Gene 593 (2016) 21–27

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Identification of nuclear localization signal within goldfish *Tgf2* transposase

Xiao-Dan Shen, Fei Hou, Jie Chen, Xia-Yun Jiang *, Shu-Ming Zou *

Key Laboratory of Genetic Resources for Freshwater Aquaculture and Fisheries, Shanghai Ocean University, 999 Huchenghuan Road, Shanghai 201306, China

ARTICLE INFO

Article history:
Received 30 May 2016
Received in revised form 12 July 2016
Accepted 24 July 2016
Available online 26 July 2016

Keywords:
Tgf2 transposase
Nuclear localization signal
DNA transposon
Cell culture

ABSTRACT

The structure of goldfish (*Carassius auratus*) *Tgf2* transposase is still poorly understood, although it can mediate efficient gene transfer in teleost fish. We hypothesized the existence of a nuclear localization signal (NLS) within *Tgf2* transposase to assist transport into the nucleus. To explore this, 15 consecutive amino acid residues (656–670 aa) within the C-terminus of *Tgf2* transposase were predicted in silico to be a NLS domain. The pEGFP-C1-Tgf2TP$^{\Delta 31C}$ plasmid encoding the NLS-domain-deleted *Tgf2* transposase fused to EGFP was constructed, and transfected into 293T cells. After transfection with pEGFP-C1-Tgf2TP$^{\Delta 31C}$, EGFP was not detected in the nucleus alone, while 67.0% of cells expressed EGFP only in the cytoplasm. In contrast, after transfection with control plasmids containing C- or N-terminal truncated *Tgf2* transposases with an intact NLS domain, EGFP was not detected in the cytoplasm alone, while approximately 40% of cells expressed EGFP only in the nucleus, and the remaining 60% expressed EGFP in both the nucleus and cytoplasm. Our results demonstrated that loss of the NLS domain results in expression in the cytoplasm but not in the nucleus. These findings suggest that 15 aa residues located from 656 to 670 aa within the C-terminus of *Tgf2* transposase can function as a NLS to assist the transfer of the transposase into the nucleus where it mediates DNA transposition.

1. Introduction

Transposable elements or transposons are discrete DNA segments that move between different, non-homologous genomic loci using a "cut-and-paste" mechanism (Zagoraiou et al., 2001; Hickman and Dyda, 2015; Atkinson, 2015). Two essential sequence elements participate in transposition; the flanking terminal inverted repeat sequences involved in specific recognition (Kunze and Starlinger, 1989; Mack and Crawford, 2001) and the encoded transposases responsible for catalyzing the DNA breakage and rejoining reactions (Kaufman and Rio, 1992; Zhou et al., 2004; Hickman et al., 2005). Previous studies have shown that DNA transposons from many different gene families are functional in diverse species; such studies are important in the development and application of DNA transposon tools in gene discovery and gene delivery (Kawakami et al., 2004; Kawakami, 2007; Kotani and Kawakami, 2008; Kong et al., 2010).

The goldfish (*Carassius auratus*) *Tgf2* transposon is a vertebrate DNA transposon that belongs to the *hAT* transposon family, which has a high bootstrap value of 96% with medaka *Tol2* transposon (Zou et al., 2010; Cheng et al., 2014). The distribution of *Tgf2*-like elements in the distantly divergent host species of goldfish and medaka indicates the taxonomic distribution of *Tgf2* in goldfish is due to horizontal transfer but not vertical inheritance (Jiang et al., 2012). In contrast, *Tgf2*-like elements are undetectable in a wide range of other species in genome sequence databases, including the fish species fugu, *Tetraodon*, stickleback, zebrafish and common carp (Jiang et al., 2012). Although active elements have been identified in vertebrates, including *Tol1* and *Tol2* from the medaka fish (Koga et al., 1995; Kawakami et al., 1998; Kodama et al., 2008), most vertebrate transposases lose their activity due to the accumulation of gene mutations in the coding region (Janicki, et al., 2011). Although the *Tgf2* transposase is active in teleost fish and has been used to generate effective genetic tools (Xu et al., 2015), it is unclear whether there is a nuclear localization signal (NLS) in *Tgf2* transposase.

Proteins with molecular weights >50 kDa usually require the interaction between a specific NLS and a transport protein that mediates transfer into the nucleus (Allen et al., 2000; Xu et al., 2015). After translation in the cytoplasm, the *Tgf2* transposase (approximately 80 kDa) is transported into the nucleus via the nuclear pore complex (NPC) (Feldherr et al., 1984). Some transposases are known to contain NLSs (Nair et al., 2003), such as *piggyBac* in the cabbage looper moth (*Trichoplusia ni*) (Keith et al., 2008), and *Hermes* in the housefly (*Musca domestica*) (Michel and Atkinson, 2003). In this study, we investigated the existence and location of NLSs within the *Tgf2* transposase.

http://dx.doi.org/10.1016/j.gene.2016.07.060
0378-1119/

2. Materials and methods

2.1. Sequence analysis and modeling of Tgf2 transposase

Functional domains of the goldfish full-length *Tgf2* transposase (686 amino acids) were predicted using Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/) (Kelley and Sternberg, 2009). A three-dimensional model of the *Tgf2* transposase monomer was generated using Phyre2 and protein structures were visualized using PyMol (www.pymol.org), based on a model of the *Hermes* transposase protein (Hickman et al., 2014). The NLS was predicted using cNLS mapper (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi).

2.2. Plasmid constructs

Plasmid pEGFP-C1, which contains EGFP driven from a CMV promoter, was obtained from Clontech (GenBank Accession No.: U55763). Fragments for the generation of *Tgf2* transposase plasmid constructs were amplified by PCR from the pET-28a-Tgf2TPase template (Xu et al., 2015) using the following primer sets: full-length *Tgf2* transposase (Tgf2TP) open reading frame (ORF), 5′-CCGCTCGAGATGTTCATTGGTCCTTTGGAAG-3′ and 5′-CGCGGATCCCTCAAAGTTGTAAAACCTCA-3′; N-terminal 120-aa deletion *Tgf2* transposase ($Tgf2TP^{\Delta 120N}$) ORF, 5′-CCGCTCGAGATGCACCCAAATTACCTCAA-3′ and 5′-CGCGGATCCCTCAAAGTTGTAAAACCTCA-3′; C-terminal 31-aa deletion *Tgf2* transposase ($Tgf2TP^{\Delta 31C}$) ORF, 5′-CCGCTCGAGATGTTCATTGGTCCTTTGGAAG-3′ and 5′-CGCGGATCCTCCTGCAGTGCTGAAAAGCC-3′; C-terminal 16-aa deletion *Tgf2* transposase ($Tgf2TP^{\Delta 16C}$) ORF, 5′-CCGCTCGAGATGTTCATTGGTCCTTTGGAAG-3′ and 5′-CGCGGATCCAAAATTGTTAGTGTCAAGCCT-3′.

The ORFs of full-length or deleted *Tgf2* transposase cDNA were amplified by PCR using *Pfu* DNA polymerase (Stratagene, La Jolla, CA, USA), digested with *Xho*I and *Bam*HI, and subcloned into the *Xho*I-*Bam*HI sites of the pEGFP-C1 vector. All plasmids were validated by DNA sequencing.

2.3. Cell transfection

293T cells were cultured in medium (M199) supplemented with 10% fetal bovine serum (FBS) and 1% antibiotics, at 37 °C in a humidified atmosphere containing 5.0% CO_2. Transfections were conducted using ViaFect Transfection Reagent (Promega, Madison, WI, USA). Cells were plated into 24-well plates ($0.5–2 \times 10^5$ cells/well) 24 h prior to transfection. For each transfection, 3 μl of TransFast™ Reagent was incubated for 5 min in 50 μl serum-free medium before the addition of 0.8 μg of plasmid DNA in 50 μl serum-free medium (total volume of 100 μl). The TransFast™ Reagent/DNA mixture was incubated at room temperature for 20 min and added directly to the cells in a drop-wise manner before agitation to mix. After incubation for 48 h to allow DNA uptake and gene expression, the cells were washed twice with phosphate-buffered saline (PBS), fixed with 4% paraformaldehyde (PFA) and then stained for 2–3 min with 800 μl Hoechst 33342 stain. Each experiment was repeated in triplicate.

2.4. Statistics

Data are expressed as mean ± S.E. Differences among groups were analyzed by one-way ANOVA followed by Fisher's post hoc tests or unpaired *t*-tests. $P < 0.001$ was considered to indicate statistical significance.

3. Results

3.1. Prediction of NLS of Tgf2 transposase

Amino acid sequence analysis based on Phyre2 prediction suggested that *Tgf2* transposase contained a N-terminal zinc finger BED domain (65–120 aa), a helix-turn-helix (HTH) binding structure (163–201 aa) and a RNase-H domain (211–683 aa) with an insertion domain (362–493 aa) (Fig. 1A). The cNLS mapper predicted the presence of a monopartite NLS (656–670 aa) of 15 amino acids (LLFSPKRARLDTNNF) within the RNase-H domain at the C-terminus of *Tgf2* transposase (Fig. 1A). This predicted NLS is located downstream of the DDE residues (D_{228}, D_{295} and E_{648}). A 3D model of the NLS and DDE signature of *Tgf2* transposase was constructed on the basis of fold recognition using PyMol (Fig. 1B). Analysis of the *Hermes* transposon cocrystal by Hickman et al. (2014) revealed that the conserved DDE residues are critical for *Hermes* transposase activity and/or DNA-binding. These conserved residues or motifs have been exploited as phylogenetic characteristics to infer evolutionary relationships among *hAT* transposases (Wicker et al., 2007). The NLS of *Tgf2* transposase is identical to that of the Tol2 transposase (Fig. 2), indicating the evolutionary importance of these sequences for NLS function (Michel et al., 2002; Yuan and Wessler, 2011).

3.2. Plasmid constructs with different deletion regions of Tgf2 transposase

To investigate the accuracy of the predicted *Tgf2* transposase NLS, we constructed the plasmid pEGFP-C1-$Tgf2TP^{\Delta 31C}$ containing a 31-aa C-terminal deletion of *Tgf2* transposase that included the predicted 15-amino acid NLS (Fig. 3). We also constructed pEGFP-C1-Tgf2TP containing the full-length *Tgf2* transposase, pEGFP-C1-$Tgf2TP^{\Delta 120N}$ containing a 120-aa N-terminal deletion of *Tgf2* transposase and pEGFP-C1-$Tgf2TP^{\Delta 16C}$ containing a 16-aa C-terminal deletion of *Tgf2* transposase (Fig. 3); these plasmids included the predicted 15-aa NLS and they were used as controls. All expression constructs were based on pEGFP-C1, which contains EGFP driven from a CMV promoter and a poly(A) sequence.

3.3. Tgf2 transposase with the NLS can be transported into the cell nucleus

When 293T cells were transfected with plasmid pEGFP-C1, the 27 kDa EGFP control was transported in and out of the nucleus by passive diffusion, which yielded an evenly dispersed fluorescence in both the cytoplasmic and nuclear compartments (Fig. 4A; Fig. 5). The pEGFP-C1-Tgf2TP plasmid expressed a 104 kDa fusion containing the full-length *Tgf2* transposase, which exceeds the molecular weight threshold into the nucleus for passive diffusion of proteins. This suggests that active nuclear transport is required for entry of the *Tgf2* transposase into the nucleus. When 293T cells were transfected with pEGFP-C1-Tgf2TP, enhanced green fluorescent protein (EGFP) expression was detected in the nucleus of 39.9% of the cells and in both the nucleus and the cytoplasm in 60.1%, while EGFP was not detected in the cytoplasm alone (Fig. 4B; Fig. 5). Similarly, when 293T cells were transfected with pEGFP-C1-$Tgf2TP^{\Delta 120N}$ or pEGFP-C1-$Tgf2TP^{\Delta 16C}$ expressing the truncated *Tgf2* transposase but including the predicted 15-aa NLS, expression of EGFP in both the nucleus and the cytoplasm was detected in 61.7% and 60.0% of cells, respectively, and in the nucleus alone in 38.3% and 39.9%, respectively (Fig. 4C, D; Fig. 5). The results showed that the *Tgf2* transposase contained a functional NLS that mediates protein transport into the nucleus.

3.4. Tgf2 transposase lacking the NLS cannot be transported into the nucleus

When 293T cells were transfected with pEGFP-C1-$Tgf2TP^{\Delta 31C}$ containing the 31-aa C-terminal deletion (including the predicted 15-aa NLS) of the *Tgf2* transposase, no EGFP fluorescence was expressed in