Gene 525 (2013) 84–91

Contents lists available at SciVerse ScienceDirect

Gene

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

Identification of multiple binding sites for the THAP domain of the *Galileo* transposase in the long terminal inverted-repeats

Mar Marzo [a,b], Danxu Liu [a], Alfredo Ruiz [b], Ronald Chalmers [a,*]

[a] School of Biomedical Sciences, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK
[b] Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

ARTICLE INFO

Article history:
Accepted 17 April 2013
Available online 4 May 2013

Keywords:
P-element
Foldback
THAP domain
DNA binding
Evolution

ABSTRACT

Galileo is a DNA transposon responsible for the generation of several chromosomal inversions in *Drosophila*. In contrast to other members of the *P-element* superfamily, it has unusually long terminal inverted-repeats (TIRs) that resemble those of *Foldback* elements. To investigate the function of the long TIRs we derived consensus and ancestral sequences for the *Galileo* transposase in three species of Drosophilids. Following gene synthesis, we expressed and purified their constituent THAP domains and tested their binding activity towards the respective *Galileo* TIRs. DNase I footprinting located the most proximal DNA binding site about 70 bp from the transposon end. Using this sequence we identified further binding sites in the tandem repeats that are found within the long TIRs. This suggests that the synaptic complex between *Galileo* ends may be a complicated structure containing higher-order multimers of the transposase. We also attempted to reconstitute *Galileo* transposition in *Drosophila* embryos but no events were detected. Thus, although the limited numbers of *Galileo* copies in each genome were sufficient to provide functional consensus sequences for the THAP domains, they do not specify a fully active transposase. Since the THAP recognition sequence is short, and will occur many times in a large genome, it seems likely that the multiple binding sites within the long, internally repetitive, TIRs of *Galileo* and other *Foldback*-like elements may provide the transposase with its binding specificity.

1. Introduction

Transposable elements (TEs) are mobile genetic components of virtually all eukaryotic species (Feschotte and Pritham, 2007; Wicker et al., 2007). These repetitive sequences make up a substantial proportion of most genomes and have a huge impact on the evolution of their hosts (Adams et al., 2000; Feschotte and Pritham, 2006; Jurka et al., 2007; Lander et al., 2001; Morgante, 2006). TEs are diverse and employ many different mechanisms for mobilization. Two major groups of transposons are recognized depending on whether they have an RNA intermediate or a DNA intermediate (Finnegan, 1989). Transposons are then further classified into numerous superfamilies and families depending on their sequence structure and similarity (Feschotte and Pritham, 2007; Jurka et al., 2007; Wicker et al., 2007).

All TE families contain autonomous and non-autonomous members. Autonomous transposons retain the ability to express the protein(s) required for their own transposition. Non-autonomous copies contain indels or point mutations that render them non-functional. The non-autonomous copies exploit the gene products of the autonomous copies, which they often outnumber (Feschotte and Pritham, 2007).

Biochemical analysis of transposition reactions helps us to understand how the elements behave in the genome, and allows the development of transposons as genetic tools. Since most of the transposon copies in higher eukaryotic genomes harbor mutations in their coding regions, different strategies have been used to reconstitute their activity. Sometimes, a simple consensus sequence constructed from different copies results in the restoration of activity e.g. *Himar1*, *Frog Prince* and *Harbinger* transposases (Kapitonov and Jurka, 2004; Lipkow et al., 2004; Miskey et al., 2003). Often, the amplification of non-autonomous transposons means that a simple consensus sequence encodes a non-functional transposase protein. In such cases the functional ancestral sequence may be reconstructed by taking account of phylogenetic information while building the consensus. For example, this approach has been used for the revival of *Hsmar1* (Miskey et al., 2007).

Abbreviations: TIR, terminal inverted repeat; bp, base pair; kb, kilobase; MBP-tag, maltose binding protein expression tag; EMSA, electrophoretic mobility shift assay; ORF, open reading frame; BS, binding site; Dbuz, *Drosophila buzzatii*; Dmoj, *Drosophila mojavensis*; Dana, *Drosophila ananassae*.

* Corresponding author. Tel.: +44 115 823 0087.
E-mail addresses: mar.marzo_llorca@nottingham.ac.uk (M. Marzo), Danxu.Liu@nottingham.ac.uk (D. Liu), Alfredo.Ruiz@uab.cat (A. Ruiz), Chalmers@nottingham.ac.uk (R. Chalmers).

0378-1119 © 2013 The Authors. Published by Elsevier B.V. Open access under CC BY license.
http://dx.doi.org/10.1016/j.gene.2013.04.050

The *P-element* was discovered in *Drosophila melanogaster* as the agent responsible for P-M hybrid dysgenesis (Kidwell, 1985; Rubin et al., 1982). It has since been studied *in vivo* and *in vitro* and is now widely used as a tool for genomic analysis of *D. melanogaster* (Rio, 2002; Ryder and Russell, 2003). The *P-element* defines a superfamily of DNA transposons, which includes *1360* and *Galileo* (see below). These elements harbor a transposase coding region flanked by TIRs, which are needed for the transposition reaction. The *P-element* transposase contains four functional domains: an N-terminal DNA binding domain, a coiled coil region presumably involved in protein–protein interactions, a GTP binding domain and a catalytic domain with four key acidic residues, which may coordinate the catalytic metal ions (Rio, 2002; Sabogal and Rio, 2010). The *P-element* catalytic domain is thought to belong to the RNase H-like superfamily of polynucleotidyl transferases (Hickman et al., 2010; Rio, 2002; Sabogal and Rio, 2010; Yuan and Wessler, 2011).

The *P-element* transposase contains a THAP domain, which is presumably involved in site-specific DNA binding. The THAP domain is an evolutionary conserved motif shared by different animal proteins, including cell-cycle regulators, pro-apoptotic factors, transcriptional repressors and chromatin-associated proteins (Clouaire et al., 2005; Quesneville et al., 2005; Roussigne et al., 2003). The domain has a long zinc finger (~90 amino acids) in which key residues are highly conserved (Roussigne et al., 2003). Crystal structures have been reported for the human THAP1 protein and the *D. melanogaster P-element* transposase (Campagne et al., 2010; Sabogal et al., 2010). These show that the THAP domain interacts with its binding sequence in a bipartite manner, through the major and minor grooves of the DNA.

The *Galileo* transposon was discovered in *Drosophila buzzatii*, where it has caused three large chromosomal inversions, which are currently segregating naturally in the population (Cáceres et al., 1999; Casals et al., 2003; Delprat et al., 2009). Although originally considered a *Foldback*-like element, it was later included in the *P-element* superfamily of cut-and-paste transposons based on the sequence of the putative transposase (Marzo et al., 2008). *Galileo* is probably widespread within the *Drosophila* genus because it has been found in species of the two main subgenera, *Sophophora* and *Drosophila* (Marzo et al., 2008). Many incomplete (non-autonomous) copies of *Galileo* have been detected in all species searched and in some cases two or more *Galileo* subfamilies have been found coexisting in the same genome (Fig. 1). For instance, three subfamilies are present in *D. buzzatii*, while *Drosophila mojavensis* harbors four subfamilies (Delprat et al., 2009; Marzo et al., 2008). To date all sequenced copies of the transposon harbor premature stop codons and/or frame-shift mutations. Nevertheless, the sequence remnants reveal that the main domains of the *P-element* transposase are present in *Galileo*.

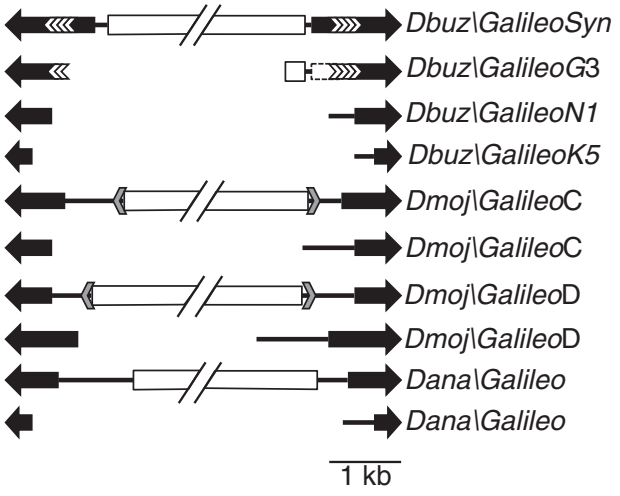

Fig. 1. Structure of representative *Galileo* copies in the different species of *Drosophila* used in this work. The black arrows are the TIR and white chevrons are tandem repeats within the respective TIRs. The white rectangles are the transposase coding regions. None of the transposase-containing copies harbor a functional ORF. The grey arrowheads are internal inverted repeats found in the *D. mojavensis* examples.

The most conspicuous features of *Galileo* are the TIRs which are 0.5 to 1.2 kb in length. This is considerably longer than other members of the *P-element* superfamily, in which the TIRs range from 30 to 50 bp. Indeed, it was the extreme length of *Galileo* TIRs that defined it as a *Foldback*-like transposon before it was recognized as a member of the *P-element* superfamily. The *Galileo* TIRs have another interesting property: namely, that the sequence conservation between elements in different species is restricted to ~40 bp at the extremities of the transposon (Marzo et al., 2008). One obvious possibility is that these regions are functional transposition sequences, and would be the equivalent to the short TIRs of the *P-element*. If true, this leaves the function of the remaining 0.5 to 1.2 kb open to question. The fact that they are not conserved between elements in different species, and that they contain internal tandem repeats in some subfamilies, has led to the suggestion that structure of the DNA may play a role in transposition (Adams et al., 2000; Ivics et al., 1997; Marquez and Pritham, 2010; Moschetti et al., 2008). The mechanism of *Galileo* transposition may therefore prove to be of considerable interest, and may explain the frequency with which this element is able to generate chromosomal inversions in *Drosophila*. In the present work we have focused on the reconstruction of an active transposase and its binding to the TIR. Although we have not succeeded in a full reconstitution of the transposition reaction, we have detected transposase binding to the extremities of *Galileo* and identified secondary binding sites in the tandem repeats of the TIR. This represents the first steps in the characterization of *Galileo* recombination. Further characterization promises to reveal fascinating details of the interactions between this transposon and its host and perhaps even the reason it promotes chromosomal inversions so frequently.

2. Results

2.1. Galileo transposase sequence reconstruction

The most complete example of the *Galileo* transposon is from *D. buzzatii* and was reconstructed from four overlapping PCR products (Marzo et al., 2008). Following the convention for *Drosophila* transposons (www.flybase.org) we will refer to this element as Dbuz*Galileo*, with the suffix 'Syn' (from synthetic) to indicate that it is a conceptual putatively complete copy (Fig. 1). This element has TIRs of 1.2 kb and an intron-less ORF encoding a 912 amino acid transposase (after correcting two stop codons and a frame shift mutation). Although there is no complete genome sequence for *D. buzzatii*, several internally deleted *Galileo* elements have been identified at the junctions of chromosomal rearrangements, and in other PCR and library screening experiments (Cáceres et al., 2001; Casals et al., 2005). Some of these elements were originally called *Kepler* and *Newton* but later assigned to different subfamilies of *Galileo*, now known as Dbuz*Galileo*-K and Dbuz*Galileo*-N, while Dbuz*Galileo*-G denotes the subfamily of the synthetic element. The various *Galileo* subfamilies have TIRs of different lengths, but share significant sequence homologies at the tips of the elements where one might expect the transposase to bind (~50 bp). Three specific examples of internally deleted G, N and K subfamily members are shown in Fig. 1. The complete genome sequences for *Drosophila ananassae* and *D. mojavensis* contained additional *Galileo* elements. In *D. ananassae* there is a single *Galileo* subfamily designated Dana*Galileo*. In *D. mojavensis* there are four subfamilies, two of which harbored transposase sequences: Dmoj*Galileo*-C and Dmoj*Galileo*-D (Marzo et al., 2008). These transposons all contain internal deletions, and two examples of members of each subfamily are shown in Fig. 1.

We were most interested in the Dbuz*Galileo*-G elements because these provide a complete transposase and have probably caused two of the three natural chromosomal deletions (Marzo et al., 2008). To