MOLECULAR FACTORS AFFECTING THE TARGET CHOICE OF BACTERIAL INSERTION SEQUENCE IS*30*

PROPOSITIONS OF PhD THESIS

MÓNIKA SZABÓ

Gödöllő
2007.

PhD School

name: Szent István University of Gödöllő

field: Biological sciences

leader: Dr. Zoltán Tuba
Chairman of department, doctor of biological science
SZIE Faculty of Agricultural and Environmental Sciences
Department of Botany and Plant Phisiology

supervisor: Dr. Olasz Ferenc
Agricultural Biotechnology Center

Dr. Hornok László
Chairman of department, member of MTA
SZIE Faculty of Agricultural and Environmental Sciences
Department of Microbiology

Approval of leader

Approval of supervisor

1. INTRODUCTION AND AIMS OF OUR STUDIES

The mobile genetic element were discovered in maize (Zea mays) by Barbara McClintock, who observed that segregation of some hybrides differs from the Mendel's heredity. Bacterial insertion sequences (IS) and transposons (*Tn*) have been described just two decads later. They were discovered by their mutagene effect and their contributions of antibiotic resistance genes. The similarity of bacterial elements to the maize elements became clear when it was recognized that these are natural resident of the bacterial genome, and that the observed insertion mutations were examples of their movement to new genetic locations. More than 1500 ISs have been so far described only in bacteria. Including the numerous eucaryotic elements, such as retroposons, retroviruses and other non-active repetitive elements we can state that mobile elements are wildly distibuted in living organisms, which may underline the importance of research work in this field.

The movement of the element, that is transposition, could be advantages and disadvantages to the host. Rearrangements caused by the transposons play an important role in the reorganization and evolution of the host genome. Numerous bacterial transposons encode such additional functions (e.g. antibiotic resistance and virulence) which could spread by the IS and may contribute to the extension of common bacterial genepool. Mobile genetic elements are capable of transferring genes horizontally using vectors (viruses, plasmids) that can be transmitted between individuals and even cross interspecies boundaries. DNA rearrangements and mutations caused by these elements increase genomic variation and may provide the basis of developing new adaptations. Their use as genetic tools is also widespread; mutagenesis-, gene insertion (knock-in) systems and gene-mapping by means of transposons are all techniques commonly used in modern biotechnology.

The number of pro- and eucariotyc elements is, however, enormous, the mechanism and regulation of transposition is not known in details. The aims of the work presented in this thesis were to extend of knowledge on IS elements and deeper understanding the molecular mechanism of transpositions. We have analysed the properties of IS*30* element, which is resident in *E. coli* K-12. As IS*30* shows a typical IS structure, it can be absolutely used to obtain more precise data about general transposition pathways.

IS*30* is 1221 bp long with one open reading frame (ORF-A) coding a 44.3 kDa transposase protein. It has 26 bp imperfect terminal inverted repeats (IRs), which contain the binding and cleavage site of transposase. Rearrangements caused by the element can be explained by two steps transposition. First is the formation of an active intermediate including a junction in which the IR ends of the element are abutted. This itermediate, which

could be a minicircle or a dimer containing two IS*30* copies, react with the target DNA in a second step leading to the formation of real transposition products (insertions, deletions, inversions and cointegrates). Insertion of IS*30* does not occur randomly, it exhibits a preference for two distinct types of target sequence: the "natural" hot spots (HS), characterized by a 24 bp long palindrome consensus, and the IRs of the element itself. One of the most perspective project was the study of target specificity of IS*30* and determine those transposase domains, which responsible for dual specificity of the element. These data may serve as a basis for the development of directed transposition systems that might become useful tools in functional genomics and applied biotechnology.

As the IRs of the element could also serve as transposition target sites, in the second part of the work we have studied the effect of IR and subterminal IS*30* sequences. The aim of this project was to understand target site selection of IS*30*, and to gain more information about the molecular mechanism of transposition.

Insertion next to IRs often results in such transposition products, which include incorrect target site duplication, while it has never been detected in HS-targeting. Mutations originally presented in the target end, frequently occurre on the donor IS element, too. These base-alterations could be considered as results of a gene conversion process. Our aim was the understanding of these sequence alterations and the establishment of a molecular model that explains the gene conversion events occurring in the course of IR targeted IS*30* transposition reactions.

Our aims were:

- detection of transposase domain(s) responsible for the dual target specificity of IS*30*.
- improvement of a directed mutagenesis system using chimera elements, in which the IS*30* transposase is joining with different DNA-binding proteins.
- study of activity and target specificity of IS*30* in zebrafish (Danio rerio).
- IR-analyses, identification of basis playing essential role in transposase binding and cleavage.
- identification of DNA elements in IS*30* affecting the efficiency of transposition.
- study of the orientation effect in transposition, identification of such genetic factors, which prevent the joining of identical IRs.
- study of transposition products containing gene conversions, improvement a molecular model, which coud explain the feature of IS*30* transposition.

2. MATHERIAL AND METHODS

In the experiments the following molecular microbiological and biochemical methods were used:

- Standard microbiological work with bacteria (*E. coli*, *S. typhimurium*) and bacteriophages (λ and R408 phages): transformation, electroporation, conjugation, transduction and Tn*10*-based transposon mutagenesis
- Microinjection of 1 or 2 cell stage zebrafish eggs
- *In vitro* mRNA synthesis by Ambion mMessage mMachine kit
- Total and plasmid DNA purification fom *E. coli*, *S. typhimurium* and zebrafish embryos
- DNA cloning, restriction cleavages and PCR amplification
- Overexpression of protein using different expression vector system and protein purification
- Agarose and poliacrylamide gelelectrophoresis
- Southern and Western hibridization
- Gel retardation assays
- *In vitro* mutagenesis using megaprimer- and OE-PCR technics
- DNA sequencing on ABI Prism 3100 Genetic Analyzer (Applied Biosystem)
- Computer analysis using the Wisconsin GCG, MultAlin, PSIPRED, WebLogo and DNA tools program packages
- Study of IS*30* transposition, isolation of fusion products using R408 transduction
- Detection of transposition intermediates (minicircles and figure of eight structure) in plasmid-based systems
- Examination of IS*30* integration in intra- and intermolecular assays
- Isolation of target sequences of IS30 on pOX38Km conjugatible plasmid
- *In vitro* transposition assay
- Primer extension experiments to detect the cleavage catalysed by IS*30* transposase

3. RESULTS

We have determined the DNA-binding motif of IS*30* transposase responsible for the dual target specificity of the element. Structural predictions of IS*30* family members revealed the presence of a conserved helix-helix-turn-helix motif (H-HTH2) in the N-terminal part of transposases which, in the case of IS*30* is preceded by an additional HTH (HTH1). We have constructed a group of site directed point and deletion mutants and determined their activities in several *in vivo* transposition assays and in binding IS*30* ends.

The residues to be changed by directed mutagenesis were selected from the secondary structure predictions. It has been shown that the unique HTH1 is responsible for the HS recognition, while the well conserved H-HTH2 is required for both the IR- and HS-targeting. Moreover, the *in vivo* transposition results of HTH2 mutants was well correlated with their ability to bind IS*30* ends. Similarly the point mutant, deletion of HTH1 destroys the HS specificity of the element, but does not affect the targeting of IRs. The target site selection of delHTH1 transposase became random, and the deletion transposase in contrast to the wt IS*30* generated 3 bp duplication of target sequences. These data provide evidence that pronounced target specificity of IS*30* could be modified by point and deletion mutant transposases.

To direct the insertions of IS*30*, the transposase was joined to unrelated DNA binding domains. Joining the transposase to the cI repressor of phage λ causes transposition into the 400 bp vicinity of the O_R1-3 operator sequences in a plasmid-based system. The cI operator in the S. *typhimurinum* genome was also attractive as transposition target site, even if the integrations next to this region was less freqently than into other sites of bacterial genome. The target specificity of the prokaryotic IS element could also modified in zebrafish. Transposase linked to the DNA-binding domain of Gli1 transcription regulator protein directs the insertions near to the Gli1 binding site located on a tester plasmid in zebrafish embryos. Our results demonstrate the possibility of fusion transposases to acquire novel target specificity in both prokaryotes and eukaryotes.

In the next part of our work we have studied the effects of IRs and the subterminal IS*30* sequences on the transposition events. The bandsift assays proved that the internal regions of IRs (positions 16-17. and 20-26.) are necessary for binding the transposase and for the formation of IR-IR junction. Deletion analyses of this junction show that the terminal part of IRs (positions 1-20.) is enough for the integration, and the binding site of transposase is not required. We have developed a cell-free recombination system and show that the transposase catalyses formation of a figure-of-eight transposition intermediate, where a 2-bp long single strand bridge holds the IRs together. Using the *in vitro* transposition assay we have proved that the 5'-GT-3' bases in the 2-3. positions of the IRs are required for cleavage at the tip of the element, which is the first step in figure-of-eight production.

We have also demonstrated that subterminal sequences adjacent to the terminal IRs, are required for optimal transposition activity of IS*30*. The absence of subterminal sequences negatively influences figure-of-eight production *in vitro*. These regions enhance both IR-IR junction formation and IR-targeting events *in vivo*. Enhancer elements were

identified within 51 bp internal to IRL and 23 bp internal to IRR. In the right end, a decanucleotide, 5'-GAGATAATTG-3', is responsible for generating wild type activity, while in the left end, a complex assembly of repetitive elements is required. According to the analysis of point and deletion mutants we presume 6 enhancer element in the left end between 27-77 bp positions, which are located in similar sequence motifs, for which a consensus AAAC can be deduced. Functioning of the 10-bp element in the right end is position dependent and the repetitive elements in the left end act cooperatively and may influence the bendability of left end. The presence of the subterminal sequences increased the figure-of-eight production *in vitro* suggesting that the subterminal enhancer effect is at least partly transposase dependent. Such enhancers may reflect a subtle regulatory mechanism for IS*30* transposition.

The deletion analyses of IS*30* ends suppose that these enhancer sequences are the main cause of orientation effect in IR-targeting events, where the targeted end determines the direction of the insertion. The target and the integrating IS*30* copies are always attached by their left and right ends leading to a head-to-tail orientation. In the absence of the subterminal enhancer sequences we could isolate junctions of two left or two right ends, while it was never detected when the two left ends were longer than 68 bp, or the right ends were at least 51 bp. At the same time the transposons with identical ends integrate into the GOHS hot spot indicating the incorrect junction formation. This was supported by the *in vitro* detection of figure-of-eight structures *in vitro* that were derived from transposons assembled with identical ends. Thus we suppose that the connection of identical IS*30* ends is significantly inhibited by the lack of the enhancer element(s) located in the other end. Moreover the incorrect junction provides 100-120 bp palindromic sequence that disturbs the replication and survival of the plasmid harbouring it. Thus the joining of long identical ends not only occurs at a reduced frequency, but causes serious replication/stability defect.

We have proved that homology between the donor and target IS*30* sequences is required for conversion events and showed that the starting point of the process is the site of integration. Conversion of wt donor sequences has been shown even if the target contains only the 26 bp IR of the element. In cases of longer homology conversion has been observed within a 200 bp range, as well. Generally the donor sequence undergoes conversion according to the target DNA, and the frequency of conversion depends on the distance of mutations from the point of insertion. Conversion takes place only in one direction even if homology is present on both sides of the insertion and can occur without formation of transposition products. Our results support the idea of formation and migration of a half

Holliday junction. The proposed model also suggests that certain components of the homologous recombination machinery are required for the resolution and repair of DNA structures formed via transpositional reactions.

4. CONCLUSIONS AND NEW SIENTIFIC RESULTS

- It has been prooved that the unique HTH1 in the N-terminal part of IS*30* is responsible for the HS recognition, while the conserved H-HTH2 is necessary for the IR-binding.
- The HS-specificity of the transposase was modified by the fusion proteins, in which the transposase was joined to heterologous DNA-binding domains.
- Targeted transposition was applied in bacteria (*E. coli* and *S. typhimurium*) and in zebrafish.
- We have determined the bindig site of transposase in the IR ends, and verified that this region is absolultely need for the first step of transposition.
- We have proved that the 5'-GT-3' bases in the 2-3. positions of the IRs are required for the transposition cleavage.
- It has been demonstrated that IS*30* sequences adjacent to the terminal IRs are required for optimal transposition activity of the element.
- We have showed that subterminal sequences are the main cause of orientation effect in the transposition of IS*30*.
- We have determined the requirements of the formation of conversion products and proposed a molecular model, which explains the their variability.

5. PUBLICATIONS

Thesis based on following publications:

International publications:

Szabó, M., Müller, F., Kiss, J., Balduf, C., Strähle, U. and Olasz, F. (2003) Transposition and targeting of the prokaryotic mobile element IS*30* in zebrafish. *FEBS Letters* **550**: 46-50.

Olasz, F., Fischer, T., Szabó, M., Nagy, Z. and Kiss, J. (2003) Gene conversion in transposition of *Escherichia coli* element IS*30*. *J. Mol. Biol.* **334**: 967-978.

Nagy, Z., Szabó, M., Chandler, M. and Olasz, F. (2004) Analysis of the N-terminal DNA binding domain of the IS*30* transposase. *Mol. Microbiol.* **54**: 478-488.

Kiss, J., Szabó, M., Chandler, M. and Olasz, F. Cis-acting element in IS30 transposition the functional analysis of the terminal sequences. Oxford Workshop on Site-specific Recombination, Transposition and DNA Dynamics. St Catherine's College, Oxford, September 19-24. 2006. (poster)

Szabó, M., Kiss, J., Nagy, Z., Chandler, M. and Olasz, F. Subterminal sequences modulating IS*30* transposition *in vivo* and *in vitro*. (előkészületben)

Hungarian publications:

Szabó Mónika, Nagy Zita, Kiss János, Olasz Ferenc: The role of IRs and subterminal sequences in IS*30* transposition. 9th Workshop of Molecular Biology Section of Hungarian Biochemical. Sopron, May 10-13. 2004. (poster)

Kohut Gábor, Szabó Mónika, Olasz Ferenc: Investigation of target specificity of chimera IS*30*. 9th Workshop of Molecular Biology Section of Hungarian Biochemical. Sopron, May 10-13. 2004. (poster)

Nagy Zita, Szabó Mónika, Michael Chandler, Olasz Ferenc: Functional analysis of the domain responsible for target specificity of IS30 transposase. 9th Workshop of Molecular Biology Section of Hungarian Biochemical. Sopron, May 10-13. 2004. (presentation)

Other publications

International publications:

Szabó, M., Kiss, J., Kótány, G. and Olasz, F. (1999) Importance of illegitimate recombination and transposition in IS*30*-associated excision events. *Plasmid* **42**: 192-209.

Kiss, J., Szabó, M., Olasz, F. (2003) Site-specific recombination by the DDE family member mobile element IS*30* transposase. *Proc. Natl. Acad. Sci. USA* **100**:15000-15005.

Hungarian publications:

Szabó Mónika, Kiss János, Olasz Ferenc: The role of illegitimate recombination in IS*30* transposition. 4th Hungarian Genetic Congress. Siófok, April 11-14. 1999. (poster)

Kiss János, Szabó Mónika, Nagy Zita, Olasz Ferenc: Site-specific and transposition recombination by the mobile element IS*30*. 9th Workshop of Molecular Biology Section of Hungarian Biochemical. Sopron, May 10-13. 2004. (presentation)

Patents

Mónika Szabó, János Kiss, Ferenc Olasz, (ABC, Gödöllő, Hungary), Ferenc Müller, Lászlo Tora, Uwe Strähle (IGBMC, Strasbourg, France). European Patent Application, N° 01402754.4. Site-directed recombinase fusion proteins and corresponding polynucleotides, vectors and kits, and their uses for site-directed DNA recombination.