Transposition and targeting of the prokaryotic mobile element IS30 in zebrafish

Mónika Szabóa,1, Ferenc Müllera,b,c,1, János Kissa, Carolin Balduf, Uwe Strähle,b,e,* and Ferenc Olasz*a,**

aEnvironmental Biosafety Research Institute, Agricultural Biotechnology Center, Szent-Györgyi Albert St. 4, H-2101 Gödöllő, Hungary
bInstitute of Toxicology and Genetics, Forschungszentrum, Karlsruhe, Postfach 3640, D-76021 Karlsruhe, Germany
cInstitut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UFLF, BP 10142, 67404 Illkirch Cedex, C.U. de Strasbourg, France

Received 2 June 2003; revised 7 July 2003; accepted 10 July 2003

First published online 30 July 2003

Edited by Giovanni Cesareni

Abstract We provide evidence that a prokaryotic insertion sequence (IS) element is active in a vertebrate system. The transposase of Escherichia coli element IS30 catalyzes both excision and integration in extrachromosomal DNA in zebrafish embryos. The transposase has a pronounced target preference, which is shown to be modified by fusing the enzyme to unrelated DNA binding proteins. Joining the transposase to the cl repressor of phage λ causes transposition primarily into the vicinity of the λ operator in E. coli, and linking to the DNA binding domain of Gli1 also directs the recombination activity of transposase near to the Gli1 binding site in zebrafish. Our results demonstrate the possibility of fusion transposases to acquire novel target specificity in both prokaryotes and eukaryotes. © 2003 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.

Key words: IS30 transposase; Gli1; λ cl repressor; Protein fusion; Zebrafish; Site-directed integration

1. Introduction

Insertion sequences (IS) are mobile DNA segments possessing the ability to insert into the genome of host organisms. The well-characterized IS30 element of Escherichia coli flanked by 26 bp terminal inverted repeats (IRs) encodes the 44.3 kDa Tpase. The Tpase interacts with the IRs and does not appear to require E. coli-specific host factors. During transposition, an intermediate, (IS30)2, composed of two abutting IS copies that are separated by 2 bp spacer is formed. The joined IRs act as a recombinogenic site and are ca. 1000 times more efficient in transposition than the single element [1,2]. The resolution of (IS30)2 during transposition results in a stable integration product. Several transposons were previously identified and isolated from fish such as sleeping beauty cI acrids Leu and Gln C-terminally to the (DBD). The ability of chimeric proteins was assessed to direct integration into the proximity of target sites of the linked DBDs in E. coli and zebrafish. Our results provide the proof-of-principle for targeting recombinations using fusion-Tpase proteins in living cells.

2. Materials and methods

2.1. Plasmids

The tester plasmid pJKI216 is described in [7]. The pCS2+-based [8] transposase producer constructs express IS30 Tpase or its derivatives fused to the 6-mer of Myc-epitope tag (MT) or a nuclear localization signal (NLS) peptide or both. The circular gfp-fg donor (Fig. 1) was generated by ligation of the purified 2.6 kb BamHI fragment of gfp-p (pMSZ198), a donor plasmid, where the gfp reporter gene was preceded by the 111 bp splice-acceptor sequence intA [9], an IS30, site and the CmR bacterial marker. The shh target plasmids carry the 11.5 kb fragment of the zebrafish sonic hedgehog (shh) locus [10]. Target plasmids shh-GOHS and shh-gli were generated by inserting the 24 bp hot spot target sequence of IS30 (GOHS) [11] or the consensus Gli binding site (GACCCACCA, [12]) into the StuI site of shh, respectively. The Tpase-cl chimeric protein was expressed from a pCS2+-based KmR vector containing the ORF-A of IS30 linked to the amino acids Leu and Gln C-terminally to the cl repressor gene of λ tss57 under the control of the lac promoter. The expression plasmid also harbors the IS30 structure and the CmR gene. In control experiments, wild-type (wt) Tpase was expressed from a similar construct without the cl repressor gene. Target plasmids were pEMBL19 and its derivatives containing GOHS or the 200 bp λ operator Oκ (κ coordinates: 37946-38107 bp, EM, PH-M17233). The Tpase-GliDBD fusion protein (expressed from a pCS2+-derived plasmid) contained the IS30 Tpase linked to the N-terminus of the 256-417 aa region of the human Gli protein [12] via Glu and Phe.

2.2. DNA/RNA procedures and microinjection

DNA techniques were performed according to [13]. Sequences of polymerase chain reaction (PCR) primers are available upon request. E. coli strains TG2 and TOP10 Electrocomp® (Invitrogen) were used for cloning and electroporation of DNA isolated from fish embryos. Tpase mRNA variants were synthesized in vitro by an Ambion
mMessage mMachine kit. Microinjection of 1 or 2 cell stage eggs was carried out as described [14]. Tpase mRNA was used in 100 ng/wl concentration in combinations with circular gfp-p or gfp-fg donor and shh targets (200 ng/wl each). Zebrafish were kept and embryos generated according to [15]. Injected embryos were grown to 1-3 somites stage and harvested for DNA preparation. 200 embryos were digested in SET buffer overnight at 55°C with 40 μg/ml proteinase K and DNA was prepared by phenol extraction [14].

3. Results and discussion

3.1. Trans-kingdom activity of IS30 in zebrafish embryos and human cells

First, IS30 Tpase-mediated excision activity was examined in the vertebrate system, zebrafish embryo. The experimental strategy included the microinjection of a two-component transposition system developed previously for E. coli [7] into fish embryos. The Tpase was expressed from a ‘producer plasmid’ or synthetic mRNA in the presence of the tester plasmid pJK1216 (Fig. 1A). Total DNA extracted from pooled 10 hpf embryos was introduced into E. coli and the in trans activity of the Tpase was monitored by recovery and analysis of the tester plasmid population. CmR KmR colonies were determined by replica-plating and plasmid DNA of some colonies were sequenced to assess the correct excision of the CmR gene. Activity of the Tpase was characterized by measuring the fraction of CmS clones among Km R transformants. The CmS frequency was 4-28-fold higher in samples exposed to Tpase than in controls (tester injected alone), showing that the prokaryotic Tpase is functional in zebrafish (Table 1). Similar frequencies were obtained using Tpase producer plasmid or Tpase mRNA. Joining of the MT or a NLS to IS30 Tpase did not significantly affect the transposition frequency indicating that the Tpase can be fused to other peptides without losing its activity.

![Fig. 1. Schematic representation of the experimental design used to detect transposition. A: The excision reaction generated by IS30 Tpase in the tester plasmid. B: The experimental design used to detect insertions in shh target plasmids.](FEBS 27526 13-8-03)

Table 1
Transpositional excision mediated by IS30 Tpase in zebrafish and Hela cells

<table>
<thead>
<tr>
<th>Components injected</th>
<th>Total KmR colonies tested</th>
<th>Number of KmR/CmS colonies</th>
<th>Correct excisions</th>
<th>Other recombinations</th>
<th>Correct excisions (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zebrafish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tester (pJK1216)</td>
<td>1949</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>tester+Tpase producer</td>
<td>1680</td>
<td>15</td>
<td>12</td>
<td>3</td>
<td>0.71</td>
</tr>
<tr>
<td>tester+NLS-Tpase producer</td>
<td>1718</td>
<td>14</td>
<td>10</td>
<td>4</td>
<td>0.58</td>
</tr>
<tr>
<td>tester+NLS-MT-Tpase producer</td>
<td>516</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0.78</td>
</tr>
<tr>
<td>tester</td>
<td>1848</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>tester+Tpase mRNA</td>
<td>907</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0.22</td>
</tr>
<tr>
<td>tester+NLS-Tpase mRNA</td>
<td>787</td>
<td>13</td>
<td>11</td>
<td>2</td>
<td>1.40</td>
</tr>
<tr>
<td>tester+MT-Tpase mRNA</td>
<td>567</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>1.41</td>
</tr>
<tr>
<td>tester+NLS-MT-Tpase mRNA</td>
<td>282</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1.06</td>
</tr>
<tr>
<td>Hela cells</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tester (pJK1216)</td>
<td>3715</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>&lt;0.026</td>
</tr>
<tr>
<td>tester+Tpase producer</td>
<td>1722</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0.12</td>
</tr>
</tbody>
</table>

*aThis excision event may represent the background Tpase activity in E. coli [7].
Capacity of Tpase for promoting the excision in human cells was also assessed with the same system. Hela cells were transfected with the tester plasmid alone or along with the Tpase producer plasmid. DNA from transfected cells was transformed into E. coli and subjected to antibiotic selection and sequencing. Correct excision of the CmR gene could not be detected (<0.026%) with the tester plasmid alone, while 0.12% of plasmids were CmS when the Tpase producer was co-transfected (Table 1). Sequencing the plasmids recovered from the CmS colonies confirmed that both recombination events detected were correct transpositional deletions (data not shown). Excisions catalyzed by IS30 Tpase require the interaction of two recombinogenic (IS30)2 structures (Fig. 1A).

Conversely, insertion of a donor DNA into a recipient target involves only one (IS30)2 structure and a target (hot spot) sequence in the recipient DNA (Fig. 1B). To test the ability of Tpase to promote insertions in fish embryos, a three-component gene trap system was constructed. The system consists of an (IS30)2 donor, a ‘hot spot’ target and the IS30 Tpase. The donor DNA contains the gfp reporter gene preceded by the splice-acceptor sequence intA and the (IS30)2 recombination site. The promotorless gfp gene is expected to be inactive unless integration of this construct occurs into an intronic sequence of a transcribed gene. Donor DNA was there-circularized gfp-fg fragment containing each functional part for gene trapping (Fig. 1B, see Section 2) but lacking the plasmid backbone. In the target plasmid shh-GOHS, the artificial IS30 hot spot sequence GOHS generated according to the consensus of E. coli genomic IS30 integration sites [11] was inserted into the first intron of the sonic hedgehog (shh) gene of zebrasfish. This construct contained also the transcriptional regulatory elements driving expression of shh in the notochord and the ventral neural tube of zebrasfish embryos [16]. When correctly oriented, the integration of gfp-fg donor into GOHS will result in gfp expression in the midline of neurula and organogenesis stage embryos.

To assess whether a sense shh-gfp fusion can express gfp faithfully, a fusion plasmid generated by transposition in E. coli was injected into embryos and gfp expression was analyzed at 24 hpf. Mosaic green fluorescent protein (GFP) activity restricted to several cells per embryo due to uneven segregation of the foreign DNA [17]. Expression of gfp in notochord cells was observed in eight out of 30 embryos (26 notochord cells out of 65 gfp-positive cells), indicating that gfp was expressed under the control of regulatory elements of the shh gene (Fig. 2B). Non-specific expression was also detected in ectodermal and other cell types, which is explained by cryptic non-specific regulatory elements in the carp β-actin intronic sequences adjacent to the splice-acceptor site (Fig. 2C).

To detect integration events in zebrasfish, the gfp-fg donor and the shh-GOHS target DNA were injected into embryos with or without synthetic mRNA encoding the Tpase. Notochord cells expressing gfp were detected in four out of 130 embryos at 24 hpf (Fig. 2D), while none of the embryos (0/195 embryos) showed notochord-specific gfp expression when Tpase mRNA was omitted (Fig. 2E) or when the gfp-fg was injected alone (data not shown). This suggests that the inser-

---

Fig. 2. Insertion of a gfp donor into the shh target induced by IS30 Tpase results in tissue-specific expression of the gfp reporter gene in zebrasfish embryos. A–E: Detection of GFP expression in one-day-old zebrasfish embryos. A: Tail region of a control non-injected embryo. B: Detection of GFP expression in notochord cells of embryos injected with the control shh-gfp fusion construct. C: Ectopic GFP activity in cells where shh is not expressed (epidermis). D: Embryo injected with gfp-fg or gfp-p (see Section 2) donors and shh-GOHS target along with Tpase mRNA expresses GFP in notochord cells suggesting regulation of the gfp gene by the shh locus. E: Embryo from the same injection showing non-specific activity in epidermal cells. Abbreviations: n – notochord, ye – yolk extension. Embryos are oriented anterior left. Lateral views of the tail region are shown. F: PCR analysis demonstrating the presence of junction fragments between microinjected gfp-fg donor and the shh-GOHS target plasmid. The experimental setup is presented over the lanes. Lanes 6 and 7 show the verification of PCR products from lanes 4 and 5 by EcoRI digestion. G: PCR reaction to detect the junction fragments produced by integration of gfp-fg donor into shh-GOHS target in sense and reverse orientations. Abbreviations: E – EcoRI digested PCR product, Mw – molecular weight marker, 5′ and 3′ – junction fragments generated at the 5′ or the 3′ end of inserted donor, respectively.
tion into shh-GOHS target was induced by the Tpase. To verify this conclusion, total DNA was prepared from microinjected embryos pooled together and the junction fragments of shh-gfp fusion were PCR amplified (Fig. 1B). Correct integration of the gfp cassette into the shh-GOHS target was obtained only when Tpase mRNA was coinjected (Fig. 2F). The junction fragments were confirmed by restriction analysis and sequencing (Fig. 2G). As expected, insertions occurred in both orientations. Frequency of integration was estimated by transformation of DNA isolated from microinjected embryos into E. coli. The frequency of shh-gfp fusion plasmids was determined as a ratio of Cm\(^R\)/Cm\(^S\)Ap\(^S\) transformants, which was \(5.1 \times 10^{-5}\) and \(5.8 \times 10^{-5}\) in two independent experiments, respectively. The correct structure of shh-gfp fusions was verified by sequencing. This is the first demonstration that a prokaryotic IS element is functional in a vertebrate system. The activity of IS30 Tpase in fish cells may indicate a theoretical possibility for horizontal gene transfer via transposon-mediated recombination between prokaryotic and eukaryotic genomes.

3.2. Modification of targeting specificity of the IS30 Tpase

We tested in E. coli whether the target specificity of IS30 Tpase can be altered by fusing it to another DNA binding protein. The Tpase was attached to the cI repressor of bacteriophage \(\lambda\). The activity of the chimeric protein was examined in a two-component transposition system consisting of an (IS30), donor plasmid that also expressed the Tpase-cI fusion protein and a target plasmid carrying the \(\lambda\) operator OR (Fig. 3A). The chimeric Tpase retained the activities of both domains as suggested by the immunity of E. coli. The chimeric Tpase fusion protein against \(\lambda\) infection, and by the similar activity as compared to wt Tpase in promoting insertions into the GOHS hot spot (2.5 \(\times\) \(10^{-2}\) and 2.0 \(\times\) \(10^{-2}\) for wt and chimera Tpase, respectively). Next, the capability of the chimeric Tpase to direct the insertions into the \(\lambda\) operator OR was investigated. The overall frequency of transposition was \(1.5 \times 10^{-4}\). Among 53 transposition products, 30 carried the inserted donor plasmid in close proximity (within 400 bp) of OR and one insertion occurred directly in the OR (Fig. 3B). Additional 22 insertions were farther from the ORs. These insertions depended also on the cI repressor domain in the chimeras as no integration into the OR target plasmid was scored when the wt Tpase was used (<2.5 \(\times\) \(10^{-5}\)). Moreover, deletion of OR from the target plasmid abolished its targeting by the Tpase-cl chimera (<1.5 \(\times\) \(10^{-5}\)). Seventeen sequenced insertion sites showed no or limited homology to either the \(\lambda\) operator or to the GOHS sequence (Fig. 3C) suggesting that the chimera has acquired a novel target specificity. Taken together, these results demonstrate that fusion of the cI repressor changes the target selection of IS30 Tpase and increases the efficiency of integration into target sites not preferred by the wt Tpase by more than 10-fold in E. coli.

We finally assessed whether the target specificity of Tpase can also be directed to new sites in zebrafish embryos by fusion with a vertebrate DBD. The three-component strategy described in Fig. 1B was applied with modifications. The DBD of the transcription factor Gli1 [12] was fused to the C-terminus of IS30 Tpase (Tpase-GliDBD). We employed shh-gli plasmid as target, which carries a Gli binding site (GACCACCCA) inserted into the first intron of the shh locus. Synthetic mRNA of Tpase-GliDBD, gfp-fg donor and the shh-gli target DNA were coinjected into 1-cell-stage zebrafish embryos. In controls, the injections were performed without Tpase mRNA or with the shh target lacking the Gli1 binding site. Transposition by Tpase-GliDBD was first analyzed by monitoring gfp expression in injected embryos. Gfp expression was detected only when all three components of the transposition system were coinjected (10 notochord cells expressed in three out of 104 embryos). No expression in the notochord was seen when IS30-GliDBD mRNA and gfp-fg donor were coinjected with a shh target without the Gli1 binding site (0/102 embryos) or when the gfp-fg donor was coinjected with shh-gli target in the absence of IS30-GliDBD (0/178 embryos). These results indicate that the Tpase-GliDBD fusion protein is able to direct insertions in the sense orientation into the shh target in a Gli1 binding site-dependent manner. To analyze the transposition events further, total DNA was extracted from gastrula-stage embryos and integration of the gfp cassette was monitored by nested PCR using primers that detect insertion of gfp in the sense orientation within 675 bp upstream and up to 2.5 kb downstream of the Gli1 binding site (Fig. 4). PCR-amplified putative junction fragments were subcloned and sequenced. Fourteen fragments representing 12 different integration events were identified out of which eight
are shown in Fig. 4. In two cases the same fragment was identified twice (e.g. clone #13 and #25). The resolution of (IS30): indicative of a legitimate transposition event was detected only in one case (clone #27). This integration occurred 36 bp away from the 3' end of the Gli1 binding site. In total, six integrations were located within 100 bases adjacent to the Gli1 binding site. In most integration, however, the (IS30): site was intact, suggesting that the insertions took place by illegitimate recombination. No junction fragments could be amplified when the Gli1 binding site was missing in the target plasmid. These results suggest that the level of both transposition may be generally achieved in any DNA target sequences but increased also the frequency of integration to a heterologous DBD did not only alter the range of development is to improve the frequency of recombination. Fusion to a heterologous DBD did not only alter the range of target sequences but increased also the frequency of integration in both E. coli and zebrafish. This suggests that the frequency may simply be increased by more efficient tethering the transposase to DNA. Fusion of transposases with general transcription factors such as the TATA box binding factors or with DNA methylases may also improve the efficiency of integration into a vast array of potential integration sites.

**Fig. 4. Directed integration in zebrafish embryos.** Analysis of the junction fragments of the gfp-fg donor and the shh-gli target plasmid detected by PCR. Thin and thick lines represent donor and target sequences, respectively. The light gray box indicates the Gli1 binding site. Squares indicate integration sites detected by PCR. In integrations within 100 bases nearby the Gli1 binding site are highlighted in dark gray. Numbers above the junction fragments show the position of integration of IS-end-containing donor sequence in the shh-gli target DNA. Black arrows indicate the annealing positions of S3, S5 and IRL oligonucleotides used in PCR reactions. Note that this strategy detects insertion in the sense orientation only. Other symbols are the same as in Fig. 1.

**References**