

Nuclear Importation of *Mariner* Transposases among Eukaryotes: Motif Requirements, and Homo-Protein Interactions

Marie-Véronique Demattei, Sabah Hedhili, Ludivine Sinzelle, Christophe Bressac, Sophie Casteret,
Nathalie Moiré, Jeanne Cambefort, Xavier Thomas, Nicolas Pollet, Pascal Gantet
and Yves Bigot

Supporting Information 5 : Sequence and expression properties of the optimized versions of the gene encoding MOS1

In silico design of optimized MOS1 ORF

Previous studies have demonstrated that the expression capacities of the eukaryotic genes can be restricted by several non-mutually exclusive factors^{1,2}. The first such factor is the ability of the mRNA to be accurately translated by the cellular machinery. Indeed, we observed that most of the MLE transposase ORFs available in databanks and in the literature had no Kozak box overlapping their start codon, thus potentially reducing the ability of the ribosomes to initiate the efficient translation of the transcript into protein. We also observed that most MLE transposase ORFs contained high levels of rare codons. For example, we found 39 rare codons in the ORF encoding MOS1 regardless of whether this protein was expressed in human or in *Drosophila* cells, the host from which it was cloned (Fig. 1, lane 1). These observations therefore supported the hypothesis that the translation of MLE transposase mRNA could be limited in many eukaryotic cells by codon usage, as previously demonstrated for other proteins³. These facts were used to define a first version of the MOS1 ORF, MOS1V1, which included a Kozak box and an optimal codon usage for protein expression in human cells (Fig. 1, lane 2).

The second kind of factor that might reduce MOS1 expression involved an intrastrand property of its ORF. We have previously shown^{4,5} that two elements belonging to the *mauritiana* sub-family, *Mos1* and *Botmar1*, both have a nucleic acid sequence that is able to form similar intrastrand DNA structures, despite their sequence divergence (25%). The increase in the sequence data available in databanks has allowed us to confirm with four other MLEs belonging to the *mauritiana* subfamily (*Momar1* (Acc No.: U15665), *Mdmar1* (Acc N°: U24436), *Ramar1* (Acc N°: DQ784570.1) and MEBOTRA1.4-9 (Acc No.: AJ781770.1)) that this property was conserved, whatever the nucleic acid sequence divergence (25-45%; data not shown). If such intrastrand

structures can be stably annealed within the genomic DNA, then they could restrict transcription into mRNA. If they occur in mRNA, they could inhibit their translation into protein by two non-mutually exclusive pathways. First, they could directly block ribosome function. Second, they could elicit the RNA-silencing, since the large intrastrand stem-loop structures might be recognized by *Dicer*-like proteins as potential substrates for making small interfering RNA molecules 20-30 nucleotides in length. Such an RNA self-silencing mechanism would be then able to erase the MLE transcripts.

To circumvent these potential problems of intrastrand annealing, we investigated the significance of the structures obtained in the wild-type *Mos1* ORF (MOS1) and the MOS1V1 using a previously devised statistical method⁵. The resulting graphs showed that all intrastrand folds obtained with MOS1 and most of those obtained with MOS1V1 were statistically consistent (Figure 2 a,b). Since a high proportion of the stem-loop structures was conserved between the folds obtained for both nucleic acid sequences, we removed them manually from the stem regions of the MOS1V1 using the degeneracy of the codon usage to design a second version of the MOS1 ORF, MOS1V2 (Fig. 1, lane 3). In this second version, we obviously avoided using rare codons to eliminate secondary structure. The statistical significance of the intrastrand folds obtained with MOS1V2 was verified. Graphs of the results (Figure 2c) showed that they were not significant, since they were all located in the middle of the space defined by the density ellipses. The nucleic acid features of the MOS1, MOS1V1 and MOSV2 ORFs are summarized in Table 1.

Table 1 – DNA sequence properties of the three versions of the MOS1 ORF

ORF	ORF length	GC content	Number of CpG dinucleotides	Percentage of sequence similarity		
				MOS1	MOS1V1	MOS V2
MOS1	1038	46.52	64	100	74.1	74.8
MOS1V1	1038	66.18	140	74.1	100	93.5
MOS1V2	1038	62.03	111	74.8	93.5	100

The third kind of factor that could affect MLE transposase expression concerned the stability of the mRNA transcripts. It was recently demonstrated that adding specific 5' and 3' untranslated terminal regions (UTRs) has the effect of markedly increasing the expression of the *Sleeping Beauty* and *Himar1* transposases^{6,7}. In an attempt to find out whether the presence of such UTRs interfered with the expression of MOS1 in HeLa cells, we constructed three expression plasmids in which MOS1 [+], MOSV1 and MOS1V2 were flanked by the β -globin 5' and 3' UTRs.

Expression of the different MOS1 ORFs in HeLa cells

In an attempt to track the expression in HeLa cells, the three MOS1 versions were fused at their C-terminal end with GFP, and then cloned in the expression vector pCS2+, as described in the core manuscript. The resulting three expression plasmids were designated MOS1-GFP, MOS1V1-GFP, and MOS1V2-GFP respectively. Three complementary constructs were made to investigate the impact of the UTR, and designated U5-MOS1-GFP-U3, U5-MOS1V1-GFP-U3 and U5-MOS1V2-GFP-U3 respectively. Finally, two expression plasmids, designated pCS2 and pCS2-GFP, were used as negative and positive expression controls respectively. The expression pattern of the various GFP fusions was investigated by western blot analyses of crude protein extracts of plasmid-transfected cell hybridized with anti-GFP antibodies. Our results indicated that the amounts of GFP fusion produced in cells transfected with the MOS1V2-GFP and U5-MOS1V2-GFP-U3 plasmids were 12 to 15-fold higher than those produced with the MOS1-GFP plasmid (Fig. 3).

Materials and Methods

Sequence analyses and statistical treatments

Analyses of the intrastrand DNA structure in the cDNA sequence used the Mfold program⁸ from the web site <http://bioweb.pasteur.fr/seqanal/tmp/mfold/>. The statistical analyses of the significance of the folds obtained for each MLE sequence were carried out by creating files of 1000 shuffled sequences with the shuffleseq program at the website <http://emboss.bioinformatics.nl/>. To calculate the folds of each shuffled sequence, the stand-alone version 3.2 of Mfold (Red Hat binaries) was downloaded. In order to match the physiological conditions of HeLa cells, the folds were computed at 37°C using default parameters. We observed that the main sequence segments found to be involved in intrastrand annealings of the MOS1 ORF at 28°C⁵ were the same whether the calculations were done at 30° or at 37°C. Computing was done on a Mac OS X version 10.4.8 equipped with 2 Go RAM and 2 Dual-Core Intel Xeon 2.66 GHz. A bash script was used to analyze all sets of 1000 sequences. The number of folds and their ΔG for each sequence were recovered, and their graphic statistical analysis was computed by calculating their nonparametric bivariate density using the JMP version 3.6 package available at <http://www.jmp.com/>.

The codon usage of the various putative hosts of the MLE was investigated using data available at <http://www.kazusa.or.jp/codon/>. It was optimized using facilities of the Jcat software at <http://www.prodoric.de/JCat/>.

DNA sources

The ORF encoding the original natural *Mos1* transposase (MOS1; Acc N°X78906) was recovered as described elsewhere⁹. The two versions of the MOS1 ORF (MOS1V1 and MOS1V2)

with the optimized nucleic acid sequences, MOS1V1 and MOS1V2, were synthesized by ATG:Biosynthetics (Germany).

Plasmid constructs

The green fluorescent protein (GFP) cassette fused at the C-terminal ends of the three transposase fragments originated from the pCAMBIA-1302 (Acc No.: AF134298) as described in the Material and Methods section of the main manuscript. It was cloned as a *EcoRI-XbaI* fragment at the *StuI* and *XbaI* sites of the multicloning site of the expression vector pCS2+ (Invitrogen; pCS2-GFP). Cloning for the fusion of the various MOS1 ORFs with GFP was done at the *SpeI* site (pCS2-MOS1-GFP, pCS2-MOS1V1-GFP, pCS2-MOS1V2-GFP). The ORFs encoding the GFP fusions with the three MOS1 versions were also cloned in a pCS2+ vector, in which they were flanked by the 5' and 3' UTRs of the *Xenopus laevis* β -globin mRNA.

Transformation of mammalian cells

HeLa cells were cultured in DMEM supplemented with 10% fetal bovine serum. About 8×10^4 cells were seeded onto a 24-well plates, one day prior to transfection. The cells were transfected with TransPEI, according to the Manufacturer's instructions (Eurogentec). Briefly, plasmid DNA (0.5 μ g) and TransPEI (1 μ l) were separately diluted into 50 μ l of 150 mM NaCl solution and then gently mixed together. After incubating for 30 min, the mixture was diluted in OPTIMEM medium to a final volume of 1 ml. The cells were then incubated with 0.1 ml complexes for 2 to 4 h. The transfection solution was then removed and replaced with fresh supplemented DMEM and the cells were incubated for 24 hours at 37°C. Cells were observed with an epifluorescence microscope (Olympus BX51). The GFP fluorescence was imaged with a blue excitation filter set (460-490 nm excitation filter, 515 nm cut-off filter).

Immunoblotting

Cells recovered from the cultures were washed three times with 1X PBS. Total protein extracts were separated by electrophoresis, adding 40 μ g of each sample to a discontinuous sodium dodecyl sulfate 8% polyacrylamide mini-gel, and then electro-blotted onto nitrocellulose filters (Bio-Rad Laboratories). After blocking with 5% skim milk in phosphate-buffered saline for 1 h, the filters were incubated overnight with rabbit polyclonal anti-GFP (1: 2000; invitrogen). The filters were then incubated with horseradish peroxidase-conjugated anti-rabbit IgG (Santa Cruz Biotechnology, Santa Cruz, CA) before being developed using enhanced chemiluminescence (Amersham Pharmacia Biotech, Sunnyvale, CA). Band intensities on the blot were measured using Image Gauge V3.45 software.

	M	S	S	F	V	P	N	K	E	Q	T	R	T	V	L	I				
[+]	tcagtgcaagtc	atg	tcg	agt	ttc	gtg	ccg	aat	aaa	gag	caa	acc	cg	aca	gta	tta	att	222		
V1	tcagtgcaagtc	atg	atg	agc	agc	ttc	gtg	ccc	aac	aag	gag	cag	acc	cg	acc	gtg	ctg	atc		
V2	tcagtgcaagtc	atg	tcc	ttc	ttc	gtg	ccc	aac	aag	gag	cag	acc	cg	acc	gtg	ctg	atc			
	F	C	F	H	L	K	K	T	A	A	E	S	H	R	M	L	V	E	A	F
[+]	ttc	tgt	ttt	cat	ttg	aag	aaa	aca	gct	gcc	gaa	tcg	cac	aga	atg	ctt	ggt	gaa	gcc	ttt
V1	ttc	tgc	ttc	cat	ctg	aag	aag	acc	gcc	gcc	gag	acc	cac	cg	atg	ctg	gtg	gag	gcc	ttc
V2	ttt	tgt	ttt	cat	ctg	aag	aag	acc	gcc	gcc	gag	tcc	cac	cg	atg	ctg	gtg	gag	gcc	ttc
	G	E	Q	V	P	T	V	K	T	C	E	R	W	F	Q	R	F	K	S	G
[+]	ggc	gaa	caa	gta	cca	act	gtg	aaa	acc	ggt	gaa	cg	tgg	ttt	caa	cg	ttc	aaa	agt	ggt
V1	ggc	gag	cag	gtg	ccc	acc	gtg	aag	acc	tgc	gag	cg	tgg	ttc	cag	cg	ttc	aag	agc	ggc
V2	gga	gag	cag	gtg	cct	acc	gtg	aag	aca	tgt	gaa	cg	tgg	ttc	cag	cg	ttc	aag	agc	gga
	D	F	D	V	D	D	K	E	H	G	K	P	P	K	R	Y	E	D	A	E
[+]	gat	ttt	gac	gtc	gac	gac	aaa	gag	cac	gga	aaa	ccg	cca	aaa	agg	tac	gaa	gac	gcc	gaa
V1	gac	ttc	gac	gtg	gac	gac	aag	gag	cac	ggc	aag	ccc	ccc	aag	cg	tac	gag	gac	gcc	gaa
V2	gac	ttt	gac	gtg	gac	gac	aag	gag	cat	ggc	aag	ccc	ccc	aag	cg	tac	gag	gac	gcc	gaa
	L	Q	A	L	L	D	E	D	D	A	Q	T	Q	K	Q	L	A	E	Q	L
[+]	ctg	caa	gca	tta	ttg	gat	gaa	gac	gat	gct	caa	acc	caa	aaa	caa	ctc	gca	gag	cag	ttg
V1	ctg	cag	gcc	ctg	ctg	gac	gag	gac	gac	gcc	cag	acc	cag	aag	cag	ctg	gcc	gag	cag	ctg
V2	ctg	cag	gct	ctg	ctg	gac	gag	gac	gac	gct	cag	acc	cag	aag	cag	ctg	gca	gaa	cag	ctg
	E	V	S	Q	Q	A	V	S	N	R	L	R	E	M	G	K	I	Q	K	V
[+]	gaa	gta	agt	caa	caa	gca	gtt	tcc	aat	cg	ttg	cg	gag	atg	gga	aag	att	cag	aag	gtc
V1	gag	gtg	agc	cag	cag	gcc	gtg	agc	aac	cg	ctg	cg	gag	atg	ggc	aag	atc	cag	aag	gtg
V2	gag	gtg	agc	cag	cag	gcc	gtg	agc	aac	cg	ctg	cg	gag	atg	ggc	aag	atc	cag	aag	gtg
	G	R	W	V	P	H	E	L	N	E	R	Q	M	E	R	R	K	N	T	C
[+]	ggt	aga	tgg	gtg	cca	cat	gag	ttg	aac	gag	agg	cag	atg	gag	agg	cg	aaa	aac	aca	tgc
V1	ggc	cg	ctg	gtg	ccc	cac	gag	ctg	aac	gag	cg	cag	atg	gag	cg	cg	aac	aac	acc	tgc
V2	ggc	cg	tgg	gtg	ccc	cac	gag	ctg	aac	gag	cg	cag	atg	gag	cg	cg	aag	aac	acc	tgc
	E	I	L	L	S	R	Y	K	R	K	S	F	L	H	R	I	V	T	G	D
[+]	gaa	att	ttg	ctt	tca	cca	tac	aaa	agg	aag	tcg	ttt	ttg	cat	cg	atc	ggt	act	gga	gat
V1	gag	atc	ctg	ctg	agc	cg	tac	aag	cg	aag	agc	ttc	ctg	cac	cg	atc	ggt	acc	ggc	gac
V2	gag	atc	ctg	ctg	agc	cg	tac	aag	cg	aag	agc	ttc	ctg	cac	cg	atc	ggt	acc	ggc	gac
	E	K	W	I	F	F	V	N	P	K	R	K	K	S	Y	V	D	P	G	Q
[+]	gaa	aaa	tgg	atc	ttt	ttt	ggt	aat	cct	aaa	cg	aaa	aag	tca	tac	ggt	cct	gga	caa	
V1	gag	aag	tgg	atc	ttc	ttc	gtg	aac	ccc	aag	cg	aag	aag	agc	tac	gtg	gac	ccc	ggc	cag
V2	gag	aag	tgg	att	ttt	ttt	gtg	aac	ccc	aag	cg	aag	aag	tcc	tac	gtg	gac	cca	gga	cag
	P	A	T	S	T	A	R	P	N	R	F	G	K	K	T	M	L	C	V	W
[+]	ccg	gcc	aca	tcg	act	gct	cca	ccc	aat	cg	ttt	ggc	aag	aag	acc	atg	ctg	tgt	ggt	
V1	ccc	gcc	acc	agc	acc	gcc	cg	ccc	aac	cg	ttc	ggc	aag	aag	acc	atg	ctg	tgc	gtg	
V2	cct	gct	aca	tcc	aca	gct	cg	cct	aac	cg	ttt	gga	aag	aag	acc	atg	ctg	tgc	gtg	
	W	D	Q	S	G	V	I	Y	Y	E	L	L	K	P	G	E	T	V	N	T
[+]	tgg	gat	cag	agc	ggt	gtc	att	tac	tat	gag	ctc	ttg	aaa	ccc	ggc	gaa	acc	gtg	aat	
V1	tgg	gac	cag	agc	ggc	gtg	atc	tac	tac	gag	ctg	ctg	aag	ccc	ggc	gag	acc	gtg	aac	
V2	tgg	gac	cag	tcc	gga	gtg	atc	tac	tac	gag	ttg	ttg	aag	ccc	gga	gag	acc	gtg	aac	
	A	R	Y	Q	Q	Q	L	I	N	L	N	R	A	L	Q	R	K	R	P	E
[+]	gca	cg	tac	caa	caa	caa	ttg	atc	aat	ttg	aac	cg	ggc	ctt	cag	aga	aaa	cca	ccc	
V1	gac	cg	tac	cag	cag	cag	ctg	atc	aac	ctg	aac	cg	gcc	ctg	cag	cg	aag	cg	ccc	
V2	gcc	cg	tac	cag	cag	cag	ctg	att	aac	ctg	aac	cg	gcc	ctg	cag	cg	aag	cg	ccc	
	Y	Q	K	R	Q	H	R	V	I	F	L	H	D	N	A	P	S	H	T	A
[+]	tat	caa	aaa	aga	caa	cac	agg	gtc	att	ttt	ctc	cat	gac	aac	gct	cca	tca	cat	acc	
V1	tac	cag	aag	cg	cag	cac	cg	gtg	atc	ttc	ctg	cac	gac	aac	gcc	ccc	agc	cac	acc	
V2	tac	cag	aag	cg	cag	cat	cg	gtg	atc	ttc	ctg	cac	gac	aat	gcc	cct	agc	cac	acc	
	R	A	V	R	D	T	L	E	T	L	N	W	E	V	L	P	H	A	A	Y
[+]	aga	ggc	ggt	cg	gac	acc	ttg	gaa	aca	ctc	aat	tgg	gaa	atg	ctt	ccg	cat	ggc	gct	
V1	cg	gcc	gtg	cg	gac	acc	ctg	gag	acc	ctg	aac	tgg	gag	gtg	ctg	ccc	cac	gcc	tac	
V2	cg	gcc	gtg	cg	gac	acc	ctg	gag	acc	ctg	aac	tgg	gag	gtg	ctg	ccc	cac	gcc	tac	
	S	P	D	L	A	P	S	D	Y	H	L	F	A	S	M	G	H	A	L	A
[+]	tca	cca	gac	ctg	gcc	cca	tcc	gat	tac	cac	cca	ttc	gct	tcg	atg	gga	cac	gca	ctc	
V1	agc	ccc	gac	ctg	gcc	ccc	agc	gac	tac	cac	ctg	ttc	gcc	agc	atg	ggc	cac	gcc	ctg	
V2	agc	ccc	gac	ctg	gcc	ccc	agc	gac	tac	cac	ctg	ttc	gcc	tcc	atg	gga	cac	gcc	ctg	
	E	Q	R	F	D	S	Y	E	S	V	K	K	W	L	D	E	W	F	A	A
[+]	gag	cag	cg	ttc	gat	tct	tac	gaa	agt	gtg	aaa	aaa	tgg	ctc	gat	gaa	tgg	ttc	gca	
V1	gag	cag	cg	ttc	gat	agc	tac	gag	agc	gtg	aag	aag	tgg	ctg	gag	cag	tgg	ttc	gca	
V2	gag	cag	cg	ttt	gat	tcc	tac	gag	tcc	gtg	aag	aag	tgg	ctg	gag	tgg	ttt	gca	gct	
	K	D	D	E	F	Y	W	R	G	I	H	K	L	P	E	R	W	E	K	C
[+]	aaa	gac	gat	gag	ttc	tac	tgg	gg	atc	cac	aaa	ttg	ccc	gag	aga	tgg	gaa	aaa	tgt	
V1	aag	gac	gac	gag	ttc	tac	tgg	gg	atc	cac	aag	ctg	ccc	gag	cg	tgg	gag	aag	tgc	
V2	aag	gac	gac	gag	ttc	tac	tgg	gg	atc	cac	aag	ctg	ccc	gag	cg	tgg	gag	aaa	tgc	
	V	A	S	D	G	K	Y	F	E											
[+]	gta	gct	agc	gac	ggc	aaa	tac	ttt	gaa	taa	atgattttttctttttccacaaaatttaacgtgttttt									
V1	gtg	gcc	agc	gac	ggc	aag	tac	ttc	gag	taa	atgattttttctttttccacaaaatttaacgtgttttt									
V2	gtg	gcc	tcc	gac	ggc	aag	tac	ttc	gag	taa	atgattttttctttttccacaaaatttaacgtgttttt									

Figure 1. Nucleic acid sequence alignment of the natural MOS1 ORF (lane [+]), the first version of MOS1 ORF, MOS1V1, in which is included a Kozak box and optimal codons for an expression in human cells, (lane V1), and the second version of MOS1 ORF, MOS1V2, that corresponds to a version of MOS1V1 in which putative secondary structures have been removed (lane V2). The amino acid sequence of MOS1 is shown above sequence alignment. Codons that are rare in the human codon usage are highlighted in gray in the MOS1 ORF. The Kozak box at the 5' end of V1, and the stop codons at the 3' ends are boxed. Sequence regions in lower-case letters at both sequence ends correspond to *Mos1* UTR regions. Putative ARE and ARE-like motifs are underlined. Numbers in italics in the right margin indicate the amino acid positions. Those in roman numerals correspond to the nucleic acid positions in the original *Mos1* transposon sequence (Acc N°: X78906).

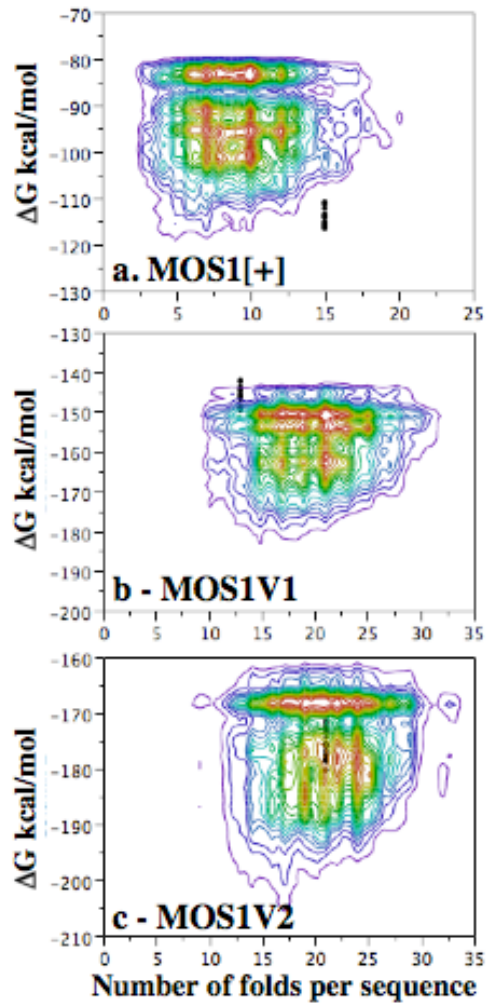


Figure 2. Graphic analysis of the statistical significance of the secondary structures calculated using the nucleic acid sequences shown in Figure 1 for the natural MOS1 ORF (a), and for MOS1V1 (b), and MOS1V2 (c). For each sequence, the number of folds (horizontal axis) was used to posit the entropy value (vertical axis) of each of its folds. Files containing 1000 scrambled DNA sequences similar to the nucleotide composition of each investigated sequence were first used to calculate the folds, using Mfold, and then the results were used to calculate the nonparametric bivariate density ellipse. 21949, 37637 and 36517 spots were obtained to calculate the ellipses in a, b and c respectively. These ellipses depicted the lack of significance of the secondary structures obtained with non-structured sequences in concentric distribution spaces ranging from 100% (red) to 0% (blue), in steps of 5%. Each of these spaces was delineated by thin or thick colored lines. Since these spaces were calculated from non-structured sequences, they also described probability spaces that can be used to locate structured or non-structured folds. The space describing an absence of significance is delineated in each graphic with a threshold of 90% by the third ellipse in blue from the outside. The folds obtained with each MLE sequence are indicated by small black squares.

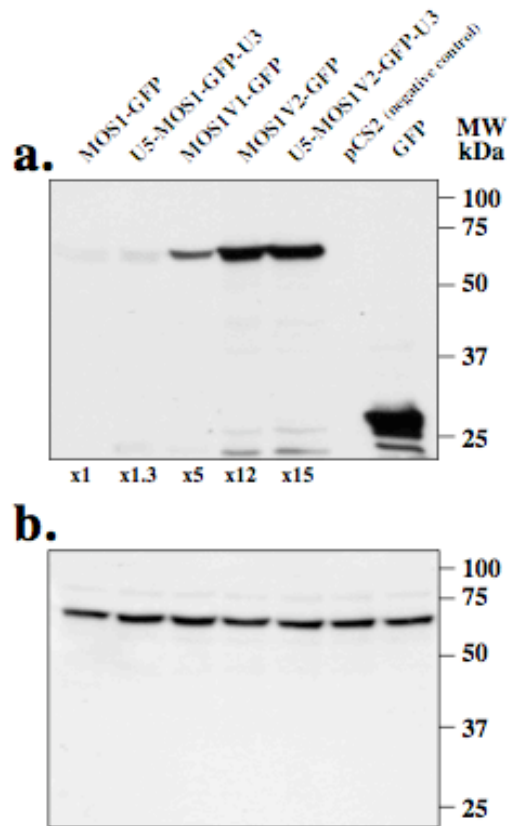


Figure 3. Western blot analysis of GFP fusion expressions in HeLa cells. GFP fusions were revealed by first hybridizing a rabbit polyclonal anti-GFP (1 : 2000) in (a). Then the filters were incubated with horseradish peroxidase-conjugated anti-rabbit IgG, followed by development using enhanced chemiluminescence. Values below the pictures indicate the increase in expression of each GFP fusion compared to MOS1-GFP. (b) shows an expression control with the housekeeping protein encoded by the MEN1 gene¹⁰ that was revealed by hybridizing a rabbit polyclonal anti-MEN1, and then incubating with horseradish peroxidase-conjugated anti-rabbit IgG, followed by development using enhanced chemiluminescence.

References

1. Cherepanov, P., Pluymers, W., Claeys, A., Proost, P., De Clercq, E. & Debyser Z. (2000) High-level expression of active HIV-1 integrase from a synthetic gene in human cells. *FASEB J.* **14**, 1389-99.
2. Guhaniyogi, J. & Brewer, G. (2001) Regulation of mRNA stability in mammalian cells. *Gene*, **265**, 11-23.
3. Yant, S.R., Huang, Y., Akache, B. & Kay, M.A. (2007) Site-directed transposon integration in human cells. *Nucl. Acids Res.* **35**:e50.
4. Rouleux-Bonnin, F., Petit, A., Demattei, M.V., and Bigot, Y. (2005) Evolution of full-length and deleted forms of the *mariner*-like element, *Botmar1*, in the Genome of the bumble bee, *Bombus terrestris* (Hymenoptera: Apidae). *J. Mol. Evol.* **60**, 736-747.
5. Petit, A., Rouleux-Bonnin, F., Lambelé, M., Pollet, N. & Bigot, Y. (2007) Properties of the various *Botmar1* transcripts in imagoes of the bumble bee, *Bombus terrestris* (Hymenoptera: Apidae). *Gene*, **390**, 52-66.
6. Keravala, A., Liu, D., Lechman, E.R., Wolfe, D., Nash, J.A., Lampe, D.J. & Robbins PD. (2006) Hyperactive *Himar1* transposase mediates transposition in cell culture and enhances gene expression *in vivo*. *Hum. Gene Ther.* **17**, 1006-1018.
7. Wilber, A., Frandsen, J.L., Geurts, J.L., Largaespada, D.A., Hackett, P.B., & McIvar, R.S. (2006) RNA as a source of transposase for Sleeping Beauty-mediated gene insertion and expression in somatic cells and tissues. *Mol. Ther.* **13**, 625-630
8. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.* **31**, 3406-3415.
9. Augé-Gouillou, C., Hamelin, M.H., Demattei MV, Periquet G, & Bigot, Y. (2001) The ITR binding domain of the Mariner Mos-1 transposase. *Mol. Genet. Gen.* **265**, 58-65.
10. Wautot, V., Khodaei, S., Frappart, L., Buisson, N., Baro, E., Lenoir, G.M., Calender, A., Zhang, C.X. & Weber, G. (2006) Expression analysis of endogenous menin, the product of the multiple endocrine neoplasia type 1 gene, in cell lines and human tissues. *Int. J. Cancer*, **85**, 877-881.