Retrotransposon *Tto1*: functional analysis and engineering for insertional mutagenesis

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Andrea Tramontano

Aus Avellino, Italien

Köln, 2011

Diese Arbeit wurde am Max-Planck-Institut für
Pflanzenzüchtungforschung in Köln,
in der Abteilung Molekulare Pflanzengenetik und
am Max-F. Perutz-Laboratories in Wien, in der Abteilung Biochemie
durchgeführt.

MFPL **Max F. Perutz Laboratories**

Berichterstatter: Prof. Dr. **George Coupland**

Prof. Dr. **Martin Hülskamp**

Tag der Disputation:

05. April 2011

"Emancipate yourselves from mental slavery:
none but ourselves can free our minds".
(Bob Marley)

Contents

Abstract

Retrotransposons are genomic parasites activated by stress conditions that can be seriously detrimental for their host. In this work I demonstrate that *Tto1*, a typical plant LTR retrotransposon with insertion preference into genes can be turned into a synthetic molecular tool for gene tagging in plants and can be used to predict models for its replication steps. Although retrotransposons have been already used in plant mutagenesis, such application always required establishing protocols for tissue cultures and regeneration in vitro. Here, I show that sequence engineering of *Tto1* provides the possibility to obtain transposition in vivo, with a simple screening method based on PCR and with the advantage to skip all in vitro manipulations. An artificial β-estradiol inducible promoter has been used to obtain transposition "on demand" in Arabidopsis plants, which generates stable unlinked insertions that follow mendelian segregation in the progeny.

Comparing serial deletions of 3' LTR of the engineered inducible *Tto1* (i*Tto1*), I have mapped its two natural terminators and identified the "minimal" R (redundant) region required to achieve the complete reverse transcription of the genomic mRNA into a new cDNA copy. Interestingly, the transcripts ending at the major "early" terminator cannot support reverse transcription, suggesting a mechanism of natural control on the expression. Transcripts with a more extended termination point contain 100 essential nucleotides that define the active nucleus of the R region. This sequence promotes the formation of a stable hairpin structure that "kisses" a complementary identical hairpin on the cDNA and determines the formation of the characteristic cDNA/mRNA heteroduplex. Since the LTR is a repeated sequence the definition of a minimal redundant region has also the important implication to reduce the only possible target for sequence-based gene silencing, which should lead to an increase of the mutagenic efficiency of i*Tto1*.

Additional investigations have been carried out in attempt to identify points of improvement of i*Tto1* performances. By sequence alignment I identified different versions of the integrase that might have influence on insertion efficiency. Furthermore I tested the *pOp6/LhGR-N* system that will provide higher expression levels in different host plants. The final goal of my work is to extend the application of i*Tto*1 to crop mutagenesis, therefore a big part of my work has been spent to develop

Tto1 constructs with activity in barley. Transgenic plants have been obtained, however the constructs still need further experimentation.

Zusammenfassung

Retrotransposons sind genomische Parasiten, welche unter Stressbedingungen aktiv werden und dadurch den Wirt schädigen können. Tto1 ist ein typisches pflanzliches Retrotransposon und insertiert bevorzugt in Gene. In dieser Arbeit konnte gezeigt werden, dass Tto1 in ein Werkzeug für Insertionsmutagenese verwandelt werden kann. Retrotransposons sind bereits zur Mutagenese von Pflanzen verwendet worden, doch verlangt dies üblicherweise Protokolle zur Gewebekultur und Regeneration. Wir konnten zeigen, dass Änderungen an Tto1 es ermöglichen, ohne Gewebekultur-Schritte *in vivo* Transpositionsereignisse herbeizuführen, welche mit einer einfachen PCR-basierenden Screening-Methode detektiert werden können. Ein β-Östradiol-induzierbarer Promotor wurde verwendet, um in Arabidopsis Pflanzen Transposition zu induzieren. Diese stabilen Neu-Insertionen werden nach Mendel´schen Gesetzen weiter vererbt.

Das veränderte Element wurde auch zur Analyse des Replikationszyklus verwendet. Es wurden serielle Deletionen in der langen terminalen Sequenz-wiederholung am 3´ Ende hergestellt. Zwei Regionen der Translationstermination wurden kartiert und eine minimale redundante Region definiert, welche für korrekte reverse Transkription notwendig ist. Transkripte, die in der ersten Terminationsregion enden, können nicht revers transkribiert werden, während die längeren Transkripte eine Kernegion von 100 Basenpaaren enthalten, welche für die reverse Transkription essentiell ist. Die Kernregion enthält eine stabile Haarnadelstruktur, die mit einer kompementären Haarnadelstruktur in der entstehenden komplementären DNA Basenpaarungen ausbilden kann, um eine DNA-RNA Heteroduplex Struktur auszubilden. Kenntnis der minimalen redundanten Region kann dazu verwendet werden, die Sequenzwiederholungen an den Enden von Tto1 zu verkürzen und so die Basis für Genstillegungen, welche oft von Sequenzwiederholungen induziert werden, zu verkleinern.

Eine Reihe von Untersuchungen wurden durchgeführt, um die Transpositions-Effizienz von Tto1 zu erhöhen. Durch Sequenzvergleiche wurden verschiedene Versionen des Retrotransposon-Enzyms Integrase identifiziert, welche Einfluss auf die Integrations-Effizienz haben sollten. Es wurde auch das pOp6/LhGR-N Induktionssystem getestet, welches höhere Expression von Tto1 in Wirtspflanzen erlauben sollte. Ein weiteres Ziel der Arbeit war es, Tto1 für Mutagenese in der Kulturpflanze Gerste heranzuziehen. Vektorkonstrukte für Gerste wurden hergestellt und zur Transformation von Gerste

herangezogen, doch stellte sich heraus, dass die Konstrukte weiterer Verbesserungen bedürfen.

Chapter 1

INTRODUCTION

At the beginning of my biological studies, I remember being told that the genes necessary for life occupy just a small portion of the whole human genome, the rest being highly condensed centromeric and telomeric sequences or simply "non gene" sequences. As I proceeded with plant genetics and as my knowledge of the development of living organisms grew, I could figure out that such sequences were something more than just "non genes", that they had a big impact on evolution and that they offered good opportunities to bring new findings in plant science. Thus, it was with big enthusiasm that I decided to undertake a PhD in this fascinating field.

1.1 Two classes TEs (Transposable elements)

The discovery of the first TE is credited to Barbara McClintock in 1950 who was awarded with the Nobel Prize in 1983 for her research. She described them as "mutable loci" (McClintock, 1950, 1953), based on the observed phenotype of the varying pigmentation of the maize kernels upon chromosomal breakage.

Ever since, the increasing number of genomes being sequenced has shown that TEs are ubiquitous and particularly abundant in eukaryotes. The only know exceptions are the protist *Plasmodium falciparum* and probably several closely related species (Wicker et al., 2007).

TEs are generally defined as mobile DNA sequences that are able to integrate at a new location into their host genome and remain intracellular during this process.

All TEs have the ability to amplify their copy number during their life cycle, but they adopt two different mechanisms of replication, which can either involve or not involve a reverse transcription step. On this basis they are conventionally divided in two classes: Class I or retrotransposons replicate via a "copy and paste" mechanism that involves an RNA intermediate; Class II adopt a "cut and paste" replication mechanism

involving a DNA intermediate (Finnegan, 1989). Nevertheless, the discovery of bacterial (Duval-Valentin et al., 2004) and eukaryotic (Lai et al., 2005; Morgante et al., 2005) TEs that copy and paste without RNA intermediate, and of MITEs (miniature inverted repeat transposable elements) that share characteristics of both classes, has challenged the two class system (Wicker et al., 2007), therefore other subdivisions have been proposed, based on enzymological categories, for instance: DDE-transposase, RT-En, Y and Y2 (tyrosine) transposase and S (serine) transposase (Curcio and Derbyshire, 2003). In the following (Fig. 1.1) I present a recent classification of TEs that follows both approaches, in a way that the enzymological criterion is applied to the two class system (Wicker et al., 2007). In this table the higher subdivision in class I and II is followed by a re-organization of the lower levels. In particular the taxon "subclass" is used to separate elements of the Class II that follow the classical "cut and paste" mechanism, therefore involving a double strand cleavage, from those which copy themselves into a new location by only cutting one DNA strand (i.e. Helitron and Maverick). Subsequently, the taxon "order" here replaces the previous "subclass", thus, i.e. the order LTR retrotransposons is used instead of the subclass LTR retrotransposons previously suggested by Casacuberta et al., 2005. To each order belong different superfamilies, such as *copia* or *gypsy* that were previously designated as "groups". The superfamilies share the same replication mechanism but they are distinguished by uniform large scale characteristics such as the protein organization, the non coding domains, the presence or absence and length of TSD (target site duplication). The families are defined by DNA sequence conservation, since in the higher taxon (superfamily) the level of protein sequence conservation is generally high. The definition of family also serves to classify non autonomous elements. Wicker and colleagues (2007) define as autonomous all those elements that appear to encode all enzymatic domains necessary for transposition, regardless of the fact that they are active or functional. They distinguish autonomous elements of one family that have been rendered defective by point mutations, insertions or deletions, from the non-autonomous elements. The non-autonomous are simply defined as elements having a highly degenerated coding region or even lack completely coding capacity, in contrast to defective autonomous elements. The authors also propose a three letter code to facilitate classification and annotation. In this system the three letters will denote respectively class, order and superfamily, therefore, to make an example that is of direct interest to me, the elements belonging to the superfamily *copia* will have the code RLC, where R denotes class: retrotransposons; L denotes order: LTR, and C denotes the superfamily: *copia.*

Classification		Structure	TSD	Code	Occurrence
Order	**Superfamily**				
Class I (retrotransposons)					
LTR	*Copia*	GAG AP INT RT RH	4–6	RLC	P, M, F, O
	Gypsy	GAG AP RT RH INT	4–6	RLG	P, M, F, O
	Bel–Pao	GAG AP RT RH INT	4–6	RLB	M
	Retrovirus	GAG AP RT RH INT ENV	4–6	RLR	M
	ERV	GAG AP RT RH INT ENV	4–6	RLE	M
DIRS	*DIRS*	GAG AP RT RH YR	0	RYD	P, M, F, O
	Ngaro	GAG AP RT RH YR	0	RYN	M, F
	VIPER	GAG AP RT RH YR	0	RYV	O
PLE	*Penelope*	RT EN	Variable	RPP	P, M, F, O
LINE	*R2*	RT EN	Variable	RIR	M
	RTE	APE RT	Variable	RIT	M
	Jockey	ORF1 APE RT	Variable	RIJ	M
	L1	ORF1 APE RT	Variable	RIL	P, M, F, O
	I	ORF1 APE RT RH	Variable	RII	P, M, F
SINE	*tRNA*		Variable	RST	P, M, F
	7SL		Variable	RSL	P, M, F
	5S		Variable	RSS	M, O
Class II (DNA transposons) - Subclass 1					
TIR	*Tc1–Mariner*	Tase*	TA	DTT	P, M, F, O
	hAT	Tase*	8	DTA	P, M, F, O
	Mutator	Tase*	9–11	DTM	P, M, F, O
	Merlin	Tase*	8–9	DTE	M, O
	Transib	Tase*	5	DTR	M, F
	P	Tase	8	DTP	P, M
	PiggyBac	Tase	TTAA	DTB	M, O
	PIF– Harbinger	Tase* ORF2	3	DTH	P, M, F, O
	CACTA	Tase ORF2	2–3	DTC	P, M, F
Crypton	*Crypton*	YR	0	DYC	F
Class II (DNA transposons) - Subclass 2					
Helitron	*Helitron*	RPA Y2 HEL	0	DHH	P, M, F
Maverick	*Maverick*	C-INT ATP CYP POL B	6	DMM	M, F, O

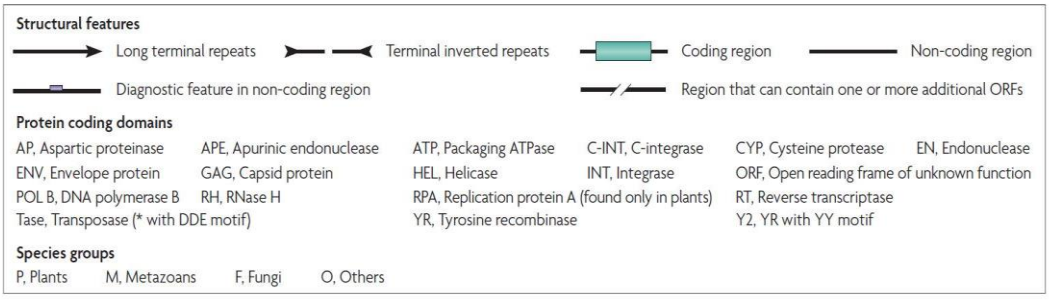


Fig. 1.1 Hierarchical classification system for TEs (taken from Wicker et al., 2007). The two main classes are subdivided in subclasses, orders and superfamilies. The superfamilies differ by functional features like protein arrangement and coding domains: The TSD, which is typical of each family, can be also used as diagnostic element.

I considered useful to describe this system of TEs classification since it differs from a commonly used one in my laboratory environment, although the difference is not substantial. This system, however, determines a rearrangement of the LTR retrotransposons group, which is the most abundant in plants, but, in particular, it contains the *copia* superfamily including the tobacco *Tto1* on which this thesis work is focused. For ease of presentation I will, as follows, first deal with Class II TEs (thus inverting a logical order), and subsequently extend more on Class I elements, particularly on LTR retrotransposons, taking *Tto1* as a model.

1.2 Class II TEs (DNA transposons)

These elements are ancient and prevalently occupy the genomes of bacteria, where they are known as IS (insertion sequences); but the also abundantly populate plants (*Ac/Ds, Mutator*) and animals, from insects to worms and humans. They are usually found in a low to moderate copy number which reflects their "cut and paste" replication mechanism. Elements of this class have no RNA as intermediate of replication, but they are subdivided in two subclasses, that are distinguished by the number of DNA strands that are cut during transposition (Wicker et al., 2007). To subclass I belong nine superfamilies of the classical transposons of the order TIR (terminal inverted repeats) characterized by TIRs of variable length. Their transposition is mediated by a self encoded DDE-transposase enzyme that excises the element from its previous locations and inserts it into a new one by cutting both DNA strands; therefore it also generates TSDs that are characteristic of each superfamily. Their insertion target sites seem to be limited to a small number of nucleotides (Kazazian, 2004); in particular *Tc1-Mariner* inserts into TA dinucleotides therefore integrating into a large number of loci. Another instance is given by *PiggyBac* that inserts into TTAA tetranucleotides (Fig.1.1). The poorly studied *Crypton* order, which is only found in fungi, is also included in subclass I and encodes a Y-recombinase, but lacks RT domain, therefore it is believed to transpose via a DNA intermediate (Goodwin et al., 2003).

The recently introduced subclass II contains the orders *Helitron* and *Maverick*.

Helitrons have been best characterized in maize, in which most are non autonomous derivatives. They transpose via a rolling circle mechanism, with only one strand cut and do not produce TSDs (Morgante et al., 2005). Interestingly, *Helitrons* have evolved the ability to capture gene fragments from the host genome; which has been suggested to be a means to evade silencing (Lisch, 2009; Morgante et al., 2005), as

will be reported later. The order *Maverick* has been found sporadically in diverse eukaryotes, but not in plants (Pritham et al., 2007). *Maverick* type elements are considered as giant elements since they can reach from 10 to 20kb and have long TIRs. Their transposition takes place via excision of a single strand followed by extrachromosomal replication and integration into a new site (Kapitonov and Jurka, 2006).

A typical aspect of DNA transposons is the "local hopping", that is the daughter copies, in most cases, insert in proximity of the parental insertion. In addition they also make "nested insertions" (Di Nocera and Dawid, 1983) in which transposition occurs into a proximal copy, which is likely the reason for the abundance of defective transposons. Although to a low level, DNA transposons can increase their copy number. *Ac* elements, for instance, excise during chromosome replication from a position that has already been replicated to another that the replication fork has not yet passed (Bennetzen, 2000; Greenblatt and Brink, 1962). Alternatively they can exploit gap repair following excision to create an extra copy at the donor site (Nassif et al., 1994).

1.3 Class I TEs (RNA transposons or retrotransposons

Also known as retroposons, retrotransposons are the most represented class of TEs, due to their "copy and paste" replication mechanism, which allowed them to reach very high copy number. They are considered to be the major contributors to the expansion of large genomes; this is particularly evident in plants were they can make up to 90% of the total DNA content (SanMiguel et al., 1996), while in animals they reach up to 45% of their genomes (Kazazian, 2004 and refs therein).

Their widespread presence has led to a debate whether they are simply genomic parasites or can also be beneficial by providing dynamic mechanisms of adaptation, which profoundly contributed to shape and re-shape the genomes of their host. This debate is also reflected by the numerous different designations that they have been given, from "selfish DNA" or "junk DNA" (Doolittle and Sapienza, 1980), to "controlling elements" (Davidson and Britten, 1979), "drivers of genome evolution" (Kazazian, 2004) or "genome's little helpers" (Symer and Boeke, 2010).

Retrotransposons are divided in five orders (Fig.1.1) comprising the well known LTR-retrotransposons and non-LTR elements LINEs, SINEs, plus the two more recent DIRS-like and Penelope like elements. Both LTR and non LTR retrotransposons are

found in all eukaryotic genomes, but LTRs are particularly abundant in plants, for example the *copia* elements BARE-1 from barley and Opie-1 and Huck2 of maize reach from 20,000 to 200,000 copies, while in humans the LINE-1 families have 100,000 copies and the SINE *Alu* counts up to 500,000 copies (Rowold and Herrera, 2000). Members of the order DIRS have been found in green algae, animals and fungi; they encode an RT, but integrate by a T-recombinase, therefore do not create TSDs. The Penelope order is found in *Drosophila virilis* and rarely in animals, fungi and plants (Evgen'ev and Arkhipova, 2005; Evgen'ev et al., 1997). They have LTR-like sequences that can be in either direct or inverse orientation; they a encode an RT and transposition is mediated by an endonuclease, with variable TSD.

1.3.1 Non LTR retrotransposons: LINEs and SINEs

By sequence analysis the LINEs (Long Interspersed Nuclear Elements) are presumably the most ancient order of retrotransposons (Xiong and Eickbush, 1990) and the most widespread in mammals. Their structure has been described in the archetype of this order: the human L1. It appears to be the integrated DNA version of an mRNA, since it contains a poly-adenylate tail at the 3' end. Two ORFs encode a nucleic acid binding protein (ORF1) with essential nucleic acid chaperone activity (Symer and Boeke, 2010 and refs therein) and an RT and an EN (endonuclease) which generates TSDs. Typically, genomic copies of LINEs are truncated from their 5' end. L1 is an autonomous element that transposes via a mechanism called target primed reverse transcription (TPRT) that appears to operate for most non LTR retrotransposons. In this mechanism, the full length transcript is exported to the cytoplasm and translated; the proteins bind the mRNA in *cis* forming the ribonucleoprotein complex that is transported back into the nucleus. Here the EN nicks a preferred genomic site, thus generating a free 3'-OH that is used by the RT to synthesize a single stranded cDNA copy (Luan et al., 1993; Symer and Boeke, 2010).

The SINEs (Short Interspersed Nuclear Elements) usually range from 80 to 500bp, are highly abundant in mammals, and generally rare in plants. SINEs are non autonomous, but did not originate from deleted class I elements. They present a poly-A tail like LINEs and rely on enzymatic activities encoded by LINEs, in particular ORF2p of L1 (Boeke, 1997) for transposition; therefore they also produce TSDs. The best known and probably most abundant of this order is *Alu*, with 500,000 copies in the human genome. *Alu*s have a dimeric structure; the 5' region contains an internal pol III promoter, which reveals their origin by accidental reverse transcription of RNApol

III transcripts, tRNA, 7SL RNA and 5S RNA. Their 3' region has unclear origin and can contain either an A or AT-rich domain, 3-5bp tandem repeats or poly-T, the Pol III termination signal (Kramerov and Vassetzky, 2005).

1.3.2. LTR retrotransposons

The LTR (Long Terminal Repeat) retroelements are less abundant in animals, but are the predominant order in the plant kingdom. They are found in all plant genomes including monocellular algae and bryophytes (Kumar and Bennetzen, 1999). They are variable in size, from a few hundred base pairs to exceptionally 25kb (*Ogre*). As showen in the TEs classification presented above (Fig. 1.1) LTR retrotransposons and retroviruses belong to the same order and share important structural and functional features. So far the investigation on LTR elements has heavily relied on retroviral models, as we will see in this work.

The LTRs can range from a few hundred bases to 5kb, and all have two conserved dinucleotides as inverted repeats, the 5' TG and the 3' CA (see Fig1.2d), that are important for the "processing" of the cDNA by the integrase (see below). The LTRs contain regulatory sequences that act as promoter (5' LTR) and as terminator (3'LTR) of transcription (Casacuberta and Santiago, 2003); this suggests that the 3' LTR might also promote the transcription of genes that are downstream of an inserted element (Kumar and Bennetzen, 1999). In the inserted copy the LTR has a structure composed by a U3 (unique 3'), an R (redundant), and a U5 (unique 5') sequence, while in the RNA only the R and U5 are present at the 5' end and R and U3 constitute the 3' end (see Figs. 1.2, 1.4 and 1.5). According to an accepted model for retroviral reverse transcription, called "LTR replication", R is necessary for the synthesis of the cDNA copy from the element's RNA genomic template (Fig. 1.4).

Other typical features also involved in reverse transcription are the PBS (primer binding site) located downstream of the 5' LTR and the PPT (polypurine tract) located upstream of the 3' LTR (see Figs. 1.2, 1.4 and 1.5b). LTR retroelement proteins are also structurally related to those of retroviruses: they also harbour a GAG and a POL domain contained in a single ORF. The GAG (Group specific antigen) codes for a coat protein (CP) involved in the maturation and packaging of the cDNA into the virus-like particles (VLPs). The POL domain encodes the enzymes necessary for the transposition: an aspartic protease (PR), a DDE transposase usually known as integrase (INT) and a reverse transcriptase with a RNaseH moiety (RT). Their role will be explained with more details in the next sections.

In the case of *Ogre* there is a second ORF, but its function is currently unknown (Neumann et al., 2003). Unlike retroviruses, LTR retrotransposons lack the ENV domain that encodes the envelope protein, therefore they do not have extracellular mobility.

Evolutionarily LTR retrotransposons and retroviruses are very close, and it has been suggested that LTR retroelements might have given rise to retroviruses by acquisition on the ENV protein and other additional and regulatory sequences (Frankel and Young, 1998; Seelamgari et al., 2004). The superfamily *Gypsy* is believed to be an ancestor of retroviruses. In support of this hypothesis the members of this superfamily have the same protein arrangement as retroviruses (Fig.1.1); and, interestingly, they can in some cases infect other individuals (Bucheton, 1995), showing therefore a possible extracellular transfer. However, it is also possible that, in a reverse process a retrovirus loses its extracellular mobility due to inactivation or deletion of the ENV domain (Capy, 2005), and generates and ERV (Endogenous Retrovirus). In another example, alternative splicing of the MLV (Murine Leukemia Virus), mRNA generate a shorter cDNA that was integrated constituting a new splice donor-associated retroelement (Houzet et al., 2003).
LTR retrotransposons are suggested to have originated from LINEs, which are the most ancient retrotransposons, by acquisition of LTRs (Bennetzen, 2000); but it has been also proposed that a fusion occurred between a DNA transposon and an LTR retrotransposon (Malik and Eickbusch, 2001). In plants, Bennetzen argued that this kind of retrotransposons might be retroviruses that were transmitted by insects feeding on gametophytic tissues. Although the cell wall would be a barrier to ENV-packaged retroviruses, these are able to replicate intracellularly and could have become LTR retrotransposons (or ERVs).

The best characterized superfamilies of this order are *Ty1-copia* and *Ty3-gypsy* (or just *Copia* and *Gypsy*, according to the new annotation system, Fig.1). They are named after the archetype of each family respectively from yeast (*Ty1* and *Ty3*) and from *Drosophila*. They share a number of features and basically differ by the protein position in the POL domain: in Copia the INT precedes RT/RH, while in Gypsy their position inverted. To the *Copia* superfamily belong many well known plant retrotransposons found in crops, such as, the barley *BARE-1*, the maize elements *Bs1* and *Opie-1*; *SIRE-1* in soybean, *Tos17* in rice and *Tnt1*, *Tto1* and *Tto2* from

tobacco. The *Athila* elements are the best known representatives of the *Gypsy* superfamily in Arabidopsis.
As already mentioned before, such elements reach an incredibly high copy number so as to constitute almost 90% of a plant genome. The genes in these plants are found like islands in a sea of repetitive sequences that might have probably also served to preserve them from the occurrence of mutations. However, it is undoubted that, at least for some very large genomes of *Triticae* (i.e maize, barley or wheat), LTR retrotransposons contribute to the "C-value paradox", that highlights the absence of correlation between the DNA content and complexity of an organism.

LTR retrotransposons of plants are generally defective and unable to transpose. Only the aforementioned *Tnt1*, *Tos17*, *Tto1* and more recently the tomato element *Rider* (Cheng et al., 2009) have been demonstrated to be able to carry out a complete transposition cycle. In contrast to such a small number, it is likely that other active retrotransposons will be discovered in the future, since some elements, even when active, transpose at a very low frequency (*Tto1*, or *Tos17*). In addition, transcripts of *Copia* retroelements have been found in a number of other species of agronomical interest (Hirochika and Hirochika, 1993), therefore this superfamily can be considered as a source for the discovery of other active retrotransposons.

1.3.3 MITEs (Miniature Inverted-repeates TEs)

A particular group of non autonomous TEs with a still indefinite evolutionary origin is that of MITEs, that share characteristics of both classes. MITEs are less than 600bp; their structure resembles that of defective DNA transposons by the presence of TIRs and the lack of coding capacity; but because of their high copy number and sequence size conservation they also seem to have a class I origin. Nevertheless, some evidence suggested that they might be a particular type of DNA transposons. In rice, for instance, tens of thousands *Stowaway* MITEs have been found to be activated by the transposases of some *Tc1-Mariner* elements (Feschotte et al., 2003) ; Other evidence comes from plants, nematodes, insects and fish where *PIF-Harbinger* control the activation of the *Tourist* element (Jiang et al., 2004).

1.4 *Tto1* (Tobacco transposon 1)

Transcription of *Tto1-1* was demonstrated for the first time in 1993 in protoplasts of the tobacco cell line BY2 (Hirochika, 1993). By RT-PCR on mRNA of the highly conserved RT (reverse transcriptase) domain, Hirochika found that the transcription of *Tto1* is highly activated during cell culture and also during tissue culture, in contrast to the first active plant retrotransposon *Tnt1* (also from tobacco) that was mainly activated in protoplasts (Grandbastien et al., 1989). *Tto1* copy number increased up to ten-fold in the cell line and only approximately two-fold in regenerated plants. Interestingly, the copy number of two other tobacco retrotransposons *Tnt1* and *Tto2* was slightly increased in the same cell line, but it was unchanged in the regenerated plants (Hirochika, 1993). The *Tto1* copy number in individuals of the same cultivars and in all tobacco cultivars analyzed is the same: approximately 30 copies per haploid genome, suggesting that *Tto1* transposition occurred very rarely during evolution. Interestingly, in yeast and *Drosophila* the number of retrotransposons can differ even between stocks of the same strain (Cameron et al., 1979; Strobel et al., 1979).

Sequence analysis showed that *Tto1* belongs to the (recently redefined) *Copia* superfamily of LTR retrotransposons and shares common features with retroviruses (Luciw, 1992) and with the other elements of the same taxon. For this reason we usually refer to retroviral models in our investigation on this retrotransposon.
Its total DNA sequence is 5.3kb long, and is flanked by two identical LTRs of 574bp. It also contains a PBS downstream of the 5' LTR and the PPT upstream of the 3' LTR. *Tto1* RNA ranges from ca 5.1kb to ca 4.7kb (Böhmdorfer et al., 2005; Hirochika, 1993), and contains one long ORF of 1338 amino acids (see Fig. 1.2). The single ORF contains the two typical domains Gag, encoding the coat protein, and Pol that encodes protease (PR), integrase (INT), and reverse transcriptase with an RNaseH moiety (RT/RH). Its transposition, mediated by INT, generally produces 5bp TSD, considered as the "footprint" of transposition events.

Tto1 life cicle

Tto1 life cycle is entirely intracellular and involves four main steps: transcription, translation, reverse transcription and integration (Fig. 1.3).
A *Tto1* pre-integrated copy is first transcribed into mRNA by the host encoded RNA polymerase II, therefore it has a cap at its 5' and a poly-A tail at its 3' end. The mRNA will serve both as a messenger and as a template for reverse transcription.

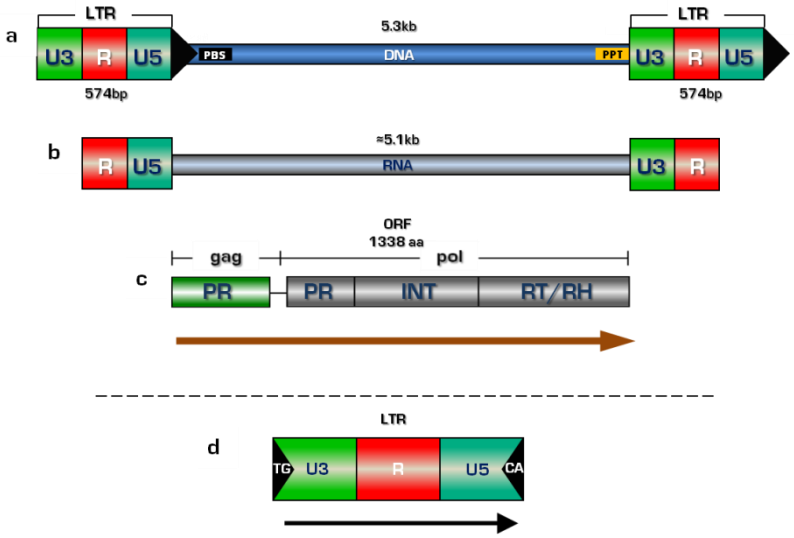


Fig. 1.2 *Tto1* nucleic acids, proteins and LTR. a) *Tto1* DNA is 5.3kb long. It is flanked by identical LTRs of 574bp, that have promoter function, at the 5', and terminator function, at the 3' end. b) The RNA is approximately 5.1 kb long, and starts at position 200 in the 5' LTR. It has a function as RNA and as genomic template. c) *Tto1* single ORF consists of 1338 aa, and is divided in Gag and Pol domains, that encode the coat protein (CP) and the poly-protein (see text).

The mRNA is transcribed starting from position 200 (Hirochika, 1993), in the 5' LTR, and ends in two major positions of the 3' LTR, 4914 and 5230 (Böhmdorfer et al., 2005). Böhmdorfer and colleagues have already well characterized the role of 5' LTR in initiation of transcription and translation; here I intended to identify which of the two termination points can give rise to transcripts that will used as a substrate for reverse transcription.

The translation of the mRNA generates the poly-protein that is assembled in the VLP (Virus Like Particle). *Tto1* can form VLPs of a different size, but the active ones measure appproximately 30nm (Böhmdorfer et al., 2008), suggesting that flexibility of VLP assembly can be a point of control on transposition. During maturation the PR cleaves the poly-protein, thus releasing Gag and the enzymes INT and RT/RH, which can then proceed to the next steps. Successively the RT will reverse transcribe the mRNA into cDNA and after disassembly of the VLPs, the PIC (Pre-Integrative Complex), constituted at least by cDNA and integrase (a dimer or a tetramer) will be transported back to the nucleus to be inserted into a new genomic location.

For the purposes of my research a particular emphasis will be only given to the reverse transcription phase and to integrase enzyme.

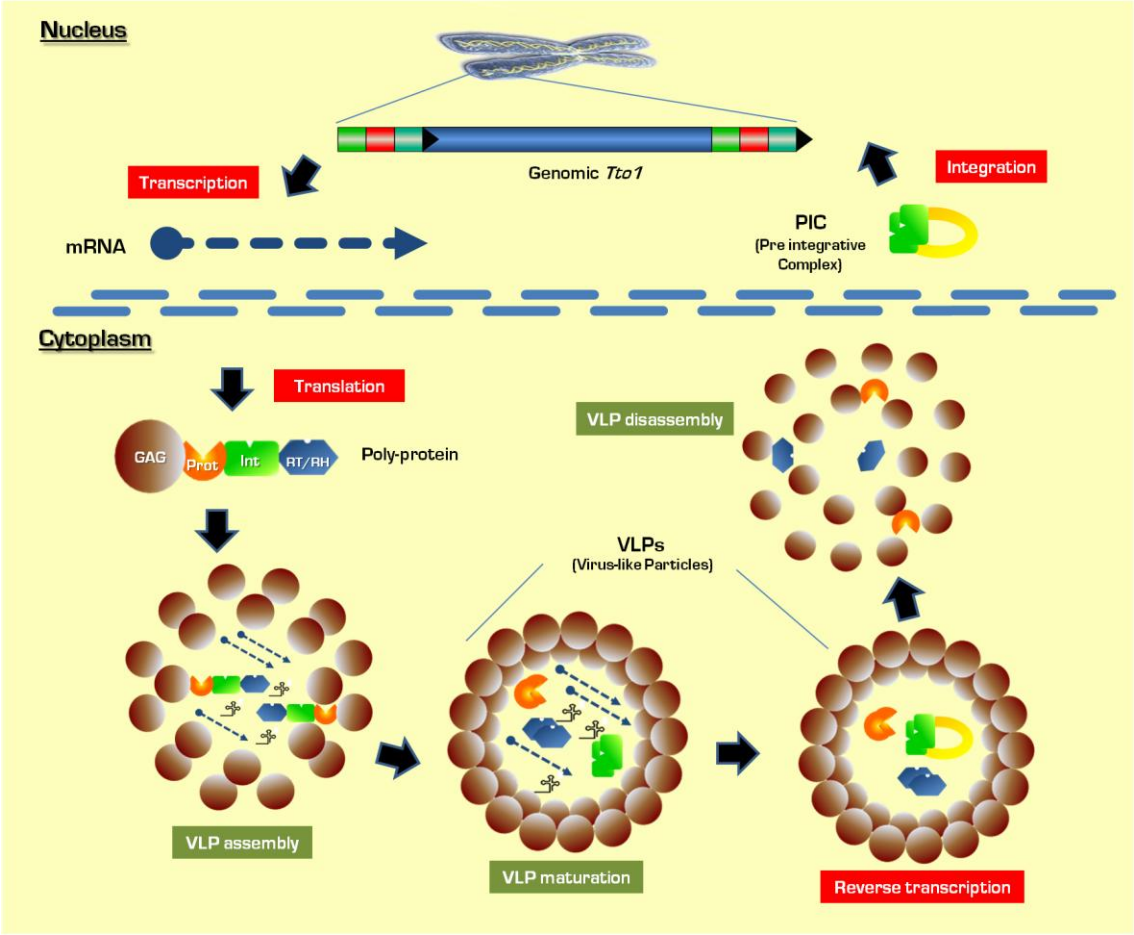


Fig. 1.3 The *Tto1* life cycle is entirely intracellular. A pre-integrated copy is first transcribed to form the RNA that will serve as a messenger for the proteins and template for genome replication. After translation, the VLPs are assembled in the cytoplasm. During maturation the PR will cleave the poly-protein in the single enzymes and RT will synthesize a new copy of cDNA. After VLP disassembly the cDNA will be transported to the nucleus to be inserted by INT in a new genomic locus.

Tto1 reverse transcription

Fig. 1.4 depicts the complex mechanism of reverse transcription as inferred from a model commonly adopted for retroviruses, named "LTR replication".

1. After maturation of the VLP, the free RT/RH initiates the cDNA synthesis from a cellular methionine-tRNA that hybridizes to the PBS on the RNA (see also Fig. 1.5a). The ensuing strand is also conventionally called leader.
2. The cDNA leader, also called strong stop DNA in the retrovirus convention, is elongated until the end of 5' LTR in the DNA. In the mean time the RNaseH moiety of the RT degrades the RNA of the heteroduplex.
3. The RNA degradation facilitates the first "jump" of the leader. The strong stop cDNA performs a template switch from the 5' end of the DNA to the 3' end of the mRNA.

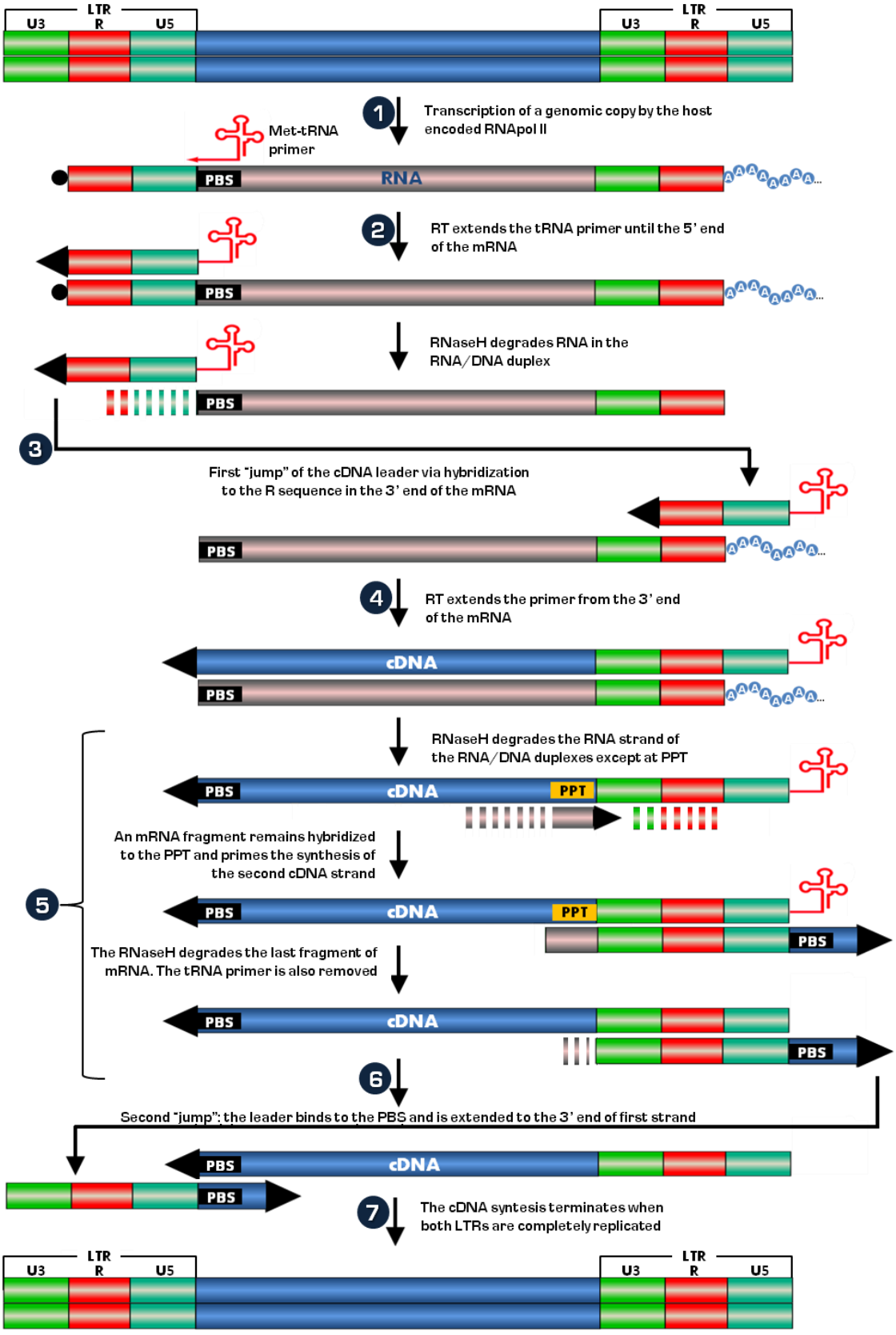


Fig. 1.4 *Tto1* reverse transcription (redrawn from Perlman and Boeke, 2004).

4. The first strand is then elongated by RT to the PBS, which constitutes now the 5' end of the RNA (U5 and R were degraded previously).
5. The RNase degrades all RNA in the duplex, except for a fragment of 13nt that binds the PPT (Fig. 1.5b) and will serve as primer for the synthesis of the second cDNA strand In this step the 3' LTR of the second cDNA step is elongated until the end and the initiator tRNA is removed.
6. The second jump takes place: the cDNA switches template again, from the 3' end of the first strand to its 5' end, hybridizing via the PBS.
7. The synthesis of cDNA is complete when the LTRs have been fully duplicated.

This model particularly emphasizes the importance of the redundant sequence (R), between 5' and 3' end of the RNA, during the process of the "first strand transfer". This step determines the synthesis of the first cDNA strand (minus strand) and consequently the production of functional new copies of the element that will be inserted into the host genome. In this work I will describe structural features of the R sequence and describe the possible dynamics involved in template switch and hybridization of the cDNA leader from the 5' end to the 3' end of the mRNA.

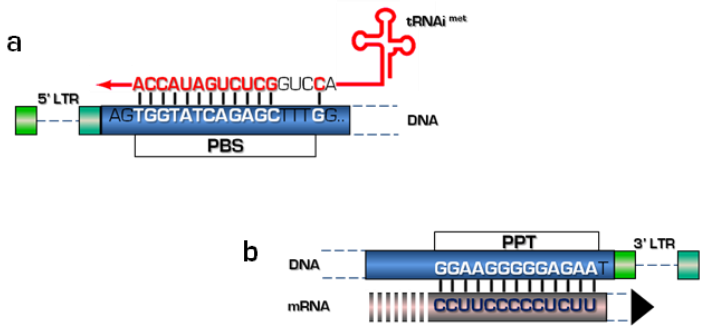


Fig. 1.5 a) PBS (Primer Binding Site), b) PPT (Poly Purin Tract)

1.5 Integrase

The integrase has the final role to deliver a new element's cDNA into the host genome. This enzyme is also responsible for the production of the target site duplications of all TEs that encode an INT. It hydrolyzes the cDNA phosphodiester backbone at the retrotranspon ends, resulting in the formation of 3'-OH, which are joined by a transesterification to the target DNA (Symer and Boeke, 2010).This reaction in *Tto1* generates staggered ends of 5 nucleotides, that are then repaired by the host machinery, thus generating the typical 5bp target site duplications (TSDs) (Katz and Skalka, 1994; Symer and Boeke, 2010).

Some structural aspects of this enzyme have been described in retroviruses, as well as its enzymatic activity. Integrase has three structural domains with a specific function. An N-terminal zinc-finger like motif (HHCC) is involved in dimerization and recognition of the LTR of the cDNA. (Katz and Skalka, 1994; Lewinski and Bushman, 2005). The recognition of the LTR is likely to be related to the "processing" of the emerging cDNA, in which the 3' ends of the linear cDNA are nicked at the TG/CA conserved dinucleotides, producing the CA-OH recessed ends that will be ligated to the chromosomal DNA. The dimerization is involved in the formation of the pre-integrative complex, but it is unclear whether two INT interact with both the cDNA and the host DNA or as in HIV-1 two additional molecules are involved in the formation of a tetramer (Li et al., 2006).

The central domain is involved in binding the cDNA and catalyzes the integration reaction itself. It is characterized by the highly conserved motif $DX_nD_{35}E$, typical of all TE transposases, that coordinates the divalent cations (Mg^{2+} or Mn^{2+}) necessary for the enzymatic activity.

A third C-terminal domain, called targeting domain (TD) has a role in directing the INT to specific genomic regions. TD of *Ty5* interacts with the heterochromatic protein Sir4, and that phosphorylation is required for this interaction (Dai et al., 2007). A single amino acid change in this region abolished targeting to silent chromatin and led to random integration of the element (Gai and Voytas, 1998).

The insights into retroviral integrase domains indicate that this enzyme is involved in different steps of the cDNA integration, which also requires an interaction with several factors. *Tto1* integrase is not yet known; moreover, previous experiments suggested that the integration step might be a point of control of transposition (Böhmdorfer et al., 2005). I therefore started investigating *Tto1* integrase, and made an attempt to obtain the purified enzyme. In addition, I was also interested in finding mutations in the region proximal to the $DX_nD_{35}E$ motif, which might influence the efficiency of this enzyme.

1.6 Control of TEs transposition

In a recent paper of outstanding interest, it has been proposed that three forces govern TEs evolution. Transposition control, TE sequence removal and population processes. (Tenaillon et al., 2010). The authors propose the analogy of a triptych, in which the lateral panels represent the first two forces, which cause mutation within

an individual, and central panel identifies the third force, which, by natural selection, determines the destiny of such mutations in the population.

It is likely that eukaryotes, after being vastly parasitized by TEs have evolved different mechanisms to control their transposition. An evidence of that seems to be that, concomitantly with their high copy number, most TEs are defective and unable to transpose. Gene silencing is apparently the most general and effective mechanism, and it generally operates a transcriptional and posttranscriptional level. TGS (Transcriptional Gene Silencing) is primarily activated by the presence of multiple copies (Casacuberta and Santiago, 2003); in *Drosophila* the severity of the repression correlates with the copy number of the element *Drosophila1* (Jensen et al., 1999). TGS is generally associated with DNA methylation. Hypermethylated promoters are a typical example of TGS of LTR retrotransposons, considering that their promoters are located in the LTRs, which are by definition repetitive. In plants, the repeated sequences of TEs are targeted by small interfering RNAs (siRNAs) that guide downstream protein complexes that initiate and maintain methylation of DNA and histones (Almeida and Allshire, 2005; Teixeira et al., 2009; Zhang, 2008) Consequently, hypermethylation increases the mutation rate rendering TEs inactive (Casacuberta and Santiago, 2003). It has also been shown that *Tto1* was specifically reactivated in an Arabidopsis *ddm1* mutant background (Hirochika et al., 2000). PTGS (Posttranscriptional Gene Silencing) is a sequence-specific RNA degradation that plants probably use against viral transcripts (Casacuberta and Santiago, 2003) or against transgenes.

According to the analogy with the triptych, the second lateral panel is constituted by the force of TE DNA removal. Evidence for this force comes specifically from the study of LTR retrotransposons. TEs removal is caused by unequal intra strand homologous recombination (UHR) between two LTRs of the same element, which leads to the so called “solo LTRs”. In addition, it has been suggested that LTR retrotransposons with sequence deletions might have been produced by illegitimate recombination (IR) (Devos et al., 2002).

Finally, Tenaillon et al. (2010) assign the central panel to population processes, which act as a sieve that determines whether the mutations produced by TEs will be advantageous or not.

The aspects of TE control involving sequence directed silencing mechanisms were of particular interest to me. In this work I have shown that the repetitive sequence contained in *Tto1*’s LTR can be reduced to an extent that does not affect reverse

transcription, but is likely to reduce repeat induced gene silencing defense mechanisms of the host.

1.7 Retrotransposons as plant mutagens

The use of TEs in plant mutagenesis is a current practice since about twenty years.
The maize DNA transposons *Ac/Ds* (Parinov et al., 1999) and *En/Spm* (Speulman et al., 1999; Tissier et al., 1999; Wisman et al., 1998) have been used in forward and reverse genetics in Arabidopsis.
In the last years LTR retrotransposons have been demonstrated to be a more powerful tool to generate mutations in plant genomes. LTR retrotransposons offer a number of advantages compared to DNA transposons. They can produce a large number of mutations, due to their "copy and paste" replication mechanism. The insertions are spread over the genome, in contrast to the typical "nested insertions" of DNA transposons, with the advantageous consequence that mutations can be easily segregated by genetic crossing to obtain single mutants.
Only a few active plant retrotransposons are known so far: *Tos17* of rice (Hirochika et al., 1996b), *Tnt1* (Grandbastien et al., 1989) and *Tto1* (Hirochika, 1993) from tobacco, which have already been used in tissue culture-induced gene mutagenesis.
Tos17 is very well studied in rice (Hirochika, 1997, 2001; Miyao et al., 2003), but *Tnt1* and *Tto1* are also active in other species. *Tnt1* has been used for insertional mutagenesis of Arabidopsis and of *Medicago truncatula* (Cheng et al., 2011; Lucas et al., 1995). *Tto1* can efficiently transpose in Arabidopsis and in rice (Hirochika et al., 1996a; Okamoto and Hirochika, 2000); in addition low reverse transcription activity of *Tto1* has been detected in barley (Böhmdorfer, 2005). Importantly, all the aforementioned retrotransposons show an insertion preference into genes and generally into euchromatic regions with high transcriptional level (Böhmdorfer et al., 2010; Okamoto and Hirochika, 2000; Yamazaki et al., 2001). It is therefore imaginable that the application of these elements will be extended to a higher number of plants including crops. Nevertheless, the necessity of tissue culture and plant regeneration is a long and tedious procedure and also has the disadvantage to induce the transposition of other TEs, otherwise silent, resulting in unwanted somaclonal variation. In this work I will show that *Tto1* is a perfect candidate to create an upgraded LTR retrotransposon plant mutagen that can be activated in the whole plant, with a very simple procedure and does not need any *in vitro* manipulation.

1.8 Different approaches to biology

In this thesis, I will also show that molecular engineering of *Tto1* can be performed following a synthetic biology approach. Two ways of approaching biology in a "synthetic" way have been followed in this work. The first proposes redesigning life, by creating "synthetic" biological systems that should be used to study biology by comparing their predicted behavior to that of natural biological systems. *Tto1* in this work has been "redesigned" to exploit its natural mutagenic potential, and to create a model for functional studies on retrotransposon replication and transposition control by the host.

The second proposes the construction of "synthetic" biological systems by assembling "interchangeable" parts with a biological origin. In my thesis, I made the attempt to identify one possible interchangeable part by analyzing and replacing *Tto1* integrase core domain. I intended to apply this new approach of bioengineering to the construction of a new mutagenic tool with improved performance.

1.9 Model plants used in this work

1.9.1 *Arabidopsis thaliana*

A. thaliana is a small angiosperm belonging to *Brassicaceae* family. Although closely related to important commercial plants as cabbage, broccoli, turnip and rapeseed, Arabidopsis is not an economically important plant, but it has become the most common model system for research in plant biology. Several characteristics made it the primary choice in compared to other model plants as rice, maize, tomato, barley, petunia and so on. Arabidopsis has a small size and is a self-pollinating plant that produces a large amount of seed in a relatively short generation time of approximately 8 weeks. All these features make it particularly practical for laboratory use. In addition it has a broad natural distribution throughout Europe, Asia and North America, so that many ecotypes have been collected from natural populations, and are available for physiological studies.

Furthermore, Arabidopsis has the smallest genome known among higher plants (Okamoto and Hirochika, 2000): 125Mb, containing approximately 25,000 genes distributed over five chromosomes. It is consequently particularly suitable for genetic engineering and for genome analysis.

Arabidopsis' genome contains a generally low number of TEs DNA, dispersed on all five chromosomes (Terol et al., 2001). It has been estimated that DNA transposons and MITEs constitute up 6% of the genome (Feschotte et al., 2002) while

retrotransposons range from 4 to 8%. (Casacuberta and Santiago, 2003). Such small percentage compared to other angiosperms suggests that Arabidopsis might have evolved an efficient mechanism of control of TEs. The transposable elements removal force, operating through homologous recombination with consequent "solo LTR" formation, seems to be quite efficient in Arabidopsis, resulting in a constant turnover of transposable elements that contributed to the small size of its genome (Pereira, 2004).

Arabidopsis is therefore a good model plant for my specific field of research. In addition, most insights in plant biology that have been obtained using this model plant seem to apply to other species. *Tto1* has been demonstrated to efficiently transpose in rice and Arabidopsis, that is, in a monocot and a dicot plant respectively, suggesting that control factors are conserved between such distant classes of plants. Thus, Arabidopsis can be a good starting point for investigations aiming at a wide use of retrotransposon mutagenesis in plants.

1.9.2 *Hordeum vulgare*

In this work I made the attempt to obtain constitutive expression of *Tto1* in barley, which represents both a monocotyledonous and an important crop plant.

Barley (*H. vulgare*) is a grass of the family Poaceae and has been one of the first cereals domesticated in the Fertile Crescent. It is widely cultivated in all temperate regions from the Arctic Circle to the tropics and is largely used in food production as well as an animal fodder. In 2007 barley ranked the fourth worldwide both in terms of quantity produced (136 million tons) and in area of cultivation (566,000 km^2) (FAOSTAT, 2007).

It is a self–pollinating species with a high degree of natural and easily inducible variation, ease of hybridization and wide adaptability to growth conditions. It has one of the largest genomes among higher plants: 5,000Mb distributed on 14 chromosomes, over 85% of which constituted by TEs (Wicker et al., 2005). Importantly, barley is a diploid species with a high level of synteny with other grass genomes especially with its hexaploid relative wheat; therefore it can be a suitable model to study the physiology of grasses. Barley is a particularly drought tolerant species, and it is already being used, by many groups to study the molecular basis of adaptation to drought.

1.10 Aims of my PhD work

In my thesis work I intended to conduct an investigation on *Tto1* life cycle from different points of view. Using sequence engineering, I wanted to gain the knowledge on its main controlling factors, in order to develop an "easy-to-handle" tool for plant insertional mutagenesis and contribute to broaden its range of applications to other plant species.

Many aspects of retrotransposon biology remain to be unraveled; therefore I mainly aimed at shedding light on some of those aspects that could be of immediate use in further investigation, in order to more extensively match to researchers' needs.

In the model plant Arabidopsis, I wanted to assess the mutagenic efficiency of an engineered *Tto1*, and use it to investigate molecular aspects of reverse-transcription. At the same time, the analysis of particular sequence features involved in reverse transcription should open the way to investigate mechanisms of control that reduce transposition rate in the host plant in the future.

I also wanted to investigate *Tto1* on the protein level, and attempted to obtain a purified integrase to perform biochemical analysis. In addition I combined sequence analysis with a synthetic biology approach, in order to identify "interchangeable parts" that should be employed to increase the integration efficiency of *Tto1*, thus contributing to create a mutagenic retrotransposon with improved performance.

These attempts were successful to a variable extent. As a consequence, several ways are left to be further explored and built anew; however I believe that the results obtained so far will give the chance to do so.

Chapter 2

RESULTS

2.1 Inducible *Tto1*

Transposition of the majority of plant retrotransposons is activated by various biotic (such as pathogen infection) and abiotic stresses (wounding, methyl jasmonate, cell culture etc. (Brookfield, 2005; Feschotte et al., 2002; Grandbastien M., 1998; Sabot and Schulman, 2006). Plant retrotransposons analyzed so far all contain a stress-inducible promoter, which links their transcription to adverse growth conditions. In tobacco cultured cells the expression of *Tto1* was indeed greatly increased (Hirochika, 1993), but consequently making tissue culture, as well as an efficient regeneration protocol, necessary to perform studies on retrotransposition and for the application in insertional mutagenesis.

Different features contained in the long terminal repeat (LTR) control however *Tto1* activity. It has been hypothesized that a complex hair-pin structure of *Tto1* mRNA 5' region, might down-regulate translation, by controlling the access of the ribosome, during non-stress conditions. The removal of control sequences and their replacement with a heterologous promoter responsive to chemical inducers made it possible to obtain an inducible *Tto1*, which allows transposition "on demand" in the whole plant, with full transposing ability and no need for regeneration.

Tto1 with a deletion until nucleotide 172 of its 5' LTR, was appended with a β-estradiol inducible promoter. Δ172 *Tto1* has also been provided with two Arabidopsis introns that interrupt the reading frame, but are lost when mRNA splicing occurs, offering a very efficient way to monitor reverse-transcription (Böhmdorfer et al. 2005, 2008, 2010). In Fig. 2.1 the features of engineered *Tto1* are shown.

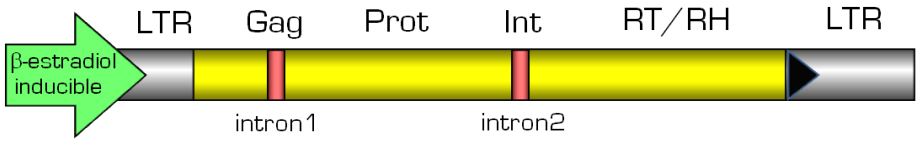


Fig. 2.1 i*Tto1*. 5' LTR was shortened until nt 172, the natural stress responsive promoter was replaced with a heterologous chemically inducible promoter. Two Arabidopsis introns have been inserted into Gag and Int domains to monitor reverse-transcription.

The engineered *Tto1,* which will be referred to as i*Tto1*, where i stands for inducible, was used in the following studies. In my work it has been the basis to investigate *Tto1*'s potential as a new plant mutagenesis tool, and in parallel to explain the role of the 3' Long Terminal Repeat in *Tto1* sequence replication and in termination of transcription.

2.2 Double nature of the 3'LTR in reverse-transcription and in termination of transcription

As mentioned above, the function of *Tto1* 5' LTR was described in previous studies and part of its sequence was replaced by sequences of interest (Fig. 2.1), showing its role in providing transcriptional and translational regulation signals. Equally, *Tto1* 3' LTR plays a crucial role, providing termination signals and a stretch of homologous sequence between the two ends of the element's mRNA, called R (redundant) region . The two functions are unavoidably linked: a full replication cycle can only take place if the mRNA is first translated into the element encoded enzymes necessary for transposition. Then RT uses the sequence redundancy between 5' and 3' RNA ends to complete the synthesis of new cDNA. In this work, I have investigated both functions

2.3 *Tto1* as a tool for mutagenesis of Arabidopsis

2.3.1 *Tto1* is active in the heterologous host Arabidopsis

In transgenic Arabidopsis regenerated plants 123 out of 165 (74%) independent *Tto1* insertions, driven by its natural LTR promoter, occurred into active coding sequences spread all over the five chromosomes, (Okamoto and Hirochika, 2000). In further experiments one insertion of a 35S promoter driven *Tto1* was detected in a gene encoding a ubiquitin fusion degradation pathway protein of the UFD1 family, which is a subunit of the ubiquitin chain binding complex CDC48 (Böhmdorfer et al., 2005). These results suggested that both the natural and the manipulated *Tto1* preferentially integrate into actively transcribed genes.

2.3.2 Transposition "on demand" of *Tto1* in Arabidopsis

The first attempt to build an inducible construct was made using a *Tto1* construct carrying a Dexamethasone responsive promoter (Böhmdorfer et al., 2005). In this

experimental case though, while Dexamethasone induction was very strong, toxic lethal effects to the plants, were described. Such toxicity was also reported in different publications (Andersen et al., 2003; Kang et al., 1999; Ouwerkerk et al., 2001). As we aim at exploiting the gene preference of *Tto1*, in a way not to kill the host, the engineered *Tto1* was linked to another inducible promoter: the β-estradiol promoter of the plasmid *pER8* (Zuo et al., 2000), to create the construct *pERnew::Tto1* (see Fig. 2.1). *pERnew::Tto1* was first electroporated into *A. tumefaciens* C58C1; then recombinant Agrobacteria were used to transfer i*Tto1* to Arabidopsis via floral dip transformation. The plants were selected on solid Ara medium containing 15mg/l Hygromycin, and the presence of the transgene was monitored by intron-PCR (Fig. 2.2). This particular application of the PCR is very efficient to detect reverse transcription events. Employing primers that bind a sequence flanking the intron – either intron1 or 2 depending on the specific case – it is possible to distinguish plants in which Tto1 mRNA is reverse transcribed and plants that still carry only the T-DNA transgene. A shorter band is amplified from the reverse transcribed cDNA that has lost the intron, compared to a longer band that is amplified from the T-DNA borne element.

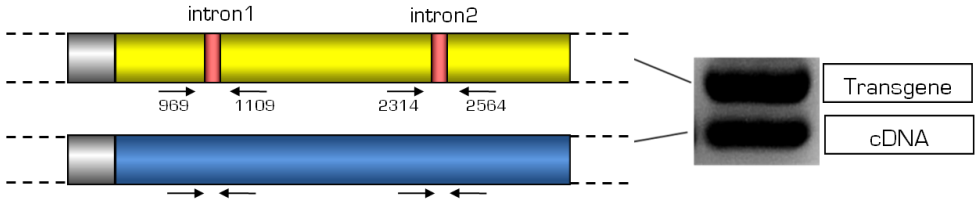


Fig. 2.2 Intron-PCR to monitor reverse transcription of 3'ΔLTR constructs. In a PCR where oligos flanking the introns are used, a double band is amplified if reverse-transcription takes place. A higher molecular band will be amplified if T-DNA mother copy is used as a template; a smaller band will be amplified from spliced and reverse transcribed new cDNA copies of the element.

T_2 seeds were germinated in liquid MS containing β-estradiol in order to optimize the exposure of i*Tto1* plants to the inducer. The seedlings were grown in liquid culture for two weeks and those which survived were transferred to soil to develop into mature plants; the induction method is illustrated in Fig. 2.3a.

A diagnostic intron-PCR was performed on DNA extracted from cauline leaves of mature plants, to test whether induction had taken place. Cauline leaves were used to detect *Tto1* copies contained in cells that originated from meristematic cells that were directly exposed to the inducer. In Fig. 2.3b a typical experiment is shown, where DNA of cauline leaves was analyzed by Intron1-PCR. Out of 70 plants tested, 3 were showing the intron-less band indicative of *Tto1* cDNA. The progeny of the 3

candidates were re-screened by intron-PCR (not shown), and crossed to wild-type Col-0 plants lacking the *Tto1*.

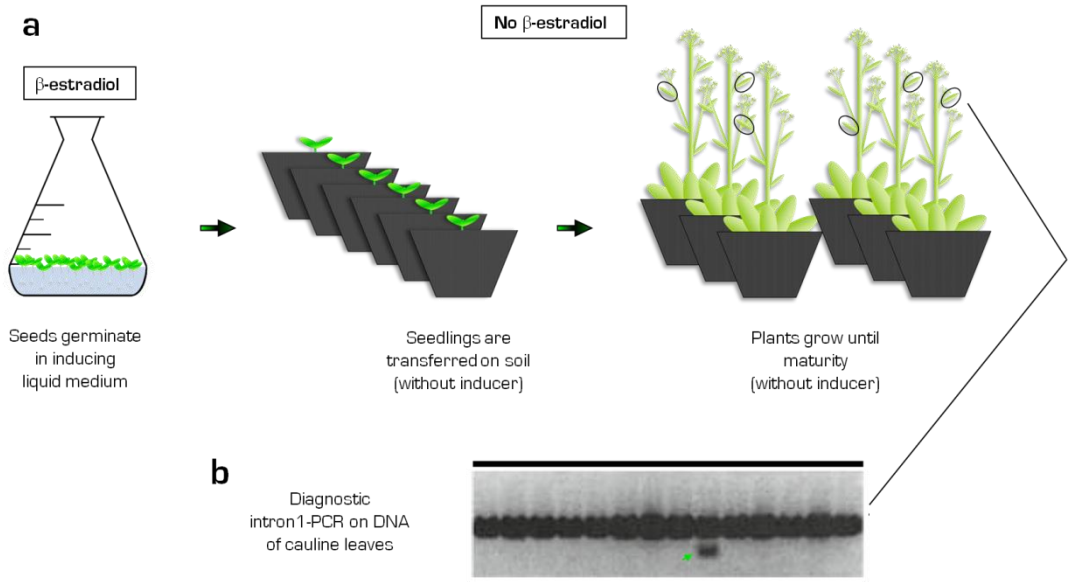


Fig. 2.3 Experimental method of i*Tto1* induction. a) Seeds were germinated in liquid culture containing β-estradiol. After two weeks the seedlings were transferred to soil, in absence of β-estradiol and grown until maturity. b) DNA from cauline leaves of mature plants was used in intron PCR to screen for the presence of intron-less bands indicative of chemically induced transposition of i*Tto1*. Progeny of the plants containing either only the spliced band or both bands was subsequently outcrossed to Col-0 and analyzed by Southern blot.

The DNA of the outcross progeny was then used in another intron-PCR to show the segregation and subsequently in a Southern blot to visualize new insertions events (Fig. 2.4).

Fig. 2.4a reports the variable heterogeneous result of the intron-PCR on the outcross progeny. While no i*Tto1* derived band was detected in lane 1 and 6, lane 3-9 only showed the band derived from T-DNA. In lane 2 only the intron-less band was detected, which can always be derived from an already integrated copy. Lanes from 10 to 12 showed the typical situation in which both transgene copy and intron-less cDNA copy are present. 1 and 3 progeny from two independent lines respectively were investigated for the presence of new insertions. The genomic DNA was digested by *EcoRI* and *HindIII*, and separated 16h on agarose gel; a probe binding to the 3' end of *Tto1* (see §4.3.2.6, §4.3.7.4 and Appendix 4.S-C) was used to detect i*Tto1* related bands. The band pattern visualized in lane 3-6 is the result of the segregation occurred in progeny of the outcross. The different bands correspond to different insertion sites in the Arabidopsis genome. Interestingly the band corresponding to the original i*Tto1* mother element was missing in some of the progeny and could be seen

only in lane 6, as a proof that insertions had segregated. No band was detected in the empty Col-0 negative control. The gel blot result is presented in Fig. 2.4b.

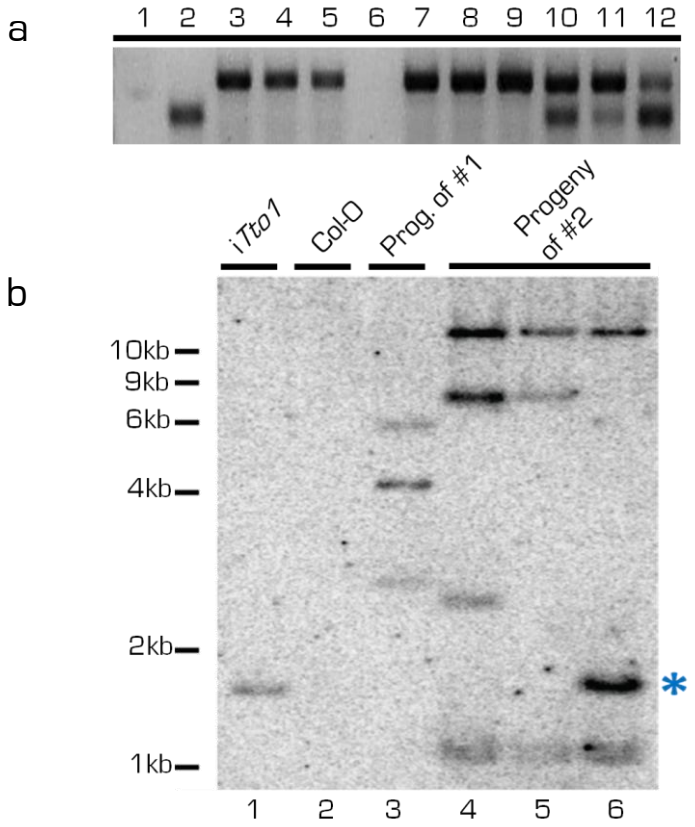


Fig. 2.4 Segregation of i*Tto1* insertions. a) An Intron-PCR on DNA of 12 independent progeny of the outcross showed segregation of *Tto1*. b) A Southern blot was performed on the outcross progeny of line #1 and #2. In lanes 3-6 a different band pattern shows that newly inserted i*Tto1* copies segregate independently. The band corresponding to the i*Tto1* donor construct in lane 1 (asterisk) is only present in lane 6, and absent from the negative control Col-0 in lane 2.

2.3.3 i*Tto1* tranposes into genes

To test then whether i*Tto1* maintained its insertion preference into genes, some of the insertions flanking sites were sequenced. The method used to isolate and sequence the insertion sites is described in more detail in §4.3.17. The DNA loaded in lanes 3-6 of Fig. 2.4b was re-run and the gel portions corresponding to the band size detected by the radioactive probe were extracted and purified from gel. They were subcloned in the plasmid *pSKII* and transformed into *E. coli* Stbl4 to make a library of clones harbouring the newly inserted *Tto1* copies and the bordering genomic regions. The library was screened by colony hybridization and two clones were bound by the *Tto1* probe. Two insertions of line #2 were identified: one occurred between the annotated genes At2g26410 and At2g26420, which correspond to a calmodulin binding protein and to the PIP5K3 (Phosphatidylinositol Phosphate 5-Kinase 3) respectively; the second one was found in the gene At3g14480, corresponding to a glycin/proline rich protein, a likely cell wall component, (see Fig. 2.5). Interestingly the two insertions are on two different chromosomes, suggesting that i*Tto1* can potentially cover the whole genome, and confirming previous results.

Both insertions also presented the typical target site duplication (TSD) that is considered the "footprint" of a retrotransposition event, as a final proof of a new insertion. TSD sequences are also reported in Fig. 2.5 right panel.

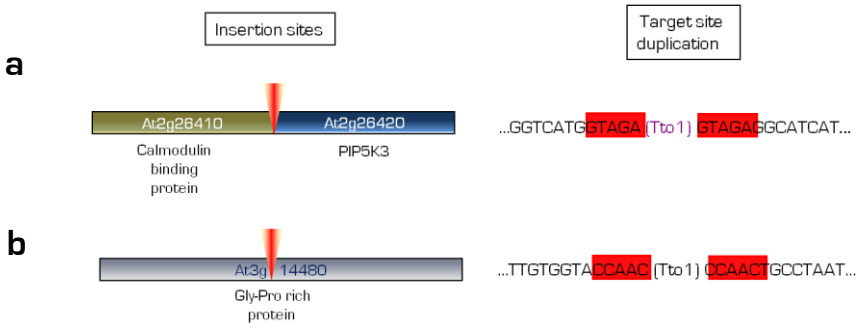


Fig. 2.5 Two new retrotransposition events were characterized. Both occurred into regions with high transcription rate. a) The first insertion occurred in the intergenic region between genes At2g26410 and At2g26420 (left). b) The second insertion was detected in the gene At3g14480 (left). The sequence of the typical TSD is also reported (right).

2.4 Analysis of the 3' Long Terminal Repeat

As mentioned above, some aspects of *Tto1* 5' LTR function initiation of transcription were described by previous studies (Böhmdorfer et al., 2005). In parallel, *Tto1* 3' LTR plays a crucial role, providing termination signals and a stretch of homologous sequence between the two ends of the element, called R region. These two functions are linked in a way that a full replication cycle is obtained when the mRNA is first translated into the enzymes necessary for transposition and then, during reverse transcription, a stretch of repetitive sequence between 5' and 3' RNA ends will be used to complete the synthesis of new cDNA. Both functions have been investigated in this work.

2.4.1 Role of 3'LTR in reverse-transcription

An accepted model for retroviral replication, (see Fig. 1.3) proposes that the R region is indispensable to achieve a complete replication of the viral/retrotransposon DNA. The redundant sequence contained in the LTR provides a hybridization site for the "strong stop cDNA" leader to the 3' end of the RNA. In order to elucidate this key point of reverse-transcription, in this work the R region of *Tto1* 3' LTR has been mapped and dynamic aspects of its mechanism of action were described.

2.4.2 *Tto1* Δ 3' LTR constructs

The first step was generating a set of five inducible *Tto1* constructs with a deletion in the 3'LTR (Fig. 2.6). The set of deletions is schematically shown in Fig. 2.6a, in comparison to the retroviral LTR model; panels b-d show sequence modifications of the 3' LTR, and of 5' end of *Tto1* respectively.

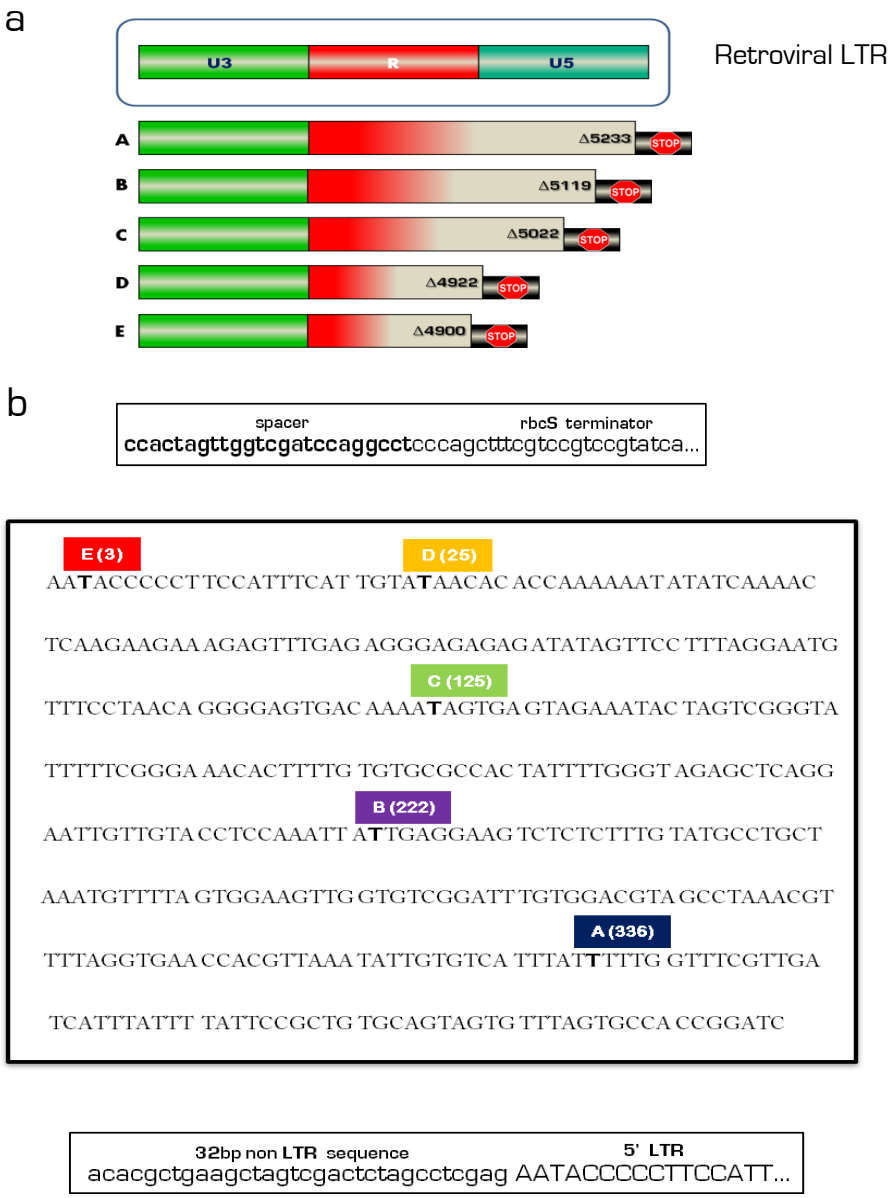


Fig. 2.6 *Tto1* Δ3' LTR constructs. a) Schematic representation of the set of deletions. The shading red color indicates uncertainty about the borders of R region. b) In all constructs, the last residue of the LTR is followed by a spacer sequence (bold, small case letters) and by the transcription termination sequence of pea rbcS-3A gene (small letters). c) Sequence of 3' LTR end. For each construct, represented by a different color, the indicative letter (A-E) is reported above the end position (bold T). In parentheses the number of redundant base pairs between the two LTRs is indicated. (d) Sequence of the 5´ end of the mRNA of all engineered constructs: a 32nt extension (small letters) precedes the LTR sequence (capital letters).

The five constructs were named A, B, C, D, and E from the least to the most extensive deletion; they all end with a T (arbitrarily chosen), in bold under each letter. The numbers in parentheses indicate how many base pairs of sequence redundancy

between 5' and 3' LTR are contained in each construct. The deletion end points were chosen accordingly to the previously mapped termination point of *Tto1* mRNA. Interesting is that the shortest termination site (nt 4914) was the most represented while the longest (ending at nt 5233), was the least abundant. The rbcS terminator was linked to each construct for providing a strong termination signal to deleted LTRs in which termination might be disrupted.

All cloning steps (described in §4.3.16.1) were carried in *E. coli*; however two different strains were used. The *pSK–* constructs were propagated in XL1-blue cells. For propagating the *pER8* constructs, which contained some redundant sequence of the LTR, the Stbl4 *E. coli* strain was used instead, in order to maximize the stability of the direct repeats (LTR) that often resulted in recombination in XL1 blue.

2.4.3 Generation of *Tto1* Δ 3' LTR transgenic Arabidopsis

Approximately 2μg of each construct were electroporated into *A. tumefaciens* strain C58C1 (Rif^+) and the recombinant clones were selected on a medium containing double antibiotic (Spectinomycin and Rifampicin). A digestion control was used to screen for the correctness of the constructs. The correct recombinant *Agrobacterium* clones were then used to insert the constructs into Arabidopsis Col-0 plants by the floral dip method.

The progeny (T_1) of floral-dipped plants (T_0) was selected on sand ¼ MS medium + Hygromycin (see §4.3.13.2), and afterwards transferred on soil, under greenhouse conditions, until next generation. T_2 plants of each line were grown in liquid Ara medium containing the chemical inducer β-estradiol for two weeks and tested by Intron1-PCR (Fig. 2.2). The plants were screened for the difference in expression level after β-estradiol induction, because the random integration of T-DNA can results in a broad range of expression level. In addition the intensity of the spliced band was used as a screening parameter to select the best expresser lines. Fig. 2.7 reports a typical test in which fifteen transgenic T_2 progeny for each construct were analyzed. The lines showing higher abundance of cDNA were chosen for next experiments (green arrow). T_3 progeny of the best expressers were then grown with β-estradiol to test *Tto1* expression, in order to select the best line for each construct.

2.4.4 "Long-PCR": a new screening approach

At this point the "Long-PCR" (see § 4.3.2.4) approach was taken to monitor *Tto1* complete reverse-transcription.

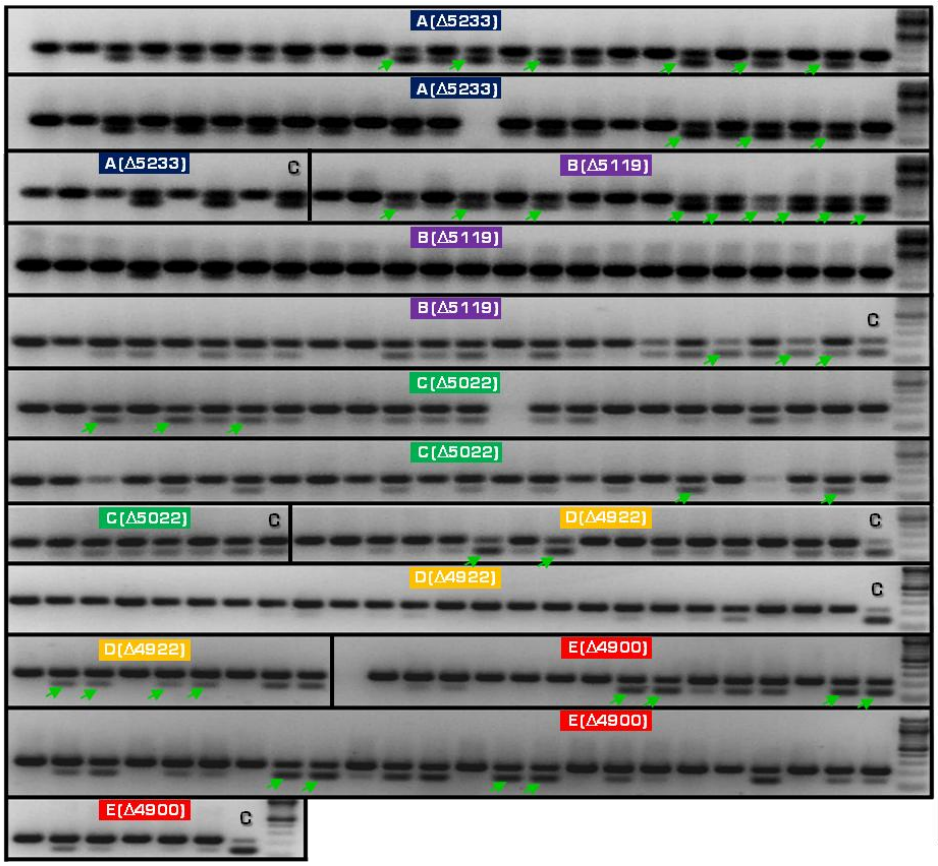


Fig. 2.7 Selection of best *Tto1* expressers from T_2 progeny. Fifteen T_2 lines for each deletion were tested by Intron1-PCR, and visualized on 2% agarose gel. The experiment was run in triplicate for constructs A, B and C and in duplicate for C and D. For each line a non induced control was done; and a positive control is also shown (indicated by C). The induced plants show the typical double band, where the lower one confirms that reverse-transcription has happened. Among the positive results we selected the one with higher ratio cDNA/T-DNA, as indicated by the green arrows. In order to distinguish the results concerning different constructs, colored boxes, containing indicative letter and the deletion end point are added to the picture.

We can hypothesize in fact that *Tto1* mRNA gives rise to aberrant priming of the mRNA 3' end with an internal sequence, downstream of the intron, so that an intron–less band might always be produced, independently of a complete reverse-transcription (Böhmdorfer et al, 2005).

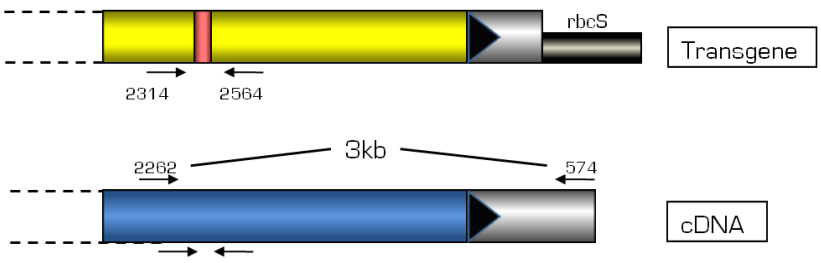


Fig. 2.8 Long-PCR to monitor complete replication of 3' end. a) Schematic representation of Long-PCR principle. In a PCR where oligos binding at position 2262 and 574 are used, a 3kb fragment is amplified if *Tto1* 3' LTR has been fully replicated during reverse-transcription. The oligo binding at position 574 reads up from the last nt of LTR therefore it can only bind if a full length new cDNA copy is synthesized.

With the Long-PCR approach we intended to monitor the complete reconstitution of *Tto1* LTR from deleted constructs by RT, into a new cDNA copy. In Fig. 2.8 the principle of this assay is schematically shown. To normalize the result of this test, an Intron2-PCR (§4.3.2.3) was first performed on an amount of starting DNA template, from the five best expressers lines assessed before, such that each construct gave the same amount of cDNA (Fig. 2.9a).

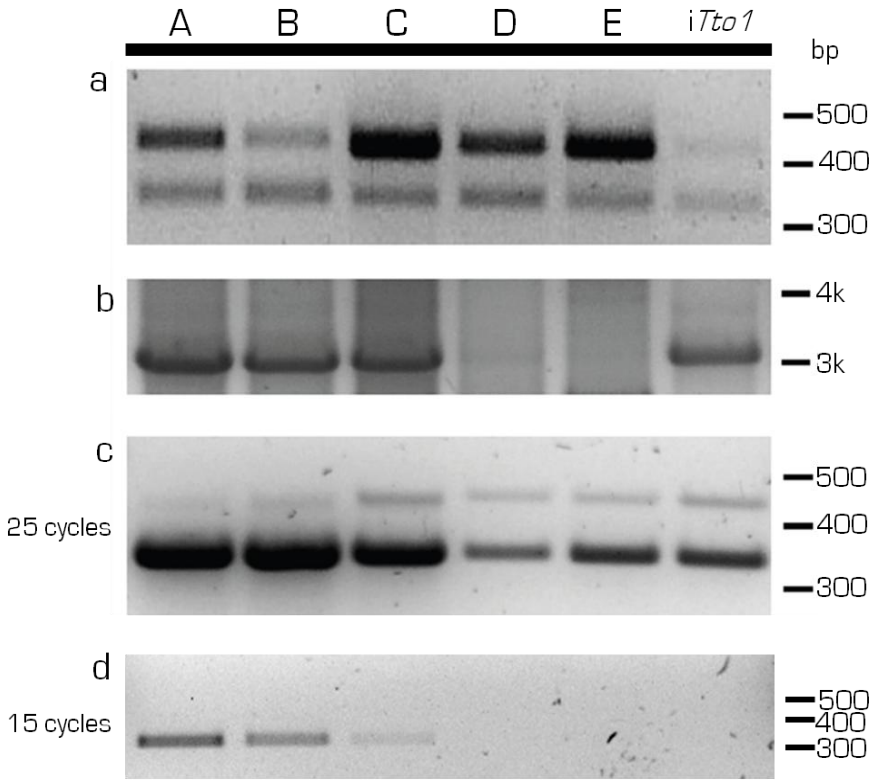


Fig 2.9 a) Intron2-PCR to normalize the test on the cDNA, b) Long-PCR testifies complete replication of 3' LTR only for constructs A, B and C, which contain the longer portions of R region. c) The 3kb band from panel c was gel purified and used in Intron2-PCR. Spliced bands are amplified more abundantly in constructs A, B and C indicating the predominance of cDNA with fully replicated LTR and confirming aberrant priming of the more extensive deletions. d) Same as c) with 15 amplification cycles, to emphasize the predominance of full length LTR cDNA. i*Tto1* indicates the positive control.

The same amount of DNA used in Fig. 2.9a was then subjected to Long-PCR using oligos T2262-2283dn and T574-548up (Fig. 2.9) and the Koncz-dip program for amplification. The oligo T574-548up is complementary to *Tto1* from the last nucleotide of the LTR and should only bind if the 3' LTR is fully reconstituted in cDNA, thus leading to the amplification of a 3kb band. According to the results of Fig. 2.9b, constructs A, B and C produced the expected band, suggesting that the redundant sequence contained in their LTR was long enough to support reverse transcription and to synthesize a full length cDNA. In contrast D and E presented respectively a

minimal amount of PCR product and no band, appearing to carry a too short sequence overlap to support complete reverse transcription (Fig. 2.9c). To confirm that the emerging cDNA was properly spliced, the 3kb bands of A, B and C, and parts of the gel corresponding to the same position of D and E were gel purified and 2.5µl of a 1:10,000 dilution of the eluate were used in another Intron2-PCR (panel c). Interestingly all constructs showed the double band. In the case of shortest constructs D and E this was attributed to the amplification of contaminating single strand cDNA emerging from aberrant fold-back priming of the mRNA, as discussed above. For the longer constructs A, B and C the result confirmed that of the previous "Long-PCR". The presence of little amounts of the larger band still suggested amplification from aberrant cDNA, but the great difference in intensity between the cDNA band and the T-DNA derived band indicated that the concentration of full length cDNA was however much predominant, and confirmed the selectivity of this PCR approach. To have a final proof of the predominance of cDNA with reconstituted LTR of construct A, B and C compared to D and E, another Intron2-PCR was done with only 15 amplification cycles. Also in this case the expectation was confirmed; the cDNA derived band was in fact only amplified from constructs A, B and C. The absence of the characteristic band from the positive control lane (i*Tto1*) was due to general low amplification profile of the control in this Intron-PCR experiments; but the observed predominance of the intron-less band (a and c) still correlated with the expected behavior.

2.4.5 Visualization of full length *Tto1* cDNA

As PCR was useful monitoring reverse-transcription, Southern blot method was applied to directly visualize the extra chromosomal newly synthesized cDNA copies of *Tto1*. 20µg of total genomic DNA of T_3 plants, used in previous PCR experiments, grown for two weeks in presence and in absence of the inducer β-estradiol, were digested with *NotI* for 4h, separated on 0.9% agarose gel, and transferred for 16h onto a nylon membrane, by the capillary method (§4.3.7.2). It is worth mentioning that *NotI* does not have any cleavage site in *Tto1* sequence, therefore it was only used in a partial digestion, in order reduce the high viscosity of the DNA samples, thus facilitating handling. The 307bp probe, homologous to *Tto1* ORF from nt 4390 to 4697, was labeled with radioactive dCTP. The probe detected both *Tto1* copies of the T-DNA integrated into the genome DNA and of the extra-chromosomal newly synthesized cDNA (Fig. 2.10). The genomic DNA migrated as a thick high molecular weight band, whereas the linear cDNA, which contains no *NotI* cleavage site,

migrated at the height corresponding to its size of 5.3kb (red arrow). The different intensity of genomic bands was due to a varying amount of genomic DNA loaded (see also Appendix 2.S-A), to the digestion grade of DNA, and to the number of insertions of the T-DNA transgene. Although for construct B the sensitivity limit of the method was almost reached, we could observe a perfect correlation with previous PCR results. Constructs A, B and C in the induced state, as well as the positive control i*Tto1* construct described in the former section, showed a 5.3kb band corresponding the complete cDNA, while no such band was detected with constructs D and E. This indicates that the 125bp of sequence redundancy, contained in construct C, are sufficient for first strand transfer, ensuring complete restoration of 3' LTR (Fig. 2.9b) and consequently replication of the element into new cDNA copies.

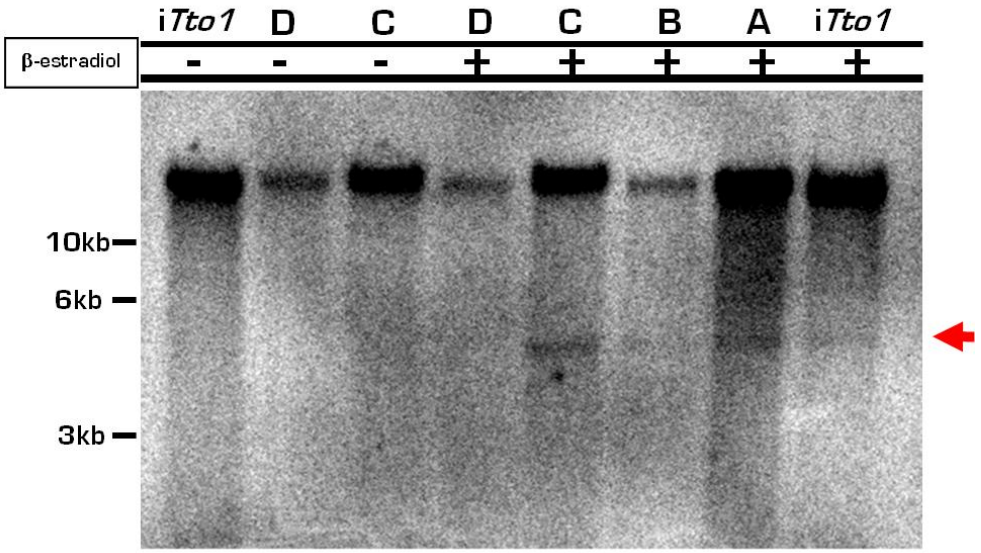


Fig. 2.10 Visualization of extra chromosomal cDNA copy of *Tto1*. The picture shows the induced (+) and non induced (-) state of 3'LTR deletion constructs and of positive control i*Tto1*. As demonstrated before, construct A, B and C can support reverse-transcription, and the full-length (5.3kb) element is synthesized (red arrow). The DNA size marker bands are indicated to the left. The different intensity of the bands relative to the genomic DNA is due to transfer efficiency and to the insertion number of T-DNA in Arabidopsis transgenic plants.

2.4.6 Mechanistic involvement of R region

These findings about the length of the R region made it more interesting to further investigate what could actually be the function of the essential stretch of 125 bases in the mechanism of the first strand transfer. We therefore, profited from the collaboration of Dr I. Hofacker, of the Institute for Theoretical Chemistry at the University of Vienna to obtain a structure prediction of the mRNA 3' end of constructs C and D, namely the shortest still active and the longer of the two constructs showing impaired strand transfer respectively.

The structures were obtained using the software "RNA Fold" that is available on the following website: http://www.tbi.univie.ac.at/~ivo/RNA/RNAfold.html.

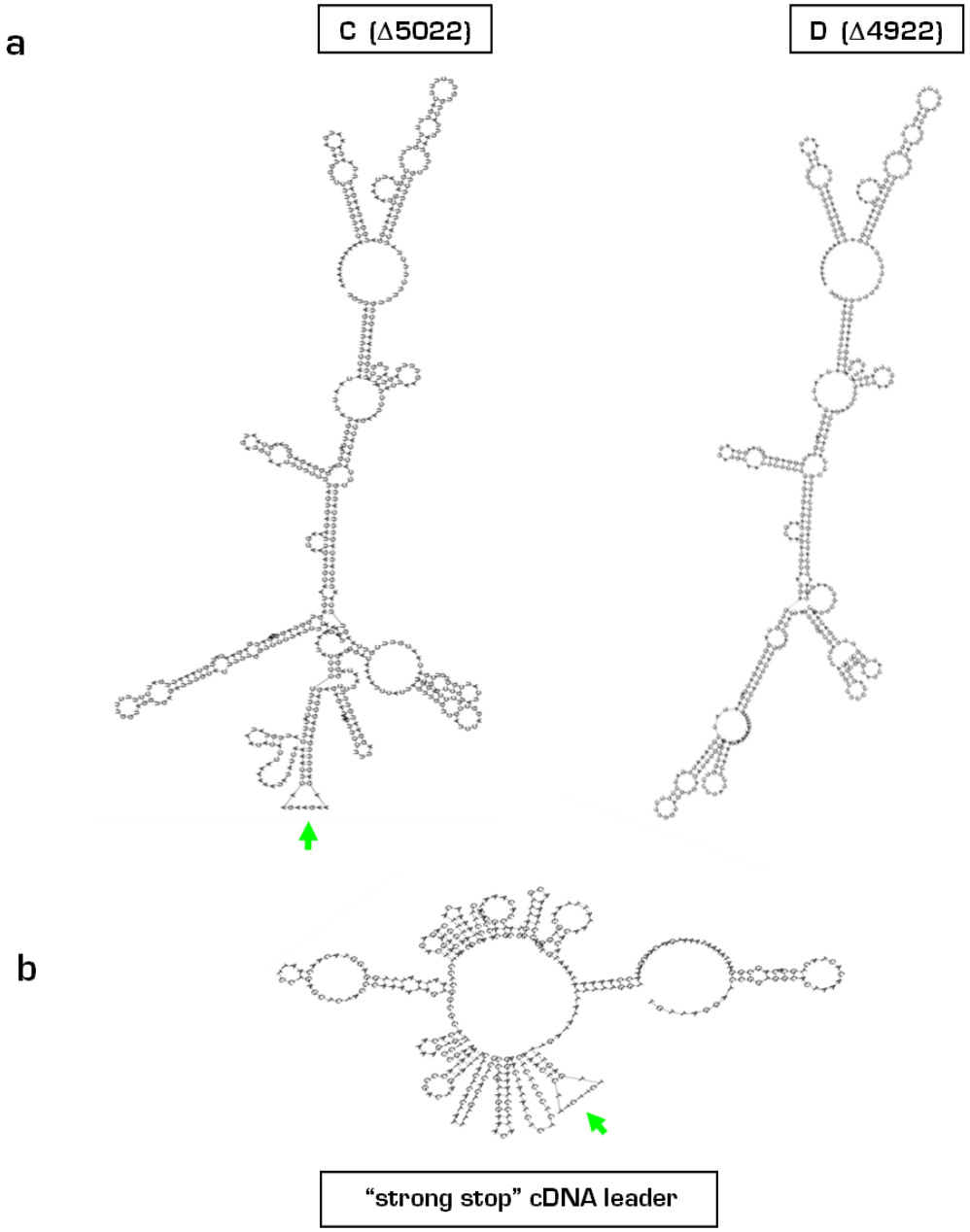


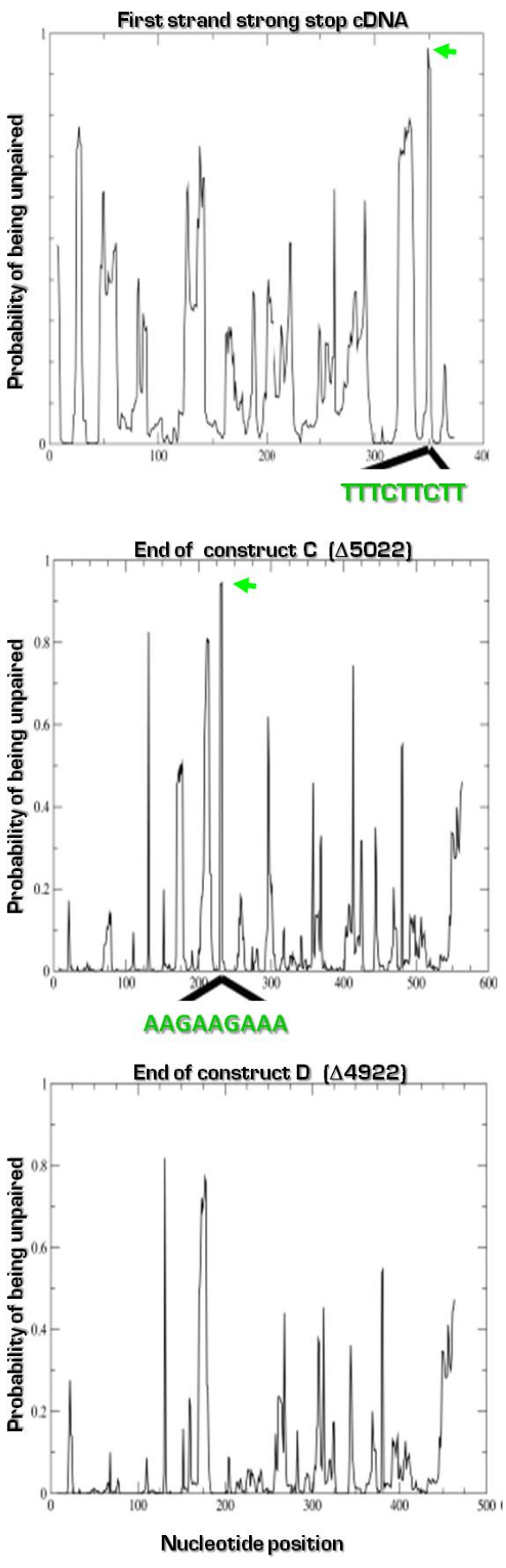
Fig. 2.11 Folding prediction of 3' end of mRNA of constructs C and D (a) and of the strong stop cDNA leader, synthesized from the 5' end of the element (b). The three structures show a complex secondary structure, in which base paired regions alternate with single stranded regions. An unpaired 9nt loop is complementary between the RNA ending at position 5022 and the cDNA leader, as indicated by a green arrow. Such structure is hypothesized to mediate the hybridization between cDNA and RNA, which determines the template switch of the strong stop cDNA.

As expected from previously published results concerning the 5' end (Böhmdorfer et al., 2005), both mRNAs form a complex and tightly base paired structure, with similar folding until the deletion point between the two constructs is reached (Fig.

2.11). The program revealed in fact that construct C contains a characteristic hairpin structure formed by the 100nt that are exclusively missing in construct D (panel a). To test whether this might be responsible for the homology search on 3' end, both structures were compared to the structure of the cDNA leader (panel b).
A 9nt loop is contained in construct C (Fig. 2.11a), which is missing in D; the same 9nt loop is present in the cDNA leader (panel b). The structure prediction data was also complemented by a diagram showing the probability of being unpaired of the three potential mRNA structures of Fig. 2.12. The probability was calculated using the software "RNAup" (Muckstein et al., 2006).
The loop in cDNA and RNA of construct C has a probability of being single stranded higher than 95%, the highest among the whole sequence. Furthermore Fig. 2.12b shows a close up on the two conserved hairpin structures: the two loops share total reverse complementarity. These findings lead then to the hypothesis that the base pairing of the 9nt hairpin of the emerging cDNA and of the RNA extends to the whole sequences that are totally complementary letting the cDNA leader extension proceed until the other end of the RNA. To support this hypothesis, a co-folding analysis of cDNA leader and mRNA ends was done by Dr. Ivo Hofacher (Institute for Theoretical Chemistry at the University of Vienna, Austria) using the software "RNAcofold" (Bernhart et al., 2006) to show whether energy parameters favor this mechanism. It was found in fact that a contribution of ca -26kJ/mol is given by the kissing of the two hairpins. The energy loss, if the emerging cDNA hybridizes to the mRNA 3' end, is -150 kJ/mol, derived by loss of the secondary structure; but the gain from forming a perfect heteroduplex is -385kJ/mol. All these results suggest that the annealing of the cDNA from the loop to its 5' end is strongly favored compared to the formation of secondary structures by the single stranded cDNA and mRNA (Tramontano et al, 2010).

2.4.7 Extension of strong stop cDNA of *Tto1* stops before the 5' end is reached

Another important feature of *Tto1* Δ3'LTR inducible constructs which might influence efficiency of first strand transfer is shown in Fig. 2.6d. All i*Tto1*-based constructs contain an extension of 32 non LTR bases at their 5' end, which is not present at the 3' end. A cDNA leader spanning this sequence would thus not be able to bind to the 3' end of RNA. In this work we wanted to elucidate another aspect of the first cDNA



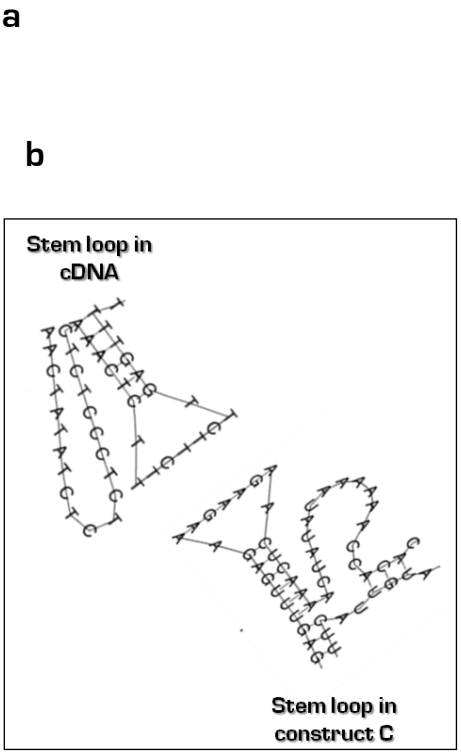


Fig. 2.12 Diagram of the probability of being unpaired of cDNA leader and RNA of constructs C and D. (a) The two highest peaks indicated by green arrow, correspond to the sequence of the two loops present only in the cDNA and in the RNA ending at position 5022. They have the highest probability of being single stranded among the whole sequence (>95%); no such peak is obtained from RNA ending at position 4922. (b). A close up of the 9nt loop is reported, to show their full complementarity. Our model implicates the kissing of the two loops, with subsequent melting of the whole two complementary sequences.

strand transfer, that is whether the template switch, from 5' to 3' end, takes place before the synthesis of the cDNA leader reaches the 5' end of the mRNA. Constructs with the 32bp extension were compared to a previously tested construct carrying an extension of only 6nt, and no difference in strand transfer efficiency was observed. This was an interesting result because it precludes the simplest accepted model reported in Fig. 1.3, according to which the 5' region of strong stop cDNA is first entirely reverse-transcribed from the mRNA and afterwards is transferred to the 3' end.

2.4.8. Role of LTR as transcriptional terminator

The natural LTR of *Tto1* contains two main termination sites (Böhmdorfer et al., 2005). The major one has been mapped to position 4914, namely between the end point of constructs D and E (Fig. 2.11a), the second is mapped to position 5230, around the deletion point of construct A (5233). An RNA ending at the major 5230 would be able to provide enough sequence redundancy to support reverse-transcription of a construct starting at position 172 of the LTR, as all engineered *Tto1* constructs used in this work. Since constructs A to D all contain the major termination site (4914), it was interesting to analyze the termination efficiency of those with shorter LTR, which do not contain the further downstream terminator. The rbcS terminator from pea appended to Δ3' LTR constructs served to this purpose. First, it had the function to replace the natural terminator in the constructs with more extensive deletions (C, D, E), so that they could always be tested by Intron-PCR, which needs self-encoded reverse-transcriptase and therefore a properly terminated mRNA. Second, the effect of a heterologous terminator linked to our engineered i*Tto1* construct was tested. Constructs C and D were analyzed in this respect.

2.4.9 RT-PCR to asses relative efficiency of Δ5022 and Δ4922 mRNA transcription

Total RNA from induced and not induced T3 plants containing construct C and D respectively were prepared as reported in §4.3.8.1. Reverse transcription was performed on 2μg of total RNA derived from construct C and D respectively, in a total reaction mix of 20μl. The oligo 292A-T2654-2634 was added to each reaction to obtain cDNA spanning the end point of both kinds of messengers. As an internal standard the oligo ubc9up binding to the gene *UBC9* (ubiquitin conjugating enzyme 9) of Arabidopsis was also added.

2µl of the ensuing cDNA were used in an Intron1-PCR with oligos 755A-T969-991 and, 756A-T1109-1086; another 2µl of the respective RNA were used in a PCR to amplify the *UBC9* control gene. Interestingly, in contrast to our expectation, the mRNA of construct D was more abundant than mRNA of C. To us this suggested that transcription itself is not influenced by the deletion, and that the higher abundance of the transcript D might be due a secondary structure more favorable to transcription, or in general we can assume that transcription might differ between fragments with different length, and as already mentioned above different T-DNA might have a different expression depending on the genomic location where they inserted. Fig. 2.13a shows this difference.

2.4.10 Mapping *Tto1* mRNA 3' ends and identification of termination signals

Poly-A RNA of both induced and non induced C and D constructs was isolated by Dynabeads® Oligo $(dT)_{25}$ (Invitrogen) as in §4.3.8.2. The mRNA linked to the beads was first extended in a PCR with the oligos 231A-T4390-4411 and 261A-dTclamp and amplified in a double nested-PCR to enrich messengers spanning the sequence from the deletion point in the LTR to the poly-A tail. The cDNA was in turn amplified with oligos 364A-T4494-4517 and 262A-clamp and subsequently with 1045-T4626-4648 and 292A-clamp, and visualized on a 1.5% agarose gel after each amplification step. Fig. 2.13b reports the result relative to the last PCR.

The four fragments shown in panel b were purified from gel and sub-cloned in the plasmid pCR® 2.1 using the TA-Cloning kit (Invitrogen) and sequenced. This experiment confirmed the expected major termination point: the poly-A tail was appended at position 4914 of *Tto1,* between the end positions of both constructs, as previously mapped. The longer band as in Fig. 2.12b was also sequenced, and showed that the transcripts end with the sequence of the rbcS promoter by which the deletion constructs were extended. The poly-A tail in fact was added to a position of the heterologous terminator, corresponding to the mapped transcription termination site of rbcS in pea, either 1658 or 1678 of accession X04333 (Coruzzi et al., 1984). In this specific case the poly-A was attached to nucleotide 249 of the rbcS terminator as present in *pER8*.

A conclusion to these results was that no termination point is present between end points of both constructs, confirming that natural termination occurs at the early end point as already described, and that in absence of the further downstream terminator, at position 5230 of *Tto1*, the transcription continues beyond the deletion points of the LTR and stops in the heterologous rbcS terminator.

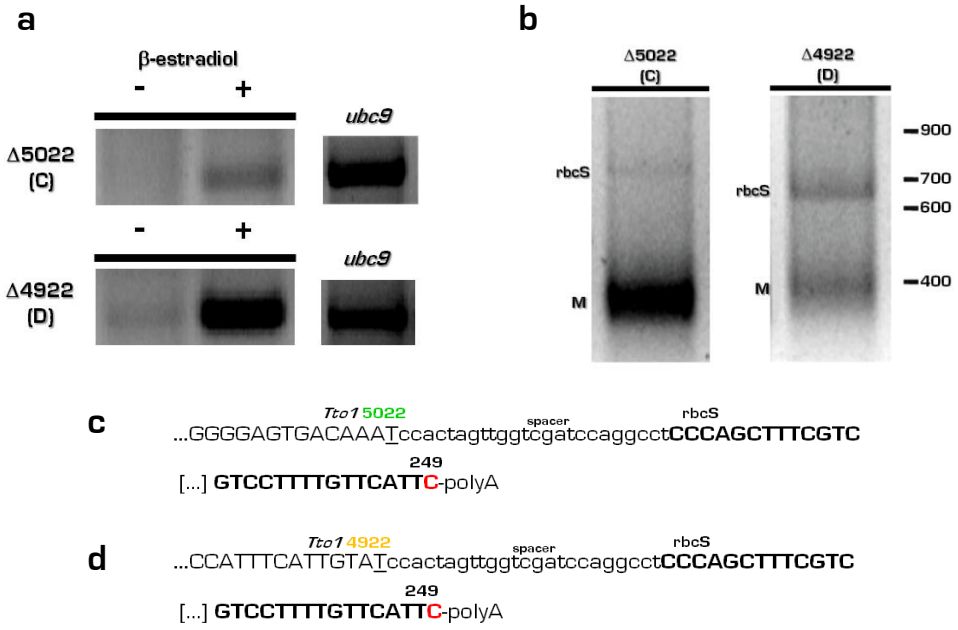


Fig. 2.13 Mapping of the 3' ends of the transcripts of constructs C and D. a) An RT-PCR shows the relative abundance of both transcripts. Interestingly the shorter mRNA (D) is more abundant than the longer mRNA derived from construct C. The *UBC9* gene was used as an internal transcription control. b) Both constructs show a common band ca 400bp long, which contains the previously mapped termination site indicated by a M, lying between the deletion point of constructs D and E (around position 4914). A longer fainter band slightly larger than 700bp for C and migrating between 600 and 700bp of the marker was also amplified from both transcripts, which corresponds to the termination site of the heterologous pea rbcS terminator. c) - d) The cDNA sequence obtained from the mRNAs ending with rbcS terminator (upper band in panel b) of constructs C and D respectively is shown. *Tto1* sequence is reported in capital letters, the end of the deletion is represented by the underlined T. The spacer (as in Fig. 2.5a) is shown in small letters; the rbcS sequence is in bold capital letters. The nucleotide 249 of rbcS, corresponding to an expected termination site, is in red.

2.5 *Tto1* integrase

As reverse transcriptase is the most conserved retrotransposon enzyme, according to its sequence the so far accepted classifications of retroelements have been made. The reverse transcriptase also appears not to be the only limiting factor to i*Tto1* retrotransposition. In fact, G. Böhmdorfer and colleagues did not register any new transposition event, in spite of the efficient production of cDNA, with Dexamethasone inducible *Tto1*. We thought consequently that integration would be worth investigating as a control key point influencing insertion frequency, choice of the target site and hot spot for sequence variability. We were interested in finding possible mechanisms of posttranscriptional and posttranslational control on the integrase. For this purpose, we first needed to know more about the *Tto1* encoded integrase.

2.5.1 Attempt to rise an αINT antibody to detect the integrase in vivo

We started investigations on the integrase on the protein level, in order to identify the protease cleavage site between RT and INT in the poly-protein. The attempt was made to raise an αINT antibody to immuno-precipitate the active integrase *in vivo*, after protease cleavage, from transgenic Arabidopsis expressing *Tto1*. The immuno-precipitated protein should be used to identify the termini of the protein and potential post-transcriptional modifications.

2.5.2 Purification of recombinant AgINT2 and immunization of rabbits

A 35kDa fragment of the integrase, called AgINT#2, where Ag stands for "antigen" was previously subcloned in the plasmid *pET19b::AgINT#2* (provided by A. Bachmair). The plasmid was transformed *E. coli* strain Rosetta(DE3) *pLysS*, and the 35kDa fragment was overexpressed as a His-tag conjugate (Fig. 2.14a and b).

A minimum induction time of 2h with 1mM IPTG was necessary to overexpress AgINT#2. Exploiting the 6x-His-tag, 800μg of AgINT#2 were purified under denaturing conditions (§4.3.11.4); the protein samples were concentrated using Centricon devices (Millipore) and sent to the company Eurogentec (Belgium) for antibody production.

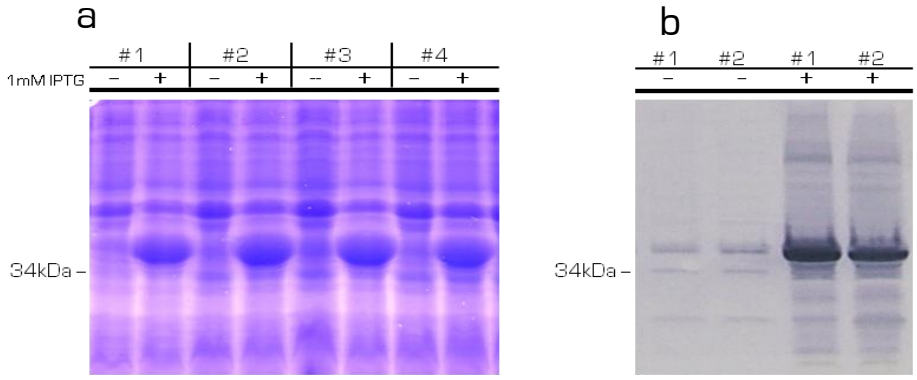


Fig. 2.14 a) A SDS-PAGE is shown, containing four of the best 6x His-AgINT#2 overexpressing Rosetta clones. For each clone the non induced state is shown, and the induced state after 2h from the addition of 1mM IPTG. b) Detection of AgINT#2 by anti His tag NI-NTA conjugate. Western blot with total proteins of clone #1 and #2 of panel a, in the induced and non induced state was performed to detect AgINT#2 from total bacterial extract.

Two rabbits, identified by the numbers #3036 and #3037, were immunized. The antisera from both rabbits were tested and the #3037 seemed not to be active, therefore only the #3036 was used in further applications. In order to enrich the specific αAgINT#2 Ab from the antiserum #3036, I subjected AgINT#2 to different native purification conditions and the best one were those reported in §4.3.11.5.

Fig. 2.15 shows a Western blot with fractions of each step of the native purification of AgINT#2. An abundant fraction of the protein was contained in inclusion bodies (Fig. 2.15 lane 3) and two major bands were revealed: the 35kDa AgINT#2 band was always accompanied by a lower unspecific and intense band migrating slightly below the 34kDa marker band.

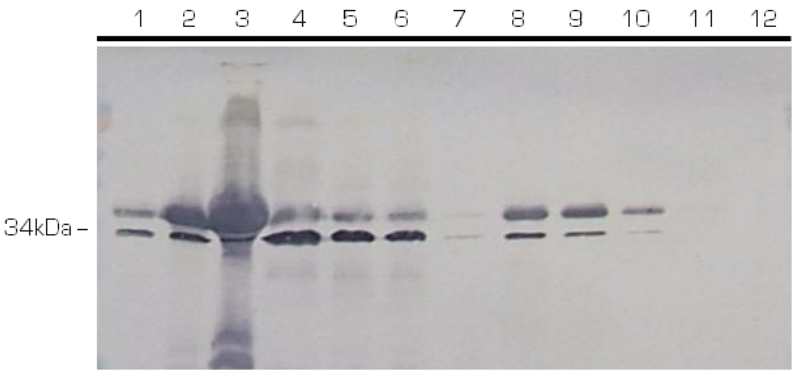


Fig. 2.15 Western blot with purification fractions of AgINT#2 protein. The protein was marked with anti His-tag Ni-NTA conjugate, and revealed by NBT-BCIP system, according to manufacturer's protocol. Lane 1: non induced total protein fraction. Lane 2: induced (2h) total protein fraction. Lane 3: crude lysate after induction. Lane 4: Supernatant after centrifugation of crude lysate as in lane 3. Lane 5: flowthrough after binding to the Ni-NTA resin. Lane 6: column wash flow-through. Lanes 7-10: samples from eluate 1 to 4. Each eluate fraction collected was 2ml, and 7.5µl were loaded on the gel. Lanes 11-12: samples from two fractions eluted with a pH4 buffer, to compare elution efficiency of the native buffer.

To find out whether it was an unwanted degradation product and to still improve the purification conditions the crude antiserum and the anti His tag Ni-NTA conjugate were compared respectively on two Western blots, both containing the four elution fractions, the crude lysate and the wash flow-trough. This experiment is reported in Fig. 2.16.

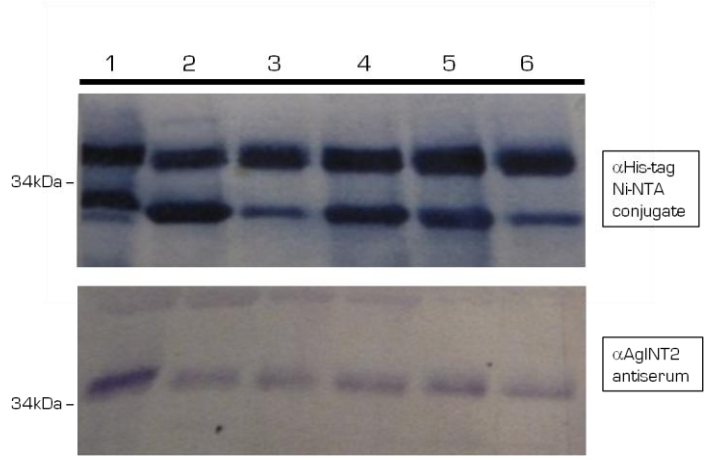


Fig. 2.16 Comparison of the efficiency of anti His-tag Ni-NTA conjugate and anti AgINT#2 Ab. Both Western blots contained: lane 1: crude lysate; lane 2: column wash flow-through; lane 3-6: eluted fractions 1 to 4. The Antibody revealed that the lower band detected by the anti His-tag conjugate was not a degradation product.

With the last experiment the higher specificity of the antibody in detecting the AgINT#2 protein compared to the anti His-tag Ni-NTA conjugate was demonstrated. We also conclude that the purification conditions used did not provoke degradation of

the overexpressed AgINT#2. Nevertheless it is noteworthy that an important fraction of the protein still localized to the insoluble fraction. Other attempts to increase the soluble fraction of AgINT#2 were unfortunately not improving the yield. The protein purified in the above mentioned conditions was however used for Ab enrichment from the 3036 antiserum. The Ab was then tested on a protein extract of induced Arabidopsis expressing *Tto1*, but unfortunately no integrase was detected (data not shown). This approach will be tried in future on plants overexpressing *Tto1* proteins.

2.5.3 The integrase from another angle

We decided then to keep investigating the integrase from another point of view, and tried to draw a picture of the genetic and natural variation of the integrase amino acid sequence. The reason for that is elucidated as follows.

In the tobacco BY2 ecotype, where it has first been isolated, there are 30 copies of *Tto1*, whereas for some other elements up to 100,000 copies have been identified in their native host. This is an interesting number as *Tto1* is one of the few known active retrotransposons, therefore a much higher copy number would be expected. Some transposons appear to be inactive due to accumulation of mutations (Ivics et al., 1997); if this is true, we believed that this could be even more the case for retrotransposons that are particularly error-prone, due to their reverse transcription step. We asked then the question what could keep a retrotransposon still active but at such a relatively low level in its native host.

2.5.4 Isolation and cloning the integrase gene from tobacco ecotypes

Looking for the answer in the aminoacid residues, the protein sequences of *Tto1-1* integrase from some of the most common tobacco cultivars and from its two progenitor species was compared to that of the BY2 cell line, namely the one being used in all previous experiments. The genomic DNA was therefore isolated from five *Nicotiana tabacum* cultivars SR1, Xanthi, Samsun NN and W38 and from the two progenitors *Nicotiana sylvestris* and *Nicotiana tomentosiformis*.

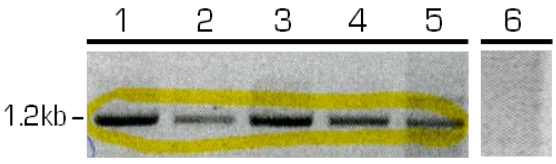


Fig. 2.17 PCR to amplify the DNA sequence of *Tto1* integrase core domain. Lane 1: SR1, L. 2 Xanthi, L.3: W38, L.4: Samsun NN, L. 5: *N. sylvestris*, L. 6: *N. tomentosiformis*. All lanes contained *Tto1* integrase, except for lane 6.

A 1.2kb fragment, spanning the whole integrase core domain coding sequence was isolated by PCR using *Pfx* DNA polymerase in combination with the oligos 912A-

H3Intdn and 913A-KpnIntup (see Fig. 2.17). While PCR produced the expected band for all the above mentioned ecotypes and for *N. sylvestris*, no integrase fragment was amplified from the other progenitor *N. tomentosiformis.* A first conclusion then was that the allotetraploid *N. tabacum* inherited its *Tto1* copy from the diploid specie *N. sylvestris.* This will be interesting in phylogenetic studies to understand the occurrence of *Tto1* and other elements in modern species.

The PCR fragments were purified from gel and subcloned right away into *SmaI* linearized *pSKII* plasmid, exploiting the blunt ends produced by the *Pfx*, and transformed in *E. coli* XL1 blue. The clones were selected by blue/white selection, screened by colony PCR and control digestion and then sequenced. From a group of 35 clones 19 were successfully sequenced and analyzed further. A variable number of clones were obtained from each ecotype, as summarized in Table 2.1.

Tobacco ecotype/species	Number of clones	Clone sequencing code
SR1	7	AA41, AA42, AA43, AA44, AA45, AA46, AA47
Xanthi	5	AA53, AA54, AA5, AA59, AA60
W38	4	AA63, AA65, AA67, AA68
Samsun NN	2	AA61, AA62
N. Sylvestris	1	AA69

Table 2.1 Nineteen clones of the integrase core domain were successfully sequenced. A different number of clones were obtained from each ecotype.

2.5.5 Natural variation in the integrase protein

The amino acid sequences of the integrase DNA clones reported in Table 2.1 were obtained by Clone Manager and an alignment was done. The whole 401 amino acids sequences were aligned and compared to *Tto1-1* (see Appendix 2.S-A). A total of 55 residues appeared not to be conserved in respect to *Tto1-1* (ca 14%). 50 are residues that change randomly, either within clones from the same cultivar or between cultivars (see bold letters in Appendix 2.S-A), that are therefore suggested putative mutations produced by replication errors, as expected from retrotransposons.

Focusing on the integrase active region, where the conserved catalytic domain $DX_nDX_{35}E$ is located, and more precisely downstream of the conserved E_{583} of *Tto1* ORF (Appendix 2.S-A), we found interestingly that the remaining 5 amino acid residues seemed to vary in a distinct manner. As summarized in Table 2.2 and shown

following in Fig. 2.18, the residues into which the K_{629}, L_{636}, E_{690}, G_{747} and L_{754} are changed allowed the distinction of the clones in two different groups. With the exception of AA53 in the case of L_{636} and of G_{748}, all the clones clearly appeared to fall in one or the other group.

aa BY2	Possible residue		
K_{629}	K	E	
	AA44, 54	AA41, 42, 43, 45, 46, 47, 53, 54 57, 59, 60, 61, 63, 65, 67, 68, 69	
L_{636}	L	S	Q
	AA44, 54	AA41, 42, 43, 45, 46, 47, 54, 57 59, 60, 61, 63, 65, 67, 68, 69	AA53
E_{690}	E	Q	
	AA44, 54	AA41, 42, 43, 45, 46, 47, 53, 54 57, 59, 60, 61, 63, 65, 67, 68, 69	
G_{747}	G	D	N
	AA44, 54	AA41, 42, 43, 45, 46, 47, 54, 57 59, 60, 61, 63, 65, 67, 68, 69	AA53
L_{754}	L	P	
	AA44, 54	AA41, 42, 43, 45, 46, 47, 53, 54 57, 59, 60, 61, 63, 65, 67, 68, 69	

Table 2.2 The 19 integrase core domain clones distributed in two distinct groups, according to the amino acid residue contained at each of the five critical positions in the active site region. 17/20 clones share the same residue (yellow), except AA53 that was considered an outlier in the case of L_{636} and G_{747}.

17 clones out of 20 analyzed (85%) presented the same residue in all five positions; they were assigned to the major group that was named "Int2" and marked with 2 green stars.

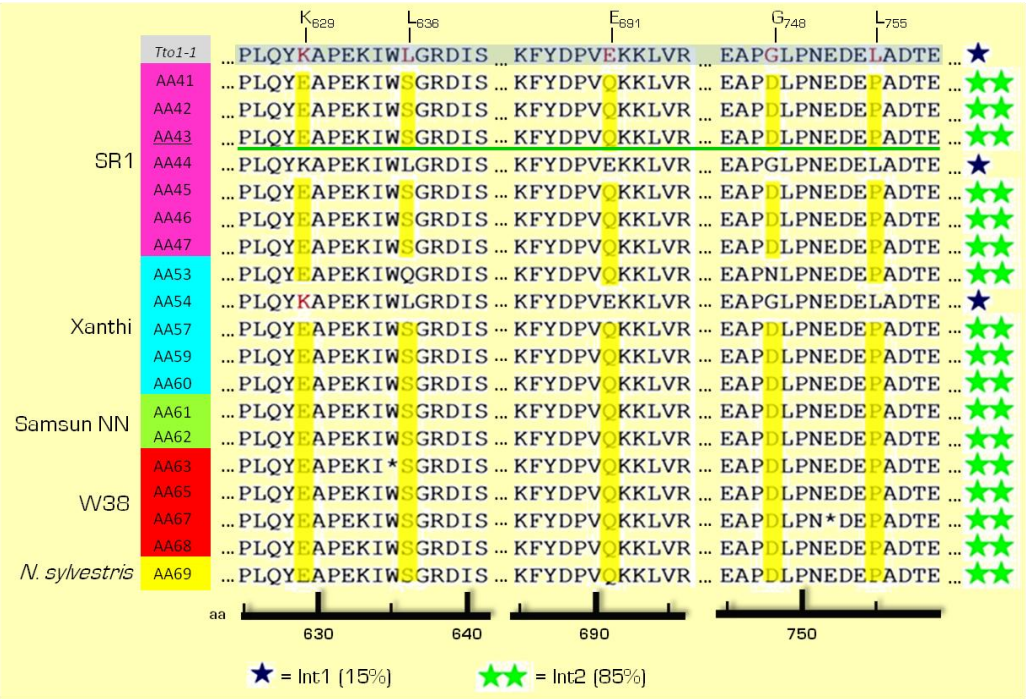


Fig. 2.18 Alignment of the integrase core domain amino acid sequences of all the clones obtained from the different tobacco ecotypes. Only the parts bearing the varying residues K_{629}, L_{636}, E_{690}, G_{747} and L_{754} are shown. The starting sequence of *Tto1-1* is highlighted in grey. The major group, called "Int2", had the same residue in 85% of the cases, and was marked with 2 green stars. The minor group, called "Int1" represented only the 15% of the cases and was marked with one blue star. The most frequent residues are highlighted in yellow. Clone AA43, which was considered to contain the consensus sequence, is underlined in green.

The second group that was only constituted by the clones AA44, AA54 and the starting sequence of *Tto1-1* (15%) was named "Int1" and marked with one blue star. In Fig. 2.18 only the sequence stretches bearing the 5 characteristic residues are shown (the whole sequences are provided in Appendix 2.S-A). The K_{629}, L_{636}, E_{690}, G_{747} and L_{754} are labeled on the *Tto1-1* sequence and highlighted in yellow in the seventeen clones that belong to "Int2". An interesting result was that *Tto1-1* happened to fall into the least represented group, as if all experiments carried on in our group and by the colleagues working on *Tto1*, had been done with a minor natural variant. The question was easy to rise in fact: what would be the difference between a *Tto1* element carrying the major or the minor variant of its integrase, and which would be more active?

2.5.6 "Re-making" *Tto1*: synthetic biology of the element

To address this question a new version of *Tto1* was created in which the integrase core domain Int1 was replaced by Int2, by cloning a 601bp DNA fragment, corresponding to *Tto1* nt 2474 to 3074, amplified from clone AA43 (see Fig. 2.19). A detailed description of all cloning steps is reported in section 4.3.16.2).

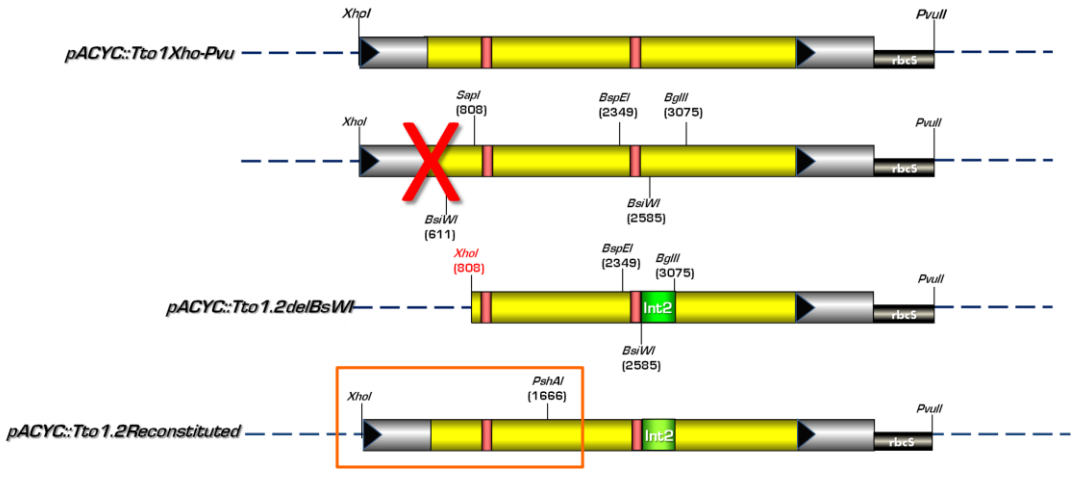


Fig. 2.19 Replacement of the integrase core domain with Int2 domain from the clone AA34

The new version of the element was named *Tto1.2*, to refer to the Int2 ore domain, whereas the *Tto1-1* version, still carrying Int1, was named *Tto1.1*.

The clone AA43 was chosen because it represented a "consensus" sequence of the integrase. In 25 of the 27 not conserved positions, occurring in the active site region, AA43 contained the most frequent residues shared by all clones, including the five residues of our specific case, except for L_{618} and A_{652} (Appendix 2.S-B). To test its activity in plants, the new synthetic retrotransposon was cloned in the plasmid

pER8new (see §2.4.2), and the ensuing vector was -named *pER::Tto 1X*, where X was chosen to give uncertainty on the function of such element. In contrast the pre-existing *Tto 1.1* was re-named *Tto 1N* where N stands for "native".

pER::Tto 1X was transferred to Arabidopsis plants via floral dip transformation and the transgenic plants were obtained as in §2.4.4. Six transgenic lines named #1, #2, #3, #6 and #7 were selected and analyzed. T_2 transgenic Arabidopsis plants of each line were induced with β-estradiol in liquid culture as in §2.4.5; their genomic DNA was isolated and used for Intron1-PCR.

Unfortunately no cDNA was detected, as only the unspliced band was amplified by PCR (data not shown). To try explaining this negative result, I isolated the mRNA (obtained as in §4.3.8.3) to check whether the expression had taken place at all, thus reducing the area of investigation. An RT-PCR was performed with Intron2 oligos on the cDNA (§4.3.9.1 and 4.3.2.3) of each line, which is shown in Fig. 2.20. The RT-PCR demonstrated that *Tto 1X* mRNA was suitable for in vitro reverse transcription. Line #4 showed no spliced band and was discarded. Lines #2 #3 and #6 were also no longer considered as they showed band intensity in favor of the unspliced T-DNA derived band. Lines #1 and #7 were instead used in experiments to follow as they showed a stronger spliced band, suggesting a higher mRNA level. The reasons for no splicing taking place in vivo might reside in an impaired formation of the VLP; but experimental evidence still remains to be found.

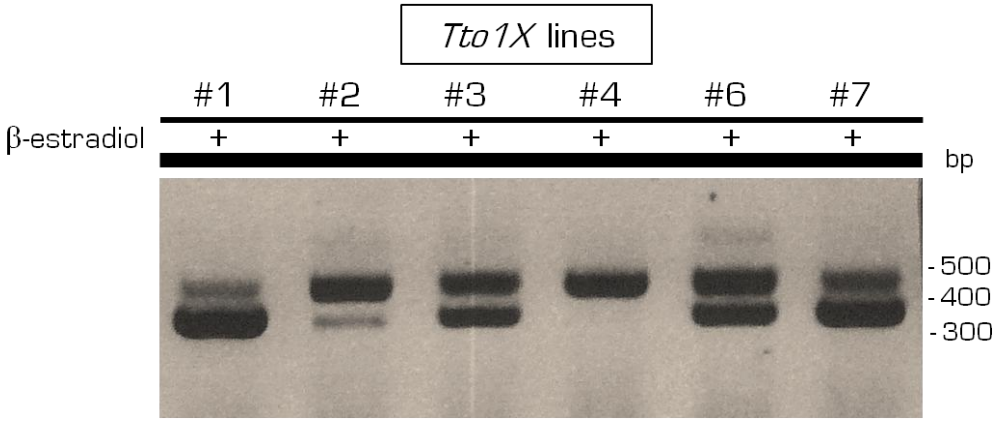


Fig. 2.20 RT-Intron2-PCR on cDNA of induced T_2 *Tto 1X* Arabidopsis lines #1, #2, #3, #6 and #7. Lines #1 and #7 showed a more abundant spliced and were used in following experiments.

2.5.7 Another synthetic *Tto 1* is being made to test

The in vitro reverse transcription rate of *Tto 1X* clearly appeared not to be higher than *Tto 1N*, the cDNA was in fact not even detected by Intron-PCR on total DNA after β-estradiol induction. Another synthetic *Tto 1* element was therefore designed. As above mentioned, *Tto 1X* exclusively contains two amino acids that are not present in

any other of the clones analyzed, namely L_{618} and A_{652}, which are mutated into a P (proline) and a D (aspartate).

To check any eventual difference in the activity compared to *Tto1N* and *Tto1X,* another synthetic element is being constructed starting from *Tto1X* sequence, which includes P_{618} and A_{652} and has been named *Tto1Y*. The different amino acid residues between the three *Tto1* versions are schematically reported and summarized in Table 2.3.

ORF	*Tto1N*	*Tto1X*	*Tto1Y*
618	L	P	L
629	K	E	E
636	L	S	S
652	A	D	A
690	E	Q	Q
747	G	D	D
754	L	P	P

Table 2.3 Distribution of the seven characteristic amino acid residues in the three different versions of *Tto1*. The respective position of each residue on *Tto1* ORF is indicated in the grey column. *Tto1N* corresponds to *Tto1-1*. *Tto1X* (derived from clone AA43) has a different residue in each position, compared to *Tto1N*. *Tto1Y* differs from *Tto1X*, by L_{618} and A_{652} that are shared with all the other clones except for AA43.

2.6 Attempts to obtain *Tto1* transposition in crops

2.6.1 *Tto1* in a monocot background

In pursuing the goal of creating a tool for crop mutagenesis, I invested a significant fraction of the time in building up a *Tto1* construct suitable for monocots. Barley was chosen for this purpose, as a plant of agronomical and research interest.

In previous experiments a Dexamethasone inducible *Tto1* was transiently expressed in barley callus, but not very high levels of cDNA were detected. In addition Intron-PCRs showed that the mRNA was only partially spliced (Böhmdorfer, 2005).

As we consider testing the activity of an element with dicot origin in a monocot plant very useful and interesting from a scientific point of view, I made a different attempt to obtain higher expression levels that would allow investigation on *Tto1* in barley.

2.6.2 Cloning of barley *Tto1*

The cloning strategy was followed in which the element starting at position 172 was cloned under the control of a constitutive promoter and its relative terminator sequence, and the pre-existing Arabidopsis introns were replaced by endogenous

introns. The synthetic biology approach was also tried with the barley constructs; therefore two versions of *Tto1* were made, containing the native integrase core domain and the Int2 domain respectively, as previously done with *pERnew::Tto1X*.

Tto1 from *pERnew::Tto1* was first subcloned using *XhoI/PvuII* cleavage sites into the backbone of plasmid *pACYC177*, to make the plasmid *pACYC::Tto1Xho-Pvu.* All intermediates of the barley vector construction were sublocned in this plasmid, as described in the Materials and Methods chapter, section 4.3.16.3 and shown in Fig. 2.21. A 601bp fragment containing the Int2 catalytic domain from clone AA34 was cloned to make a second synthetic version of *Tto1* called *Tto1.2* (see §2.5.5). The Int2 was cloned using the same strategy as for the Arabidopsis *Tto1X* , which is reported in Fig. 2.19. The first pre-existing integrase intron (intron2) of Arabidopsis was replaced with the 86bp barley xylose isomerase intron 18 (Xyl18). A 148bp fragment containing the Xyl18 intron (86nt) was excised by *BspEI -BsiWI* digestion from the plasmid *pUCBlint+*, in which it was previously subcloned, and cloned into the integrase domain between the *BspEI* and *BsiWI* sites at position 2349 and 2585 respectively (Fig. 2.22).

In the next step the nos terminator was added: a 300bp fragment containing the nos terminator was PCR amplified from the *HindIII* linearized plasmid *pWBVec8*, the purified fragment was then ligated to both versions of the plasmid. The pre-existing rbcS terminator was also removed in this step. Intron1 (Gag intron) was then cloned: the xylose isomerase intron 13 (xyl13) of 190bp replaced the pre-existing Arabidopsis intron. A 1kb fragment containing the Barley intron, previously subcloned in the plasmid *p2RT172Bla*, was excised by digestion with *BsiWI* and *BglII* and inserted into both *Tto1.1* and *Tto1.2* vectors.

The last feature added to complete the active part of barley constructs was the constitutive 1517bp *Mub1* promoter from maize. It was amplified from the plasmid *p6U* in two PCR steps that responded to a specific and speculative need that is explained as following.

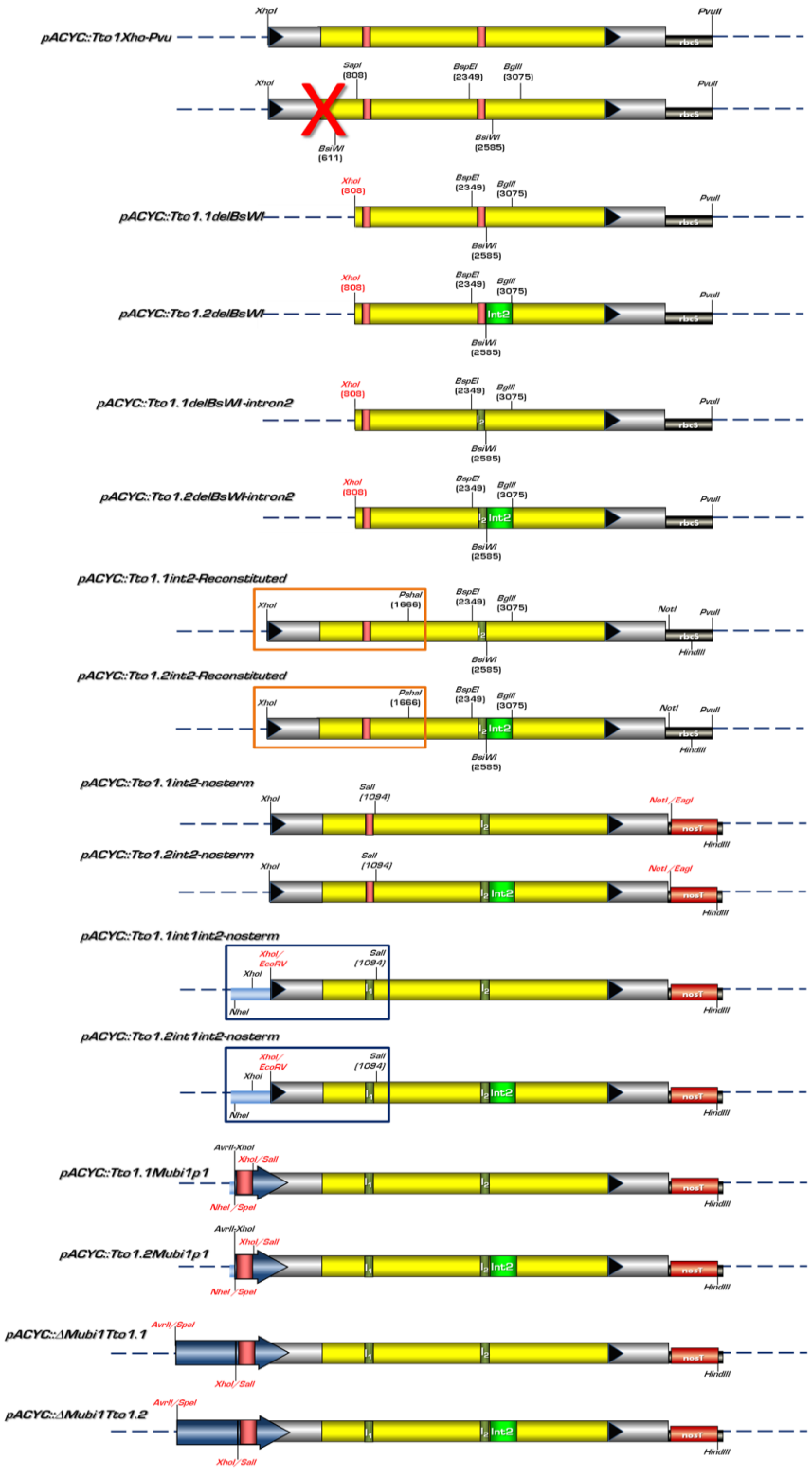


Fig. 2.21 Cloning strategy of *Tto1N* and *TtoX* construct for barley. All modifications were carried out sublocned in *pACYC* backbone. After the last step, both constructs were ligated to the plant vector *pWBVec8* as shown in Fig. 2.23.

The *Mub1* promoter contains an intron which has the function of transcriptional enhancer (Fig. 2.22), and produces an mRNA with an untranslated region slightly longer than 100bp. Since we already demonstrated (§2.4.8) that the expression

efficiency of *Tto1* in *pER8* is not affected by an extension of the mRNA up to 32bp, we wanted to maintain the same sequence features for the barley *Tto1* vector. Therefore the decision was made not to take the risk of having a 100bp leader and to shorten it to 30bp but without affecting on the enhancer. This was obtained by deleting the 75 nucleotides from 919 to 994 of the *Mub1* promoter sequence (see Fig. 2.21).

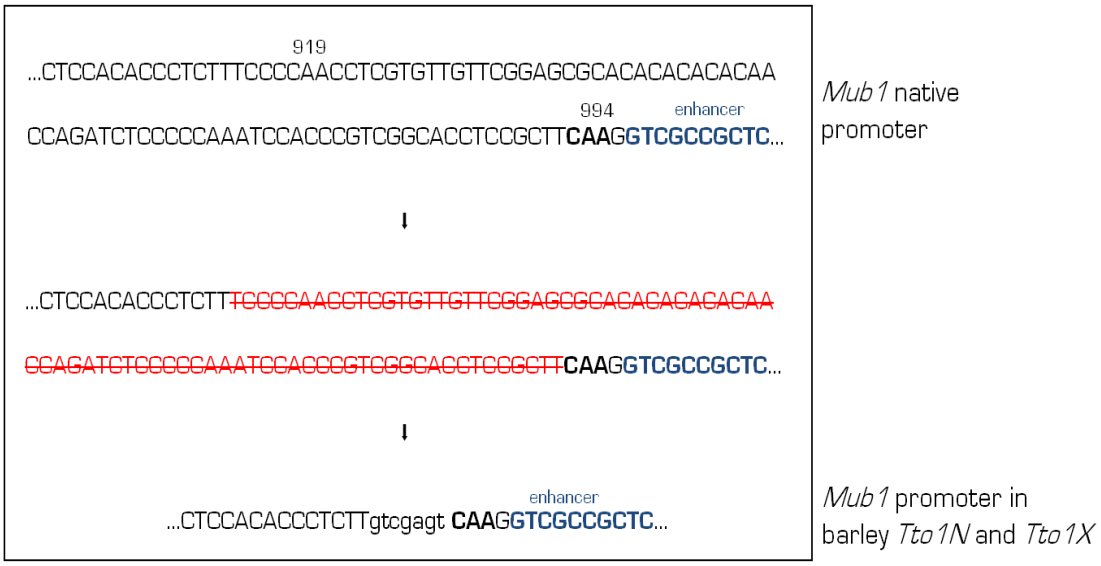


Fig. 2.21 Deletion of 75bp from *Mub1* promoter, to shorten the 100bp untranslated region of *Mub1* promoter to 30bp, Nucleotides from 919 to 994 are deleted, but the intron having enhancer function is not affected. A *Sal/Xhol* combination (see small letters) was used to re-ligate the two ends ensuing from the deletion, The CAA triplet upstream of the enhancer was chosen for cloning reasons.

After each cloning step (Fig. 2.21) the constructs were sequenced to check correctness of the sequences. After appendage of the terminator, the constructs had all control elements and were cloned into the barley expression vector *pWBVec8*, which confers Hygromycin resistance to the plants, to make *pVec8::Tto1N* and *pVec8::Tto1X*, using the same nomenclature as for Arabidopsis synthetic constructs. The *Tto1* barley expression cassette is illustrated in Fig. 2.23.
All *pACYC* based intermediated were propagated in the *E. coli* strain Sbtl4. In order to transform barley plants 2µg of *pVec8::Tto1N* and *pVec8::Tto1X* were electroporated to the strain AGL10 (as in 4.3.15.4) of *A. tumefaciens*, which specifically infects monocots. The putative transgenic *Agrobacterium* clones were grown on selective medium, screened by colony PCR and inoculated in soft agar, and sent to the lab of Dr. J. Kumlehn. at the Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK) in Gatersleben (Germany), in order to obtain *Tto1* transgenic barley plants.

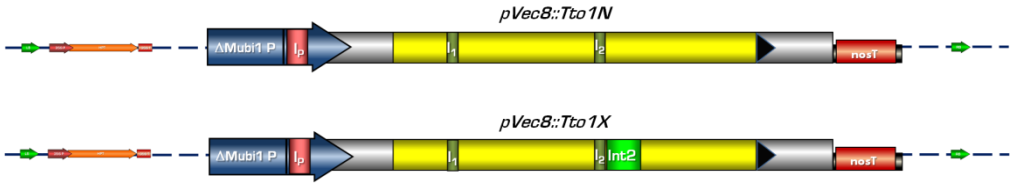


Fig. 2.23 Schematic representation of constructs *Tto1N* and *Tto1X* for barley. The barley *Tto1* is preceded by the constitutive ubiquitin 1 promoter and followed by nopaline synthase terminator. Introns xyl13 (I1) and xyl18 (I2) from barley xylose isomerase are contained in GAG and Integrase domain respectively *Tto1X* contains the integrase2 core domain as from clone AA34.
The *Tto1* part is magnified in respect to the other features of the expression cassette. On the left side: LB (left border), 35S promoter, *hpt* (Hygromycin resistance) gene, 35S terminator. On the right side: RB (left border).

2.6.3 *Tto1* transgenic barley

Our colleagues in Leipzig where able to obtain 4 transgenic lines. Among these 3 contained *Tto1N* and 1 contained *Tto1X*. The lines were named B01N-B03N for *Tto1N* and B01X for *Tto1X*. To check the activity of *Tto1* Intron1 and Intron2-PCR were performed and a control PCR on *HPT* (Hygromycin resistance) gene was also done. The PCR result on all lines was unfortunately not promising, as shown in Fig. 2.24. First of all no *Tto1* derived band was amplified from line B02N and B01X, which looked like the negative control untransformed barley ecotype Golden Promise (GP). More importantly though we could observe no *Tto1* activity in either the line as no intron-less band was amplified. On the other side lines B01N and B03N only produced the T-DNA derived unspliced 426bp band expected from Intron1-PCR and the 330bp unspliced band expected from Intron2-PCR. In contrast the positive control *HPT* gene produced the expected 1.3kb band (Fig. 2.24). One possible conclusion would be that some truncations might have occurred on the right border side of the cassette, as no amplification defect was encountered with the resistance gene that is cloned upstream of *Tto1*, closer to the left border (Fig. 2.23). To check then if transcripts were produced at all, the RNA was isolated from all the 4 lines, but unfortunately after RT-PCR neither *Tto1N* nor *Tto1X* mRNA could be detected.
We had to conclude that such constructs do not work in barley, one possible reason being the deletion of the 75bp in the promoter region that might have dramatically reduced its activity. The use of another promoter either constitutive or inducible, for instance β-estradiol or ethanol is to be considered.

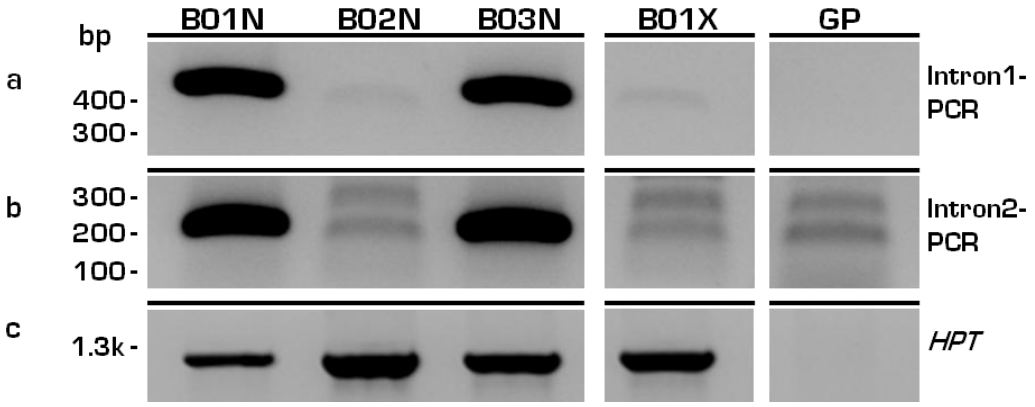


Fig. 2.24. Intron-PCRs to check expression of *Tto1* in transgenic barley. 4 transgenic lines were analyzed, B01- B03 containing *Tto1N* and B01 containing *Tto1X*, and compared to GP (Golden Promise) wt control. a) Intron1-PCR detected unspliced 426bp band only in B01N and B03N lines, and no intron-less band. b) The same result was obtained with Intron2-PCR. Only the 330bp unspliced band was amplified. c) PCR on *HPT* gene resulted in the expected 1.3kb band, demonstrating that the cassette was integrated.

2.7 A binary inducible system to improve i*Tto1*

2.7.1 Dexamethasone vs β-estradiol inducible system

In pursuing the aim of improving i*Tto1* as a tool for gene tagging of crops, the use of another inducible system was preliminarily tested. Although the *pER8*-based construct i*Tto1* proved very useful and successful and results have been published (Böhmdorfer et al., 2010), its β-estradiol inducible promoter provides a relatively weak induction, compared to the efficiency of the strong "double 35S" constitutive promoter tested in previous experiments (Böhmdorfer et al., 2005). Furthermore, it has been demonstrated that, despite of their strong activation, Dexamethasone responsive promoters can be lethal to the plant (Andersen et al., 2003; Kang et al., 1999; Ouwerkerk et al., 2001) therefore we wanted to test the *iTto1* technology in combination with an improved Dexamethasone inducible system. In this two component system one transgenic plant, called "donor", carries an integrated copy of i*Tto1* derived from the plasmid *pBIB::pOp6-Tto1*, while the second plant, called "activator", containing the strong Dexamethasone responsive regulon LhGR-N, provides the transcription factors, necessary for the expression of the transgene. For more details please refer to §4.3.16.4 and 4.3.18. This system is designed in a way that the two plants are crossed and subsequently the hybrid progeny is treated with Dexamethasone to monitor the transgene expression (Craft et al., 2005; Moore et al., 1998; Samalova et al., 2005).

2.7.2 *pOp6-Tto1*/LhGR-N appears to promote transcription at higher efficiency

In this section a preliminary experiment of induction of *pBIB::pOp6-Tto1/LhGR-N* in Arabidopsis is shown. The F_2 progeny of the crosses between donor lines 2-1, 3-1 and 4-1, and activator line S5 respectively, was induced for two weeks with Dexamethasone and tested by Intron1-PCR. Line 3-1 did not result in any detectable cDNA band, therefore only the results of lines 2-1 and 4-1 are described. In Fig. 2.25 the Intron1-PCR results on *pOp6-Tto1/LhGR-N* F_2 progeny are compared to an Intron1-PCR performed on F_2 carrying the Δ5119 and Δ5022 *Tto1* deletion construct respectively, in order to visualize differences in expression efficiency between the two inducible systems.

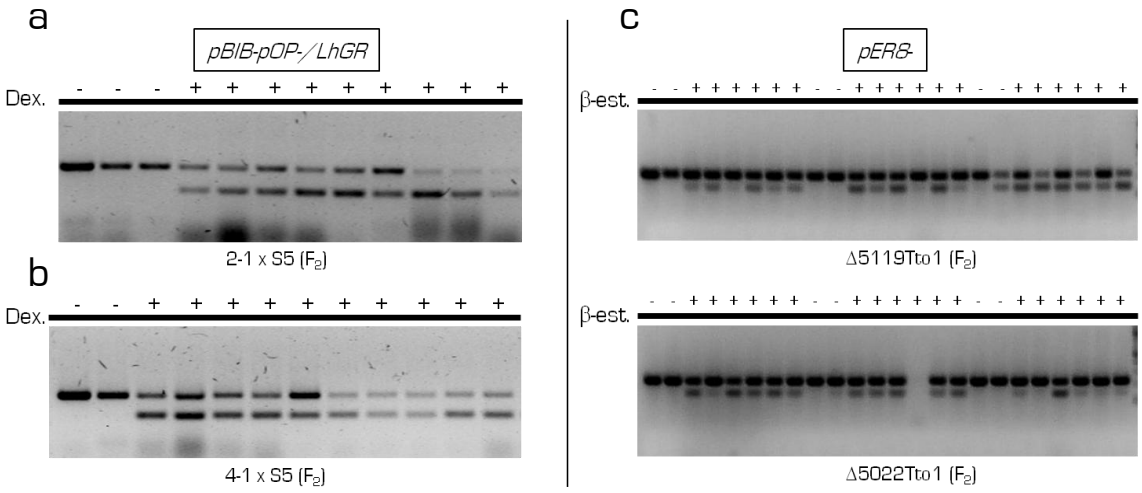


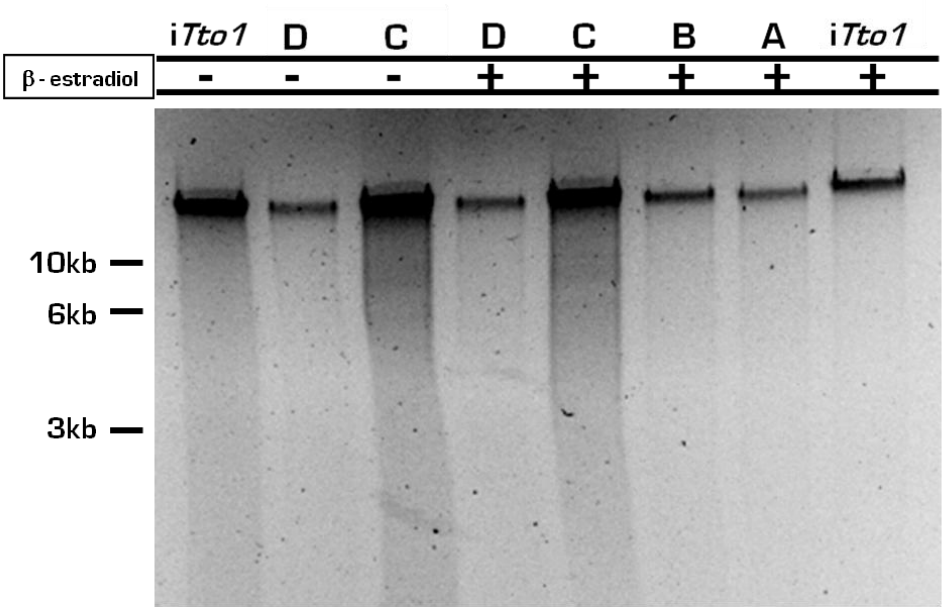
Fig. 2.25. a) - b) Intron1-PCR on *pOp6-Tto1/LhGR-N* F_2 progeny to check Dexamethasone induced *Tto1* expression. c) The result was compared to a similar experiment performed on plants contained *pER8* based *Tto1* deletion construct,

Although the absolute abundance of cDNA bands expressed by *pOp6-Tto1* (panel a) was not dramatically increased compared to *pER8* based *Tto1* constructs, an encouraging result was that in all cases its relative abundance appeared higher in *pOp-Tto1* progeny than in *pER8-Tto1*. To be more precise, the cDNA band was at least equal in intensity to the unspliced band, even when the latter was amplified at low efficiency, see for example lane 12 of panel a or lanes 9 and 10 of panel b; whereas for *pER8-* based *Tto1* constructs (panel b) the cDNA band was never more intense than the unspliced band. Moreover, in some cases (panel a, lane 10 and 11) when the T-DNA band was poorly amplified the cDNA band resulted to be more intense.

Another important feature for an inducible system is that it should be efficiently repressed in absence of the inducer. As repression in absence of β-estradiol did not

appear to be always efficient with *pER8* based constructs- as indicated by the dot in panel b and observed in other cases not reported - the *pOp6/LhGR* system suggested being more reliable in this respect.

Appendices to Results



Preparative agarose gel for Southern Blot presented in Fig. 2.10.

Alignment of 20 sequences of *Tto1* integrase core domain protein sequence

➔ Insert in 61 between Q9 and V10: STRYSPQVPGPKLQASISQQ; clone 62 has 2 frame shifts that were corrected prior to translation.

➔ Tobacco ecotypes color code : ...., BY2; ...., SR1; ...., Xanthi; ...., Samsun NN; ...., W38; ...., *N. sylvestris*.

```
Tto1-1  SFQASISQQVINVAENDSNIKLWHRRLGHMSEKSMARLVKKNALPGLNQIQLKKCADCLAGKQNRVSFKRFPPSRRQNVLDLV
41      SFQASISQQVINVAENDSNIKLWHRRLGHMSEKSMARLVKKNALPGLNQIQLKKCADCLAGKQNSVSFKRFPPSRRQNVLDLV
42      SFQASISQQVINVAENDSDIKLWHRRLGHMSEKSMARLVKKNALPGLNQIQLKKCADCLAGKQNRVSFKRFPPSRRQNVLDLV
43      SFQASISQQVINVAENDSNIKMWHRRLGHMSEKSMARLVKKNALPGLNQIQLKKCADCLAGKQNRVSFKRFPPSRRQNVLDLV
44      SFQASISQQVINVAENDSNIKLWHRRLGHMSEKSMARLVKKNALPGLNQIQLKKCADCLAGKQNRVSFKRFPPSRRQNVLDLV
45      SFQASISQQVINVAENDSNIKLWHRRLGHMSEKSMARLVKKTALPDLNQIQLKKCADCLAGKQNRVSFKRFPPSRRQNVLDLV
46      SFQASISQQVINVAENDSNIKLWHRRLGHMSEKSMARLVKKNALPGLNQIQLKKCADCLAGKQNRVSFKRFPPSRRQNVLDLV
47      SFQASISQQVINVAENDSNIKLWHRRLGHMSEKSMARLVKKNALPGLNQIQLKKCADCLAGKQNRVSFKRFPPSRRQNVLDLV
53      SFQASISQQVINVAENDSNIKLWHRRLGHMSEKSMARLVKKNALPGLNQIQLNKCADCLAGKQNRVSFKRFPPSRRQNVLDLV
54      SFQASISQQVINVAENDSNIKLWHRRLGHMSEKSMARLVKKNALPGLNQIQLKKCADCLAGKQNRVSFKRFPPSRRQNVLDLV
57      SFQASISQQVINVAENDSNIKLWHRRLGHMSEKSMARLVKMNALPGLNQIQLKKCADCLAGKQNSVSFKRFPPSRRQNVLDLV
59      SFQASISQQVINVAENDSNIKLWHRRLGHMSEKSMARLVKKNALPGLNQIQLKKCADCLAGKQNRVSFKRFPPSRRQNMLDLV
60      SFQASISQQVINVAENDSNIKLWHRRLGHMSEKSMARLVKKNALPGLNQIQLKKCADCLAGKQNRVSFKRFPPSRRQNVLDLV
61      SFQASISQQVINVAENDSNIKLWHRRLGHMSEKSMARLVKKNALPGLNQIQLKKCADCLAGKQNRVSFKRFPPSRRQNMLDLV
62      SFQASFSQQFINVAENDSNI*LWHRRLGLMSEKSMARLVKKNALPCLNQIHLIKCADCLPGKQNRVSFKRFPPSRRKNVLDLV
63      SFQASISQQVINVAENDSNIKLWHRRLGHMSEKSMAHLVKKNALPGLNQIQLKKCADCLAGKQNRVSFKRFLPSRRQNVLDLV
65      SFQASISQQVINVAENDSNIKLWHRRLGHMSEKSMARLVKKNALPGLNQIQLKKCADCLAGKQNRVSFKRFPPSRRQNVLDLV
67      SFQASISQQVINVAENDSNIKLWHRRLGHMSEKSMVRLVKKNALPGLNQIQLKKYADYLAGKQNRVSFKRFPPSRRQNVLDLV
68      SFQASISQQVINVGENDSNIKLWHRRLGHMSEKSMARLVKKNALPGLNQIQLKKCADCLAGKQNRVSFKRFPPSRRQNVLDLV
```

D546;
conserved residue
in catalytic site
D366 – D367;
either one can be part
of the catalytic site
tol-1
41
42
43
44
45
46
47
53
54
57
59
60
61
62
63
65
67
68
69

Ess:
conserved residue
in catalytic site
Tto1Y
Tto1N
Tto1N
Tto1-1
HQFTPPKTPQLNGLAERMNRTLIERTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVPLQYKAPEKIWLGRDISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRTLIERTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVPLQYEAPEKIWSGRDISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRTLIERTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVSLQYEAPEKIWSGRDISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRTLIERTRCLLSHSKLPKAFWGEALVTAAYVVPNHSPCVPLQYEAPEKIWSGRDISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRTLIERTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVPLQYKAPEKIWLGRDISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRTLIERTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVPLQYEAPEKIWSGRDISYDQLRVFCC
HQFTPPKTPQLNGLAERMNRTLIERTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVPLQYEAPEKIWSGRDISYDQLQVFGC
HQFTPPKTPQLNGLAERMNRTLIERTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVPLQYEAPEKIWSGRGISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRTLIERTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVPLQYEAPEKIGQGRDISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRTLIERTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVPLQYKAPEKIWLGRDISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRTLIERTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVPLQYEAPEKIWSGRDISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRTLIERTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVPLQYEAPEKIWSGRDISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRTLIEKTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVPLQYEAPEKIWSGRDISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRTLIERTRCLLSHSKLPKAFWGEALVTTAYVLNHSPCVPLQYEAPEKIWSGRDISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRTLIEKTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVPLQYEAPEKIWSGRDISYDQLRVFGC
HQFTPPKTPQLNGLVERMNRTLIERTRCLLSHSKLPKAFWGEALVTATYVLNHSPCVPLQYEAPEKI*SGRDISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRTLIERTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVPLQCEAPEKIWSGRDISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRTFIERTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVPLQYEAPEKIWSGRDISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRILIERTRCLLSHSKLPKDFWGEALVTAAYVLNHSPCVPLQYEAPEKIWSGRDISYDQLRVFGC
HQFTPPKTPQLNGLAERMNRTLIERTRCLLSHSKLPKAFWGEALVTAAYVLNHSPCVPLQYEAPEKIWSGRDISYDQLRVFGC

Tto1 Y
Tto1 IN
Tto1-1
41 KAYVHVPKDERSKLDVKTRECVFIGYGQDMLGYKFYDPVQKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVVPRQ
42 KAYVHVPKDERSKLDVKTRECVFIGYGQDMLGYKFYDPVQKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVVPRQ
43 KDYVHVPKDERSKLDVKTRECVFIGYGQDMLGYKFYDPVQKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVVPRQ
44 KAYVHVPKDERSKLDVKTRECVFIGYGQDMLGYKFYDPVEKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVVPRQ
45 KAYVHVPKDERSKLDVKTRECVFIGYGQDMLGYKFYDPVQKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVVPRQ
46 KAYVHVPKDERSKLDVKTRECVFIGYGQDMLGYKFHDPVQKKLVRSRDVVFVEDQTIEDIVEKSTDDSAEFELPPTVVPRQ
47 KAYVHVPKDERSKLDVKTRECVFIGYGQDMLGYKFYDPVQKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVGPRQ
53 KAYVHVPKDERSKLDVKTRECVFIGYGQDMLGYKFYDPVQKKLVRS*DVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVVPRQ
54 KAYVHVPKDERSKLDVKTRECVFIGYGQDMLGYKFYDPVEKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVVPRQ
57 KAYVHVPKDERSKLDVKTRECVFIGYGQDMFGYKFYDPVQKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVVPRQ
59 KAYVHVPKDERSKLDVKTRECVFIGYGQDMLGYKFYDPVQKKLVRSRDVVFVEDQTIEYIDKVEKSTDDSAEFELPPTVVPRQ
60 KAYVHVPKDERSKLDVKTRECVFIGYGQDMIGYKFYDPVQKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVVPRQ
61 KAYVHVPKDERSKLDVKTRECVFIGYGQDMLGYKFYDPVQKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVVPRQ
62 KAYVHVPKDERSKLDIKTRECVFIGYGQDMLGYKFYDPVQKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVVPRQ
63 KAYVHVPKDERSKLDVKTRECVFIGYGQDMLGYKFNDPVQKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVVPRQ
65 KAYVHVPKYERSKLDVKTRECVFIGYGQDMFGYKFYDPVQKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVVPRQ
67 KAYVHVPKDERSKLDVKTRECVFIGYGQDMLGYKFYDPVQKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVVPRQ
68 KAYVHVPKDERSKLDVKTRECVFIGYGQDMLGYKFYDPVQKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSVEFELPPTVVPRQ
69 KAYVHVPKDERSKLDVKTRECVFIGYGQDMLGYKFYDPVQKKLVRSRDVVFVEDQTIEDIDKVEKSTDDSAEFELPPTVVPRQ

- ➔ The 55 not conserved residues are in bold on *Tto1-1* sequence, apart from K_{628}, L_{636}, E_{690}, G_{747} and L_{754} that are in red.
- ➔ The sequence of the "consensus" clone AA43 in underlined in green.
- ➔ The residues present in AA43/ *Tto1X*, which fall in the major group, are highlighted in yellow.
- ➔ The residues that different between *TtoY1* and *Tto1X* are labeled on *Tto1-1* sequence and highlighted in red on AA43 sequence.

Chapter 3

DISCUSSION

3.1 From *Tto1-1* to i*Tto1*: engineering of a retrotransposon

The investigation on plant transposable elements has been so far always linked to establishing *in vitro* culture and regeneration protocols to activate transposition. This is due to the fact that all plant elements studied so far have a stress responsive promoter that restricts their activity only to stress conditions.

Tto1-1 transcripts were first isolated from tobacco protoplasts (Hirochika, 1993), where mRNA synthesis was regulated by its natural 5' LTR promoter. Subsequently it was demonstrated that *Tto1* is activated by a number of defense-related biotic and abiotic stresses in addition to tissue culture, such us viral infections, treatment with methyl jasmonate and fungal elicitors chitin oligomer and xylanase (Hirochika et al., 1996a; Hirochika et al., 1996b; Takeda et al., 1998, 1999). In later years Sugimoto et al. also demonstrated that the tobacco transcription factor NtMYB2, which is transcriptionally regulated by wounding and treatment with fungal elicitors, activates *Tto1* LTR promoter by binding the 13bp *cis*-regulatory elements, the L-Box and H-Box-like motif, contained in the LTR, and that these are sufficient to induce transcription of the element (Sugimoto et al., 2000). As all other plant retrotransposons, *Tto1* has been studied always by tissue culture induced retrotransposition. When the *Tto1* LTR natural promoter was replaced by the double CaMV 35S promoter, transposition was detected in regenerated plants (Böhmdorfer et al., 2005). In general the transcriptional profile and thus the properties of cultured plants differs significantly from differentiated tissues (Böhmdorfer et al., 2010). Tissue culture is, additionally, a long and tedious work and during the prolonged *in vitro* manipulations other elements can also be induced and transpose (Hirochika, 1992, 1993), thus leading to unwanted somaclonal variation of the regenerated plant. Furthermore, the continuous expression provided by the 35S promoter is expected to switch on RNA-based transcriptional and post-transcriptional defense mechanisms, that "turn down"

the retrotransposon by promoter silencing and/or RNA degradation (Cheng et al., 2006; Ding et al., 2007; Matzke and Birchler, 2005; Miura et al., 2001).
In sum tissue culture seems not to offer control on element replication that would support biochemical analysis. Dissecting the retrotransposon life cycle in its different steps, is a necessary condition to investigate the various aspects of transposition mechanisms; therefore molecular engineering of *Tto1* aimed at providing the element with features that allow easier handling and analysis in the whole plant.

3.2 i*Tto1* as a molecular tool for new gene isolation

The employment of transposable elements in genetic analysis is not a new concept. Prokaryotic IS elements have been used in the past for mutagenesis of bacteria (Kleckner, 1977). The mobile T-DNA from *Agrobacterium* (Azpiroz-Leehan and Feldmann, 1997; Krysan et al., 1999) as well as Class II elements such as maize transposons *Ac/Ds* (Parinov et al., 1999) and *En/Spm* (Speulman et al., 1999; Tissier et al., 1999; Wisman et al., 1998) have been used in forward and reverse genetics in Arabidopsis. Although many genes have been isolated using these elements, there are several limitations associated with their mechanism of replications, for example that a high number mutations are not tagged by *Ac/Ds* (Bancroft et al., 1993) or T-DNA (Castle et al., 1993), possibly due to imprecise excision of *Ac/Ds* and abortive integration of T-DNA, which is usually not found with retrotransposons (Hirochika, 1997). Another bottleneck to the use of DNA transposons is represented by their “cut and paste” replication mechanism, which induces unstable mutations; in addition DNA transposons tend to generate “nested” insertions (Bancroft and Dean, 1993), therefore a very large number of plants would be necessary to obtain unlinked mutations, that are distributed on all the chromosomes.
A few active plant retrotransposons are known, among which *Tos17* of rice (Hirochika et al., 1996b), *Tnt1* (Grandbastien et al., 1989) and *Tto1* (Hirochika, 1993) are the only elements whose transcriptional and translational activities have been demonstrated. *Tos17* has been largely used for tissue culture-induced gene mutagenesis of rice (Hirochika, 1997, 2001; Miyao et al., 2003), showing preference for low copy sequences and for genes (Yamazaki et al., 2001). *Tnt1* and *Tto1* can also transpose in heterologous host plants. *Tnt1* is in fact activated by tissue culture in Arabidospsis and *Medicago truncatula* (Cheng et al., 2011; Lucas et al. 2005); *Tto1* transposes in Arabidopsis and rice (Hirochika et al., 1996a; Okamoto and

Hirochika, 2000) and both to insert preferentially into genes. It is noteworthy that, although they all belong to the *Ty1/copia* group that is ubiquitous in plants (Hirochika and Hirochika, 1993; Voytas et al., 1992), and could be theoretically applied to a large number of plants, only *Tto1* so far has been shown to have activity in a monocot plant (rice) (Hirochika et al., 1996a) and to be transcribed in barley (G. Böhmdorfer unpublished). Considering that monocots and dicots diverged 200 million years ago, we were prompted to put efforts in understanding crucial steps of *Tto1* replication in order to broaden its application range in saturation mutagenesis of crops.

3.3 Technical and scientific advances of i*Tto1* in plant mutagenesis

3.3.1 "Transposition on demand"

The use of an inducible promoter makes it possible to obtain a complete cycle of transposition in the whole plant, skipping the callus and regeneration procedure, and also to separate the transposition cycle in its different steps in order to investigate on specific aspects.

Deletion studies of *Tto1* 5' leader provided information that allowed replacing the natural LTR promoter with a heterologous promoter. The first attempt to obtain inducible transposition of *Tto1* was done with a Dexamethasone responsive promoter (Böhmdorfer et al., 2005), which on one hand gave high transcription levels, but on the other hand turned out to be very toxic to plant. To overcome this problem *Tto1* was linked to a β-estradiol inducible promoter, which created the basis for the experiments carried out in this work. With this system we were able to switch on and off the expression of *Tto1*; the transcription of the element was in fact interrupted when the plants were transferred on soil devoid of the inducer. We wanted to induce transposition in the apical meristem, because this tissue differentiates both in somatic and in gamete cells. To obtain an optimal exposure of the apical meristem to the β-estradiol, the seeds were germinated in liquid medium containing the inducer; this could nevertheless represent a bottleneck for recalcitrant seeds that do not germinate *in vitro*. Other methods for the β-estradiol treatment can be imagined however, for example hydroponic culture where the inducer is provided through the roots and systemically transported to the apical meristem, - in the laboratory I actually use a variation of this method in small scale with Arabidopsis seedlings. Alternatively the inducer could be supplied, by spraying or addition of drops of the inducer directly on the apical meristem of adult plants in a pre-reproductive phase.

However, the technical advance of i*Tto1* has provided an interesting approach in which biology and synthetic biology can proceed simultaneously and in a mutually beneficial way.

In my work one of the very few active plant retrotransposons has been engineered and used for plant mutagenesis, and at the same time has been used to investigate important aspects of its replication cycle. "On demand" transposition of i*Tto1* is a big step forward in this field of research, because the researchers can keep the element silent under normal conditions and induce it according to their experimental need, by simply supplying a chemical to the plant, without any regeneration step. In addition to the technical advance of such methodology, my results demonstrate that *Tto1* replication can be made independent from the plant inducing factors. This has two important consequences from a scientific point of view. First, *Tto1* life cycle can be dissected into single steps, thus focusing on specific aspects of its replication; second, the transposition process can be studied *in vivo*, using a wild type background for a model. The i*Tto1* technology has been successfully employed by my colleagues to explain key aspects of the element translation (Böhmdorfer et al., 2008). As will be discussed in the next sections, the same approach, combined with sequence prediction informatic tools, allowed us to understand the role of LTR in reverse transcription, and to build a mechanistic model for its role.

Considering that *Tto1* is active in different host plants, both dicots and monocots, I think that the "transposition on demand" approach can be applied to investigate other plant transposable elements and the cellular factors that control them. In combination with its gene preference it can be used as an insertional mutagen in a wide range of plants.

3.3.2 Intron-PCR, a powerful screening method

One of the important modifications on *Tto1*, which also provides a powerful and simple genotyping method, was the insertion of two Arabidopsis introns, in the GAG and in the INT domain respectively. The two introns are used as labels to monitor reverse-transcription by a simple PCR assay (Böhmdorfer et al., 2008; Böhmdorfer et al., 2010). Intron-PCR is a method that employs primers flanking the intron and allows, in one step, the distinction of the extra-chromosomal spliced copies, which have been properly reverse-transcribed and have therefore lost the intron, from transgene copies carried by the T-DNA. Another important aspect is that Intron-PCR can also identify plants with new transposition events. As shown in Fig. 2.4a lane 2, the presence of the single intron less band indicated unambiguously that the PCR

product resulted from amplification of an inserted copy of *Tto1*. It is also possible to screen for the abundance of non integrated transcripts: the decrease of the intracellular non integrated cDNA can be seen during various steps of cell divisions. Intron-PCR also provides indirect proof of the formation of the VLP that is a prerequisite for the cDNA synthesis; this methods is more sensitive than the immunological detection of GAG protein.

3.3.3 i*Tto1* preferentially inserts into genes

The stress activated *Tto1* has inserts into genes. In transgenic Arabidopsis regenerated plants, 74% of independent *Tto1* insertions, driven by its natural LTR promoter, occurred into active coding sequences spread all over the five chromosomes, (Okamoto and Hirochika, 2000). Using a previous 35S promoter, *Tto1* also inserted into a constitutively expressed gene (Böhmdorfer et al., 2005). In this work two insertions have been characterized, one of which occurred between two metabolic genes and the second one occurred into a structural gene, suggesting that the engineered *Tto1* also maintains its preference for genes. Notably, the three mentioned insertions are localized respectively on chromosome 4, 2 and 3 of Arabidopsis, showing a propensity of the element to spread all over the genome. This fact is consistent with the "copy and paste" replication mechanism of retrotransposons, where the high number of new sequences increases the probability to cover a large part of the genome. Preference for genomic regions with a high transcriptional profile also seems to be characteristic of retrotransposons of the *Ty1/copia* superfamily (Cheng et al., 2011; Lucas et al. 2005; Yamazaki et al., 2001; Okamoto and Hirochika, 2000; Hirochika et al., 1996a).

3.3.4 i*Tto1* induces stable and unlinked mutations

Tto1 integrated after induction of i*Tto1* with β-estradiol is stable in successive generations. So far we have not observed any secondary transposition of the element. Considering also that in previous experiments *Tto1* transcripts were detected in cultured but not in cells of intact normal plants (Hirochika, 1993; Wessler, 1996), we believe that this absence of transposition should be maintained in other species under normal life conditions. The restriction to stress related conditions and the apparent selection against the germinal transposition can also be seen as an advantage to the element. As suggested by Moreau-Mhiri et al., 1996, in plants where germ cells derive from somatic cells that continue dividing throughout the whole development, the potential insertion of new retrotransposon copies (usually

into a gene) in some lineages that will proceed to production of gametes, would increase the germinal mutation frequency to such an extent to reduce the fertility of the host. Thus, elements that adopt such replication cycle are expected to be quickly eliminated from the population (Moreau-Mhiri C and H, 1996). Circumventing this is an advantage of the application of i*Tto1* in plant mutagenesis as well as the fact that once a new element copy is inserted in the genome, it will segregate like a normal gene, as demonstrated by the Southern blot experiment in Fig. 2.4; it should therefore be reasonably easy to select single mutants. In addition, i*Tto1* seems also suitable for saturation mutagenesis, in light of the fact that the insertions are unlinked and that *Tto1* integrates into genes, therefore not a large number of plants will be needed to cover a whole genome. Handling big populations can be a real problem with plants larger than Arabidopsis. As Hirochika already estimated for example, a population of 50,000 mutant lines of rice would be required to provide 99% of probability of finding a mutant of any one gene (Hirochika, 2001).
Nevertheless, several features of the element can be improved by further studies.

3.4 Possible improvements of i*Tto1*

We have obtained a 4% insertion rate (3/70 plants) of i*Tto1*, with up to four new insertions per plant in the progeny, while with previous double 35S promoter 2/13 regenerated plants showed transposition events. Thus, it might seem then that the new system is not as efficient, but there are additional considerations. First of all, with i*Tto1* there is no need for callus culture because transposition is induced in the whole plant. Second, it is not clear whether this number depends on the generally low activity of the β-estradiol inducible promoter or on our screening method. It is likely that the number of plants with transposed copies would be higher (2-fold), because only one cauline leaf per plant was used for the diagnostic Intron-PCR. That is, since we induced transposition in the apical meristem, and since this tissue has a minimum number of two genetically effective cells (Rédei and Koncz, 1992), such number would be correspondingly higher if the meristem consists of more cells at the time of transposition (Böhmdorfer et al., 2010). Consequently, another advantage of the inducible promoter is that transposition can be induced at a later developmental stage, when the meristem is larger. Furthermore, I think that for this purpose the method for inducer supply could also be improved, in a way to carry out the chemical treatment on plants with a larger size compared to those that are grown *in vitro* in

our common experimental settings. However, the performance of i*Tto1* can be improved by using other expression systems like the Dexamethasone inducible pOp6/LhGR that is already being tested in Arabidopsis (see below).

Another factor that can influence the number of insertions and the transposition in general might be the occurrence of silencing mechanisms in the host cell.
It is known that DNA methylation is a common mechanism to suppress transposable elements, and it directly affects expression of native *Tto1* in Arabidopsis (Hirochika et al., 2000); but it also possible that extra-chromosomal copies can trigger transposon suppression mechanisms (Böhmdorfer et al., 2010). The first way to overcome this inconvenient is represented by the inducible promoter, so that the element is normally kept silent and "active" copies of *Tto1* will arise only after addition of the β-estradiol. RNA- based or posttranscriptional silencing can also dramatically reduce the activity of the retrotransposon. In relation to that it should be considered that Arabidopsis' small genome, compared to the vast majority of the higher plants, reflects its highly reduced number of TE DNA, and this could be the result of a very efficient defense mechanism evolved by this plant. In this regard the T-DNA mediated transformation ensured that only one or two copies of *Tto1* are inserted into Arabidopsis genome, thus reducing the amount of homologous sequences in the nucleus. In other words, I believe that modifications on the LTR, which could represent a typical target for homology-dependent gene silencing (Jordan, 2009; Matzke and Birchler, 2005; Tijsterman et al., 2002) should be effective. For example reducing the length of the redundant sequence between the 5' and the 3' ends to a minimum will lead to an improvement of the element (see below).
Genetic investigation is in addition being carried out on the Arabidopsis mutant *ddm1* (defective in DNA methylation), that has been transformed with i*Tto1*. The native *Tto1* was already shown to be re-activated in such mutant in Arabidopsis (Hirochika et al., 2000), showing the clear role of methylation in TGS of the retrotransposon, therefore it will be of undoubted interest to see whether also the low copy engineered i*Tto1* is subjected to the same kind of repression.
Integration is another crucial step with probable influence on the copy number, since the integrase is involved in reactions with both the element and the host cell DNA. In light of the fact that, in spite of the high levels of Dexamethasone induced *Tto1* transcription, no transposition was detected, we think that integrase might specifically be a target of posttranslational modifications. Therefore different

investigation approaches have been tried in this work to gain information about this enzyme that should be used to improve the insertion efficiency of i*Tto1*.

3.5 Application of i*Tto1* based constructs in functional analysis

A Dexamethasone inducible *Tto1* construct was already successfully used to investigate translation of the element encoded proteins (Böhmdorfer et al., 2008).

As discussed previously, the β-estradiol inducible i*Tto1* maintains all the features of the native element; therefore it has been employed in a functional analysis of *Tto1* reverse transcription. i*Tto1* is silent when introduced to the plant and due to its chemically inducible promoter, we could induce it at our experimental need to focus on one single cycle of cDNA synthesis. In addition, for this specific part of my work, since all secondary possible effects due to integration were to be avoided, all constructs with deletions of the 3' LTR, contained an E_{583} to A change to inactivate the integrase (Böhmdorfer et al., 2008), with no "side effect" on any other replication intermediate. The Intron-PCR was again routinely employed to monitor reverse transcription. Moreover, the upgraded "Long PCR", proved an efficient method to screen the active constructs, and its results correlated perfectly with those of Southern blots. The molecular and biochemical approach was then complemented by an *in silico* analysis of the characteristic constructs that led to the creation of a mechanistic model for the strand transfer process. These results, with my particular satisfaction, have been recently published (Tramontano et. al., 2011) in the journal "Virology".

3.6 The multiple roles of LTR

As described in the first chapter, and supported by experimental data, LTRs play multiple roles. On one hand, the 5' LTR carries out the role of transcriptional promoter, as well as of guidance of translation. On the other hand, the 3' LTR has transcriptional terminator function and mediates the template switch by the cDNA leader during reverse transcription. An accepted model called "LTR replication" in retroviruses (Fig. 1.4), proposes that this process is mediated by the "R" region contained in the LTR, nevertheless the functional dynamics of this sequence were not known to date.

3.6.1 Termination sites in the LTR

Previous results (Böhmdorfer et al., 2005) showed that termination can occur at different points in the LTR. Most transcripts had the poly-A attached to position 4914, while the second most frequent class of mRNA ended at position 5230. Previously it was shown that the naturally occurring mRNA starts at position 200 (Hirochika, 1993). It follows that an mRNA with these features would not be able to support reverse transcription for two reasons. First, a *Tto1* transcript ending at position 4914, which corresponds to position 188 of the 5' LTR, does not have the sequence redundancy that is necessary to carry out the template switch of the strong stop cDNA leader. Second, RNA folding prediction suggested that an mRNA starting at position 200 would be not accessible to the ribosome, due to a tightly base paired conformation of its 5' region, thus preventing translation (Böhmdorfer et al., 2005). In contrast, mRNAs ending with nucleotide 5230 that corresponds to 504 in the 5' LTR will have 333bp overlap of R sequence for the first strand transfer. Nevertheless, we had no direct evidence about the sequence length required for this process, so the constructs with serial deletions (from 4900 to 5233) were specifically designed to span the two termination points and try to define the R region. The employment of the strong rbcS terminator from pea, made it sure to obtain functional transcripts and therefore the translation of all the constructs in which the natural "late" terminator (5230) was deleted (practically all except for the deletion construct A). In this way I could isolate the mRNA of both constructs C and D, and confirm, (Fig. 2.13b), that the two natural termination sites are conserved in inducible constructs and that the "early" terminator still remains predominant. More importantly, the transcripts ending in the rbcS promoter supported the correct *Tto1* protein synthesis, thus allowing the analysis of the reverse transcription.

3.6.2 Role of the R region and mechanistic model

By the "Long-PCR" (Fig. 2.9b) and the Southern blot (Fig. 2.10) we proved that *Tto1* transcripts with the "early" termination point do not support reverse transcription. In Fig 2.9b we can see that only constructs A, B, C gave rise to the expected band, indicating that the 100bp sequence differing between constructs C and D might be essential to reconstitute a full-length LTR. The result of the Southern blot confirmed that of Long-PCR, showing that full-length (5.3kb) cDNA was made by constructs A, B and C and by the undeleted i*Tto1* control only. In sum, mRNAs with termination point at position 4914 are not templates for *Tto1* sequence replication. For structure prediction analysis constructs C and D were of particular interest because the "Long-

PCR" results (Fig. 2.9b) showed that reverse transcription was clearly compromised in construct D, while in C it was still properly carried out. The subsequent *in silico* analysis pointed at elucidating the role of the 100bp sequence stretch between the constructs C and D.

The 3' mRNA ends of constructs C and D fold similarly, except for one additional hairpin structure formed by the 100nt, exclusively present in C (Fig. 2.11). Most interestingly a complementary hairpin is also formed in the 5' end of the strong-stop cDNA leader, by a sequence having perfect homology with that of construct C (Fig. 2.12a-b). This finding suggested a model for a sequence redundancy search between the two LTRs of the mRNA. The formation of the 9nt loop was confirmed by "RNAup" that predicted for this sequence the highest probability to be single stranded (over 90%, Fig. 2.12a-b) in the whole sequence of both mRNA 3' end of construct C and in 5' end of the cDNA leader. The hypothesis was made that the cDNA/mRNA hybridization is a kinetically favored mechanism in which the formation of a perfect heteroduplex is mediated by "kissing hairpins" (Chang and Tinoco, 1994). Energetic parameters further supported this model, in fact "RNAcofold" calculated a sharp gain of -235kj/mol for the formation of the heteroduplex, against the formation of separate secondary structures of the single stranded cDNA leader and mRNA. In other words, the two hairpins come in close contact in the VLP and base pair, due to their perfect complementarity, thus extending the melting of the secondary structures of cDNA leader and mRNA along the whole sequence. In this way, a very stable heteroduplex is formed, and the cDNA will be extended before the 5' end of the mRNA template is reached (see also Fig. 1.3). The model is summarized in Fig. 3.1, drawn by A. Bachmair and recently published (Tramontano et al., 2011).

Nevertheless, some considerations still need to be made. Usually retroviruses include two copies of RNA in their capside, which form dimers. The dimerization is functionally linked to packaging and not to the reverse transcription process; in addition it is not demonstrated that during dimerization the 5' and 3' ends of two mRNAs are aligned (Jewell and Mansky, 2000; Paillart et al., 1996). It has also been demonstrated though, that in the related Ty3 retrotransposon of yeast, a complex consisting of two mRNAs and two tRNA primers is formed before reverse transcription starts (Gabus et al., 1998); and in HIV it has been proposed that the binding of the tRNA primer to the 3'end of the mRNA may facilitate the strand transfer (Brule et al., 2000). In my case "RNAcofold" did not predict any dimer structure for *Tto1*, which might be a relatively simpler model. The model shown in Fig 3.1 indicates that two mRNA

templates are involved in the strand transfer, basically that the strong stop cDNA is transferred from the 5' end of one molecule to the 3' end of another molecule. Although we do not know whether for *Tto1* the R sequence of LTR of one or two mRNA molecules mediates the cDNA/mRNA hybridization, this is irrelevant for the explanation of the mechanistic interaction of the kissing hairpins.

In sum, using *Tto1* as a model, and supported by different methods we have shown the crucial role played by secondary structures assumed by emerging cDNA and mRNA template during reverse transcription, specifically with regard to the first strand transfer process. The absence of such secondary structure impairs cDNA synthesis and consequently leads to no transposition.

However, we still want to provide more direct evidence to improve our model, by testing constructs in which the DNA region that was predicted to form the characteristic loop is deleted. In addition, since a *Tto1* with an inactivated integrase was used, the next experiments will try to demonstrate that the constructs with deletions in the LTR can complete the transposition cycle by effectively making new insertions into the genome

3.7 i*Tto1* adopts an "invasion strand transfer" mechanism

Another important feature to be discussed is the terminal extension of 32bp only present at the 5' LTR (see Fig. 2.6d) of all i*Tto1* based constructs. In a model proposed for the first strand transfer of HIV, called "terminal transfer", the template switch takes place once the synthesis of the strong stop cDNA leader has reached the mRNA 5' end (Basu et al., 2008). My results show that this is not the case for *Tto1*, because according to this model the cDNA leader containing such a sequence would not be able to base pair with the 3' LTR that does not contain it, thus blocking the reverse transcription. Since reverse transcription of *Tto1* was not affected by such unspecific extension, it in contrast seems to follow another model, called "invasion transfer", in which the cDNA leader is transferred to the other template before the 5' end is reached. It is however not excluded that this is only characteristic of the engineered element, and that a native *Tto1* might adopt both modes. Since the *Tto1* is still a different model compared to retroviruses and since the engineered *Tto1* seems to efficiently transpose the "invasion transfer" might be predominant.

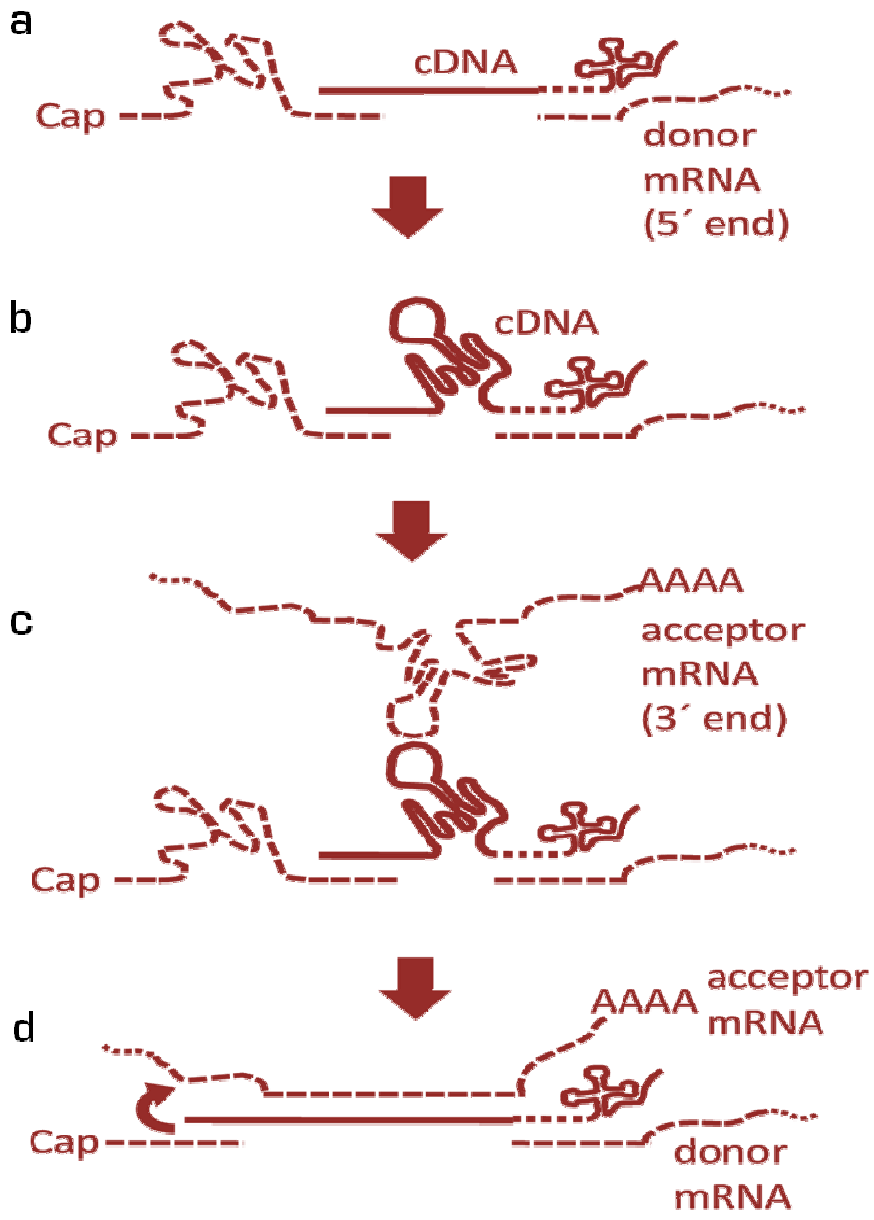


Fig. 3.1. Model for first strand transfer of *Tto1* reverse transcription. (a) Reverse transcription starts with alignment of the tRNA primer. The RNase H function of reverse transcriptase degrades the mRNA of the emerging RNA/DNA duplex, starting after the RNA-RNA hybrid formed by primer and mRNA, ending at a sliding window of short heteroduplex that is bound to the reverse transcriptase. The emerging single-stranded cDNA engages in secondary structure formation (b), but at least one characteristic sequence (identified in this work) remains single-stranded. (c) The 3´ end of the same or of another mRNA molecule contains a complementary loop, which starts base pairing. Perfect complementarity between cDNA and mRNA 3´ end favors formation of heteroduplex, thereby replacing secondary structures (d). In this complex, donor and acceptor mRNA are closely aligned, facilitating template switch of reverse transcriptase (curved arrow) before RT reaches the mRNA 5´ end.

3.8 Implications of a "shorter active" redundant region

A sequence redundancy between 5' and 3' end of the mRNA is essential, for the completion of cDNA synthesis.

With my results I have demonstrated that an mRNA with only 125bp overlap instead of the canonical 574 can still be reverse transcribed. Provided the direct evidence that constructs with deletion of the 3' LTR can actually transpose, an interesting consideration has to be done concerning the significance of defining a "shorter active" redundant region. Generally accepted models for homology dependent gene-silencing mechanisms (also called co-suppression), indicate that the severity of the repression correlates with the copy number and does not require translatable sequences (Baulcombe, 2004; Jensen et al., 1999; Jorgensen, 1995; Matzke and Birchler, 2005; Meyer and Saedler, 1996; Reuter et al., 2009; Tijsterman et al., 2002). A key point in this process is the arising of antisense RNA species whose probability to occur from different copies is higher than from a single one. Retrotransposons are a perfect example of repetitive sequences, considering their mode of replication and their incredibly high abundance in eukaryotic genomes. Notably, the LTR can be a template for antisense transcript production and induce silencing also when it is the only repetitive sequence present, like in the case of a single copy retrotransposon.

Therefore I think that reducing the repetitive sequences in *Tto1* to a minimum, should reduce gene silencing levels *Tto1* transgenes thus contributing to increase its efficiency in gene targeting applications.

3.9 Integrase (or a DNA tailor)

3.9.1 Integrase might influence transposition rate

The Integrase (INT) plays a crucial role in the transposition mechanism as well, since it ensures the orderly linkage of the retrotransposon DNA with the host genome. It is indeed the integrated DNA that is transcribed by the host machinery, to give rise to the RNA that will serve as genome template and as messenger for the element proteins. Biochemical studies have demonstrated that INT is involved in the "processing" of the emerging cDNA, which is the preparatory step before insertion. Data suggest that INT is also indirectly involved in the targeting of the insertion site in the host cell genome, as revealed by studies on yeast retrotransposons. Ty5 usually integrates in heterochromatin at the telomers and at the silent mating loci. The

target specificity is mediated by the interaction of the TD (targeting domain), at the C-terminus of the INT, with the heterochromatic protein Sir4, and that phosphorylation is required for this interaction (Dai et al., 2007). A single amino acid change in this region abolished targeting to silent chromatin and led to random integration of the element (Gai and Voytas, 1998), suggesting that if not directly with the DNA, INT interacts with DNA associated proteins. Another possibility that determines the targeting of the integration site, could be a concerted action with the reverse transcriptase. For instance in *Ty1* of yeast, which targets upstream regions of RNAPol III promoters, RT and IN remain closely associated after reverse transcription, until the formation of the pre-integrative complex (Wilhelm and Wilhelm, 2006). In this case it would be interesting to identify possible replication intermediates after the first strand transfer that has been previously described.
In another case in *Ty1* it was found that mutations in RTT genes, that regulate repairing of double strand breaks, determined an over 100-fold increase of transposition (Scholes et al., 2001).
In a few words, INT interacts with retrotransposon DNA and either directly or indirectly with the host cell genomic DNA.
Concerning INT's modes of interaction with DNA and DNA associated proteins, structural studies would be of help in this regard, but there will be still time to wait, since crystals of the HIV IN have been obtained only last year (Maertens et al., 2010).

In any case we want to know about *Tto1* integrase and as a first step I tried to define the termini of the protein, as a single moiety after protease cleavage. Since gene bank sequence search provided partial information, I tried the immune-precipitation of the enzyme. The antibody was raised, but it did not prove effective in detecting the integrase in a protein extract of Arabidopsis expressing *Tto1*. Considering that the *pER8* based expression is relatively low, strategies to overexpress the integrase in planta are already under evaluation, and another expression system has already been preliminarily tested (*pOp6/LhGR*, §3.10). We will still try to obtain a purified integrase and proceed to mass spectrometry, to identify the protease cleavage sites on both termini and possible posttranscriptional modifications that the protein might undergo in the host cell. We also wish to carry out cellular localization studies on the integrase, via conjugation it with GFP or other tags, therefore identification of the protein boundaries will be a priority in the next future.

In the meantime, in my lab new transgenic Arabidopsis lines are being assessed to gain information about the relation between integrase and DNA and DNA associated proteins. In fact i*Tto1* has been transformed in two lines with mutations in orthologs loci of human genes *Ku70* and *FancD2* that can be considered as "caretakers" of DNA damages (Hays et al., 2008). The first gene encodes a protein that together with Ku80 is known to make up the Ku heterodimer, which binds to DNA double-strand break ends and is required for the non-homologous end joining (NHEJ) of DNA repair pathway. Mutants for the second gene also show defects in DNA repair with hypersensitivity to DNA cross-linking agents, increased chromosomal breakage that determine chromosomal instability. We think that the analysis of *Tto1* replication in these mutant backgrounds should definitely contribute to help shedding light on the retrotransposon integration process, and provide the information that might be exploited to make the i*Tto1* plant molecular mutagen more efficient.

3.9.2 "Molecular reconstruction" of the integrase

Most transposons are inactive due to accumulation of mutations (Ivics et al., 1997); consequently retrotransposons should be affected by mutations to a even greater extent, as their replicate through a reverse-transcription step. It also known that reverse transcriptase is the most conserved enzyme among retroviral enzymes, therefore we expected the integrase to be a possible source of "interesting" mutations. On the basis of these observations I also conducted a genetic analysis of the integrase protein sequence, which was combined with a synthetic biology approach of "molecular reconstruction" of the integrase domain.

In an outstanding work, the authors where able to revive a "Sleeping Beauty" transposon, which contained a consensus sequence obtained by alignment of related DNA transposons of the *Tc1/mariner*-like superfamily from fish. The consensus eliminated the inactivating mutations, thus "awakening" a primordial form of the transposon that resulted in a perfectly functional and hyperactive element. (Ivics et al., 1997; Mates et al., 2009).

Tobacco contains circa thirty copies of *Tto1* might undergo different modifications by the host. The different genomic location of each copy might, for example, affect the methylation level. Considering that different modifications might result in different activity, I was particularly interested in finding a possibly more active variant of INT that should be used to improve i*Tto1* insertion efficiency. Following a similar approach to that of the above mentioned "Sleeping Beauty", I made an alignment of twenty

clones of the integrase core domain of different cultivars of tobacco and of two progenitors.

The alignment revealed that 14% of the residues are not conserved, and seem to change randomly between clones of different cultivars and of the same cultivar. This fact is consistent with the occurrence of replication errors produced by the reverse transcriptase, and with the fact that methylation makes retrotransposons hot spots for random mutations (Hirochika and Hirochika, 1993). In addition, mutations seemed to be also randomly distributed along the whole *Tto1* sequence (data not shown). In contrast to the random distribution of 50 of the 55 not conserved residues, five amino acids, located in the region proximal to the conserved $DX_nD_{35}E$ integrase catalytic domain, clearly identify two different groups according to INT sequence. 85% and 15% of the clones respectively, contain the same residue in the five characteristic positions, thus identifying two variants of INT (Fig. 2.18). Strikingly, by analyzing the sequence of the two other least conserved *Tto1* regions (data not shown), that is the promoter and the GAG domain, we could not observe any such neat distribution of not conserved amino acids.

More interestingly, the major group "Int2" (85%), does not include *Tto1-1*'s INT variant that is carried by all previous constructs and especially by the i*Tto1* used for gene tagging in Arabidopsis in the first part of this work. Therefore, this data brought to raise the question whether *Tto1-1* carries a more or less active INT, which I wanted to answer.

The molecular reconstruction approach that lead to building *Tto1X,* was tried to test possible differences between Int1 and int2. The clone AA43 represented the consensus sequence, which contained the five differential amino acids, E_{629}, S_{636}, Q_{690}, D_{747} and P_{754} (identifying int2), and the most conserved residues in all the remaining 50 variable positions (Appendix 2.S-A). A little divergence from the consensus was nevertheless represented by the fact AA43 also contains two mutations L_{618} to P and A_{652} to D that are absent from all other clones and that we decided to consider in a second round. The attempt of integrase reconstruction did not unfortunately bring to the expected result: no activity of *Tto1*X was detected *in vivo* by Intron-PCR. However, the RT-PCR indicated the mRNA had been transcribed by the host encoded RNApol II, suggesting therefore that the formation of the VLP might be impaired. This hypothesis will be confirmed in future experiment, by trying an immunological detection of the VLPs by a αGAG antibody on a protein extract of lines expressing *Tto1X*.

The five mutated amino acids might affect both the activity of INT and the protein-protein interaction in the context of the VLP; nevertheless, negative effects on INT cannot be studied yet, due to lack protein activity, and the low specificity of the αINT antibody *in vivo*.

However, it is interesting that the five mutations are in most cases found in most cases associated together, as to suggest an either positive or negative effect of this combination. To this matter it could be useful to analyze the clones AA53 has a Q at position 636 in place of the S, and an N at position 747 in place of the P. Of particular interest can be the two mutations specific of AA43, namely the P_{618} and D_{652}. If AA43 represents an inactivated retrotransposon, then they could be two additional mutations that contributed to a stronger inactivation; we can for example consider the two proline P_{618} and P_{754}, and that proline is known to have a negative effect as "α-helix breaker", therefore such a combination might be particularly detrimental for the element that carries them. On the other hand one could also speculate that AA43 represents a recently active integrase, and that the P_{618} and D_{652} mutations identify a sort of "initial inactivation procedure" by the host on that specific copy of *Tto1*, which has a negative effect on folding of the poly-protein or protease cleavage into the single enzymatic activities. These observations however remain hypothesis, since the exact boundaries of the protein are still not known. In the near future a *Tto1Y* synthetic element will be tested, which will contain the consensus sequence (without P_{618} and D_{652}). I also think that another modified retrotransposon should be constructed, which contains P_{618} and D_{652} in a *Tto1N* background.

3.10 i*Tto1* in barley

In the future we want to extend i*Tto1* application to crop plants. Experimental evidences is encouraging, since *Tto1* is active in heterologous hosts. Among these, rice (Ouwerkerk et al., 2001) was of particular interest since it is a member of the monocotyledonous class, therefore it is imaginable that control/expression mechanisms are conserved with the considerably distant dicotyledonous class.

Barley is both a monocot and an important crop for Europe; in addition it is yet poorly studied, and therefore it represented a good choice to test *Tto1* for more extensive applications. Inducible expression was already tried in transiently transformed calli (G. Böhmdorfer PhD thesis), but inspite of the promising expression levels obtained with the Dexamethasone inducible system in rice (Ouwerkerk et al., 2001), the mRNA level was low. One of the reasons behind the poor expression of the *pTA* based vector in

callus could be the inefficient splicing of its mRNA. Consequently we reasoned that the replacement of both with endogenous introns, from the xylose isomerase gene should be effective.

We made the attempt to obtain *Tto1* expression, starting so to say from a step backward, with a constitutive promoter from maize. For this purpose I made a construct containing the strong promoter of the ubiquitin 1 gene from maize, in combination with nopaline synthase terminator sequence of T-DNA of *Agrobacterium*. The promoter was engineered in order remain below the limit of 32bp of non LTR bases at the mRNA 5' end, which were previously demonstrated not to have a negative influence on *Tto1* reverse transcription (see Fig. 2.6d and §2.4.7).

Unfortunately the result was in contrast to our expectations. Strikingly, Intron PCRs revealed that neither the Gag nor the Int intron were spliced, suggesting failure in the cDNA synthesis both for *Tto1N* and *Tto1X* (Fig. 2.24). Nevertheless, a following RT-PCR also proved the absence of the transcripts, which demonstrated a defect in the mRNA production. In principle it should be considered that the transcription might have been affected by the modification of the promoter sequence, although the deleted 75bp did not seem to contain known domains for control of transcription. I propose that the expression should be tried using the Mub1 promoter in its native form, which will make clear whether the enhancer is affected or not. Alternatively an inducible system could be again tried either with a β-estradiol or an ethanol responsive promoter, or with the more advanced Dex inducible *pOp6/LhGR-N* system (see following). The latter could be an interesting option because it has already been shown to work efficiently in maize (Segal et al., 2003), and because, in spite of the very poor efficiency, a certain expression of *Tto1* could however be seen in transgenic plants that contained another Dex inducible system (*pTA*).

The splicing of the intron/enhancer contained in the promoter might also be affected, thus abolishing its enhancer effect to an extent that not even a minimal transcription is detected. As a further hypothesis, the occurrence of other factors such as truncations of the T-DNA expression cassette should not be excluded. A PCR to detect the *HPT* resistance gene, that is cloned upstream of *Tto1* (right downstream of the T-DNA left border), produced the expected band (Fig. 2.24), suggesting that a possible deletion might affect the other end of the expression cassette in proximity of the T-DNA right border. However a direct proof of this hypothesis has not yet been obtained.

It also cannot be ruled out that, due to the high expression level of the Mub1 promoter, homology dependent silencing might have abolished the expression of the

construct. Such observation is supported by the fact that over 90% of the barley genome consists of TE DNA, and that the LTR retrotransposons comprise the largest fraction of repetitive elements in Triticeae (SanMiguel et al., 2002; Shirasu et al., 2000; Wicker et al., 2001). Therefore, it is likely that *Tto1* LTRs have been the target of silencing mechanisms in barley cells.

3.11 Possible advantages of the *pOp/LhGR* binary system

The pOp6 promoter has already been successfully applied to species like Arabidopsis, and tobacco (Craft et al., 2005; Samalova et al., 2005). Previously, the *pOp/LhG4* system was used in the monocot maize (Segal et al., 2003). As a consequence, I am convinced that the inducible *pOp/LhGR* system could be applied to barley as well, to obtain *Tto1* transposition.

In general the Dexamethasone inducible transactivation *pOp/LhGR* system offers several advantages for both *in vitro* and on soil experimentation. First of all, *pOp/LhGR* can be maximally induced with approximately 1μM dexamethasone and induction to 50% of maximum requires about 0.2μM (Samalova et al., 2005). Importantly, the transcription of a gene under the control of *pOp6/LhGR* system can be easily modulated by the concentration of Dexamethasone (Samalova et al., 2005), offering therefore a better possibility to escape from gene silencing mechanisms of the host, by avoiding a critical copy number of transcripts.

Moreover, with this system Dexamethasone can be applied to plant tissues at many stages using a variety of application procedures, and could be considered useful to induce apical transposition if applied at a developmental stage when the meristem has reached larger number of active cells, with consequent increase of transposition rate. Samalova and colleagues also showed that Dexamethasone can be delivered through the roots in older plants on soil, with consequent activation of the promoter throughout the plant especially in the vascular tissue.

My preliminary results showed that *Tto1* expression with this system is generally higher than *pER8* based constructs in Arabidopsis. In the next future, the efforts will be made to develop an i*Tto1* gene tagging methodology with enhanced performance and with a broad application range to crop plants.

3.12 Synthetic biology

I think that something must be said about the intriguing approach made during my thesis work to "reconstruct" *Tto1*. Since I referred to it as "synthetic biology", I believe that a section of my thesis should be dedicated to this particular concept of bioengineering.

The expression "synthetic biology" appeared the first time in Stéphane Leduc's publication of La Biologie Synthétique (Leduc, 1912). In 1974, the Polish geneticist Waclaw Szybalski used the term "synthetic biology", writing:

"Let me now comment on the question: "what next". Up to now we are working on the descriptive phase of molecular biology. [...] But the real challenge will start when we enter the synthetic biology phase of research in our field. We will then devise new control elements and add these new modules to the existing genomes or build up wholly new genomes".(Szybalski, 1974).

Since that time molecular biology has revolutionized the classical biology based on the observation of living organisms by modifying their genetic code and engineering them according to our need. In my personal opinion, synthetic biology is a different concept of bioengineering or biotechnology, although the two arts overlap to a great extent. The idea of cells like "biofactories" holds *in se* the concept of reprogramming life. We can consider all those micro-organisms that are now commonly used in production of drugs, such as insulin, in the bioremediation, to eliminate hazardous environmental contaminants or cells employed in the production of monoclonal antibodies for research as forms of synthetic life. They in fact either did not exist in nature or have their physiological peculiarities "re-programmed" or "re-designed" (Benner, 1987; Benner, 2003; Szostak, 2001) to obtain performances that match our specific needs.

Synthetic biology does aim at creating new biological systems, but with a different approach and different scope. Chopra and Kamma in 2006 summarized them as follows:

1. Engineering of biological systems, which can be assembled together to create biological circuits that behave in a predicted way.

They also evoked the concept of biological interchangeable components, already proposed by Gibbs et al. in 2004.

2. Redesigning life, by constructing biological systems that should be used to increase our understanding of biology by comparing the differences between observed and predicted behavior, that is, between "natural" biology and synthetic biology.

3. Creating alternative life, by involving molecules that are unnatural in living systems. Although Living organisms are composed by molecules such as DNA, RNA, proteins etc,, there are certain unnatural molecules which can perform the same function as the above mentioned ones (Benner and Sismour, 2005).

In think that, in my work, the engineering and the attempts of molecular reconstruction of *Tto1* cover the first two approaches.

The retrotransposon, a biological system, has been provided with interchangeable parts, the heterologous promoters, which gave and will give the chance to understand important steps of its reverse transcription, and previously did about its mRNA transcription and translation. The insertion preference of i*Tto1* for genes has been exploited to turn it into a possible broad range plant mutagen, therefore we can say that its natural mutagenic potential has been re-designed into an advantageous feature for the plant community. In addition, our efforts point at a more extensive "re-design" in a way to improve its natural gene preference and its performance as "genome intruder".

This is, to some extent, what I have tried to do by creating the construct *Tto1X*, and, in the next future, *Tto1Y* : re-designing with interchangeable parts. In this specific case the interchangeable part was the "consensus" integrase core domain (Int2) isolated from one clone, and cloned into the *Tto1* backbone. This perfectly reflects what was written by Benner and Sismour in 2005: "The parts come from natural living systems (that is, they are biological)"; their assembly is, however unnatural", that is synthetic.

Independent of the result, (my experience suggests that it can be quite unpredictable), I find this approach of using proteins as interchangeable parts particularly intriguing, and the majority of synthetic biologists also does. In fact, the synthetic biology of nucleic acids is successful, due to the relative simplicity of the molecules compared to proteins, and in particular to the repeating charge of its interchangeable parts: nucleotides. With proteins it is way more difficult since they do not have a repeating charge. The proposals of engineering proteins, by replacement of their interchangeable units, amino acids, is as old as recombinant DNA technology, but amino acids interact in a too complex way to make prediction easy for the researchers. However, even though good results have been obtained by amino acids replacement, for instance DNA polymerases used for sequencing (Tabor and Richardson, 1995), enzymes in commercial detergents (Igarashi, 2003) and so on, they do not match the synthetic biology concept to capture the emergent properties

of living systems (Benner and Sismour, 2005). It was also said that today the technology of amino acid replacement is done using a combination of calculation, design, screening, selection and luck (Arnold, 2001); therefore, on the basis of what I have done already, I believe that *Tto1* can be a good field on which to combine the above mentioned factors, provided that the luck variable will not be too predominant on the others.

3.13 Conclusions

The main conclusion of my thesis is that the retrotransposon *Tto1*, has been turned from a potential harm to the plant into a controllable element that inserts into genes "on demand".

The direct corollary of the first conclusion is that, due the ease of handling provided by genetic manipulations, it can be used as a diagnostic tool for functional analysis of its reverse transcription.

However, a more elaborated picture of what has been done and discussed here should be of help for the reader and for comprehension from a wider angle.

This work is based on the possibility to perform sequence manipulation of *Tto1* and to still obtain an active element that maintains its natural ability to transpose in plant. It has been the continuation of previous research in which it was demonstrated the role of the 5' LTR in control of transcription and translation initiation. In this work I elucidated the function of 3' LTR in transcription termination and in reverse transcription. A consequent general remark is that *Tto1* transposition undergoes control at transcriptional, posttranscriptional and posttranslational level, and I think that I could confirm that with my results.

The first level of TEs repression is DNA methylation and this has not been investigated in my work, but experimental evidence will be obtained by analyzing *Tto1* in the *ddm1* mutant background. I instead demonstrated that transcriptional control acts by combining transcript initiation and termination points. Notably, I found that two major terminators are in the LTR and the most frequent terminator determines no transposition of the element. Although a complete transposition is followed by successful integration into the genome, and the direct proof of that is still missing since I used constructs with an inactivated integrase, I have clearly demonstrated that shorter *Tto1* mRNAs do not support reverse transcription, and therefore there is no cDNA that would be inserted into the host DNA.

As previously demonstrated and published (Böhmdorfer et al., 2008) *Tto1*'s translation appears to be controlled by the presence of different ORFs and different ATGs, only one of which leads to the formation of active VLPs. These data suggest a general mechanism in which the choice of the right start and the right end, among different wrong combinations, determines whether the element would replicate or not. The choice in this case is made by the host cell.

I would personally describe, in a slightly facetious way, the Long Terminal Repeats of retrotransposons as a switch board with many buttons, and the host cell as the operator that pushes the right ones in response to environmental stimuli. While the genetic information is contained between the two LTRs, the two terminal repeats contain the information necessary for their activation. In addition to date I have to say that the accumulation of mutations by retrotransposons may be another way to control them; apart from mutations derived by reverse transcription errors, methylation usually increases the occurrence of mutations that can contribute to inactivate the elements. Furthermore other posttranslational modifications have an influence on the enzyme activity. With my analysis of integrase sequence and my synthetic biology approach, I have tried identifying this sort of mutations, which would probably confirm this further level of control.

All these points of control give the possibility to maintain a tight equilibrium between the host and its "nuclear parasites". My facetious interpretation of this long relationship might also stand on the basis of other definitions such as "The take and give between retrotransposable elements and their hosts" (Beauregard et al., 2008), to mean that both receive an advantage from their cohabitation. In fact, as on one side the elements appear to be activated mainly during adverse conditions to the host, acting like perfect opportunists that sneak in to a heterologous genome, on the other side they have been maintained because of their ability to bring advantages to the cell by re-shaping genomes, activating or inactivating genes, contributing to telomere integrity and conferring selective traits to the host.

From an evolutionary angle this can be the manner by which both the retrotransposon and the host play their role but "respect" each other, in a way that this equilibrium allows the mutual existence of both the cell and the parasite, by exchanging each other a favor. This could be a reason why the plants have kept potentially harmful genome invaders: they can exploit TEs ability to reshape the genome, activate genes or also offer hot spots for sequence mutations. At the same time, the host has evolved different mechanisms to maintain TEs invasiveness at a

low level, so that they would not interfere with the physiology of the cell except for specific cases.

Chapter 4

MATERIALS & METHODS

4.1 MATERIALS

The reagents used in this PhD thesis were bought from Duchefa, Fluka, Gibco, Invitrogen Merck, Pierce, Roche, Roth, Serva, Sigma. The enzymes were purchased from Fermentas, Invitrogen, New England Biolabs (NEB), Promega, Roche, Takara.
Primary antisera were produced by EUROGENTEC, Belgium.
Secondary antibody Goat Anti-Rabbit was provided by Sigma.

NCBI accession number for *Tto1* sequence used in this work: D83003.

4.1.1 Bacteria

4.1.1.1 *Escherichia coli* strains

- DH5α: F⁻ *gyrA96* (Na1^r) *recA1 endA1 thi-1 hsdR17* ($r_k^-m_k^+$) glnV44 *deoR D(lacZYA argF) U169* [f80dD(*lacZ*)M15]
- ElectroMAX Stbl4: electro-competent cells (Invitrogren, cat. no. 11635-018). Used to maximize the stabilization of direct repeats (LTR) and retroviral sequences.
- XL1-blue: *supE44 hsdR17 recA1 endA1 yrA46 thi relA1 lac F´* [*proAB+ laclq lacZDM15 Tn10* (tet^r)]
- OneShot TOP10: chemically competent cells (Invitrogen, cat. no. 44-0012)
- Rosetta(DE3) pLysS: *F ompT hsdS$_B$(R$_B^-$ m$_B^-$) gal dcm λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5*]) pLysSRARE (Cam^r)

4.1.1.2 *Agrobacterium tumefaciens* strains

- C58C1: pCV2260 (Rif^r); used for transformation of Arabidopsis

- AGL10: pVec10 (Rif^r); used for transformation of barley

4.1.2 Plants

4.1.2.1 Arabidopsis ecotypes

Columbia (Col-0)

4.1.2.2 Tobacco ecotypes

- *Nicotiana tabacum* BY2, SR1, Xanthi, Samsun NN, W38.
- *Nicotiana Sylvestris*
- *Nicotiana tomentosiformis*

4.1.2.3 Barley ecotypes

- *Hordeum vulgare* cv Golden Promise.

4.2 Media

4.2.1 Media for *Arabidopsis thaliana*

Seedlings were incubated in growth chambers under long day conditions, (LD=16h light). Plants were instead grown in greenhouse, usually under LD.

4.2.1.1 1% Ara medium.

This medium is prepared using 1/3 of the concentration of the common MS (Murashige-Skoog). The composition for 1l is: 4.3g MS salt, 10g sucrose, 0.5g MES), pH 5.7 with KOH, dH_2O up to 1l. For plates 8g plant agar Merck, were added before autoclaving.

After the autoclave step 2ml 500x vitamin-mix per 1l medium were added, (final concentration 2x). Composition of 500x vitamin-mix for 100ml: 5g myo-inositol, 1g thiamine, 50mg nicotinic acid, 50mg pyridoxine, 10mg biotin, dH_2O up to 100ml. The medium was then filter-sterilized and stored at -20°C.

Modifications to Ara1%:

Hygromycin and kanamycin were added to the media to allow the selection of

transgenic plants. Claforan (Cefotaxime) was added to inhibit the growth of *Agrobacteria* during sT1-selection. β-estradiol was used to induce the expression of genes under the control of responsive promoters. The concentrations used were the following:

- Hygromycin: 15–20mg/l
- Kanamycin 50mg/l
- Claforan: 100-200mg/l
- β-estradiol: 5-10μM (from a 5mM stock in DMSO).

4.2.1.2 Gamborg B5 medium

This medium was only used in some cases for liquid culture, and with ½ of the sucrose concentration. The composition for 1l is: 1,54g Gamborg B5 salts, 10g sucrose, 1x Gamborg vitamins (to be added just before use). The pH was adjusted to 5.7 with KOH, and the medium filter-sterilized and autoclaved for 10mins.

4.2.2 Media for *Hordeum vulgare*

Barley plants were grown and transformed by Dr. Götz Hensel. at Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK). Media and methods used for barley are published by Hensel et al, *International Journal of Plant Genomics,* vol. 2009, article ID 835608.

4.2.3 Media for *Escherichia coli*

E. coli cells were usually grown at 37 °C, and sometimes at room temperature (25°C).

4.2.3.1 LB (Luria-Bertani)

LB is a standard medium for the cultivation of *E. coli*.

Composition for 1l: 10 g bacto-tryptone, 5 g yeast extract, 10 g NaCl, pH 7.0 with NaOH, dH_2O up to 1l.

4.2.3.2 TSS medium

TSS is a modified LB used for the production of chemically competent XL1blue or DH5α *E. coli* cells to be transformed with plasmids.

Composition for 95ml: 10% PEG 4000, 50mM $MgCl_2$, in 90ml LB, adjust the pH to 6.7, LB was added up to 95ml. Filter-sterilized 9.5ml aliquots were then prepared and stored at -20°C and after thawing 0.5ml DMSO were added to the medium, that was kept on ice until further use.
For plates, 1.5% (15g per 100ml) Bacto-Agar were added to the medium before autoclaving. Antibiotics to select bacteria containing the plasmid of interest were added to the media after autoclaving and cooling down to a temperature lower than 50 °C. The concentrations used were the following:
- Kanamycin: 25-50 mg/l
- Ampicillin: 100 - 200mg/l
- Chloramphenicol: 25 mg/l
- Spectinomycin: 50 mg/l.

4.2.4 Media for *A. tumefaciens*

Agrobacteria were grown at either 28° or 30°C, with no significant difference.

4.2.4.1 YEB medium: for the strain C58C1

Composition for 1l: 5g beef extract, 1g yeast extract, 5g Peptone, 5g sucrose, pH7.2 - 7.3 with NaOH; 2mM $MgSO_4$ (sterile filtrated) were added after autoclaving. Antibiotics were added sterile-filtrated to the medium after autoclaving. The concentrations used were the following:
- Kanamycin: 50mg/l
- Rifampicin: 25-50mg/l
- Spectinomycin: 50mg/l.

4.2.4.2 AGL10 medium: for the strain AGL10

Composition for 1l: 5g mannitol, 1g L-glutamic acid or 1.15g Na-Glutamate, 0.25g KH_2PO_4, 0.1g NaCl, 0.1g $MgSO_4+7H_2O$, 5g Tryptone, 2.5g yeast extract, 10µl Biotin (from a 0,1g/l stock), dH_2O up to 1l.

4.3 METHODS

4.3.1 DNA isolation methods

4.3.1.1 Plasmid DNA small scale preparations (Mini-preps)

Plasmid DNA was usually prepared from saturated bacterial cultures, 16h for *E. coli*. Rarely plasmid DNA was also isolated from *A. tumefaciens* cultures - as the yield is usually low: agros were generally screened by PCR.

For high copy number plasmids 1.5ml of the ON culture were centrifuged down at 16000g speed, while at least 4ml were necessary for low copy number plasmids. After the centrifugation step the supernatant was completely removed and the pellet treated with the following kits according to the manufacturer's instruction.

- Macherey Nagel Nucleospin Plasmid Kit (#740588.50).
- Promega SV Minipreps kit (#A1460).

4.3.1.2 Plasmid DNA Large Scale preparations (Midi/Maxi-preps)

Large plant vectors for cloning are often amplified in low numbers in *E. coli* (low-copy plasmids). To get the necessary amount of such plasmid DNA, larger bacterial culture needed to be used, than those used for mini-preps.

The bacterial pellet was in fact obtained from 15-50ml of bacterial saturated culture (after 16h growth) by Nucleobond-AX (P100) kit (Macherey-Nagel cat. no. 740511) according to manufacturer´s protocol, with small modifications: the lysate was filtrated through Miracloth (Calbiochem) before loading on the DNA column. After precipitation, (as in §4.3.1.5), the samples were dissolved in dH_2O and the concentration determined by a photometer, as explained in §4.3.1.6.

4.3.1.3 Quick and Dirty (QND) small scale plant genomic DNA isolation

For genotyping, DNA was crudely isolated from *Arabidopsis* plants with the following method. 30–50mg of plant material (leaves, flower buds, or entire plantlets approximately 2 weeks old) were frozen in liquid nitrogen. After adding one micro spoon of quartz sand and 300-400µl (depending on sample size) isolation buffer (200mM Tris-Cl pH7.5, 250mM NaCl, 0.5% SDS, 25mM EDTA), the samples were homogenized using a glass rod and a IKA-Mixer. After centrifugation (14000rpm at room temperature for 5mins), the supernatant was mixed

with 1vol isopropanol at room temperature for 5mins, followed by another centrifugation step (14000rpm at room temperature for 5mins). The pellet was washed with ice-cold 70% ethanol and dried. Next the DNA was dissolved in 30–50µl 1x TE (10mM Tris-Cl pH7.5, 1mM EDTA) by incubating the sample at 65°C for 5mins (and gently stirring the pellet in between). 2–5µl of this DNA were used in ensuing PCR reactions.

4.3.1.4 Large scale plant genomic DNA isolation.

For more advanced downstream applications, larger preps of plant genomic DNA were made. The plants were ground to a fine powder using a SamplePrep Freezer/Mill 6870 (SPEX) and stored at -80°C. The genomic DNA was then isolated from 1g of powder using the kit: Illustra™ DNA Extraction Phytopure, (GE Healthcare, Cat. no. RPN-8511), according to the manufacturers' protocol.The same kit was also used to obtain high quality genomic DNA mini-preps, and also used for genotyping.

4.3.1.5 Precipitation of DNA

To concentrate DNA or to change buffer conditions (e.g. after restriction), the DNA was precipitated. To each DNA sample 1/10 vol. of 3M KOAc (pH 4.9–5.3) and 2.5 volumes of 96% EtOH were added, the tube gently inverted to obtain a uniform mix and incubated at -20°C for at least 30mins (or ON in the case of plant genomic DNA). Next, the DNA was collected by 10mins centrifugation at 14000 rpm at 4°C) and washed with 70% cold EtOH. The dried pellet was dissolved in sterile dH_2O.

4.3.1.6 Determination of DNA concentration

The concentration of DNA samples was estimated by measuring the absorbance of the sample with a photometer. Two different photometers were used.
1- At MPIZ (Cologne) the measurements were done in the classic way: 5µl of DNA were added in 995µl dH_2O, in a quartz cuvette, and the 260nm absorbance was read. The concentration was estimated following the rule that an A260=1 corresponds to a concentration of 50µg/µl.
2- At MFPL Vienna) a NanoDrop 2000 photometer was used. With this machine it was enough to add 1µl of the sample in the appropriate place, and after selecting the function "DNA" the machine would retrieve the concentration value. The two methods were however comparable.

4.3.2 PCR (Polymerase Chain Reaction)

The PCR allows the distinct amplification of a sequence by choosing specific oligomers that act as reaction primers. In this thesis the PCR was used either as an analytical or as a preparative method. Analytical PCR was usually performed in a smaller volume and with a lower number of cycle, compared to the preparative one, which was performed to produce fragments to be processed in following experiments.

The sequences of all the oligos used in this work are reported in Appendix A.

4.3.2.1 Analytical PCR

The reaction was performed with a home-made *Taq*-polymerase provided by Dr. Iain Searle, at MPIZ in Cologne. At MFPL in Vienna the Promega GoTaq-polymerase (cat. no. M3175) was used for routine analytical PCR. This method was used to check for the presence of a specific DNA sequence using starting from different templates: plasmids, genomic DNA, bacterial colonies in the procedure called Colony-PCR (§4.3.2.2). Here is a general PCR reaction mix, in a total volume of 25µl:

2–5µl DNA + 2µM primer 1 and primer 2 (usually 0.5µl of a 100µM stock solution) + 0.25-0.5mM dNTPs (usually 2-5µl of a 2.5mM stock solution) + 1x Brown-Taq buffer (2.5µl of a 10x stock provided by Dr. Iain Searle) or 1x GoTaq Buffer (5µl of the 5x stock solution) + 0.5µl Brown-Taq (5U/µl) or 0.125µl GoTaq (5U/µl) + dH_2O up to 25µl.

A standard PCR program was used:

4min. 94°C, 45x (30sec. 94°C, 30sec. 50°C, 30-40sec 72°C), 10min 72°C, 4°C forever.

4.3.2.2 Colony-PCR

Colony-PCR is an analytical PCR used to quickly identify positive clones after transformation. In this procedure the template DNA is replaced by a bacterial colony, directly taken from a Petri dish by a toothpick or a pipette tip and dissolved in the PCR reaction mix. The same toothpick is then used to pre-inoculate the clone in 200µl of medium, at 37°C. This method allows the analysis of many clones at once, and gives the possibility to start an ON culture of the positive ones from the correspondent pre-inoculum. The PCR was carried out as in

§4.3.2.1, but with a longer denaturation step (up to 10mins) and only 25 amplification cycles, to prevent the formation of too many unspecific products.

4.3.2.3 Intron-PCRs

This PCRs were performed according to the reaction described in §4.3.2.1.
2-5µl of genomic DNA isolated as in §4.3.1.2.1 4.3.1.2.2 were used as a template. The two Intron-PCRs are distinguished, depending on the intron that had to be detected.
Intron1-PCR was performed on the intron inserted into the GAG domain (intron1), employing the following oligos: T969-911 (AGC TCG AAG AGT TGT ATG CCT CT) and T1109-1086 (CCG ACA ACT GGT CGA CAA TCC CTT).
Intron2-PCR was performed on the intron interrupting the INT domain (intron2), employing the oligos T2314-2334 (GGT GGA AAG AGA GAC TGG TAA) and T2654-2634 (CCC GTA ATT GAT CAT AAG AGA).
The program used was the same as in §4.3.2.1. After the run 15µl of the reaction mix were loaded and separated on a 2% agarose gel.

4.3.2.4 "Long-PCR"

This PCR method was used (§2.4.4) to amplify a 3kb cDNA from the 5.3kb full length cDNA that should arise after the reverse-transcription of *Tto1*.
To obtain such a product the oligos T2314-2334 (GGT GGA AAG AGA GAC TGG TAA) and T574-548 (TGT TAG GAT CCG GTG GCA CTA AAC ACT), were used in a 50µl reaction mix that is as follows: 5µl template DNA + 1µM primer 1 and primer 2 (usually 0.5µl of a 100µM stock solution) + 0.4mM dNTPs (usually 8µl of a 2.5mM stock solution) + 5µl of 10X LA buffer + 2.5U of LA Taq polymerase (0.5µl of (5U/µl stock solution) + dH_2O up to 50µl. The amplification program used, named Koncz-dip, contained an extended amplification step: 7min 95°C, 35X (30sec 95°C, 30sec 50°C, 8min 68°C), 10min 68°C, forever 4°C.

4.3.2.5 Preparative PCR

The preparative PCR was carried out in a volume from 50 to 100µl. LA Taq polymerase (Takara cat. no RR002M) or Platinum Pfx DNA polymerase (Invitrogen cat. no. 11708-013) that have a proof reading activity, were used. The PCR program was the standard program (§4.3.2.1) with changes of the length of the elongation step (from 30sec to 2mins).

4.3.2.6 PCR to amplify Southern Blot's probe

The probe directed to *Tto1* sequence was produced by PCR using the Promega GoTaq-polymerase (cat. no. M3175) and purified from the gel using the Nucleospin-Extract Kit II (Macherey-Nagel 740609.050). The PCR program was the standard one as from §4.3.2.1.

The oligos were: (T4390-4411 (CAT CGC AGC AAC GGA GGC TTG C) and T4697-4677 (CCA CTA TCC CTG CAG CTT CTC). The template was *Tto1* subcloned into the plasmid *pACYC177* backbone. The sequence of the probe is reported in Appendix B.

4.3.3 Agarose Gel Electrophoresis

Agarose gel electrophoresis is the most common method to distinguish nucleic acids fragments by size. The samples are usually loaded on a gel submersed in a saline "running buffer" that allows the separation of the fragments when electrical energy is applied. The running buffer used in all my experiments was 1x TAE, which is generally prepared as follows: 50x TAE stock: 242g Tris base, 57.1ml glacial acetic acid, 37.2g Na_2EDTA x $2H_2O$, and diluted before use. Low concentration gels (i.e. 0.8%) were used for a better resolution of large DNA fragments (2kb and higher). High concentrated gels (e.g. 2-2.5%) were used to separate fragments from 100 to 250bp (including RNA fragments). The needed amount of agarose was molten in 1x TAE, by boiling the mixture in a microwave. After cooling down to a temperature of ca 50°C, ethidium bromide to a final concentration of 5mg/l was added, which is used for the visualization of the nucleic acids under UV-light. The gel was poured in a tray of a different size according to the need, applying a comb to produce loading slots of the wanted size. The samples were mixed with 1/6 vol. of 6x loading buffer (50% glycerol, 0.2M EDTA pH 8.0, 0.005% bromophenol Blue or Orange G) before loading.

DNA size markers were: GeneRuler 1kb ladder, Fermentas; 1kb ladder, NEB. GeneRuler 100bp DNA ladder, Fermentas; 100bp ladder, NEB. 6µl of 1X dilution of the markers were routinely used in the gels.

4.3.4 Purifcation of DNA from agarose gel

With this method DNA fragments with a certain length and origin were isolated and purified for further applications, like cloning and sequencing. The DNA was

separated on an agarose gel and the fragments of choice isolated using the Nucleospin-Extract kit II (Macherey-Nagel cat. no. 740609.050), or the Wizard SV Gel and PCR Clean-Up System (Promega, Cat. No. A9282) according to the manufacturer's protocol. The above mentioned kits were also used to elute DNA fragments from a PCR reaction mix.

4.3.5 Restriction of DNA

4.3.5.1 Analytical restriction

Analytical restriction of plasmid DNA was used to identify positive bacterial clones after transformation with a construct of interest. After isolation of the plasmid (see §4.3.1.1, 4.3.1.2), this restriction was usually performed with 10U of enzyme to 1μg of plasmid DNA for 1 to 3h incubation, or with 0.2U of enzyme to 1μg of DNA for 16h, in a volume of 20-30μl. The temperature was depending on the enzyme.

4.3.5.2 Preparative restriction

For cloning purposes, a larger amount of plasmid DNA was used in comparison to analytical digestions. In such cases, from 3 to 10μg of purified plasmid DNA were digested. The incubation times and the temperature were the same as for analytical digestions, but reaction volume was always at least 50μl.

4.3.6 Ligation of DNA

The ligase catalyzes the formation of phosphodiester bonds between adjacent 3´-OH and 5´-P in double strand, and it is used in the common cloning protocols to join an insert to a vector backbone.

Ligations were usually performed in a volume of 10μl, with 2U (excess) of T4-Ligase (Roche, Cat. No. 481 22) at 14-16°C for 16h. The ratio was usually 3 (ng of insert) : 1 (ng of plasmid) for sticky ends ligations and from 5 to 10:1 for blunt ends ligations.

4.3.6.1 Shrimp Alkaline Phosphatase (SAP) Dephosphorylation

This treatment results in a dephosphorylation of 5´-ends of dsDNA to prevent self-annealing of vector DNA cut with a blunt end enzyme, thus reducing the number of back-ligated (empty) vector backbone. Digested and purified plasmid backbone up to 1µg was incubated with 1U SAP in 1x SAP buffer (Roche cat. no. 1 758 250). After 30 minutes at 37°C, the enzyme was deactivated by incubation at 65°C for 15mins. For subsequent ligations, the plasmid was precipitated as described in §4.3.1.5.

4.3.7 Southern Blots

In a Southern blot, gene DNA sequences can be detected by radioactively labeled probes. In this work, two Southern approaches were carried on. One was the classical method that detects genomic DNA sequences, in order to identify new transposition events of *Tto1*; the second one aimed at the identification of extra chromosomal copies of *Tto1* in order to detect newly reverse-transcribed cDNA copie of the element, before the integration. The two methods differ in practical steps, which will be described as following.

4.3.7.1 Genomic DNA Digestion

From 5-10µg DNA (isolated as in §4.3.1.4) were digested with 160U of *HindIII* and *EcoRI*, (both Roche) in buffer B at 37°C overnight. The next day, 1/10 of the sample was loaded on an agarose gel to check if the DNA was digested properly and if the concentrations were equal. Finally, the DNA was loaded on a 0.9% agarose gel and separated overnight at 45V and 120mA.

In the Southern blot approach aiming at the detection of extra chromosomal *Tto1* cDNA at least 20µg of genomic DNA was used. A DNA digestion was performed overnight (16h) using 0.2U *NotI* (Roche) per µg DNA, and the day after separated for 5h on 0.9 % agarose gel, at 130V and 330mA. The *NotI* digestion was considered useful to help the transfer of the DNA onto the nylon membrane; in order to eliminate the high viscosity of the high molecular weight genomic DNA, maintaining *Tto1* intact, as *NotI* does not have any cleavage site in *Tto1* internal sequence.

4.3.7.2 Blotting of DNA Gels

After separation of the samples on a 0.9% agarose gel, all parts of the gel without DNA were removed and a picture was taken with a ruler placed next to the gel. After rinsing in distilled water, the gel was shaken softly in 0.25N HCl, then in denaturing solution (1M NaCl, 0.5M NaOH) and, finally, in neutralizing solution (1.5M NaCl, 0.5M Tris-Cl pH 7.0) each time for 30min. Between each incubation step, the gel was rinsed with distilled water. The dimension of the gel were measured and a piece of nylon membrane Hybond™-N, (RPN 1520 N, Amersham) was cut the size of the gel and dipped in 20x SSC (see following text).

Finally, the DNA was blotted on the membrane in 20x SSC (3M NaCl, 0.3M Na-Citrate x $2H_2O$; pH7.0 with 1M NaOH) using capillary force overnight. A Tray was half filled with 20x SSC; then 3 filter papers were soaked with SSC and arranged on a plastic support to make a "bridge", with two edges always immersed in the buffer. The gel was put on the triple layer of filter paper and the membrane was laid on top of the gel, trying to remove all air bubbles that might form in the interface, with a glass pipette On top of the membrane another 5 filter papers, the size of the gel, were laid down taking care to avoid the contact with the surrounding surface. A stack of absorbent paper was then put on 5 filter papers, to maintain the capillary force. A glass plate then was put on top to keep the structure together and surmounted by maximum 300g weight. The weight to apply depends however on the surface of the membrane. The next day, the DNA was cross-linked to the membrane by UV Stratalinker, using the "Auto-Crosslink" program.

4.3.7.3 DNA blotting check with methylene blue stain

This method was used to check the efficiency of Southern blots before the hybridization step.

After the UV cross-linking step, the membrane was soaked in a solution containing 0.04% methylene blue in 5% acetic acid for 10-15mins at room temperature, under gently shaking. Afterwards the membrane was rinsed with distilled water 3 times for 10sec, or until the bands appeared blue against the white-bluish background.

The methylene blue stain could be completely removed from the membrane by washing with 0.1-1% SDS; nevertheless in my case the stain was removed with the pre-hybridization buffer containing SDS. The liquid stain could be also re-used in case of necessity.

4.3.7.4 Radioactive Labelling of the Probe

Labeling was performed using the random-primer method:

+ 200–400ng DNA

+ x µl H_2O (to volume)

➔ denatured for 5mins at 95°C, then kept on ice

+ 6µl 10x random hexa-nucleotide mix (Roche), on ice

+ 6µl 5mM dNTPs-dCTP, on ice

+15µl α32-P dCTP (150µCi/µl)

+ 2µl Klenow-polymerase (2U/µl), Roche

+ ddH_2O to a volume of 60µl

➔ incubated at 37°C for 1h at least.

The labeled probe was purified using either the Nucleo-Spin extract kit II (Macherey-Nagel cat. no. 740609.050) or the Wizard SV Gel or PCR Clean-Up System (Promega, cat. no. A9282) as in §4.3.4. The volume of the probe was increased to 300µl with 1x TE (10mM Tris-Cl pH7.5, 1mM EDTA) or dH_2O. After denaturing (95°C for 5 minutes) the probe was added to the membrane (see §4.3.7.5).

4.3.7.5 Southern Blot Probe

Two different probes were tested in this work:

1- Probe directed against the *Tto1* ORF. The probe was PCR amplified from the construct *pACYC::Tto1* as in §4.3.2.6 and the standard PCR program as in § 4.7.1.The 307bp fragment was isolated from agarose gel as in §4.3.4. The sequence of the probe is reported in Appendix B.

2- A second long probe was also used, that corresponded to the full length 5.3Kb *Tto1 XhoI/PvuII* excised from the construct *pACYC::Tto1* and purified from gel as above (§4.3.4).

4.3.7.6 Hybridization

The nylon membrane was first pre-hybridized in ca 20ml Church-and-Gilbert buffer (1% BSA, 1mM EDTA, 0.5M Na-phosphate buffer pH 7.2, 7 % SDS) at 65°C for 1h. After exchanging the buffer with a fresh 12ml aliquot, the denatured probe (5mins at 95 °C) was added, and the blot hybridized at 65°C overnight under constant shaking.

The next day, the blot was rinsed with ca 25ml and then incubated with 12ml of the first washing solution (2x SSC, 0.1 % SDS) at 65°C. After 20mins, it was replaced by the same volume of the second washing solution (1x SSC, 0.1% SDS), for washing at 65°C for 15mins. Next, the membrane was incubated with the third washing solution (0.5x SSC, 0.1% SDS) at the same temperature for 5mins. Finally, the blot was sealed in a plastic bag and exposed to a phosphor image screen for 3days at least.

4.3.7.7 Stripping of Blots

A radioactive probe can be removed from the membrane once hybridized. In this way a membrane can be hybridized with different probes and thus used in different experiments.

Approximately 250ml of boiling 0.1% SDS were poured over the membrane and left cool down to room temperature, adding four blue pipette tips to the liquid to ensure that the membrane would be always submerged, thus preventing it to dry out. After ca 5h the membrane was dipped a few times in 2x SSC to wash it and sealed still wet in a plastic bag until next use. The membrane was sealed wet to prevent it to dry: this is crucial point for further use.

4.3.8 RNA Isolation methods

In this work RNA was isolate from Arabidopsis and barley (*Ordeum vulgare* L.) in order to perform expression studies, analyze splicing of the sequence of interest and clone the mRNA 3' ends of *Tto1* deletion constructs.

4.3.8.1 DEPC-treatment of solutions for RNA work

RNases are very stable enzymes that are not destroyed by autoclaving. For RNA work, all solutions and equipment were made RNase-free by DEPC (Diethylpyrocarbonate) treatment.

450µl DEPC were added to 1l of solution (ddH_2O or buffer); the bottle was shaken and left in the fume hood for at least 1h (leaving the lid slightly opened). With a subsequent autoclaving the DEPC was destroyed. All solutions were treated with DEPC except from Tris that was sterile-filtrated.

4.3.8.2 Total RNA isolation

Up to 100mg of plant material were ground to a fine powder in liquid nitrogen

and used for RNA isolation using the RNeasy Plant Mini Kit (Qiagen, cat no. 74104) according to the manufacturers' protocol. The RNA was usually eluted in 50µl RNase-free water, and its concentration was estimated as described in §4.3.1.6 .The average yield resulted to be 0.8µg/µl. The RNA was visualized on 2% agarose gels. It was crucial to work quickly and change gloves at every step, to avoid RNase contamination.

4.3.8.3 mRNA Isolation

The total RNA extracted as in §4.3.8.2 was loaded onto Dynabeads® Oligo $(dT)_{25}$ (Invitrogen), and mRNA was isolated according to manufacturers' protocol. The concentration was estimated as in §4.3.8.5 and 2µg of each preparation were used in subsequent in vitro reverse transcription.

4.3.8.4 DNase Digest of RNA Samples

In my working case this step was not necessary because the RNeasy Plant Mini Kit was used and because the oligos used in for following RT-PCR flank an intron, therefore the cDNA and the genomic DNA have a different size.
The possible DNA contaminations were however eliminated using the RNase-Free DNase Set (Qiagen cat. no. 79254) performing the "Optional On-Column DNase Digestion" according to manufacturer's protocol.

4.3.8.5 Determination of RNA concentration.

The concentration of the RNA samples was determined using a NanoDrop 2000 photometer, as for the DNA (§4.3.1.6).

4.3.8.6 Visualization of total RNA on agarose gel

The total RNA preps (see §4.3.8.2) were visualized on 2% agarose gels as explained in §4.3.3, to check for the yield and the integrity of the RNA.

4.3.9 Sequencing of *Tto1* mRNA 3´ends

4.3.9.1 Amplfication of the mRNA 3' ends constructs Δ5022 and Δ4922

The seedlings of T3 Δ5022 and Δ4922 lines were grown on liquid 1% ara/hygromicin (15mg/l) medium, with the addition of β-estradiol, for 15 days in a growth chamber. After harvesting and grinding in liquid nitrogen, the total

RNA was isolated from 80mg of plant powder using RNeasy Plant Mini Kit, Qiagen, cat no. 74104) as in §4.3.8.2.

The total RNA preps, being 50µl, were increased to 100µl with RNase-free dH_2O and loaded on Dynabeads® Oligo $(dT)_{25}$ (Invitrogen) (§4.3.8.3) for the isolation of the mRNA, following manufacturer's protocol.

15µl of RNA bound to the beads and the rest was stored at 4°C. The cDNA-synthesis was performed using the AMV Reverse Transcriptase (Roche, Cat. no. 11 495 062 001). The Oligo $(dT)_{25}$ part of the beads was used as an upward reading primer. The Protector RNase Inhibitor (Roche, Cat. No. 03 335 399 001) was added to the reverse transcription reaction mix.

The reverse transcription reaction mix was:

15µl $(dT)_{25}$-bound mRNA

0.6µl RNase Inhibitor (25U) +

0.5µl 10mM dNTPs (diluted in DEPC-dH_2O) +

1µl AMV Reverse Transcriptase (25U/µl) +

DEPC-dH_2O (to a total volume of 20µl)

➔ incubation at 42°C for 3h

The cDNA was then amplified in three PCR steps, in which the LA *Taq* polymerase was used. The first PCR reaction mix was:

2µl cDNA, 0.5µl of oligo 231A-T4390-4411 (CAT CGC AGC AAC GGA GGC TTG C) (100pmol/µl), 0.5µl of oligo 261A-dTclamp (CGG ACG CTC AGC CAG G $(T)_{25}$) (100pmol/µl), 5µl 10x LA Taq Buffer, 8µl 2.5 mM dNTPs, 0.5µl LA*Taq* (5 U/µl), ddH_2O up to 50µl.

The first PCR program was:

94°C 4mins, 5x (94°C 30sec, 42°C 30sec, 72°C 90sec), 35x (94°C 30sec, 55°C 30sec, 72°C 90sec), 72°C 10min, 4°C forever.

The second PCR mix was almost identical to the first PCR. 1-2µl of the product of the first PCR were used as a template and the two oligos T4494-4517 (GTG CTA TCC ACC TTG CGA AGA ATG C) and 262A-clamp (CGG ACG CTC AGC CAG GTT T) were used at a concentration as in the first PCR.

The second PCR program was:

94°C 4min, 20x (94° 30sec, 55°C 30sec, 72°C 90sec), 72°C 8min, 4°C forever.

5µl of the second PCR were loaded on a 1.5 % agarose gel to check for the presence of the fragment of the expected length. After seeing the expected product, the third PCR step was performed. 1µl of the product of the second PCR were used as a template and, maintaining the same concentrations as the first two steps, the oligos 1045-T4626-4648 (GAT CGG ACA TGT TGA CCA AGA CT) and 262A-clamp were used. 5µl of the third PCR were loaded on a 1.5% agarose gel to check for the presence of the fragment of the expected length. The last were then purified and cloned as explained in the next section.

4.3.9.2 Cloning of the mRNA 3' ends of constructs Δ5022 and Δ4922

The approximately 400bp and 300bp more intense fragments and the fainter fragments of ca 750bp and 650bp given by Δ5022 and Δ4922 respectively were purified from the gel (see §4.3.4). To verify their specificity, the eluted DNA fragments were re-amplified with a PCR reaction mix and a program identical to the third amplification step.

After the elution the fragments were directly cloned in the pCR2.1 vector (Invitrogen, Cat. no. K2000-01) in a standard 10µl ligation mix, as suggested by the manufacturer:

1µl 10X ligation buffer (Roche)

2µl pCR2.1 vector (25ng/µl)

6µl PCR fragment

1µl T4-ligase (1U/µl) (Roche)

➔ incubation at 14 °C for 16h.

5/10µl of the ligation were then transformed into *E. coli* OneShot TOP10 chemically competent cells (Invitrogen, cat. no. 44-0012). The recombinant clones were selected on kanamycin, then the plasmid DNA was isolated (§4.3.1.1) and sequenced.

4.3.9.3 RT-PCR of the mRNA of constructs Δ5022 and Δ4922

This method was used to assess the relative expression of Δ5022 and Δ4922 mRNAs. First the cDNA of the two transcripts was synthesized by an in vitro reverse transcription.

In this case the specific upward reading oligo 292A-T2653-2634 (CCC GTA ATT GAT CAT AAG AGA) was used to amplify the mRNA part encompassing the integrase intron (intron2). The oligo AtUBC9TGAup (GGG GTA CCA GAT CTC AGC CCA TGG CAT ACT TTT GGG T) was also added to amplify the constitutive *UBC9* gene as an internal expression control. To normalize the test, 2µg of induced and non induced total RNA of Δ5022 and Δ4922 were used as a template for cDNA synthesis. The volume of the RNA sample was adjusted to 5.95µl and incubated at 70°C for 5min to disrupt secondary structures and immediately transferred on ice. It was added in the following reaction mix:

4µl 5x AMV Buffer +
0.65µl RNase Inhibitor (40U/µl) +
8µl dNTPs (2.5mM) +
0.2µl oligo 292A-T2653-2634 (100pmol/µl) +
0.2µl oligo AtUBC9TGAup (100pmol/µl) +
1µl AMV Reverse Transcriptase (25U/µl)
➔ incubation at 42°C for 3h.

Afterwards 2µl of cDNA product (from each RT mix) were used as a template; the downward reading oligo 291A-T2314-2334 (GGT GGA AAG AGA GAC TGG TAA) was also added in an Intron2-PCR reaction mix for amplification.
This experiment was not successful, therefore the cDNA amplification, without investigation on the reasons, was performed with Intron1-PCR oligos combination: 755A-T969-991 (AGC TCG AAG AGT TGT ATG CCT CT) + 756A-T1109-1086 (CCG ACA ACT GGT CGA CAA TCC CTT) in the following reaction mix:

2µl cDNA +
5µl 5x LA Buffer (containing $MgSO_4$) +
0.5µl RNase Inhibitor (40U/µl) +
8µl dNTPs (2.5mM) +
0.5µl oligo 755A-T969-991 (100pmol/µl) +
0.5µl oligo 756A-T1109-1086 (100pmol/µl) +
0.5µl LA *Taq* polymerase (5U/µl)
23µl ddH_2O (up to the volume of 50µl).

The oligo AtUBC9TGAup was combined with the downward reading oligo AtUBC9TGAdn (TCC CCC GGG AGA TCT AGG ATG GCA TCG AAA CGG ATT TTG AAG) for amplification of the control gene, in an identical reaction mix.
The PCR program was the following:
94°C 5mins, 35x (94°C 30sec, 50°C 30sec, 68°C 1min), 68°C 10min, 4°C forever.

4.3.10 Protein Isolation Methods

In this work proteins were isolated from Arabidopsis and from *E.coli*, the purpose being different. Proteins isolated from plants were only used as a control, in experiments, whereas *E. coli* proteins were isolated to analyze the overexpression of *Tto1* proteins of interest.

4.3.10.1 Protein isolation from Arabidopsis

A fresh leaf (ca150mg) was harvested, and homogenized with a micro spoon of sand and 200µl protein isolation buffer (50mM Tris-Cl pH 6.8, 4% SDS, 10% β-mercaptoethanol) in a 1.5ml Eppendorf tube.
Alternatively the plant material was ground to a fine powder with liquid nitrogen. Then 200µl of the above mentioned isolation buffer were added to 150mg of the powder and mixed by vortexing. In both cases, after centrifugation (1min, 10000rpm), the supernatant was transferred to a new 1.5ml tube and incubated at 95°C for 10mins. In between, the samples were briefly vortexed two or three times. Finally, they were centrifuged again (14000 rpm, 10mins), and the supernatant was mixed with 1volume of 2x LSB (50% glycerol, 20mM DTT, 2% SDS, 125mM Tris-Cl pH 6.8, 0.003% bromophenol-blue) in a fresh 1.5ml tube. 20µl of the extract were loaded on a 12 or 15% poly-acrylamide gel for separation, depending on the protein mass.

4.3.10.2 Protein isolation from bacterial cultures

The samples were usually prepared from 0.5ml of bacterial culture, before or after overexpression. The 0.5ml were added to a 1.5ml Eppendorf tube, centrifuged at 14krpm for 1min and the supernatant carefully removed. The pellet was resuspended in 200µl 1x LSB and thoroughly mixed by vortexing. The de-

naturation of the protein was obtained by boiling the samples at 95°C (a thermal block was routinely used) for 10mins.

4.3.11 Protein Overexpression and Purification Methods

The methods hereby reported were used to obtain the overexpression and subsequent purification of the integrase protein fragment AgINT#2 and antibody purification.

4.3.11.1 Small scale AgINT#2 protein induction

First a little amount of a glycerol stock (§4.2.14.1) of Rosetta(DE3) p*LysS* containing the plasmid p*ET19::AgINT#2* was streaked on a fresh LB + 100-200 mg/l Ampicillin + 25 mg/l Chloramphenicol plate and incubate at 37°C for 16h. The day after some of the single colonies that had grown were pre-inoculated in 2ml of selective medium and grown over-night. The day after 0.4ml of the pre-inoculum were inoculated in 4ml of selective medium and let grow for ca 2h to physiologically adapt to the fresh medium. After the 2h the OD_{600} of a 1:10 dilution of the culture was measured: he cells must be in a log phase to be induced, (an OD_{600} from 0.6 to 0.9 is usually used). At this point the non induced control was made, taking a 0.5ml aliquot of the pre-cultures. A fresh 4ml selective LB culture was then prepared and inoculated with 0.4ml of pre-culture; the OD_{600} was routinely adjusted 0.7 for induction. The overexpression of AgINT#2 was induced adding 44µl of 1mM IPTG (to a final concentration of 1mM) in the 4.4ml total volume. The T_0 control sample was taken at this time. The induced cultures were then grown for 2h under shaking at 250rpm. The protein samples were prepared from 0.5ml of culture (an OD_{600}=1 was usually reached at this point), by centrifuging and resuspending in 50µl of 1x LSB buffer, treated as described in §4.3.10.2 and separated on SDS-PAGE gel (§4.3.12.1).

4.3.11.2 Big scale AgINT#2 protein induction

In order to get an appropriate amount of AgINT#2 protein for purification, larger cultures were induced. The procedure was basically the same as for small scale induction and the differences are reported below.

An overnight pre-inoculum of 15ml was first made from a single colony of the best expressers assessed by small scale induction experiments. The day after

10ml were inoculated in 500ml of selective LB, in a 2l Erlen Meyer flask and let grow under shaking at 37°C. The OD_{600} was measured every hour to obtain the value of 0.7; the time was slightly longer in respect to small cultures. 5ml of 100mM IPTG were then added for the induction and the culture was grown 2h to obtain overexpression of the protein. Afterwards the culture was split in two 300ml centrifuge bottles and the cells were harvested by centrifuging at 5krpm for 20min. For purification under denaturing conditions the centrifugation was performed at room temperature, whereas for native conditions the centrifuge was pre-cooled at 4°C. The supernatant was carefully removed and the pellet was weighted stored at -80°C until further use.

4.3.11.3 OD_{600} determination of bacterial cultures

Usually a 1:10 dilution was used to measure OD_{600} of a bacterial culture. 0.1ml of the culture of interest were diluted in 0.9ml of fresh LB medium added to a transparent plastic cuvette. After setting the "blank" with 1ml of fresh medium the OD_{600} was measured with a classical spectrophotometer.

4.3.11.4 Batch purification of AgINT#2 under denaturing conditions

This method was used to purify the appropriate amount of AgINT#2 protein, ca 800µg, to raise the antibody.

The bacterial lysate was prepared from 2.3g of cells (pellet). The pellet was thawed on ice for 15min and resuspended in 11.5ml of lysis buffer, containing 6M GuHCl, 0.1M Na-phosphate and 0.01M Tris-Cl; it was than vortexed for 15min trying to avoid foaming to obtain the lysis and a 0.5ml sample was taken to check lysis efficiency on SDS-PAGE. The cellular debris was then separated by a centrifugation at 10000g for 30min, and the supernatant was saved for the following purification. A sample was also taken from the supernatant and from the pellet (after resuspending) to compare the amount of protein present in the lysate and in the pellet.

The protein was then purified on a Ni-NTA column exploiting the 6x His-tag, following the protocol 17 reported on the booklet "The QIA*expressionist*™" published by Qiagen. The protein was eluted at pH 4.1 and the fractions were first neutralized (to pH 7.0) with 1.5M Tris pH 8.8 and checked on SDS-PAGE to assess the quality of the purified samples. The protein preps were subsequently concentrated using Centricon Filter Device YM-30 (Millipore); their concentration was estimated as in §4.3.11.5, and loaded on a large preparative poly-

acrylamide gel (§4.3.12.5) in order to obtain a single pure band that was sent to the company Eurogentec for antibody production.

4.3.11.5 Batch purification of AgINT#2 under native conditions

The protein AgINT#2 was purified under native conditions in order to use it in the subsequent antibody enrichment from the antiserum #3036. The protein overexpression was induced as in §4.3.11.2.

The frozen bacterial pellet of ca 3g was thawed for 20min on ice and resuspended in 15ml of lysis buffer (50mM NaH_2PO_4, 300mM NaCl, 0.1% Triton, 10mM imidazole, pH 8). To help resuspending, the pellet was pipetted up and down for a few times. The mixture was then added with lysozyme to a final concentration of 1mg/ml and incubated on ice for 30min. In the next step the sample was further lysed using a sonicator equipped with a microtip; 6 intermittent bursts of 30sec at 200W with a 30sec cooling period between each burst (the sample was kept on ice) were given. After the sonication the lysate was centrifuged at 1000g for 30min at 4°C to separate the cellular debris. The supernatant was saved and used in the following purification. After each step a sample was taken to be checked on SDS-PAGE. The purification of the protein was performed following protocol 12 from the booklet "The QIA*expressionist*™" published by Qiagen, with some modifications. 0.1% Triton was in fact added to each buffer and the wash buffer had the same composition as the lysis buffer. Furthermore a second elution step was performed with a low pH buffer (4.1), which contained 300mM NaCl, 0.1% Triton, 50mM acetic acid, 10mM NaAc, to test the efficiency of the elution with the standard buffers.

4.3.11.6 Determination of protein concentration

The concentration of protein samples was estimated performing a classical Bradford assay in which the BCA™ Protein Assay Kit (Pierce, Cat no. 23225) was used.

4.3.11.7 Dialysis of AgINT#2 in preparation of Ab affinity purification

In order to enrich the anti AgINT#2 antibody, the protein was dialyzed in order to be made free of the elution buffer used for purification (§4.3.11.5). This was done by dialysis. 13ml of eluate were dialyzed were loaded into a 3.5cm long membrane with a 10kDa cut off, in order to eliminate also the unwanted smaller sized contaminating proteins. The dialysis was performed against 1.5l (over

100 times the starting volume) of coupling buffer containing 0.1M MOPS pH 7.5 and 2% glycerol, which was used later on in affinity purification. The procedure was run for 36h at 4°C, and buffer was exchanged every 12h; the last time 0.5mM DTT was added. The solution was continuously agitated with a magnetic stirrer.

4.3.11.8 Anti AgINT#2 Ab affinity purification.

The Ab was purified using Affi-Gel® 10 Gel (Biorad, Cat. no. 153-6046) with little modifications on the manufacturer's protocol.

0.8ml of the Affi-Gel resin were first equilibrated with 5ml (>5vol) of coupling buffer without DTT. Then 1.5mg of protein in a volume of 3ml were added to the resin and coupled over-night at 4°C in a chromatography column under gentle inversion shaking. The day after 100µl of 1M ethanolamine-Cl pH 8 were added to obtain the saturation of all ester groups that might still be active; this was done by incubation at 4°C for 1h under gentle shaking. The flow-through was the collected and a sample was taken to check coupling efficiency by SDS-PAGE. The column was then washed with 4vol of 1x PBS pH 7.2 (8g NaCl, 0.2g KCl, 1.44g Na_2HPO_4, 0.24g KH_2PO_4 for 1l; pH is adjusted with HCl).

In the next step the binding of antiserum #3036 (third bleed) was performed. 50ml of crude antiserum were slowly loaded on the column and let pass trough by gravity. The flow-through was kept, in order to re-use it in case of need. The column was washed with 4vol PBS, and another sample was taken.

The elution was finally performed with 7ml of 0.1M glycine pH 3.0 + 10% glycerol. 7 1ml fractions were collected in 2ml Eppendorf tubes and immediately neutralized with 150µl of 1M Tris-Cl pH 8.0, (pH 7.5) and stored at 4°C.

Aliquots of each elution fraction and each purification step were then checked on a 12% poly-acrylamide gel.

4.3.12 Protein Visualization Methods

4.3.12.1 Poly-acrylamide gels and SDS-PAGE

Poly-acrylamide gels were used to obtain the separation of proteins according to the size. The denatured protein samples (see §4.3.10.2) were usually separated on 12% (or 15%) poly-acrylamide gels (Ausubel et al., 2002) that had the following composition:

- separating gel **12%**: 1.6ml dH_2O, 2ml 30% acrylamide mix (SERVA), 1.3ml 1.5M Tris-Cl pH 8.8, 50µl 10% SDS, 50µl 10% APS (ammonium-persulfate), 4µl TEMED.
- separating gel **15%**: 1.1ml dH_2O, 2.5ml 30% acrylamide mix, the concentration of the other components being the same as for the 12% gel.
- stacking gel (low concentration): 1.4ml dH_2O, 330µl 30% acrylamide mix, 250µl 1M Tris-Cl pH 6.8, 20µl 10% SDS, 20µl 10% APS, 2µl TEMED.

The composition for 500ml of the 5X running buffer was: 7.55g Tris, 36g glycine, 2.5g SDS, dH_2O added to volume.

The protein size marker used was the PageRuler Prestained Protein Ladder, (Fermentas, #SM0671). 6µl were generally used; but the amount was doubled for gels that were further used in Western blots.

The samples (usually 12µl of 1:1 protein and 2x LSB) were loaded on a gel and separated for 2h using two different power settings. The first part of the run was performed at 90V for ca 30mins to collect all samples at the border between the stacking gel and the separating gel. For the second part (the separation) the voltage was turned to 120-130V (ca 15mA) for 1.5h.

After the gel run, the stacking gel, containing the loading slots, was removed and the gel was stained with Roti-Blue (Roth) and distained by washing 3 times for 10min with distilled water under gentle shaking, according to Roth's suggestions.

4.3.12.2 Western blots

In a Western blot the proteins of interest are transferred to a solid support, like a PVDF or nitrocellulose membrane, and subsequently visualized by antibodies coupled to enzymatic reactions.

In this work, a primary antibody specific for *Tto1* INT protein (precisely the AgINT#2 fragment) was used. In turn, the primary Ab was detected by a secondary Ab coupled to alkaline phosphatase. After separation on a polyacrylamide gel (§4.3.10.2), the gel was incubated and gently agitated in Transfer Buffer (190mM glycine, 25mM Tris, 20% methanol, 0.05% SDS) for approximately half an hour. In the meantime, a PVDF membrane (Immobilon-P, Millipore) was soaked in 100% methanol and shaken in Transfer Buffer for at least 10mins. In the next step the proteins were transferred to the membrane by electro-blotting (Biorad) at 50V at 4°C for 1h or 15V and 60mA at 4°C for 16h.

The blotting unit consisted in the typical "sandwich": a sponge covered by two layers of 3MM paper, the gel, the membrane, another two layers of 3MM paper and a second sponge. After transfer, the membrane was washed twice in 1x ANT (150mM NaCl, 50mM Tris-Cl pH8.0, 0.002% NaN3) for 5mins and then blocked in 1x ANT/20% NCS (new-born calf serum) at room temperature for at least 1h.

4.3.12.3 Revelation of the Western blot with antibodies

In the next step the membrane was exposed to the primary Ab (rabbit anti-INT antibody #3036) at a 1:3'000 dilution in 1x ANT/20% NCS, either at 4°C overnight or at room temperature for 3h. After this incubation the membrane was washed with 1x ANT/0.05% Tween-20 at room temperature for 10mins for 3 times to remove the unbound primary Ab. Subsequently, it was incubated with the secondary goat anti-rabbit Ab (Sigma A-3812) at a 1:1000 dilution, in 1 x ANT/20% NCS, for 1.5h. After three washing steps (1x ANT/0.05% Tween-20) at room temperature for 10mins, the membrane was developed with 90µl NBT (110mM in 70% DMF) and 70µl BCIP (90mM in DMF) in 1x TE for 10mins up to 16h. Finally, the blot was washed with water, dried on filter paper and sealed in a plastic bag for storage.

4.3.12.4 Revelation of the Western blot with Anti His Ni-NTA AP Conjugate

This method was used to detect the AgINT#2 integrase fragment, overexpressed and purified as a 6x Histidine-tag chimeric protein. The Ni-NTA (Nickel-NitriloTriaceticAcid) AP conjugate (Qiagen, Cat. no. 34510) specifically binds to the His-tag. It can be revealed like an antibody as in a common Western blot, exploiting the enzymatic reaction of the AP (alkaline phosphatase), that produces a colored precipitate on the site of binding to the protein.

The samples were generally separated on SDS-PAGE and blotted as described in §4.3.11.1 and 4.3.11.2. The binding of the Ni-NTA resin to the chimeric protein was then obtained following the manufacturer's protocol. The Qiagen protocol was also followed for detection, which actually matched the same detection method with NBT and BCIP already described in §4.3.11.3.

4.3.12.5 Large preparative poly-acrylamide gel (for antibody production)

A large SDS-PAGE with a preparative purpose was done in order to obtain a well distinct band containing the AgINT#2 that was subsequently sent to Eurogentec

for antibody production. The gel had the volume of 50ml and the following composition:

- separating gel 12%: 16ml dH_2O, 20ml 30% acrylamide mix (SERVA), 13ml 1.5M Tris-Cl pH 8.8, 500µl 10% SDS, 500µl 10% APS, 40µl TEMED.
- stacking gel: 7ml dH_2O, 1.65ml 30% acrylamide mix, 1.25ml 1M Tris-Cl pH 6.8, 100µl 10% SDS, 100µl 10% APS, 10µl TEMED.

The composition of the running buffer was the same as reported in §4.3.12.1.

8 samples, containing from 120 to 200µg of protein after the concentration step with Centricon, were adjusted to 90µl and added with 90µl of 2X LSB, to have a total of 180µl. They were normally denatured by boiling 10min and separated at 60V for 10h. The day after the gel was first washed with distilled water for 15min and stained with 4M NaAc under gentle shaking. With this peculiar method of staining the SDS would precipitate and the bands would appear white. The bands of 35KDa corresponding to AgINT#2of interest were cut from the gel, placed in a 2ml Eppendorf tube and stored at -80°C until being send to Eurogentec.

4.3.13 Methods for Arabidopsis

In this work the Arabidopsis ecotype Colombia was used as a plant model. The plants were routinely reproduced on soil in greenhouse or on agar MS medium in a growth chamber, for specific applications.

4.3.13.1 Seed sterilization

In order to prevent contamination from the fungi and bacteria that might infect the plant, the seeds were usually made sterile before being plated on agar medium and before being sown on soil in the greenhouse. In case of T1 generation plants, the seed sterilization was also used to prevent the growth of *Agrobacterium* cells that were still leaving under the bacterial coat.

A number of methods were used, which involved Cl, either as a gas or a solution, and the most effective one is reported following.

A small amount of seeds (a very small spoon) was placed in an 1.5ml Eppendorf tube and added with 0.5ml of sterilizing solution (5g $Ca(OCl)_2$, 0.1ml 20% Triton X-100 for 100ml solution) normally stored at 4°C. The tube was then shaken for 15min at room temperature at 1.2krpm. Afterwards the tube was spun

down at top speed for 30sec and the supernatant was discarded. The seeds were then washed 3 times with 0.5ml of sterile distilled water. The tube was each time vortexed and spun down at top speed for 30sec and the supernatant (water) was removed. The tube containing the wet seeds was then placed under a sterile hood for an overnight for drying (sometimes up to 24h were necessary). The dry seeds were then stored at room temperature and used within 3 weeks.

4.3.13.2 Floral-dip transformation of Arabidopsis plants

This method was used to obtain transgenic Arabidopsis plants with the constructs of interest.

Preparation of the plants

The plants were grown for about three weeks under SD (short day = 8h light) condition, until the primary shoot reached 4-5cm. The shoot was then cut and the plants were let grow for another 7-10 days. During this time several secondary shoots developed and reached the length of 10-15cm. This procedure was made to obtain a higher number of flowers for the later transformation.

Preparation of the bacteria

A single colony of Agrobacterium containing the construct of interest was inoculated in 20ml of YEB medium (§4.2.4.1) containing the selective antibiotic and 2mM $MgSO_4$ and grown for 2 days (until saturation) at 28°C, under shaking at 200rpm. 2ml of the pre-culture were then added in 100ml of fresh medium and grown for 16h in the same conditions as above.

The following day the bacteria were collected by centrifuging for 20min at 5krpm.

After removing the supernatant, that normally still contained some bacterial cells, the pellet was resuspended in 2vol (200ml) of 5% Sucrose + 0.05% Silwet and poured into an appropriate beaker. Silwet is a surfactant agent that prolongs the permanence of the Agrobacterium on the surface of the plant, thus increasing the probability of transformation.

Transformation

The beaker containing the bacterial solution was put on plate with a rotating magnetic stirrer. After removing the eventual already mature siliques, the flowers were dipped for about 30sec while stirring. The plants were laid on the side and immediately covered with a lid and transferred to the greenhouse. The lid was kept for 1 day in order to maintain a high humidity percentage thus helping

the plants recover from the stress derived by the multiple in Agrobacterium infection. After removing the lid, the plants were grown normally and the seed harvested after approximately one month.

4.3.13.3 Selection of recombinant plants on solid 1% Ara medium

Recombinant Arabidopsis plants were routinely on solid 1% Ara added with 15mg/l hygromycin (§4.2.1.1). Approximately 200 sterilized seeds (§4.3.13.1) were uniformly distributed on the plate; the last was sealed with Parafilm and incubated 2 days at 4°C for vernalization. The plate was then transferred to a growth chamber under LD conditions. After ca 15 days –the seed had germinated- it was possible to distinguish the resistant seedling from the non resistant ones, as they were generally larger, showing longer roots and hypocotyls and a deeper green color of the cotyledons. The resistant (or putative) seedlings were then transferred on soil and grown under LD conditions in a greenhouse or in a cabinet with more controlled parameters.

4.3.13.4 Alternative selection method of recombinant plants on SiO_2 sand

Another method was also used mainly for the selection of Δ3'LTR *Tto1* recombinant Arabidopsis plants, which used SiO_2 sand instead of solid MS medium. Bacterial Petri dishes with vents, of the diameter of 20cm were filled with ca 50g of sand; and soaked with 20ml of ¼ MS medium, 1x Vitamix and 15mg/l hygromycin. Non sterile seeds were then distributed on the surface; the plates incubated 2days at 4°C for vernalization, and transferred to a growth chamber, for germination and growth.

The plates were not sealed, in order to increase air exchange thus drastically reducing the risk of fungal contamination. The seed were left to grow 2 weeks, being supplied every week with the same volume of medium. This medium resulted to be very efficient against fungal and also bacterial contamination, due to the increased air exchange, achieved by no sealing of the plates, and less time consuming, as no sterilization of the seed was required and no autoclaving of the medium that was filter-sterilized.

4.3.14 Methods for *E. coli*

4.2.14.1 Glycerol stocks of bacteria

This method was applied to store clones of interest of *E. coli* and *A. tumefaciens*. 4ml of medium were inoculated with a single bacterial colony in a 20ml test tube. The cultures were normally grown for 16h at either 37° or 28-30°C under at least 200rpm shaking. Afterwards (the day after) 0.75ml of the culture were mixed with sterile autoclaved 50% glycerol in a 2ml cryo-vial, mixed thoroughly and stored at -80°C.

4.2.14.2 Preparation of chemically competent *E. coli*

E. coli cells needed to be made competent in order to be transformed with the constructs of interest. The following protocol was used to prepare cells to heat-shock transformation.

A 5ml preculture was inoculated with a single colony from plate and grown overnight at 37°C and 200rpm. On the following day 2ml of the pre-culture were inoculated in 100ml of fresh medium and grown at 37°C and 200rpm until they reached and OD_{600} of 0.3–0.4. The cells were harvested by centrifuging at 5000rpm for 10min at 4°C. The pellet was resuspended in 10ml of cold TSS medium (§4.2.3.2) solution and frozen in liquid nitrogen in 100-200µl aliquots, which were then stored at -80°C.

4.3.14.3 Heat Shock Transformation of *E. coli*

Chemically competent *E. coli* cells were transformed with a plasmid using the following protocol.

A 100µl aliquot of cells was thawed on ice and mixed with 50–150ng of DNA (usually 1µl of a plasmid miniprep or 5/10µl of ligation mix were used) in a 1.5ml Eppendorf tube. The tube was incubated for approximately 30min on ice, heat-shocked for 2min at 37°C and again transferred on ice for another 30sec. 750µl of fresh LB medium without antibiotic were then added and the cell were incubated for 1h at 37°C with a shaking speed of 900rpm. Afterwards the bacteria were harvested by centrifuging at 3'000rpm (ca1000g) for 3min. The supernatant was removed, but 200µl were left in the tube, which were used to resuspend the pellet. Finally the cells were plated on LB plates with the appropriate antibiotics and grown for 16h at 37°C.

4.3.14.4 Electroporation of *E. coli*

In this thesis work the *E. coli* cells ElectroMAX Stbl4 were also used, which are electro-competent, therefore they were transformed by electroporation. Such cells are already provided competent by Invitrogen and the electroporation protocol is reported on the product user manual (Invitrogen, cat. no. 11635-018).

4.3.15 Methods for *A. tumefaciens*

4.3.15.1 Preparation of chemically competent *A. tumefaciens* C58C1 cells

Agrobacteria were made transformation competent using the following protocol.

10ml YEB (§4.2.4.1) containing 50mg/l Rifampicin and 2mM $MgSO_4$ were inoculated with a single Agrobacterium colony and incubated at 28°-30°C for two days, with a shaking speed of 200rmp. The day after 4ml of this pre-culture were transferred to 100ml of fresh YEB and let grow for 3.5h, when an OD_{600}=0.25 approximately was reached. Cells were harvested in 50ml tubes by centrifugation at 5000rpm at 4°C for 20mins; the pellet was kept on ice and resuspended in 1ml of cold YEB (+ 50mg/l Rifampicin and 2mM $MgSO_4$). 200µl aliquots were made and immediately frozen in liquid nitrogen and stored at -80°C until further use.

4.3.15.2 Heat Shock Transformation of *A. tumefaciens* C58C1

This method was used to introduce plasmid DNA in Agrobacteria by heat shock.

A 200µl aliquot of competent cells of the strain C58C1 was thawed on ice and added with 2–5µg DNA in the same tube. The mix was frozen in liquid nitrogen for 1min then it was transferred at 37 °C for 5min and immediately added with 1ml of fresh YEB (+50mg/l Rifampicin and 2mM $MgSO_4$). The tube was then incubated at 28-30°C for 2h under shaking.

Afterwards the cells were harvested by spinning down; the supernatant was removed, but 200µl were left and used to resuspend the pellet. 180µl were plated and the remaining 20µl were first diluted 1:10 and also plated on solid YEB containing 10mM MgSO4, 50mg/l Rifampicin and the selective antibiotic. The plates were incubated at 30°C for 2 days.

4.3.15.3 Preparation of electro-competent *A. tumefaciens* AGL10 cells

The Agrobacterium strain AGL10 was specifically used to obtain transgenic barley plants. These cells were transformed with plasmids of interest by electroporation and beforehand they were made electro-competent using the following protocol.

A pre-culture was prepared by inoculating a single colony of AGL10 cells in 10ml of AGL10 medium (§4.2.4.2) containing 25mg/l Rifampicin and 0.5mM $MgSO_4$ and grown for 2 days at 30°C under agitation (200rpm). 4ml of the pre-culture were then inoculated in 100ml of fresh AGL10 medium and grown at 30°C for at least 16h under vigorous shaking (>200rpm), until an OD_{600} of 1-1.5 was reached. The culture was then cooled on ice and transferred to 250ml centrifuge bottles on ice. The cells were collected by centrifuging at 5000rpm for 10min 4°C. After removing the supernatant (working on ice and under a sterile hood) the pellet was washed 3 times with 50ml of cold sterile 1mM HEPES pH 7 and once with 50ml of cold sterile 10% glycerol. The bacteria were then resuspended in 4ml of 10% glycerol, divided in 200µl aliquots in 1.5ml Eppendorf tubes and immediately frozen in liquid nitrogen. The aliquots were finally stored at -80°C until use.

4.3.15.4 Electroporation of *pVec8::Tto1N* and *pVec8::Tto1X* in AGL10 cells

Two 50µl aliquots of electro-competent AGL10 cells were thawed on ice and mixed with 250ng of the barley vectors *pVec8::Tto1N* and *pVec8::Tto1X* respectively (on ice). The mix was transferred into a pre-cooled (-20°C) electroporation cuvette with an electrode distance of 0.2cm. A single pulse of 2.5kV using the 25µF capacitor was applied righter after removing the cuvette from ice. Immediately after the pulse 800µl of SOC medium or just AGL10 medium (also YEB can be used) were added and the mix was transferred into a 15ml Falcon tube. The tube was then incubated at 28-30°C for 1.5h with vigorous shaking. After this period 200µl of each mix were spread on a solid AGL10 medium containing 25g/l Rifampicin and 25g/l Spectinomycin and grown for 2days at 30°C, until single colonies would appear.

4.3.16 *Tto1* constructs cloning strategies

4.3.16.1 Cloning of 3' LTR deletion constructs

In this section the cloning strategy to obtain the Δ3'LTR *Tto1* construct A, B, C, D and E is described. Technical parameters such as amount of insert versus plasmid, PCRs or fragment purification are fully described in former sections of this chapter.

The plasmid *pTAs::Tto1Swa* (provided by Andreas Bachmair) was digested with *XhoI-PvuII* to excise the 5.3kb *Tto1*. The 5.3kb full-length *Tto1* was then cloned into *XhoI-PvuII* digested vector *pER8* (Spec$^+$), giving the vector *pERnew::Tto1SwaI.* The pea rbcS terminator sequence was in the meantime PCR amplified from a *SpeI* linearized *pER8* vector and blunt cloned into *SmaI* site of the plasmid *pSKII*, thus giving the construct *pSKII::rbcS*. *pER8new::Tto1SwaI* was used as a template to generate the fragments containing five deleted versions of the LTR, by combination of the oligo 1045-T4626-4648 with the "shortening" oligos 782A-T5233up (for construct A), 783A-T5119up (B), 784A-T5022up (C), 786A-T4922up (D), 787A-T4900up (D) respectively. The PCR fragments were gel purified and cloned using *SwaI* and *PflMI* restriction sites into *pSKII::rbcS*. The ensuing plasmids *pSK::Δ5322-rbcS, pSK::Δ5119-rbcS pSK::Δ5022-rbcS, pSK::Δ4922-rbcS and pSK::Δ4900-rbcS* were digested by *SwaI-PvuII* and cloned into *SwaI-PvuII* digested *pERnew::Tto1Swa,* to give the constructs *pERnew::Tto1Δ5322* (A), *pERnew::Tto1Δ5119* (B), *pERnew::Tto1Δ5022* (C), *pERnew::Tto1Δ4922* (D) and *pERnew::Tto1Δ4900* (E). Each fragment generated by PCR was sequenced. All cloning steps were carried in *E. coli*, nevertheless two different strains were used. The *pSK*—backbone constructs were propagated into XL1-blue cells, while for propagating the pER8 based constructs, it was necessary to use the strain ElectroMAX Stbl4, in order to maximize the stability of the direct repeats (LTR) that often resulted in recombination of the LTR, in XL1 blue.

4.3.16.2 Construction of inducible *Tto1N* and *Tto1X* for Arabidopsis

The full-length 5.3kb *Tto1* from *pERnew::Tto1* was first subcloned using *XhoI/PvuII* cleavage sites into a modified *pACYC177* (Amp$^+$), to make the plasmid ***pACYC::Tto1Xho-Pvu.*** As *Tto1* contains 2 *BsiWI* sites (position 611 and 2585), in order to clone the integrase2 (int2) core domain and the barley Xyl18 intron (intron2) the *BsiWI* site at position 611, had to be temporarily removed.

A 808bp *XhoI-SapI* fragment, containing the first *BsiWI* site, was excised from *pACYC::Tto1Xho-Pvu* and the 2 ends were filled-in to re-create *XhoI* site and make ***pACYC::Tto1delBsiWI***.

The filling in was performed using the Klenow Enzyme (Roche, Cat. no. 11008404001) following manufacturer's suggestion.

Subsequently a 601bp *BsiwI–BglII* fragment containing the int2 domain from the clone AA34, was cloned in *pACYC::Tto1delBsiWI* to make the plasmid ***pACYC::Tto1.2delBsiWI***. The previously named ***pACYC::Tto1Xho-Pvu*** construct was then renamed ***pACYC::Tto1.1Xho-Pvu,*** because it still contained the native integrase.

At this point the deleted 5´ region (*XhoI-SapI* 808bp fragment), containing the first *BsiwI* site (position 611), was then restored by ligating a *Tto1* 1661bp *XhoI-PshaI* fragment, excised from *pACYC::Tto1Xho-Pvu,* with ***pACYC::Tto1.2delBsiWI,***, to make the plasmid ***pACYC::Tto1.2reconstituted***.

In a last step the full length *Tto1.2* was excised by a *XhoI-PvuII* digestion and cloned in the *XhoI-PvuII* digested *pER8new* to make the vector *pERnew::Tto1.2*. In the end *pERnew::Tto1* was renamed *pERnew::Tto1N, pERnew::Tto1.2* was renamed *pERnew::Tto1X* in order to distinguish between the native (N) while *pERnew::Tto1.2* was renamed *pERnew::Tto1X* where X was used to refer to uncertainty of the function of the Int2 version. A schematic representation of the cloning steps is represented in Fig. 2.19.

4.3.16.3 Cloning of constitutive *Tto1N* and *Tto1X* for barley

Starting from the previous cloning steps in which a synthetic version of an inducible *Tto1* construct (*pERnew::Tto1X*) was made, the construction of constructs suitable for barley were carried out.

To replace the pre-existing Arabidopsis intron, a 148bp fragment containing the barley Xyl18 intron (intron2) was excised by *BspEI* and *BsiWI* digestion from the plasmid *pUCBIint+* (obtained from A. Bachmair) and cloned in *BspEI* site at position 2349 and *BsiWI* site at position 2585, to make ***pACYC::Tto1.1delBsiWI-intron2*** and ***pACYC::Tto1.2delBsiWI-intron2***. The deleted 5´ region (*XhoI-SapI* of 808bp) containing the first *BsiwI* site (611), was then restored by ligating a *Tto1* 1661bp *XhoI-PshaI* fragment excised from the *pACYC::Tto1Xho-Pvu,* in the two previous constructs, to make plasmids ***pACYC::Tto1.1int2reconstituted*** and ***pACYC::Tto1.2int2reconstituted***. In the next step the nos terminator was added to the developing barley construct. A 300bp fragment containing the nos

terminator was PCR amplified from the *HindIII* linearized plasmid *pWBVec8*, using the oligos 1113-nosTBarleydn and 1114-nosTBarleyup. Such purified product was then *EagI* and *HindIII* digested and ligated with the *NotI-HindIII* digested *pACYC::Tto1.1reconstituted* and *pACYC::Tto1.2reconstituted* to make the vectors ***pACYC::Tto1.1int2-nosterm*** and ***pACYC::Tto1.2int2-nosterm***. The pre-existing rbcS terminator was also displaced in this step.

Intron1 (Gag intron) was then cloned in *pACYC::Tto1.1intron2nosterm* and *pACYC::Tto1.2intron2nosterm.* The *XhoI-SalI* (1kb) fragment containing the Arabidopsis Gag intron was excised and replaced with a 1kb *EcoRV-SalI* sequence excised from the plasmid *p2RT172Bla*. The *XhoI* cleavage site was blunted by a Klenow Enzyme reaction and ligated to the *EcoRV* site of the 1kb insert and to *SalI.* The 2 kinds of *Tto1* vectors, containing both introns and nos terminator were named ***pACYC::Tto1.1int1+int2-nosterm*** and ***pACYC::Tto1.2int1+int2-nosterm***.

To complete the barley construct the constitutive ubiquitin1 promoter (*Mubi1*) of maize was added to the previously made constructs. The promoter was obtained by PCR amplification from the plasmid *p6U* in 2 PCR steps.

The first part was a 500bp fragment spanning from positions 992 to 1510 of the promoter, containing the enhancer, and it was amplified with oligos 1136-Mubpintrondn and 1132-Mubpup. The fragment was then digested with *SpeI* and *SalI* and ligated with the *NheI-XhoI* digested *pACYC::Tto1.1int1+int2-nosterm* and *pACYC::Tto1.2int1+int2-nosterm*, to make plasmids ***pACYC::Tto1.1Mub1p1*** *and* ***pACYC::Tto1.2Mub1p1***. The second part was a 900bp fragment spanning from positions 1 to 911 and was amplified with the oligos 1131-Mubpdn and 1137-Mubpintronup. It was subsequently *SalI* and *SpeI* digested and ligated to *AvrII* and *XhoI* digested *pACYC::Tto1.1Mub1p1* and *pACYC::Tto1.2Mub1p1* (*AvrII* and *XhoI* cleavage sites were contained in the oligo 1136-Mubpintrondn). In this way the plasmids ***pACYC::ΔMub1Tto1.1*** and ***pACYC::ΔMub1Tto1.2*** were made. After each cloning step the ensuing constructs were sequenced to check correctness of the sequence. Both constructs were then renamed *pACYC::Tto1N* barley and *pACYC::Tto1X* barley respectively. The last *Tto1* constructs were finally transferred to the plasmid *pWBVec8*, by digestion of both *Tto1* inserts and of the plasmid with *NotI* and *HindIII*, to *make* ***pVec8::Tto1N*** and ***pVec8::Tto1X***. All previously described cloning steps are schematically represented in Fig. 2.21 in the results chapter.

4.3.16.4 Construction of the *pBIB-pOP6::Tto1* vector

In order to test *Tto1* with the transactivated system pOP6/LhGR-N, (see §2.7 and 4.3.18) the donor vector *pBIB::pOP6-Tto1* had to be made.

A modified version of the plasmid *pBIB-Hyg* (Becker, 1989) containing the strong chimerical enhancer sequence pOP6 from the expression cassette pOP6-GUS (Craft et al., 2005; Samalova et al., 2005) was provided by Andreas Bachmair. The *pBIB*-based plasmid also had the UTR region shortened, in order to have the TATA box starting right after the enhancer region. The modified *pBIB* vector was digested with *SmaI* and *SalI*, and a *PvuII -XhoI Tto1* fragment *Tto1* carrying the rbcS terminator *pCYC::Tto1Xho-Pvu* (§4.316.2) was inserted. The ensuing plasmid *pBIB::pOp6-Tto1* was transformed in *E. coli* for propagation, successively electroporated into *A. tumefaciens* C58C1 and finally transferred to Arabidopsis. The construct was then tested as described in §4.13.18.

4.3.17 Sequencing of *Tto1* insertion flanking sites

This method was used to identify the flanking sites of i*Tto1*insertions in Arabidopsis.

The plant tissue of lines showing the spliced band after a diagnostic intron-PCR was ground to a fine powder using a SamplePrep Freezer/Mill 6870 (SPEX). The genomic DNA was isolated using a commercial kit as reported in §4.3.1.4. The DNA was digested with *EcoRI* and *HindIII* (§4.3.7.1) subjected to DNA gel blot analysis as described in section 4.3.7.and published by Böhmdorfer et al. 2005).

To isolate DNA regions flanking *Tto1* insertions, the DNA was digested with *EcoRI* and *HindIII* and separated on a 0.9% agarose gel. Size fractions of 500–1000 bp and of 1000–2000bp were purified from the gel using the kit Wizard SV Gel and PCR Clean-Up System (see §4.3.4).

The fractions were ligated to the *EcoRI* and *HindIII* digested vector *pBluescript SK+* and used to transform *E. coli* strain Stbl4 (Invitrogen).

A colony hybridization was carried out using the methods published by Ausubel et al. 1987 and a 900bp probe was used, which spanned *Tto1* sequence from nucleotides 4390 to 5290. (see Appendix B for complete sequence of the probe). After 2 rounds of sub-cloning 2 positive clones were found. The plasmid DNA of the positive clones were sequenced, identifying the 3' region of 2 *Tto1*

insertions. Such information was used to design PCR primers for amplification of the 5' second *Tto 1*—genome junction.
The PCR reactions to identify the 5' end junction performed as described in §4.3.2.1. The primer combination included sequence specific oligos:
First insertion: *Tto 1* specific oligo 234A-T189-165 (GAA ATG GAA GGG GGT ATT TAT AGT T) and 1254-At3g14480dn1 (ATG TCG GAT TTG GTT ATT GCA CTC).
Second insertion: oligo 234A-T189-165 (GAA ATG GAA GGG GGT ATT TAT AGT T) and 1256-At2g26420up1 (CCC TCG AGG CAC CTA TTT GGC ACC T).
The PCR fragments were directly subjected to sequencing.

4.3.18 The Dexamethasone inducible *pBIB::pOp6Tto1/ LhGR-N* binary system

4.3.18.1 Assessment of hybrid Arabidospsis *pBIB-pOp6::Tto1/ LhGR-N* lines

pOp6/LhGR-N is a binary transactivated system that needs to components: a donor line carrying the gene of interest, and an activator line carrying *LhGR-N* expression cassette, which is responsive to the glucocorticoid dexamethasone. Plants carrying the two components respectively have to be crossed and the hybrid line is treated with dexamethasone, to obtain expression of the desired gene (Craft et al., 2005; Moore et al., 1998; Samalova et al., 2005).
The activator line S5 Colombia was crossed to the donor containing *pBIB::pOp6-Tto 1*, carrying the selective resistance to kanamycin, was provided by Andreas Bachmair. The donor line was obtained introducing *pBIB::pOp-Tto 1*, carrying the selection marker of resistance to hygromycin B, into wild type Columbia and using the methods described in §4.3.13.2 and 4.3.13.3. Three donor lines, 2-1, 3-1 and 4-1 were isolated and used crossed with S5.
The pollen of S5 was spread on the stigma of pistils of premature flowers of the lines from 2-1 to 4-1 and the developing siliques were let grown isolated with a transparent film, to distinguish from the non crossed siliques on the same plant. The seed of the hybrid siliques were sterilized and germinated on solid 1% Ara plates supplemented with the double antibiotic hygromicin and kanamycin. The F1 seeds were selected (4.3.13.3) and the progeny was used for induction experiments.

4.3.18.2 Preliminary Induction experiment of F1 progeny *pBIB::pOp6-Tto1/LhGR-N*

The F2 seed of the cross S5 x 2-1 to 4-1 were used to test the expression of *Tto1* after chemical induction. The seeds were germinated on selective solid 1% Ara plates, and the best seedlings were transferred to liquid medium supplemented with 5µM dexamethasone. Some seedlings were grown on medium lacking dexamethasone for negative control. After two weeks of induction the genomic DNA was prepared (§4.3.1.3) and the expression level was tested by Intron1-PCR (§4.3.2.3).

4.3.19 Softwares for RNA structure prediction

All secondary structure predictions were carried out using tools of the ViennaRNA suite (Hofacker et al., 1994), using standard folding parameters including dangling end energies for the bases adjacent to a helix (-d2 option). Cofold analyses were performed using "RNAcofold" (Bernhart et al., 2006). For prediction of the secondary structure of strong stop cDNA, special energy parameters for DNA folding were used (Mathews et al., 2004). For the computation of the energies of cDNA/mRNA hybrid stacks, parameters of Wu et al. (2002) were used. "RNAup" (Mückstein et al., 2006) was used to compute the probability that regions of the strong stop cDNA, or of mRNAs, remain unpaired.

INTRONS USED IN ARABIDOPSIS CONSTRUCTS

Intron1 (**GAG**) – 94bp - **intron RNA Polymerase II large subunit**

5'- GTAGCAATTT CATCAGCGGT TGCTTTTTAA AGTTATCTAT GATATTGATG AGAGGATAAA
CCAACTCTTA TGCTGATATC TTATCCATTT TCAG - 3'

Intron2 (**Integrase**) – 85bp - **intron RNA Polymerase II large subunit**

5'- GTAGCAATTT CATCAGCGGT TGCTTTTTAA AGTTATCTAT GATATTGATG AGAGGATAAA
CCAACTCTTA TGCTGATATC TTATCCATTT TCAG - 3'

Introns used in barley constructs

Intron1 (**GAG**) – 190bp - **intron barley Xylose Isomerase 13**

5'- GTATTTTTTG AAATGTGGCA AGAAAATCAT TTAGTACAAC CCTTCCTGTT TTTGTATGTT
ATGTGTGTAT TTGCAACAAT TCTATGCAAA ATGAATCATG TGAATTAAAC ATGTTTTCC
TGACTATCAA CTTAGAACGT TCTATTTTGA AATAAAATGT ATTCTTTTTT CTAATAATAC
GGTGATGCAG-3'

Intron2 (**Integrase**) – 86bp: **intron barley Xylose Isomerase 18**

5'- GTAACTGGAG ACATCATTAG TTCATCACTC GGTAAAATTT GCACATGCCT TTACCTAAAT
GTAAGGTTTG TTTTCTATGT TATTAG-3'

Tto1 probe used in Southern blots - bp 4390 to 4697

5'- CATCGCAGCA ACGGAGGCTT GCAAAGAATT GATATGGATG AAGAAGTTCT TAACTGAACT
TGGATTTTCG CAAGACGGTT ATCAGTTATT TTGTGATAGT CAAAGTGCTA TCCACCTTGC
GAAGAATGCC TCATTCCATT CCAGATCCAA ACATATTGAT GTGAGATATA ATTGGATCAG
GGATGTGTTG GAGAAGAAGA TGTTGCGGCT TGAAAAGATC CATACAGACG AAAATGGATC
GGACATGTTG ACCAAGACTT TACCGAAAGG GAAGTTTGAG TTCTGTAGAG AAGCTGCAGG
GATAGTGG - 3'

Tto1 probe used in colony hybridization - bp 4390 to 5290

5'- CATCGCAGCA ACGGAGGCTT GCAAAGAATT GATATGGATG AAGAAGTTCT TAACTGAACT
TGGATTTTCG CAAGACGGTT ATCAGTTATT TTGTGATAGT CAAAGTGCTA TCCACCTTGC
GAAGAATGCC TCATTCCATT CCAGATCCAA ACATATTGAT GTGAGATATA ATTGGATCAG
GGATGTGTTG GAGAAGAAGA TGTTGCGGCT TGAAAAGATC CATACAGACG AAAATGGATC
GGACATGTTG ACCAAGACTT TACCGAAAGG GAAGTTTGAG TTCTGTAGAG AAGCTGCAGG
GATAGTGGAT CCACCATATA GTTGGAAGGG GGAGAATTGT TAGTTTTTCC AACAATTATG
GTGATGTATG GGAGAGGTAA GCAATGATGC AACTCTCCTT AGTGAGATAA GCAATGATGC
AATGAGTGGT AGGTGAGATG AGAATATTGC AAATTGATCC TTCTTGGTAG GTGAGATTTG
CACTTTTGGT CCCTATCTCA AAACTATAAA TACCCCCTTC CATTTCATTG TATAACACAC
CAAAAAATAT ATCAAAACTC AAGAAGAAAG AGTTTGAGAG GGAGAGAGAT ATAGTTCCTT
TAGGAATGTT TCCTAACAGG GGAGTGACAA AATAGTGAGT AGAAATACTA GTCGGGTATT
TTTCGGGAAA CACTTTTGTG TGCGCCACTA TTTTGGGTAG AGCTCAGGAA TTGTTGTACC
TCCAAATTAT TGAGGAAGTC TCTCTTTGTA TGCCTGCTAA ATGTTTTAGT GGAAGTTGGT
GTCGGATTTG TGGACGTAGC CTAAACGTTT TAGGTGAACC ACGTTAAATA TTGTGTCATT
TATTTTTGGT TTCGTTGATC ATTTATTTTA TTCCGCTGTG CAGTAGTGTT TAGTGCCACC
G - 3'

LIST OF OLIGONUCLEOTIDES USED IN THIS WORK

Oligo ID	Sequence
	Oligos with complementarity to *Tto1*
231A-T4390-4411	CAT CGC AGC AAC GGA GGC TTG C
234A-T189-165	GAA ATG GAA GGG GGT ATT TAT AGT T
291A-T2314-2334	GGT GGA AAG AGA GAC TGG TAA
292A-T2653-2634	CCC GTA ATT GAT CAT AAG AGA
427A-T4697-4677	CCA CTA TCC CTG CAG CTT CTC
755A-T969-991	AGC TCG AAG AGT TGT ATG CCT CT
756A-T1109-1086	CCG ACA ACT GGT CGA CAA TCC CTT
782A-T5233up	CGG CCC CCA ACT AGT GGA ATA AAT GAC ACA ATA TTT AAC GT
783A-T5119up	CGG CCC CCA ACT AGT GGA TAA TTT GGA GGT ACA ACA ATT CC
784A-T5022up	CGG CCC CCA ACT AGT GGA TTT TGT CAC TCC CCT GTT AGG AA
786A-T4922up	CGG CCC CCA ACT AGT GGA TAC AAT GAA ATG GAA GGG GGT ATT
787A-T4900up	CGG CCC CCA ACT AGT GGA TTT ATA GTT TTG AGA TAG GGA CC
872A-KpnIntdn	GCC CAA GCT TCC AGG CGA GCA TCT CCC AAC AAG TT
873A-KpnIntup	GCC CGG TAC CTG TGG AGA ATA TCT GGT AGA CTG TT
887A-T2055dn	GCG GCT AGC CTA AAC CAG ATC CAG TTG AA
888A-T2434up	GCG AAG CTT AGG AGG TGT GAA CTG ATG TC
889A-T2243dn	GCG GCT AGC CGA AAG ACA TGG GTA TAC AC
890A-T2629up	GCG AAG CTT TCT TCC TAA CCA AAT CTT TTC TGG A
891A-T2459dn	GCG GCT AGC GCT GCG AGG ATG AAC CGT AC
892A-T2804up	GCG AAG CTT ACT CCT GAC GAG CTT CTT TT
912A-H3Intdn	GCC CAA GCT TCC AGG CGA GCA TCT CCC AAC AAG TT
913A-H3Intup	GCC CGG TAC CTG TGG AGA ATA TCT GGT AGA CTG TT
931A-T2044-2019	GGC GTT CTT CTT TAC CAA ACG CGC CAT T
932A-T3038-3065	CAA CCA CCA ATC CTC AAT AAC CCT CCT T
933A-TSwa4732-4705	CGG ATT TAA ATC TAA CAA TTC TCC CCC TTC CAA CTA TAT
953A-2479-2501	AGG GCG TAC GTT GAT CGA GAG AAC CAG ATG TTT
962A-T3478-3501	GAT GGA TGT AAA AAC CGC CTT CCT
1045-T4626-4648	GAT CGG ACA TGT TGA CCA AGA CT
1072-T1162-1185	ACT CCC AGA GTC ATG GGA AAC CTT
1073-T3808-3830	CAA GTT CTT CGC CAT GAA AGA CT
1074-T4111-4134	CGC TGT AGG AGT GGT AAG TAG ATT
1077-T1838-1863	CCT TCC ATA ATG GCC AAT GGA AGC T

1078-T49-71	GAT GCA ACT CTC CTT AGT GAG AT
1119-T574-548up	TGT TAG GAT CCG GTG GCA CTA AAC ACT
1233-T4674-4700	GTA GAG AAG CTG CAG GGA TAG TGG ATC
T2262-2283	CGC TGA AGA CCA AGG ATC AAG T
T4494-4517	GTG CTA TCC ACC TTG CGA AGA ATG C

Oligos used in barley work

1113-nosTBarleydn	CCC CCG GCG GCC GAT CGT TCA AAC ATT TGG CAA TAA AGT T
1114-nosTBarleyup	CCC CCG GAA GCT TCC CGG GCG ATC TAG TAA CAT AGA TGA CAC
1131-Mubpdn	CCG GAC TAG TAA GCT TGC GGC CGC TGC AGT GCA GCG TGA CCC
1132- Mubpup	GGC CGT CGA CTG CAG AAG TAA CAC CAA ACA ACA
1133-Mubseqdn1	CGA GTA GAT AAT GCC AGC CTG TT
1134-Mubseqdn2	GGA TGC GAC CTG TAC GTC AGA
1135-CYC177Nhedn	GAC ACC CTC ATC AGT GCC AAC AT
1136-Mubpintrondn	CCG GAC TAG TCC TAG GCC GCT CGA GTC AAG GTA CGC CGC TCG TCC TCC
1137-Mubpintornup	GGC CGT CGA CAA GAG GGT GTG GAG GGG GTG TCT
1344-ubp5nosplice	TTG TCG AGT CAA GGT ACG CCG CT
1345-ubp5splicedn	TTG TCG AGT CAA GTC GAG AAT ACC
1365-SCE113dn2	AGG GAC TGA TTG GGA AGG TGG AAA
1366-SCE113up2	CTC TCA GAC CAA TGC AGG ATA CTG CT

Oligos for sequencing iTto1 insertoin flanking sites

1254-At3g14480dn1	ATG TCG GAT TTG GTT ATT GCA CTC
1255-At3g14480up1	TCC ACA ACC GCC GCC TCC GCA ACC
1256-At2g26420up1	CCC TCG AGG CAC CTA TTT GGC ACC T
1257-At2g26420dn1	TTT GTC CAC CGG CAG CTT GCA CAT

Oligos used in RNA work

261A-dTclamp	CGG ACG CTC AGC CAG GTT TTT TTT TTT TTT TTT TTT TTT TTT
262A-clamp	CGG ACG CTC AGC CAG GTT T
AtUBC9TGAdn	TCC CCC GGG AGA TCT AGG ATG GCA TCG AAA CGG ATT TTG AAG
AtUBC9TGAup	GGG GTA CCA GAT CTC AGC CCA TGG CAT ACT TTT GGG T

Oligos with complementarity to *pER8* backbone

263A-RBCSterm	GAT TCT GGT GTG TGG GCA ATG AAA CT

278A-pNOSdn	GGG TTT CTG GAG TTT AAT GAC
780A-pERTtermup	GGTGCGGGCCTCTTCGCTTAT
781A-SwaIPfltermdn	CGG CCC ATT TAA ATC GCC CAC TAG TTG GTC GAT CCA GGC CTC CC
818A-BstBlinker	ATG CCC TTC GAA GGG CAT
819A-BstBRBCup	ATG CCC TTC GAA CAG CTG GCG AAA GGG GGA TGT GCT
849A-HPTup	GAA CAT CGC CTC GCT CCA GT
1120-RBpER8in	GCA GGA TAT CGT GGA TCC AAG CTT
1121-RBpER8out	TCA GGA TCT CTA GCC AGG CAC ATT

Oligos with complementarity to *pACYC177* backbone

987A-H3XAarPtop	AGC TTC TCG AGC CTA GGG ACC TGC CGG TGT TCC CGG GCA GCT GAA C
988A-H3XAarPbot	CAG CTG CCC GGG AAC ACC GGC AGG TGC CTA GGC TCG AGA
989A-CYC177seq	GAT TCC GGT ATC GGT CTG CGA TT
1079-CYC177dradn	GCC ACG TTG TGT CTC AAA ATC TC

Bibliography

A

Almeida, R., and Allshire, R.C. (2005). RNA silencing and genome regulation. Trends Cell Biol 15, 251-258.

Andersen, S.U., Cvitanich, C., Hougaard, B.K., Roussis, A., Gronlund, M., Jensen, D.B., Frokjaer, L.A., and Jensen, E.O. (2003). The glucocorticoid-inducible GVG system causes severe growth defects in both root and shoot of the model legume Lotus japonicus. Mol Plant Microbe Interact *16*, 1069-1076.

Arnold, F.H. (2001). Combinatorial and computational challenges for biocatalyst design. Nature *409*, 253-257.

Azpiroz-Leehan, R., and Feldmann, K.A. (1997). T-DNA insertion mutagenesis in Arabidopsis: going back and forth. Trends Genet *13*, 152-156.

B

Bancroft, I., and Dean, C. (1993). Transposition pattern of the maize element Ds in Arabidopsis thaliana. Genetics *134*, 1221-1229.

Bancroft, I., Jones, J.D., and Dean, C. (1993). Heterologous transposon tagging of the DRL1 locus in Arabidopsis. Plant Cell *5*, 631-638.

Basu, V.P., Song, M., Gao, L., Rigby, S.T., Hanson, M.N., and Bambara, R.A. (2008). Strand transfer events during HIV-1 reverse transcription. Virus Res *134*, 19-38.

Baulcombe, D. (2004). RNA silencing in plants. Nature *431*, 356-363.

Becker, J.D. (1989). Binary vectors which allow the exchange of plant selectable markers and reporter genes. Nucleic Acids Research, *18*, 203.

Beauregard, A., Curcio, M.J., and Belfort, M. (2008). The take and give between retrotransposable elements and their hosts. Annu Rev Genet *42*, 587-617.

Benner, S., A. (1987). Redesigning life. Organic chemistry and the evolution of protein. Chimia *42*, 142-148.

Benner, S.A. (2003). Act natural. Nature *421*, 118.

Benner, S.A., and Sismour, A.M. (2005). Synthetic biology. Nat Rev Genet *6*, 533-543.

Bennetzen, J.L. (2000). Transposable element contributions to plant gene and genome evolution. Plant Mol Biol *42*, 251-269.

Bernhart, S.H., Tafer, H., Muckstein, U., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2006). Partition function and base pairing probabilities of RNA heterodimers. Algorithms Mol Biol *1*, 3.

Boeke, J., D., Stoye, J., P. (1997). Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In: Coffin, J, M, Hughes,, S, H, Temin, H (eds) Retroviruses Cold Spring Harbor Laboratory Press, Plainview, 335-435.

Böhmdorfer, G. (2005) Functional analysis of the retrotransposon Tto1. PhD thesis.

Böhmdorfer, G., Hofacker, I.L., Garber, K., Jelenic, S., Nizhynska, V., Hirochika, H., Stadler, P.F., and Bachmair, A. (2005). Unorthodox mRNA start site to extend the highly structured leader of retrotransposon Tto1 mRNA increases transposition rate. RNA *11*, 1181-1191.

Böhmdorfer, G., Luxa, K., Frosch, A., Garber, K., Tramontano, A., Jelenic, S., Weber, M., and Bachmair, A. (2008). Virus-like particle formation and translational start site choice of the plant retrotransposon Tto1. Virology *373*, 437-446.

Böhmdorfer, G., Tramontano, A., Luxa, K., and Bachmair, A. (2010). A synthetic biology approach allows inducible retrotransposition in whole plants. Syst Synth Biol *4*, 133-138.

Brookfield, J.F. (2005). The ecology of the genome - mobile DNA elements and their hosts. Nat Rev Genet *6*, 128-136.

Brule, F., Bec, G., Keith, G., Le Grice, S.F., Roques, B.P., Ehresmann, B., Ehresmann, C., and Marquet, R. (2000). In vitro evidence for the interaction of tRNA(3)(Lys) with U3 during the first strand transfer of HIV-1 reverse transcription. Nucleic Acids Res *28*, 634-640.

Bucheton, A. (1995). The relationship between the flamenco gene and gypsy in Drosophila:

how to tame a retrovirus. Trends Genet *11*, 349-353.

C

Cameron, J.R., Loh, E.Y., and Davis, R.W. (1979). Evidence for transposition of dispersed repetitive DNA families in yeast. Cell *16*, 739-751.

Capy, P. (2005). Classification and nomenclature of retrotransposable elements. Cytogenet Genome Res *110*, 457-461.

Casacuberta, J.M., and Santiago, N. (2003). Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. Gene *311*, 1-11.

Castle, L.A., Errampalli, D., Atherton, T.L., Franzmann, L.H., Yoon, E.S., and Meinke, D.W. (1993). Genetic and molecular characterization of embryonic mutants identified following seed transformation in Arabidopsis. Mol Gen Genet *241*, 504-514.

Chang, K.Y., and Tinoco, I., Jr. (1994). Characterization of a "kissing" hairpin complex derived from the human immunodeficiency virus genome. Proc Natl Acad Sci U S A *91*, 8705-8709.

Cheng, C., Daigen, M., and Hirochika, H. (2006). Epigenetic regulation of the rice retrotransposon Tos17. Mol Genet Genomics *276*, 378-390.

Cheng, X., Zhang, D., Cheng, Z., Keller, B., and Ling, H.Q. (2009). A new family of Ty1-copia-like retrotransposons originated in the tomato genome by a recent horizontal transfer event. Genetics *181*, 1183-1193.

Cheng, X., Wen, J., Tadege, M., Ratet, P., and Mysore, K.S. (2011). Reverse genetics in medicago truncatula using Tnt1 insertion mutants. Methods Mol Biol *678*, 179-190.

Chopra P. and Kamma A. (2006). Engineering life through Syntethic Biology. In Silico Biology *6*, 0038.

Coruzzi, G., Broglie, R., Edwards, C., and Chua, N.H. (1984). Tissue-specific and light-regulated expression of a pea nuclear gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase. EMBO J *3*, 1671-1679.

Craft, J., Samalova, M., Baroux, C., Townley, H., Martinez, A., Jepson, I., Tsiantis, M., and Moore, I. (2005). New pOp/LhG4 vectors for stringent glucocorticoid-dependent transgene expression in Arabidopsis. Plant J *41*, 899-918.

Curcio, M.J., and Derbyshire, K.M. (2003). The outs and ins of transposition: from mu to kangaroo. Nat Rev Mol Cell Biol *4*, 865-877.

D

Dai, J., Xie, W., Brady, T.L., Gao, J., and Voytas, D.F. (2007). Phosphorylation regulates integration of the yeast Ty5 retrotransposon into heterochromatin. Mol Cell *27*, 289-299.

Davidson, E.H., and Britten, R.J. (1979). Regulation of gene expression: possible role of repetitive sequences. Science *204*, 1052-1059.

Devos, K.M., Brown, J.K., and Bennetzen, J.L. (2002). Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. Genome Res *12*, 1075-1079.

Ding, Y., Wang, X., Su, L., Zhai, J., Cao, S., Zhang, D., Liu, C., Bi, Y., Qian, Q., Cheng, Z.*, et al.* (2007). SDG714, a histone H3K9 methyltransferase, is involved in Tos17 DNA methylation and transposition in rice. Plant Cell *19*, 9-22.

Di Nocera, P.P., and Dawid, I.B. (1983). Interdigitated arrangement of two oligo(A)-terminated DNA sequences in Drosophila. Nucleic Acids Res *11*, 5475-5482.

Doolittle, W.F., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature *284*, 601-603.

Duval-Valentin, G., Marty-Cointin, B., and Chandler, M. (2004). Requirement of IS911 replication before integration defines a new bacterial transposition pathway. EMBO J *23*, 3897-3906.

E

Edwards, K.J., Lee, M., Avramova, Z.*, et al.* (1996). Nested retrotransposons in the intergenic regions of the maize genome. Science *274*, 765-768.

Evgen'ev, M.B., Zelentsova, H., Shostak, N., Kozitsina, M., Barskyi, V., Lankenau, D.H., and Corces, V.G. (1997). Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in Drosophila virilis. Proc Natl Acad Sci U S A *94*, 196-201.

Evgen'ev, M.B., and Arkhipova, I.R. (2005). Penelope-like elements–a new class of retroelements: distribution, function and possible evolutionary significance. Cytogenet Genome Res *110*, 510-521.

F

FAOSTAT (2007). Food and Agriculture Organization of the United Nations. http://faostatfaoorg/.

Feschotte, C., Jiang, N., and Wessler, S.R. (2002). Plant transposable elements: where genetics meets genomics. Nat Rev Genet *3*, 329-341.

Feschotte, C., Swamy, L., and Wessler, S.R. (2003). Genome-wide analysis of mariner-like transposable elements in rice reveals complex relationships with stowaway miniature inverted repeat transposable elements (MITEs). Genetics *163*, 747-758.

Finnegan, D.J. (1989). Eukaryotic transposable elements and genome evolution. Trends Genet *5*, 103-107.

Frankel, A.D., and Young, J.A. (1998). HIV-1: fifteen proteins and an RNA. Annu Rev Biochem *67*, 1-25.

G

Gabus, C., Ficheux, D., Rau, M., Keith, G., Sandmeyer, S., and Darlix, J.L. (1998). The yeast Ty3 retrotransposon contains a 5'-3' bipartite primer-binding site and encodes nucleocapsid protein NCp9 functionally homologous to HIV-1 NCp7. EMBO J *17*, 4873-4880.

Gai, X., and Voytas, D.F. (1998). A single amino acid change in the yeast retrotransposon Ty5 abolishes targeting to silent chromatin. Mol Cell *1*, 1051-1055.

Gibbs, W., W. (2004). Synthetic life. Sci Am *290*, 74-81.

Goodwin, T.J., Butler, M.I., and Poulter, R.T. (2003). Cryptons: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi. Microbiology *149*, 3099-3109.

Grandbastien, M.A., Spielmann, A., and Caboche, M. (1989). Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature *337*, 376-380.

Grandbastien, M.A., and Caboche, M. (1995). RNA-mediated transposition of the tobacco retrotransposon Tnt1 in Arabidopsis thaliana. EMBO J *14*, 2364-2373.

Grandbastien M., A. (1998). Activation of plant retrotransposons under stress. Trends Plant Sci *3*, 181-187.

Greenblatt, I.M., and Brink, R.A. (1962). Twin Mutations in Medium Variegated Pericarp Maize. Genetics *47*, 489-501.

H

Hays, L.E., Zodrow, D.M., Yates, J.E., Deffebach, M.E., Jacoby, D.B., Olson, S.B., Pankow, J.F., and Bagby, G.C. (2008). Cigarette smoke induces genetic instability in airway epithelial cells by suppressing FANCD2 expression. Br J Cancer *98*, 1653-1661.

Hilbricht, T., Varotto, S., Sgaramella, V., Bartels, D., Salamini, F., and Furini, A. (2008). Retrotransposons and siRNA have a role in the evolution of desiccation tolerance leading to resurrection of the plant Craterostigma plantagineum. New Phytol *179*, 877-887.

Hirochika, H. (1992). [Retroposons of plants]. Tanpakushitsu Kakusan Koso *37*, 1060-1066.

Hirochika, H. (1993). Activation of tobacco retrotransposons during tissue culture. EMBO J *12*, 2521-2528.

Hirochika, H., and Hirochika, R. (1993). Ty1-copia group retrotransposons as ubiquitous components of plant genomes. Jpn J Genet *68*, 35-46.

Hirochika, H., Otsuki, H., Yoshikawa, M., Otsuki, Y., Sugimoto, K., and Takeda, S. (1996a). Autonomous transposition of the tobacco retrotransposon Tto1 in rice. Plant Cell *8*, 725-734.

Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H., and Kanda, M. (1996b). Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci U S A *93*, 7783-7788.

Hirochika, H. (1997). Retrotransposons of rice: their regulation and use for genome analysis. Plant Mol Biol *35*, 231-240.

Hirochika, H., Okamoto, H., and Kakutani, T. (2000). Silencing of retrotransposons in arabidopsis and reactivation by the ddm1 mutation. Plant Cell *12*, 357-369.

Hirochika, H. (2001). Contribution of the Tos17 retrotransposon to rice functional genomics. Curr Opin Plant Biol *4*, 118-122.

Houzet, L., Battini, J.L., Bernard, E., Thibert, V., and Mougel, M. (2003). A new retroelement constituted by a natural alternatively spliced RNA of murine replication-competent retroviruses. EMBO J *22*, 4866-4875.

I

Igarashi, K., Hagihara, H. & Ito, S. (2003). Protein engineering of detergent α-amylases. Trends Glycosci Glycotechnol *15*.

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

J

Jensen, S., Gassama, M.P., and Heidmann, T. (1999). Taming of transposable elements by homology-dependent gene silencing. Nat Genet *21*, 209-212.

Jewell, N.A., and Mansky, L.M. (2000). In the beginning: genome recognition, RNA encapsidation and the initiation of complex retrovirus assembly. J Gen Virol *81*, 1889-1899.

Jiang, N., Feschotte, C., Zhang, X., and Wessler, S.R. (2004). Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). Curr Opin Plant Biol *7*, 115-119.

Jordan IK, M.W. (2009). Genome defense against transposable elements and the origin of regulatory RNA. In: Lankenau D-H, Volff J-N (eds) Transposons and the dynamic genome. Springer, Berlin., 77-94.

Jorgensen, R.A. (1995). Cosuppression, flower color patterns, and metastable gene expression States. Science *268*, 686-691.

K

Kang, H.G., Fang, Y., and Singh, K.B. (1999). A glucocorticoid-inducible transcription system causes severe growth defects in Arabidopsis and induces defense-related genes. Plant J *20*, 127-133.

Kapitonov, V.V., and Jurka, J. (2006). Self-synthesizing DNA transposons in eukaryotes. Proc Natl Acad Sci U S A *103*, 4540-4545.

Katz, R.A., and Skalka, A.M. (1994). The retroviral enzymes. Annu Rev Biochem *63*, 133-173.

Kazazian, H.H., Jr. (2004). Mobile elements: drivers of genome evolution. Science *303*, 1626-1632.

Kleckner, N. (1977). Translocatable elements in procaryotes. Cell *May;11*, 11-23.

Kramerov, D.A., and Vassetzky, N.S. (2005). Short retroposons in eukaryotic genomes. Int Rev Cytol *247*, 165-221.

Krysan, P.J., Young, J.C., and Sussman, M.R. (1999). T-DNA as an insertional mutagen in Arabidopsis. Plant Cell *11*, 2283-2290.

Kumar, A., and Bennetzen, J.L. (1999). Plant retrotransposons. Annu Rev Genet *33*, 479-532.

L

Lai, J., Li, Y., Messing, J., and Dooner, H.K. (2005). Gene movement by Helitron transposons contributes to the haplotype variability of maize. Proc Natl Acad Sci U S A *102*, 9068-9073.

Leduc, S. (1912). La Biologie Synthétique A Poinat Editeur, Paris.

Lewinski, M.K., and Bushman, F.D. (2005). Retroviral DNA integration–mechanism and consequences. Adv Genet *55*, 147-181.

Lisch, D. (2009). Epigenetic regulation of transposable elements in plants. Annu Rev Plant Biol *60*, 43-66.

Luan, D.D., Korman, M.H., Jakubczak, J.L., and Eickbush, T.H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell *72*, 595-605.
Lucas, H., Feuerbach, F., Kunert, K.,

Luciw, P.A.a.L., N. J. (1992). Mechanisms of retrovirus replication. Chapter 5 in: The Retroviridae,
Volume 1 Editor Jay A Levy Plenum Press, NY

M

Maertens, G.N., Hare, S., and Cherepanov, P. (2010). The mechanism of retroviral integration from X-ray structures of its key intermediates. Nature *468*, 326-329.

Malik, HS, and Eickbush, TH. Phylogenetic Analysis of Ribonuclease H Domains Suggests a Late, Chimeric Origin of LTR Retrotransposable elements and Retroviruses. Genome Research 11, 1187 – 1197 (2001)

Mates, L., Chuah, M.K., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J.*, et al.* (2009). Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nat Genet *41*, 753-761.

Matzke, M.A., and Birchler, J.A. (2005). RNAi-mediated pathways in the nucleus. Nat Rev Genet *6*, 24-35.

Meyer, P., and Saedler, H. (1996). Homology-Dependent Gene Silencing in Plants. Annu Rev Plant Physiol Plant Mol Biol *47*, 23-48.

McClintock, B. (1950). The origin and behaviour of mutable loci in maize. Proc Natl Acad Sci U S A *36*, 344-355.

McClintock, B. (1953). Induction of Instability at Selected Loci in Maize. Genetics *38*, 579-599.

Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., and Kakutani, T. (2001). Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature *411*, 212-214.

Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K., and Hirochika, H. (2003). Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. Plant Cell *15*, 1771-1780.

Moore, I., Galweiler, L., Grosskopf, D., Schell, J., and Palme, K. (1998). A transcription activation system for regulated gene expression in transgenic plants. Proc Natl Acad Sci U S A *95*, 376-381.

Moreau-Mhiri C, M.J.-B., Audeon C, Ferault M, Grandbastien M-A,, and H, L. (1996). Regulation of expression of the tobacco Tnt1 retrotransposon in heterologous species following pathogenrelated stresses. Plant J *409-419*.

Morgante, M., Brunner, S., Pea, G., Fengler, K., Zuccolo, A., and Rafalski, A. (2005). Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet *37*, 997-1002.

Muckstein, U., Tafer, H., Hackermuller, J., Bernhart, S.H., Stadler, P.F., and Hofacker, I.L. (2006). Thermodynamics of RNA-RNA binding. Bioinformatics *22*, 1177-1182.

N - O

Nassif, N., Penney, J., Pal, S., Engels, W.R., and Gloor, G.B. (1994). Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol Cell Biol *14*, 1613-1625.

Neumann, P., Pozarkova, D., and Macas, J. (2003). Highly abundant pea LTR retrotransposon Ogre is constitutively transcribed and partially spliced. Plant Mol Biol *53*, 399-410.

Okamoto, H., and Hirochika, H. (2000). Efficient insertion mutagenesis of Arabidopsis by tissue culture-induced activation of the tobacco retrotransposon Tto1. Plant J *23*, 291-304.

Ouwerkerk, P.B., de Kam, R.J., Hoge, J.H., and Meijer, A.H. (2001). Glucocorticoid-inducible gene expression in rice. Planta *213*, 370-378.

P

Paillart, J.C., Marquet, R., Skripkin, E., Ehresmann, C., and Ehresmann, B. (1996). Dimerization of retroviral genomic RNAs: structural and functional implications. Biochimie *78*, 639-653.

Parinov, S., Sevugan, M., Ye, D., Yang, W.C., Kumaran, M., and Sundaresan, V. (1999). Analysis of flanking sequences from dissociation insertion lines: a database for reverse genetics in Arabidopsis. Plant Cell *11*, 2263-2270.

Pereira, V. (2004). Insertion bias and purifying selection of retrotransposons in the Arabidopsis thaliana genome. Genome Biol *5*, R79.

Pritham, E.J., Putliwala, T., and Feschotte, C. (2007). Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. Gene *390*, 3-17.

R

Rédei, G.P., and Koncz, C. (1992). Classical mutagenesis. In: Koncz C,Chua N-H, Schell J (eds) Methods in Arabidopsis research World Scientific, Singapore, 16-82.

Reuter, M., Chuma, S., Tanaka, T., Franz, T., Stark, A., and Pillai, R.S. (2009). Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. Nat Struct Mol Biol *16*, 639-646.

Rowold, D.J., and Herrera, R.J. (2000). Alu elements and the human genome. Genetica *108*, 57-72.

S

Sabot, F., and Schulman, A.H. (2006). Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. Heredity *97*, 381-388.

Samalova, M., Brzobohaty, B., and Moore, I. (2005). pOp6/LhGR: a stringently regulated and highly responsive dexamethasone-inducible gene expression system for tobacco. Plant J *41*, 919-935.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards KJ, Lee M, Avramova Z, Bennetzen JL. (1996). Science. Nov 1; *274* (5288):737-8.

SanMiguel, P.J., Ramakrishna, W., Bennetzen, J.L., Busso, C.S., and Dubcovsky, J. (2002). Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5A(m). Funct Integr Genomics *2*, 70-80.

Scholes, D.T., Banerjee, M., Bowen, B., and Curcio, M.J. (2001). Multiple regulators of Ty1 transposition in Saccharomyces cerevisiae have conserved roles in genome maintenance. Genetics *159*, 1449-1465.

Seelamgari, A., Maddukuri, A., Berro, R., de la Fuente, C., Kehn, K., Deng, L., Dadgar, S., Bottazzi, M.E., Ghedin, E., Pumfery, A.*, et al.* (2004). Role of viral regulatory and accessory proteins in HIV-1 replication. Front Biosci *9*, 2388-2413.

Segal, G., Song, R., and Messing, J. (2003). A new opaque variant of maize by a single dominant RNA-interference-inducing transgene. Genetics *165*, 387-397.

Shirasu, K., Schulman, A.H., Lahaye, T., and Schulze-Lefert, P. (2000). A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res *10*, 908-915.

Speulman, E., Metz, P.L., van Arkel, G., te Lintel Hekkert, B., Stiekema, W.J., and Pereira, A. (1999). A two-component enhancer-inhibitor transposon mutagenesis system for functional analysis of the Arabidopsis genome. Plant Cell *11*, 1853-1866.

Strobel, E., Dunsmuir, P., and Rubin, G.M. (1979). Polymorphisms in the chromosomal locations of elements of the 412, copia and 297 dispersed repeated gene families in Drosophila. Cell *17*, 429-439.

Sugimoto, K., Takeda, S., and Hirochika, H. (2000). MYB-related transcription factor NtMYB2 induced by wounding and elicitors is a regulator of the tobacco retrotransposon Tto1 and defense-related genes. Plant Cell *12*, 2511-2528.

Symer, D.E., and Boeke, J.D. (2010). An everlasting war dance between retrotransposons and their metazoan hosts In: Kurth, R, Bannert, N (Eds), Retroviruses Caister Academic Press, 1-33.

Szostak, J.W., Bartel, D. P. & Luisi, P. L. (2001). Synthesizing life. Nature *409*, 387–390.

Szybalski, W. (1974). In Vivo and in Vitro Initiation of Transcription. In: A Kohn and A Shatkay (Eds), Control of Gene Expression, pp 23–4, and Discussion pp 404–5 (Szybalski's concept of Synthetic Biology), 411–2, 415–7 New York: Plenum Press, 405.

T

Tabor, S., and Richardson, C.C. (1995). A single residue in DNA polymerases of the Escherichia coli DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. Proc Natl Acad Sci U S A *92*, 6339-6343.

Takeda, S., Sugimoto, K., Otsuki, H., and Hirochika, H. (1998). Transcriptional activation of the tobacco retrotransposon Tto1 by wounding and methyl jasmonate. Plant Mol Biol *36*, 365-376.

Takeda, S., Sugimoto, K., Otsuki, H., and Hirochika, H. (1999). A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in

responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. Plant J *18*, 383-393.

Teixeira, F.K., Heredia, F., Sarazin, A., Roudier, F., Boccara, M., Ciaudo, C., Cruaud, C., Poulain, J., Berdasco, M., Fraga, M.F., *et al.* (2009). A role for RNAi in the selective correction of DNA methylation defects. Science *323*, 1600-1604.

Tenaillon, M. I., Hollister, J. D., Gaut, B. S. (2010). A triptych of the evolution of plant transposable elements. Trends in Plant Science *15,* 471-478.

Terol, J., Castillo, M.C., Bargues, M., Perez-Alonso, M., and de Frutos, R. (2001). Structural and evolutionary analysis of the copia-like elements in the Arabidopsis thaliana genome. Mol Biol Evol *18*, 882-892.

Tijsterman, M., Ketting, R.F., and Plasterk, R.H. (2002). The genetics of RNA silencing. Annu Rev Genet *36*, 489-519.

Tissier, A.F., Marillonnet, S., Klimyuk, V., Patel, K., Torres, M.A., Murphy, G., and Jones, J.D. (1999). Multiple independent defective suppressor-mutator transposon insertions in Arabidopsis: a tool for functional genomics. Plant Cell *11*, 1841-1852.

Tramontano, A., Donath, A., Bernhart, S.H., Reiche, K., Böhmdorfer, G., Stadler, P.F., and Bachmair, A. Deletion analysis of the 3' long terminal repeat sequence of plant retrotransposon Tto1 identifies 125 base pairs redundancy as sufficient for first strand transfer. Virology.

V - W

Voytas, D.F., Cummings, M.P., Koniczny, A., Ausubel, F.M., and Rodermel, S.R. (1992). copia-like retrotransposons are ubiquitous among plants. Proc Natl Acad Sci U S A *89*, 7124-7128.

Wessler, S.R. (1996). Turned on by stress. Plant retrotransposons. Curr Biol *6*, 959-961.

Wicker, T., Stein, N., Albar, L., Feuillet, C., Schlagenhauf, E., and Keller, B. (2001). Analysis of a contiguous 211 kb sequence in diploid wheat (Triticum monococcum L.) reveals multiple mechanisms of genome evolution. Plant J *26*, 307-316.

Wicker, T., Zimmermann, W., Perovic, D., Paterson, A.H., Ganal, M., Graner, A., and Stein, N. (2005). A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley Hv-eIF4E locus: recombination, rearrangements and repeats. Plant J *41*, 184-194.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., *et al.* (2007). A unified classification system for eukaryotic transposable elements. Nat Rev Genet *8*, 973-982.

Wilhelm, M., and Wilhelm, F.X. (2006). Cooperation between reverse transcriptase and integrase during reverse transcription and formation of the preintegrative complex of Ty1. Eukaryot Cell *5*, 1760-1769.

Wisman, E., Hartmann, U., Sagasser, M., Baumann, E., Palme, K., Hahlbrock, K., Saedler, H., and Weisshaar, B. (1998). Knock-out mutants from an En-1 mutagenized Arabidopsis thaliana population generate phenylpropanoid biosynthesis phenotypes. Proc Natl Acad Sci U S A *95*, 12432-12437.

X - Y - Z

Xiong, Y., and Eickbush, T.H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J *9*, 3353-3362.

Yamazaki, M., Tsugawa, H., Miyao, A., Yano, M., Wu, J., Yamamoto, S., Matsumoto, T., Sasaki, T., and Hirochika, H. (2001). The rice retrotransposon Tos17 prefers low-copy-number sequences as integration targets. Mol Genet Genomics *265*, 336-344.

Zhang, X.Y. (2008). Perspective - The epigenetic landscape of plants. Science *320*, 489-492.

Zuo, J., Niu, Q.W., and Chua, N.H. (2000). Technical advance: An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. Plant J *24*, 265-273.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit –einschließlich Tabellen, Karten und Abbildungen – die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Andreas Bachmair und von Prof. Dr. George Coupland (Max-Planck-Institut für Pflanzenzüchtungsforschung) betreut worden.

Köln,
den 21. Februar 2011

Andrea Tramontano

CURRICULUM VITAE

- **NAME:** Andrea Tramontano
- **ANSCHRIFT:** Landstrasser Hauptsrasse, 34/2/29 - 1030 - Wien (A)
- **GEBURTSDATUM und -ORT:** 21.02.1981 - Avellino (I)
- **TELEFON:** +43 650 2489375, +43 1 4277 42810
- **E-mail:** **andrea.tramontano@univie.ac.at; tramonta@mpipz.mpg.de**

AUSBILDUNG

Okt. 2008-April 2011:	**Weiterführung der Doktorarbeit** in der Gruppe von Dr. A. Bachmair in den Max F. Perutz Laboratories, Universität Wien.
Juli 2006–Sept. 2009:	**Beginn der Doktorarbeit** am Max-Planck-Institut für Pflanzenzüchtungsforschung (MPIPZ, Köln) in der Arbeitsgruppe von Dr. A. Bachmair mit dem Thema: "Analyse des Retrotransposons *Tto1*".
21.03.2005:	**Master Diplom in Pflanzen Biotechnologie** an der Università degli Studi di Napoli "Federico II".
Okt. 2003 -März 2005:	**Diplomarbeit** mit dem Abschlußthema entitled: "Expression and purification in *E. coli* strains of two natural proteins with *in vitro* antimicrobial activity", unter der Betreuung von Prof. Rosanna Capparelli, an der Facoltà di Agraria, Portici.
Okt. 1999-März 2005:	**Studium** an der Università degli Studi di Napoli "Federico II".
J uli 1999:	**Abitiur**, am Liceo Classico "Vittorio Imbriani", Pomigliano D'arco, Napoli.

ARBEITSERFAHRUNGEN

Feb. 2010-Nov. 2010:	Er hat bei der Veranstaltung Vienna BioCenter PhD Symposium 2010 "The Origin Of Life" teilgenommen.
Jan. 2006-Apr. 2006:	Hat seine Arbeit über Antimikrobiale Peptide unter der Leitung von Prof. R. Capparelli und Prof. D. Iannelli, in der Dipartimento di Zootecnìa e Ispezione degli Alimenti, at Faculty of Agriculture, Portici, of Università degli Studi di Napoli "Federico II" weitergeführt.
Jan. 2004-Apr. 2006:	Hat an einem regional Projekt über die Rettung von Typischem Germoplasm, die zur Charakterisierung von *Prunus armeniaca* L. und *Prunus avium* L- Sorten, durch Molekulare Marker der DNA in Kampanien beitragen, teilgenommen.

PUBLIKATIONEN

- **Tramontano, A.**, Donath A., Bernhardt S. H., Reiche K., Böhmdorfer, G Stadler P. F., Bachmair A. Deletion analysis of the 3′ long terminal repeat sequence of plant retrotransposon Tto1 identifies 125 base pairs as sufficient for first strand transfer. Virology, in press.

- Böhmdorfer, G., **Tramontano, A**., Luxa, K., and Bachmair, A., 2010. A synthetic biology approach allows inducible retrotransposition in whole plants. Syst. Synthet. Biol., 2010 Jun; 4(2):133-8

- Böhmdorfer, G., Luxa, K., Frosch, A., Garber, K., **Tramontano, A**., Jelenic, S., Weber, M., Bachmair, A., (2008). Virus-like particle formation and translational start site choice of the plant retrotransposon Tto1. Virology 373, 437-446.

.

AUSGEWÄLTE PRÄSENTATIONEN

- **Tramontano, A**. and Bachmair, A.: Transposon tagging with engineered retrotransposon *Tto1*. Plant Gene Discovery Technologies 2011, Vienna. (Vortrag)

- **Tramontano, A**., Böhmdorfer, G., and Bachmair, A.: Biology and synthetic biology of retrotransposon *Tto1* in *Arabidopsis thaliana.* TNAM 2010 -7th Tri-National Arabidopsis Meeting, Salzburg. (Poster)

Acknowledgments

The last page is to express my gratitude to my dear boss, **Prof. Andreas Bachmair** for his support and encouragement all over my PhD experience, from Cologne to Vienna. Thanks to his special ability in sharing with me his enthusiasm and optimism in pursuing the objectives, I developed a focus and became more eager for scientific research. He always provided me with technical support, direction and became more of a mentor, than a professor.

Thanks to my dearest colleagues, **Karolin**, **Konstantin**, **Prabha** and **Rebecca** for having shared with me the alternating fortunes of the PhD student life, for understanding my daily tirades against everything and for their precious friendship.

Thanks to **Gudrun**, whose collaboration added considerably to my research experience in Vienna.

A very special thanks goes out to **Kerstin Luxa**, for the time and energy that she invested in supporting me and my work in Cologne, and for her "useful" practical advices in critical moments.

I would like to thank **Prof. Alexander Donath** at Institute of Bioinformatics University of Leipzig, **Dr. Stephan Bernhart**, **Prof. Peter Stadler** and **Prof. Ivo Hofacher** at Institute for Theoretical Chemistry and Structural Biology University of Vienna and **Dr. Kristin Reiche** at Fraunhofer Institute for Cell Therapy and Immunology for their precious contribution to the publication of my work. Equally I would like thank **Dr. G. Hensel** and **Dr. J. Kumlehn** at the Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK) in Gatersleben (Germany), for their work on barley.

Finally, I would like to extend my gratitude to all the people who contributed in making my experience abroad, first in Cologne and then in Vienna, unforgettable: **Matteo**, **Betina**, **Chiarina**, **Kerstin R.**, **Sara**, **Federica**, **Elisa** and **Francesca**.

Vorrei calorosamente rigraziare i miei amici di sempre, con i quali spero di continuare a condividere i miei e i loro successi, e **Mary**, per esserci sempre stata.

Dulcis in fundo, sento il bisogno di ringraziare **la Mia Famiglia** per il suo appoggio incondizionato e presente, capace di colmare ogni distanza.

Grazie