dc_67_10

Relics from the Past: Molecular Biology and Genetic Applications of Resurrected DNA Transposons in Vertebrates

Theses for the Degree of

Doctor of the Hungarian Academy of Sciences

Zoltán Ivics

Max Delbrück Center for Molecular Medicine

Berlin

2010

BACKGROUND

Transposons are discrete segments of DNA that have the distinctive ability to move and replicate within genomes. Transposons were discovered in the 1940's by Barbara McClintock (who later was awarded with the Nobel Prize for this discovery) in the maize genome, and have since been found ubiquitous in essentially all living organisms. The process of element movement is generally called transposition, and can contribute to insertional mutagenesis, altered gene expression and recombination. Transposons make up significant fractions of genomes; for example, about 45% of the human genome is composed of sequences of a variety of different elements. Transposons are best viewed as molecular parasites that propagate themselves using resources of the host cell. The transposition process is under regulation by both element- and host cell-encoded mechanisms and factors; these include transcriptional regulation of transposase expression, regulation of synaptic complex assemby, modulation of both transposon excision as well as integration by components of the chromatin and various factors that contribute to the target site selection properties of a given element. Transposition also has profound effects on the cell's life by modulating cellular pathways and gene functions by interacting with host proteins and by insertional mutagenesis. Despite their parasitic nature, there is increasing evidence that transposable elements are a powerful force in gene evolution. Indeed, about 50 human genes are derived from transposable elements, among them genes that are responsible for immunoglobulin gene recombination in all vertebrates.

Transposons are natural gene delivery vehicles, and have been revolutionizing genomic manipulations in diverse model systems. Transposons have been developed as useful tools for genomic manipulations, including transgenesis and insertional mutagenesis, in invertebrate animal systems as well as in plants, but similar technologies have been impossible in vertebrate systems for a long time, for the simple reason that the vast majority of DNA transposons are extinct in vertebrate genomes.

AIMS

1 Relics from the past: molecular biology of resurrected transposons and transposase-derived cellular genes in vertebrates

In this aim we set out to investigate the molecular biology of DNA transposons in vertebrate genomes. For this purpose we aimed at molecular reconstruction of active elements, mainly from the *Tc1/mariner* superfamily. The main question that we addressed was, first and foremost, can we reactivate „dead“ elements at all? Can we collect sufficient sequence as well as phylogenetic information of these transposons that would be required for their functional reactivation? Second, once we had these elements, we aimed at using them as experimental systems to address fundemental questions related to their mechanism of transposition, regulation by both element- as well as host cell-encoded factors and mechanisms and their interaction with their cellular environment. Finally, by using the reconstructed transposons as molecular references for transpositional activities, we set out to probe the functional properties of domesticated, transposase-derived cellular genes.

2 DNA transposons as a gene delivery platform for genetic manupilations in vertebrates

In this aim we set out to establish the resurrected transposons as molecular tools for genomic manipulations of vertebrate genomes. We specifically aimed at developing hyperactive transposase mutants as well as transposon-based vector system of enhanced utility for gene transfer in vertebrate cells. We further aimed at validating transposon vectors for delivering shRNA cassettes into cells for stable knockdown of gene expression. Finally, we aimed at improving the safety profile of transposon-based gene vectors by employing chromatin insulators to shield transcriptional activities of transgene cassette and by investigating technologies for target-selected transposon insertion for the purpose of human gene therapy.

RESULTS

We made the following major discoveries and conclusions:

1. Molecular phylogenetic data were used to construct a synthetic transposon, *Sleeping Beauty (SB)*, which could be identical or equivalent to an ancient element that dispersed in fish genomes in part by horizontal transmission between species. A consensus sequence of a transposase gene of the salmonid subfamily of elements was engineered by eliminating the inactivating mutations. *SB* transposase binds to two sites (the DRs) within the inverted repeats (IRs) of salmonid transposons in a substrate-specific manner, and mediates precise, cut-and-paste transposition in fish as well as in mouse and human cells.

2. We investigated transcriptional activities of *SB* in order to assess its potential to alter host gene expression upon integration. The untranslated regions (UTRs) of the transposon direct convergent, inwards-directed transcription. Transcription from the 5'-UTR of *SB* is upregulated by the host-encoded factor HMG2L1, and requires a 65-bp region not present in commonly used *SB* vectors. The *SB* transposase antagonizes the effect of HMG2L1, suggesting that natural transposase expression is under a negative feedback regulation. *SB* transposon vectors lacking the 65-bp region associated with HMG2L1-dependent upregulation exhibit benign transcriptional activities, at a level up to 100-times lower than that of the MLV retrovirus long terminal repeat.

3. We established that the DNA-bending, high mobility group protein, HMGB1 is a host-encoded cofactor of *SB* transposition. Transposition was severely reduced in mouse cells deficient in HMGB1. This effect was rescued by transient overexpression of HMGB1, and was partially complemented by HMGB2, but not with the HMGA1 protein. Overexpression of HMGB1 in wild-type mouse cells enhanced transposition, indicating that HMGB1 can be a limiting factor of transposition. *SB* transposase was found to interact with HMGB1 *in vivo*, suggesting that the transposase may recruit HMGB1 to transposon DNA.

HMGB1 stimulated preferential binding of the transposase to the DR more distant from the cleavage site, and promoted bending of DNA fragments containing the transposon IR. We propose that the role of HMGB1 is to ensure that transposase-transposon complexes are first formed at the internal DRs, and to subsequently promote juxtaposition of functional sites in transposon DNA, thereby assisting the formation of synaptic complexes.

4. We used the *SB* element as a tool to probe transposon-host cell interactions in vertebrates. The Miz-1 transcription factor was identified as an interactor of the *SB* transposase in a yeast two-hybrid screen. Through its association with Miz-1, the *SB* transposase downregulates cyclin D1 expression in human cells, as evidenced by differential gene expression analysis using microarray hybridization. Downregulation of cyclin D1 results in a prolonged G1 phase of the cell-cycle and retarded growth of transposase-expressing cells. G1 slowdown is associated with a decrease of cyclin D1/cdk4-specific phosphorylation of the retinoblastoma protein. Both cyclin D1 downregulation and the G1 slowdown induced by the transposase require Miz-1. A temporary G1 arrest enhances transposition, suggesting that *SB* transposition is favored in the G1 phase of the cell-cycle, where the nonhomologous end joining (NHEJ) pathway of DNA repair is preferentially active. Because NHEJ is required for efficient *SB* transposition, the transposase-induced G1 slowdown is probably a selfish act on the transposon's part to maximize the chance for a successful transposition event.

5. We have shown that DNA CpG methylation upregulates transposition of IR/DR elements in the *Tc1/mariner* superfamily. CpG methylation provokes the formation of a tight chromatin structure at the transposon DNA, likely aiding the formation of a catalytically active complex by facilitating synapsis of sites bound by the transposase.

6. We established that the distribution of experimentally induced *SB* insertions in the human genome can be considered fairly random,

because most chromosomes can serve as a target; no obvious hotspots with multiple insertions were found, and no preference for coding versus non-coding DNA was observed. We further showed that the *SB* element displays a certain degree of specificity in target site utilization at the DNA sequence and structural level. A palindromic AT-repeat consensus sequence with bendability and a symmetrical pattern of hydrogen bonding sites in the major groove of the target DNA define preferred sites for integration.

7. A novel open reading frame-trapping method was used to isolate uninterrupted transposase coding regions from the genome of the frog species *Rana pipiens*. The isolated clones were about 90% identical to a predicted transposase gene sequence from *Xenopus laevis*. None of these native genes was found to be active. Therefore, a consensus sequence of the transposase gene was derived. This engineered transposase and the transposon inverted repeats together constitute the components of a novel transposon system that we named *Frog Prince* (*FP*). *FP* has only about fifty percent sequence similarity to *SB*, and catalyzes efficient cut-and-paste transposition in fish, amphibian and mammalian cell lines. We demonstrate high-efficiency gene trapping in human cells using *FP* transposition. *FP* is the most efficient DNA-based transposon from vertebrates described to date, and shows about 70% higher activity in zebrafish cells than *SB*. *Frog Prince* can greatly extend our possibilities for genetic analyses in vertebrates.

8. *Hsmar1*, one of the two subfamilies of *mariner* transposons in humans, is an ancient element that entered the primate genome lineage ~50 million years ago. Although *Hsmar1* elements are inactive due to mutational damage, one particular copy of the transposase gene has apparently been under selection. This transposase coding region is part of the SETMAR gene, in which a histone methylatransferase SET domain is fused to an *Hsmar1* transposase domain. A phylogenetic approach was taken to reconstruct the ancestral *Hsmar1* transposase gene that we named *Hsmar1-Ra*. The *Hsmar1*-Ra transposase efficiently mobilizes

Hsmar1 transposons by a cut-and-paste mechanism in human cells and zebrafish embryos. *Hsmar1*-Ra can also mobilize short inverted-repeat transposable elements (MITEs) related to *Hsmar1* (*MiHsmar1*), thereby establishing a functional relationship between an *Hsmar1* transposase source and these MITEs. *MiHsmar1* excision is two orders-of-magnitude more efficient than that of long elements, thus providing an explanation for their high copy number. We show that the SETMAR protein binds, and introduces single-strand nicks into *Hsmar1* inverted repeat sequences *in vitro*. Pathway choice for DNA break repair was found to be characteristically different in response to transposon cleavage mediated by *Hsmar1*-Ra and SETMAR *in vivo*. Whereas nonhomologous end-joining plays a dominant role in repairing excision sites generated by the *Hsmar1*-Ra transposase, DNA repair following cleavage by SETMAR predominantly follows a homology-dependent pathway. The novel transposon system can be a useful tool for genome manipulations in vertebrates, and for investigations into the transpositional dynamics and contribution of these elements to primate genome evolution.

9. Ancient, inactive copies of transposable elements of the *PIF/Harbinger* superfamily have been described in vertebrates. Based on a predicted consensus sequence, we reconstructed the functional components of the *Harbinger3_DR* transposon in zebrafish, including a transposase and a second, transposon-encoded protein of unknown function that has a Myb-like trihelix domain. The reconstructed *Harbinger* transposon shows efficient cut-and-paste transposition in human cells, and preferentially inserts into a 15-bp consensus target sequence. The Myb-like protein is required for transposition, and physically interacts with the transposase. The Myb-like protein enables transposition in part by promoting nuclear import of the transposase, and by binding to the transposon ends. We investigated the functions of two, transposon-derived human proteins: HARBI1, a domesticated transposase-derived protein and NAIF1 that contains a trihelix motif similar to that described in the Myb-like protein. Physical interaction, subcellular localization and DNA-binding activities of

HARBI1 and NAIF1 suggest strong functional homologies between the *Harbinger3_DR* system and their related, host-encoded counterparts.

10. The *SB* transposon is a promising vector for transgenesis in vertebrates, and is being developed as a novel, nonviral system for gene therapeutic purposes. A mutagenesis approach was undertaken to improve various aspects of the transposon, including safety and overall efficiency of gene transfer in human cells. We constructed a “sandwich” transposon, in which the DNA to be mobilized is flanked by two complete *SB* elements arranged in an inverted orientation. The sandwich element has superior ability to transpose >10 kb transgenes, thereby extending the cloning capacity of *SB*-based vectors. We derived hyperactive versions of the *SB* transposase by single-amino-acid substitutions. These mutations act synergistically, and result in an almost 4-fold enhancement of activity when compared to the wild-type transposase. We also created a library of mutant *SB* transposase genes by using an *in vitro* evolution paradigm. One particular mutant, called *SB100X*, exhibited a ~100-fold enhancement of transposition as compared to the wild-type transposase, and showed robust gene transfer efficiencies in mouse embryos as well as human hematopoietic stem cells. The improved vector system should prove useful for efficient gene transfer in vertebrates.

11. We have developed a stable RNA interference (RNAi) delivery system that is based on the *FP* transposon. This plasmid-based vector system combines the gene silencing capabilities of H1 polymerase III promoter-driven short hairpin RNAs (shRNA) with the advantages of stable and efficient genomic integration of the shRNA cassette mediated by transposition. We show that the *FP*-based shRNA expressing system can efficiently knock down the expression of genes in human cells. Transposon-mediated genomic integration ensures that the shRNA expression cassette and a selectable marker gene within the transposon remain intact and physically linked. We demonstrate that a major advantage of our vector system over plasmid-based shRNA delivery is both its enhanced frequency of intact genomic integration as well as

higher target suppression in transgenic human cells. Due to its simplicity and effectiveness, transposon-based RNAi is an emerging tool to facilitate analysis of gene function through the establishment of stable loss-of-function cell lines.

12. Transposon-based gene vectors have become indispensable tools in vertebrate genetics for applications ranging from insertional mutagenesis and transgenesis in model species to gene therapy in humans. The transposon toolkit is expanding, but a careful, side-by-side characterization of the diverse transposon systems has been lacking. Here we compared the *SB*, *piggyBac* and *Tol2* transposons with respect to overall activity, overproduction inhibition (OPI), target site selection and transgene copy number as well as long-term expression in human cells. *SB* was the most efficient system under conditions where the availability of the transposon DNA is limiting the transposition reaction including hard-to-transfect hematopoietic stem/progenitor cells, and the most sensitive to OPI, underpinning the need for careful optimization of the transposon components. *SB* and *piggyBac* were about equally active, and both more efficient than *Tol2*, under nonrestrictive conditions. All three systems provided long-term transgene expression in human cells with minimal signs of silencing. *SB*, *Tol2* and *piggyBac* constitute complementary research tools for gene transfer in mammalian cells with important implications for fundamental and translational research.

13. Incorporation of chicken beta-globin HS4 insulator sequences in *SB*-based vectors reduces transactivation of model promoters by transposon-borne enhancers, and thus may lower the risk of transcriptional activation of host genes situated close to a transposon insertion site.

14. Random chromosomal transposition is clearly undesired for human gene therapeutic applications due to potential genotoxic effects associated with transposon integration. We demonstrated targeted chromosomal insertion of the *SB* transposon in human cells. We established a

successful strategy based on targeting proteins that can bind both transposon and target DNA to direct *SB* element transposition into the vicinity of a specific DNA sequence in the human genome. Furthermore, transposon targeting based on protein-protein interactions between the *SB* transposase and a targeting fusion containing the N-terminal protein interaction domain of *SB* is a successful strategy to direct SB integrations into a given locus in the human genome. This approach was found to enable a ~10^7-fold enrichment of transgene insertion at a desired locus. Our results provide proof-of-principle for directing chromosomal insertion of an otherwise randomly integrating genetic element into preselected sites. Targeted transposition could be a powerful technology for safe transgene integration in human applications.

LIST OF PUBLICATIONS THAT FORM THE BASIS OF THIS THESIS

I. **Ivics, Z**., Izsvák, Zs., Minter, A. and Hackett, P.B. (1996). Identification of functional domains and evolution of Tc1-like transposable elements. *Proc. Natl. Acad. Sci. USA* 93:5008-5013.

II. **Ivics, Z**., Hackett, P.B., Plasterk, R.H. and Izsvák, Zs. (1997). Molecular reconstruction of *Sleeping Beauty*, a Tc1-like transposon in fish, and its transposition in human cells. *Cell* 91:501-510.

III. Walisko, O., Schorn, A., Rolfs, F., Devaraj, A., Miskey, C., Izsvák, Z. and **Ivics, Z**. (2008). Transcriptional activities of the *Sleeping Beauty* transposon and shielding its genetic cargo with insulators. *Mol. Ther*. 16:359-69.

IV. Zayed, H., Izsvák, Zs., Khare, D., Heinemann, U. and **Ivics, Z**. (2003). The DNA-bending protein HMGB1 is a cellular cofactor of *Sleeping Beauty* transposition. *Nucleic Acids Res*. 31:2313-2322.

V. Walisko, O., Izsvák, Z., Szabó, K., Kaufman, C.D., Herold, S., and **Ivics, Z.** (2006). Sleeping Beauty transposase modulates cell-cycle progression through interaction with Miz-1. *Proc. Natl. Acad. Sci. USA* 103:4062-4067.

VI. Jursch, T., Izsvák, Z. and **Ivics, Z.** (2010). Regulation of DNA transposition by CpG methylation and chromatin structure in human cells. *Mobile DNA*. (to be submitted)

VII. Vigdal, T., Kaufman, C., Izsvak, Z., Voytas, D. and **Ivics, Z.** (2002). Common physical properties of DNA affecting target site selection of *Sleeping Beauty* and other Tc1/*mariner* transposable elements. *J. Mol. Biol*. 323:441452.

VIII. Miskey, Cs., Izsvák, Zs., Plasterk, R.H. and **Ivics, Z**. (2003). The *Frog Prince*: a reconstructed transposon from *Rana pipiens* with high activity in vertebrates. *Nucleic Acids Res*. 31:6873-6881.

IX. Miskey, C., Papp, B., Mátés, L., Sinzelle, L., Keller, H., Izsvák, Z. and **Ivics, Z**. (2007). The ancient *mariner* sails again: Transposition of the human *Hsmar1* element by a reconstructed transposase and activities of the SETMAR protein on transposon ends. *Mol. Cell. Biol*. 27: 4589-600.

X. Sinzelle, L., Kapitonov, V.V., Grzela, D.P., Jursch, T., Jurka, J., Izsvák, Z. and **Ivics, Z.** (2008). Transposition of a reconstructed *Harbinger* element in human cells and functional homology with two transposon-derived cellular genes. *Proc. Natl. Acad. Sci. USA* 105:4715-20.

XI. Zayed, H., Izsvák, Zs., Walisko, O. and **Ivics, Z.** (2004). Development of hyperactive *Sleeping Beauty* transposon vectors by mutational analysis. *Mol. Ther.* 9:292-304.

XII. Kaufman, C.D., Izsvák, Z., Katzer, A. and **Ivics, Z**. (2005). *Frog Prince* transposon-based RNAi vectors mediate efficient gene knockdown in human cells. *Journal of RNAi and Gene Silencing* 1:97-104.

XIII. Grabundzija, I., Irgang, M., Mátés, L., Belay, E., Matrai, J., Gogol-Döring, A., Kawakami, K., Chen, W., Ruiz, P., Chuah, M.K., VandenDriessche, T., Izsvák, Z. and **Ivics, Z**. (2010). Comparative analysis of transposable element vector systems in human cells. *Mol. Ther*. (in press)

XIV. **Ivics, Z**., Katzer, A., Stüwe, E.E., Fiedler, D., Knespel, S. and Izsvák, Z. (2007). Targeted *Sleeping Beauty* transposition in human cells. *Mol. Ther.* 15:1137–1144.