plasmid DNA sequence (E126-L). The migrating position of the free probe (F) and the different retarded bands (B1–B3) is shown on the right. A scheme representing Emi126 TIR2 (E126) and E126-L, used as probe and/or competitors is shown on the bottom. Emi126 sequences are shown as a closed boxes whereas genomic and plasmid DNA sequences are shown as open boxes.

**Figure 7.** Footprinting analysis of the Lemi1 proteins binding to Emi126. Increasing concentrations of the GST–Orf1-2 were incubated with radioactively labelled Emi126 TIR1 (A) and TIR2 (B) probes. The position of the TIR sequence is shown by a grey arrow and the position of the TIR and subterminal repeated motifs are indicated by solid arrows. The position of G nucleotides of the upper strand (TIR2) or the bottom strand (corresponding to C nucleotides in the upper strand) (TIR1), revealed by DMS reactions, are shown on the right of each gel. G (or C) positions, numbered from the first nucleotide of the TIR, as in Figure 5, are 1, 9, 10, 12, 34, 37 and 80 (TIR1), and −23, −22, −13, −12, −6, 1, 9, 10, 12, 25, 27, 41, 42, 44, 61 and 64 (TIR2). A closed triangle indicates the DNase I hypersensible position flanking the protected region in TIR2. A sequence comparison of the TIRs and the repeated motifs is shown at the bottom. Identical nucleotides appear as white letters on black boxes.
binding of the transposase to subterminal repeated motifs, is a key regulatory step in transposition [reviewed in (25)]. Moreover, in some cases, such as for the \textit{Sleeping Beauty} transposon, transposase binding to a subterminal repeat greatly enhances transposition (15). Interestingly, the subterminal repeat of \textit{Sleeping Beauty} consists of the 3' half of the TIR, similarly to what we describe here for the \textit{Lemi}1/\textit{Emigrant} elements. We thus propose that the conservation of the subterminal repeated motifs could modify the potentiality of a particular \textit{Emigrant}-related element to be mobilized by the \textit{Lemi}1-encoded transposase.

CONCLUSIONS

The results presented here show that the \textit{Arabidopsis Emigrant} MITE has transposed in the recent past and that the protein(s) encoded by the \textit{pogo-like Lemi}1 element specifically and cooperatively bind \textit{Emigrant} elements. This suggests that \textit{Lemi}1 could mobilize \textit{Emigrant} elements and makes the \textit{Lemi}1/\textit{Emigrant} couple an ideal system to study the transposition mechanism of MITEs. Our results show that the sequence of the TIR itself is the primary determinant for binding. But our results also show that, once the binding to the TIR is accomplished, \textit{Lemi}1 proteins can also bind subterminal repeated motifs. These results thus show that, at least for the \textit{Emigrant}/\textit{Lemi}1 system, the final protein/DNA structure depends on transposase binding to subterminal repeated motifs, and suggests that the conservation of internal sequences could influence the ability of a particular MITE to be mobilized by a related transposase.

ACKNOWLEDGEMENTS

The authors wish to thank Michael Joulie for his help in the analysis of \textit{Lemi}1 copy number. We are grateful to Elena Casacuberta, Lluís Espínàs, Inmaculada Hernández-Pinzón, Paloma Mas and Soraya Pelaz for their critical reading of the manuscript. This work was funded by the Spanish Ministerio de Ciencia y Tecnología (Grant BIO2003-01211) and the Centre de Referència en Biotecnologia (CeRBa) from the Generalitat de Catalunya. C.L. is recipient of an I3P Postdoctoral contract from the Spanish Ministerio de Ciencia y Tecnología. Funding to pay the Open Access publication charges for this article was provided by Ministerio de Ciencia y Tecnología (Grant BIO2003-01211).

Conflict of interest statement. None declared.

REFERENCES