Interactions between a yeast retrotransposon and its host that influence target specificity

Jiquan Gao

Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/etd

Part of the Cell and Developmental Biology Commons, and the Genetics and Genomics Commons

Recommended Citation

Gao, Jiquan, "Interactions between a yeast retrotransposon and its host that influence target specificity" (2009). Graduate Theses and Dissertations, 10754.

http://lib.dr.iastate.edu/etd/10754

This Dissertation is brought to you for free and open access by the Graduate College at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Interactions between a yeast retrotransposon and its host that influence target specificity

by

Jiquan Gao

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Microbiology

Program of Study Committee:
Daniel Voytas, Major Professor
Allen Miller
Kristen Johansen
Michael Shogren-Knaak
Steve Whitham

Iowa State University
Ames, Iowa
2009

Copyright © Jiquan Gao, 2009. All rights reserved.
# TABLE OF CONTENTS

## ABSTRACT

### CHAPTER 1. GENERAL INTRODUCTION
- TRANSPOSABLE ELEMENTS: DEFINITION AND CLASSIFICATION 1
- INTERPLAY BETWEEN RETROTRANSPOSON AND HOST 3
- RETROTRANSPOSITION AND IMPACTS ON HOST GENOME 3
- THE REGULATION OF RETROTRANSPOSITION BY HOST CELLS 7
- RETROTRANSPOSITION IN RESPONSE TO ENVIRONMENTAL STRESS CONIFERS GENETIC VARIATION ON HOST CELLS 10
- YEAST RETROTRANSPOSON TY5 AS A MODEL TO STUDY THE RELATIONSHIPS AMONG RETROELEMENT, HOST, AND ENVIRONMENTAL STRESS 13
- DISSERTATION ORGANIZATION 15
- REFERENCES 17

## CHAPTER 2. PHOSPHORYLATION REGULATES INTEGRATION OF THE YEAST TY5 RETROTRANSPOSON INTO HETEROCHROMATIN
- ABSTRACT 22
- INTRODUCTION 23
- RESULTS 24
- DISCUSSION 34
- EXPERIMENTAL PROCEDURES 38
- REFERENCES 42

## CHAPTER 3. THE RIS1 UBIQUITIN LIGASE MEDIATES SIR4 TRUNOVER UPON INTERACTION WITH THE TARGETING DOMAIN OF THE YEAST RETROTRANSPOSON TY5
- ABSTRACT 45
- INTRODUCTION 46
- RESULTS 48
- DISCUSSION 57
- EXPERIMENTAL PROCEDURES 62
- ACKNOWLEDGEMENTS 65
- REFERENCES 65

## CHAPTER 4. DNA DAMAGE BLOCKS INTEGRATION OF THE YEAST RETROTRANSPOSON TY5
- ABSTRACT 69
- INTRODUCTION 70
- RESULTS 73
- DISCUSSION 80
- EXPERIMENTAL PROCEDURES 82
ACKNOWLEDGEMENTS
REFERENCES

CHAPTER 5. GENOME-WIDE ANALYSIS OF INTEGRATION SITES OF THE YEAST
RETROTRANSPOSON TY5
ABSTRACT
INTRODUCTION
RESULTS
DISCUSSION
EXPERIMENTAL PROCEDURES
REFERENCES

CHAPTER 6. GENERAL CONCLUSIONS
GENERAL DISCUSSION
REFERENCES

ACKNOWLEDGEMENTS
ABSTRACT

Transposable elements have great potential to restructure the genome of their hosts by mutating genes through insertion, imposing regulation on adjacent genes, and causing recombination-based genome rearrangements. The movement of transposable elements, therefore, could have two consequences to the host genome: under normal conditions, high mobility of transposable elements will be harmful to the integrity of the host genome, whereas under stress conditions, transposition could benefit host cells by creating genetic variations that offer selective advantages. McClintock hypothesized that the host cell and transposable elements have co-evolved diverse mechanisms to mutually benefit their survival.

The yeast retrotransposon, Ty5, has proven to be a great model to understand the integration site preference of retrotransposons. Ty5 preferentially integrates into heterochromatin at the telomeres and silent mating loci. It has been shown that Ty5 integration preference is directed by the interaction between a small domain of Ty5 integrase (the targeting domain, TD) and Sir4, a heterochromatic protein. In this thesis, we demonstrated that serine 1095 in TD is phosphorylated by the host cell and that this phosphorylation is essential for interacting with Sir4 and subsequently mediating Ty5 integration into heterochromatin. In the absence of phosphorylation, created for example by substituting serine 1095 with alanine, Ty5 becomes a potent mutagen. More than half of the insertions generated by phosphorylation-defective TD mutants occur in coding sequences. Therefore, the yeast cell actively controls Ty5's mutagenic potential through posttranslational modification, and this minimizes deleterious consequences of transposition. Furthermore, TD phosphorylation is reduced under stress conditions, specifically starvation for amino acids, nitrogen, or fermentable carbon, allowing the Ty5 integrations to occur in euchromatin. This result is consistent with McClintock's hypothesis that mobile elements reshape host genomes as an adaptive response to environmental stress.

The crosstalk between Ty5 and the host cell is further evidenced by the regulation of heterochromatic silencing by Ty5. It has been shown that the expression of Ty5 TD disrupts silencing at telomeres and HM loci, and this anti-silencing activity depends on TD’s ability to interact with Sir4. Single amino acid substitutions that abolish TD interaction with Sir4 no longer break transcriptional silencing. We showed that TD breaks silencing by mediating
Sir4 turnover, which is dependent on 1) Ubc4, an ubiquitin conjugating enzyme, 2) Ris1, a RING type ubiquitin ligase and 3) the proteasome pathway. We, for the first time, demonstrate that the RING finger domain of Ris1 has direct ubiquitination activity, and this activity requires the intact form of the RING finger. Interestingly, we found that a domain located in the C-terminus of Esc1, which is functionally equivalent to TD, also possesses Ubc4- and Ris1-dependent anti-silencing activity. Like TD, this Esc1 domain interacts with Sir4 and directs Ty5 integration when it is substituted for TD within integrase. Our results suggest that Ty5 has the potential to regulate the dynamics of heterochromatin in a fashion similar to Esc1, a host cellular protein that regulates heterochromatin under certain events such as cell cycle progression and stress.

We further assessed the regulation of Ty5 transposition by the cell cycle and DNA damage to understand the interrelationship between Ty5, the host cell and environmental conditions. We found that Ty5 integration is not inhibited in yeast cells arrested in G1 phase by the mating pheromone alpha factor, although transposition of Ty1 and Ty3 is repressed by alpha factor treatment. In addition, we demonstrated that Ty5 cDNA integration is restricted upon induction of DNA damage by HU and MMS, but that Ty5 cDNA recombination is not affected. By performing genetic analysis, we confirmed that depletion of Sml1, which results upon DNA damage, is the cause for the Ty5 integration deficiency. The mechanism by which Sml1 regulates Ty5 integration is under investigation.

When yeast cells lack Esc1 or Ris1, or experience DNA damage, Ty5 still appears to predominantly integrate into heterochromatin. These conclusions are based on the characterization of a handful of Ty5 integration sites. To explore the possibility that Ty5 integration patterns are modestly changed under these conditions, we developed an important new approach for studying global Ty5 integration patterns that uses linker mediated PCR followed by 454 sequencing and bioinformatic analysis. We sequenced more than 300,000 PCR amplicons and mapped 20,000 unique insertions and 85,000 multi-hit insertions onto the yeast genome. These insertions were derived from both haploid and diploid strains. Our mapping data demonstrate that Ty5 integrates throughout domains of heterochromatin at the telomeres, silent mating loci and rDNA. This integration pattern, however, is not random:
most insertions fall within intergenic regions. Furthermore, the rare Ty5 insertions that occur in euchromatin also occur in intergenic regions in both haploid and diploid strains, demonstrating that the bias is not due to selection. This distribution pattern supports the safe haven hypothesis, namely that Ty5 is very-well adapted to life in the yeast genome because under normal growth conditions, it preferentially inserts into gene-poor region to minimize the potential deleterious effects on host genome integrity.
CHAPTER 1. GENERAL INTRODUCTION

TRANSPOSABLE ELEMENTS: DEFINITION AND CLASSIFICATION

Transposable elements (TE), or mobile elements, are genetic fragments that can move from one location in the chromosomes of their host cell to another (Biemont and Vieira, 2006; Kazazian, 2004). Over the course of evolution, TEs have arguably become the most successful gene families, in that TEs are ubiquitously distributed in nearly every organism, including bacteria, fungi, plants, animals, and humans. The presence and translocation of TEs have broad impacts on host genomes, such as causing insertional mutations, mediating chromosomal rearrangement, regulating gene expression, and producing gene duplications. Their significant impact on the host genome and great potential to become valuable research and therapeutic tools make TEs the target of intensive study.

Transposable elements can be divided into two classes: DNA transposons and retrotransposons (Kazazian, 2004). DNA transposons are prevalent in bacteria, and also found in many eukaryotic genomes, including those of insects, worms, and humans. DNA transposons encode transposase, which mediates their mobilization via a “cut and paste” mechanism; the DNA transposon is excised and inserted into a new genomic location through the transposon-encoded transposase. Transposases typically recognize the terminal inverted repeats (TIRs) flanking the transposon and enzymatically excise the transposon DNA away from its genomic context by breaking the phosphodiester backbone of DNA. The reverse enzymatic activity of transposase, namely phosphodiester bond formation, allows the transposon to insert into another DNA location.

Retrotransposons are more abundant in eukaryotes than in prokaryotes. They mobilize themselves through RNA intermediates, which are reverse-transcribed into cDNA and inserted into the genome by the retrotransposon-encoded integrase protein. Retrotransposons undergo duplicative transposition. As a result, their copy numbers are amplified after each transposition. For instance, the long interspersed nucleotide elements-1 (LINE1) retrotransposon constitutes 17% of the total human genome (Lander et al., 2001).
Retrotransposons can be classified into long-terminal repeat (LTR) retrotransposons and non-LTR retrotransposons based on whether they have long terminal direct repeat sequences at their ends. Non-LTR retrotransposons (e.g., LINE1) of mammals typically have two ORFs, one of which encodes a nucleic acid binding protein and the other an endonuclease and reverse transcriptase. The 3’ end of non-LTR retrotransposons is commonly polyadenylated. Non-LTR retrotransposons transpose through target primed reverse transcription (TPRT). To begin the retrotransposition, the element-encoded endonuclease nicks the target genomic DNA, leaving a 3’-OH. The 3’-OH then can prime reverse transcription using the non-LTR retrotransposon RNA as a template. The transposition is finished up by gap repair to fill the cleavage site, forming a target-site duplication.

![Fig1. Life cycle of LTR retroelements (retrotransposons and retroviruses). IN, integrase; RT, reverse transcriptase; VLP, virus-like particle; VP, virus particle. Blue triangles represent LTRs.]
LTR retrotransposons structurally resemble retroviruses, with the exception that they do not encode an env gene. Both contain gag and pol genes, which encode the viral coat protein (Gag) and reverse transcriptase (RT), protease, ribonuclease H, and integrase (IN). The env gene of the retroviruses encodes the envelope protein essential for extracellular transmission. The typical life cycle of LTR retrotransposons can be divided into four steps: transcription, particle formation, reverse transcription, and integration (Fig. 1).

LTR retrotransposons are first transcribed into mRNA by the host RNA pol II transcription machinery. The mRNA is exported into the cytoplasm, where it is translated into element proteins. The element proteins can be packaged together along with the mRNA to form virus-like particles (VLPs). Within the VLP, the mRNA is reverse transcribed to a linear, double-stranded cDNA by the element-encoded RT using a host tRNA as the primer. cDNA associates with element-encoded integrase to form a pre-integration complex (PIC), which is imported into the nucleus and mediates cDNA integration into the chromosome.

INTERPLAY BETWEEN RETROTRANSPOSON AND HOST

Retrotransposons make up a large fraction of the host genome; for example, they comprise about 50% of mammalian genomes (Biemont and Vieira, 2006). Retrotransposons’ large contribution to and movement in the host genome suggest that retrotransposons and their host cell must co-evolve mechanisms that mutually enable their survival. Four areas of study contribute to the understanding of interactions between retrotransposons and host cells and have drawn much attention of scientists: the mechanisms of retrotransposition, the regulation of retrotransposition, the biological function of retrotransposition, and the use of transposable elements as research tools.

Retrotransposition and impacts on host genome

Although they reside in the host cell and their proliferation completely relies on the host cell, retrotransposons can have broad impacts on genome dynamics of the host genome by mobilizing within it. The mobility of retrotransposons can influence the host genome in
many distinct ways, including (but not limited to) mutating genes, regulating gene expression, mediating chromosomal rearrangement, and influencing gene transduction and duplication (Feschotte and Pritham, 2007). Recent study has increasingly elucidated the impact of retrotransposition on host genome dynamics.

**Retrotransposons cause insertional mutations**

The most straightforward and direct effect of retrotransposition, when occurring at a coding sequence, is to generate mutants. Like DNA transposons, many retrotransposons have been successfully applied as a tool for making mutants for genetic research. The L1 elements, the most abundant autonomous non-LTR retrotransposons in humans, were developed as a useful tool for in vivo mutagenesis in mice (An et al., 2006).

**Retrotransposons regulate the transcription of adjacent genes**

It is of note that retrotransposons, as a source of regulatory elements, also are able to regulate gene expression in many indirect ways at both transcriptional and post-transcriptional levels. At the transcriptional level, a retrotransposon inserted upstream of a gene may cause multiple effects, including: (i) insertion of promoter sequences and alteration of the transcription start site; (ii) disruption of existing upstream cis-regulatory element(s); and (iii) introduction of a new cis element such as a transcription factor binding site to change the gene regulation. In addition, a retrotransposon inserted within an intron may form antisense RNA that will potentially interfere with sense transcription. Finally, a retrotransposon may be transcriptionally silenced and subsequently repress the transcription of adjacent gene(s).

At the post-transcriptional level, a TE inserted downstream of a gene may introduce an alternative polyadenylation site or the binding site for a microRNA. A retrotransposon inserted within an intron can change the normal splicing pattern of a pre-mRNA, conferring a variety of alternative splicing events (intron retention, exon skipping, etc.). A retrotransposon inserted within an intron containing cryptic splice sites may be incorporated as an alternative
exon (‘exonized’) to form a new protein when expressed. Expression of an unstable mRNA that has a premature stop codon introduced by a TE can cause the RNA to be degraded via the nonsense mediated decay (NMD) pathway.

Retrotransposons mediate chromosomal rearrangement

Since retrotransposons account for a large fraction of the host genome and each retrotransposon has multiple copies across the different chromosomes, they are potentially a driving force to cause homologous recombination-based rearrangement. The host’s homologous recombination machinery forms the basis of retrotransposon-mediated rearrangements, which could happen both within the same chromosome and between different chromosomes (Mieczkowski et al., 2006).

When two identical retrotransposons flanking a nonrepeat sequence are located in a chromosome in direct orientation, an unequal crossover will result in a coupled deletion-duplication. Alternatively, the internal fragment could be circularized and popped out as a result of a crossover between the two flanking retrotransposon repeats. Subsequently, the circular internal fragment could reintegrate into a sister chromatid containing identical retrotransposons at a non-allelic position, and then a duplication could be generated. When two identical retrotransposons flanking a nonrepeat sequence are located in a chromosome in inverted orientation, crossover between inverted retrotransposon repeats will generate inversion of the internal fragment located between the two retrotransposons. When two identical retrotransposons are located on two different chromosomes in the same orientation relative to the centromere, the crossover will generate a translocation. However, when two identical retrotransposons on two different chromosomes are in opposite orientation relative to the centromere, the crossover will generate one acentric and one dicentric chromosome.

All of the above events were observed in wild type cells under normal conditions. For instance, it was reported that a URA3 gene inserted near the MAT locus on one of the two copies of chromosome III in a diploid yeast strain was lost at a frequency of about 2x10^{-4} (Umezu et al., 2002). Accounting for 8% of the total URA3-loss in this strain, a variety of
chromosome rearrangements were detected: 1) a 94Kb deletion resulting from recombination between the MAT locus and HMR; 2) a 30Kb deletion resulting from recombination between two Ty elements on the right arm of III; 3) a deletion of sequence on the right arm and a duplication of sequence on the left arm resulting from recombination between a Ty element on the right arm and a Ty element on the left arm; 4) translocation between Ty elements on chromosome III and Ty elements on a non-homologous chromosome; and 5) a complicated rearrangement in which the internal portion of chromosome III was duplicated and a fragment from the left arm of chromosome III was transferred to the right arm of III, presumably resulting from the recombination between two inverted Ty elements on the right arm of chromosome III and a subsequent Ty-Ty recombination event in the resulting dicentric product.

Retrotransposons contribute to gene transduction

When retrotransposons are transcribed and incorporated into VLPs, in some cases the transcript of adjacent genes is also captured in the VLPs. Within VLPs, host gene transcripts are also reverse transcribed to form cDNA. The cDNA derived from the host transcript either becomes part of retrotransposons via recombination with retrotransposon cDNA, or replaces the retrotransposon cDNA in the VLP. Subsequently, the cDNA is introduced back into the nucleus through integrase-mediated trafficking or recombined with the host genes to cause intron loss or generate various forms of pseudogenes. Given the abundance of retrotransposons in eukaryotic genomes, this process is thought to be an evolutionarily potent mechanism for the duplication of host genes. There are several reports of retrotransposable elements mediated gene transduction. For instance, L1, a non-LTR retrotransposon in humans, is observed to mediate this process (Ejima and Yang, 2003; Xing et al., 2006). Ty1, an LTR retrotransposon in budding yeast, also showed the ability to form pseudogenes (Maxwell and Curcio, 2007).
The regulation of retrotransposition by host cells

Since the movement of retrotransposons can potentially mutate host genes or activate or repress adjacent genes under normal conditions, it is tightly controlled by host cells. To combat the potentially harmful effects of active retrotransposons, the host cell has evolved multiple regulatory mechanisms to suppress their activity.

Transcriptional regulation of retrotransposition

At the transcriptional level, the silencing of retrotransposons is often associated with DNA methylation, a hallmark of transcriptional repression. DNA methylation is presumably triggered by two sources: (1) direct interactions between identical or nearly identical retrotransposons in the genome; and (2) small RNA resulting from cleavage of retrotransposon transcripts via the RNAi pathway (Feschotte et al., 2002). So far, transcriptional silencing of retrotransposons has only been observed in plants through mutant analysis. In the absence of DNA methyltransferases Met1 and Cmt3, or the chromatin remodeling factor Ddm1 in Arabidopsis, retrotransposons such as Tar17, Ta3, and Athila were activated (Feschotte et al., 2002; Hirochika et al., 2000). Recently, a novel mechanism was described involving the recruitment of CENP-B homologues, heterochromatic proteins that repress Tf2, LTR retrotransposons in Schizosaccharomyces pombe, was described. It was proposed that during evolution the CENP-B homologues Abp1, Cbh1, and Cbh2 acquired the ability to bind the LTRs of Tf2 and subsequently recruited histone deacetylases Clr3 and Clr6 and thus resulted in Tf2 silencing (Cam et al., 2008; O'Donnell and Boeke, 2008).

At the post-transcriptional level, RNAi-mediated gene silencing was shown to suppress the mobility of plant and animal retrotransposons. RNAi-mediated gene silencing is triggered by double strand RNA (dsRNA) produced by the pairing of sense and antisense RNAs of identical retrotransposons. Formation of dsRNA might also involve RNA-dependent RNA polymerase (RdRP). The dsRNA is recognized and cleaved by the RNase III helicase Dicer (Dcr), generating small interfering RNAs (siRNAs) of 21–25 nucleotides. siRNAs may guide the RNA-induced silencing complex (RISC) to degrade the homologous
mRNAs. The siRNAs may also enter into the nucleus and mediate DNA methylation, possibly through RNA-DNA interaction. Mutations in genes in the RNAi pathway result in mobilization of families of transposable elements in *Drosophila melanogaster* (Aravin et al., 2004), in *Arabidopsis thaliana* (Zilberman et al., 2003), in *Caenorhabditis elegans* (Ketting et al., 1999; Tijsterman et al., 2002), and in humans (Chicas et al., 2004; Wu-Scharf et al., 2000). RNA-dependent gene silencing is another mechanism to suppress the transposition of the yeast Ty1 retrotransposon, even in a species in which traditional components of the RNAi pathway do not exist. This mechanism is triggered by the pairing between a Ty1 sense RNA transcript and a Ty1 antisense noncoding RNA transcript initiating within the first Ty1 ORF and continuing into the 5’ LTR. The resulting silencing of Ty1 is partially mediated by histone deacetylase and Set1-dependent histone methylation, showing an analogy with heterochromatin gene silencing in other species (Berretta et al., 2008). Retrotransposons’ RNA can also be destabilized by host factors such as the APOBEC3 family of cytidine deaminases. APOBEC3 was shown to restrict the replication of the HIV, L1 retrotransposon in humans and Ty1 retrotransposons in yeast by removing an amino group from cytidine in the RNA transcript, resulting in the product being degraded by Uracil-N-Glycosylase (UNG) (Dutko et al., 2005; Harris and Liddament, 2004; Kinomoto et al., 2007; Schumacher et al., 2005).

Retrotransposons’ cDNA has been shown to be a limiting factor for transposition efficiency. The host cell has developed one mechanism to regulate cDNA synthesis or stability once mRNA is made. For example, Ty1 is the most abundant retrotransposon in yeast, having 32 full-size elements in the genome (Kim et al., 1998). Ty1 mRNA accounts for as much as 0.1–0.8% of total RNA in yeast, and is highly stable. However, only one out of 14,000 mRNA molecules can make Ty1 cDNA (Curcio and Garfinkel, 1999). Global surveys for host regulators of Ty1 transposition screened out many host factors presumably suppressing and facilitating cDNA synthesis and stability. A large number of genes involved in yeast genome maintenance (e.g. DNA repair, recombination, and replication and telomere maintenance) negatively regulate Ty1 cDNA stability, while the *PMR1* gene (encoding a calcium/manganese exporter), the *DBR1* gene (encoding a debranching enzyme) and several components of the deadenylation-dependent mRNA degradation pathway stimulate the
synthesis of Ty1 cDNA, although the corresponding mechanism is not well defined. Interestingly, some host factors regulating Ty1 transposition also affect the mobility of Ty3, another LTR retrotransposon in budding yeast (Lesage and Todeschini, 2005).

The global survey for factors regulating Ty1 transposition also reveals many proteins involved in the regulation of Ty1 integration. The DNA end-binding protein Ku, involved in many processes that ensure genome integrity (e.g., chromatin assembly, double strand break repair via the end-joining pathway, telomere maintenance, and transcriptional gene silencing) was required for Ty1 integration, and Ty1 cDNA levels and integrity were unaffected in absence of Ku (Downs and Jackson, 1999). Similarly, human Ku protein was controversially suggested to be essential for HIV integration (Ariumi et al., 2005; Daniel et al., 1999).

Integration site preference into “safe havens”

The recent study of several LTR retrotransposons, including Ty1, Ty3 and Ty5 in S. cerevisiae and Tf1 in S. pombe, showed strong integration biases over the genome. Ty1 and Ty3 preferentially integrate into 750bp upstream and 1–3bp upstream of PolIII-transcribed genes, whereas Ty5 integration favors Sir complex–mediated silent chromatin. Tf1 shows strong preference to integrate upstream of PolII genes. Although LTR retrotransposon integration hot spots vary, integration into their favorite sites is commonly much less harmful than integration into active genes (which lentivirus strongly prefers). Based on this observation, scientists hypothesized that transcriptionally inactive regions serve as “safe havens” to bear the integration of retrotransposons in order to minimize the deleterious effects caused by insertions (Boeke and Devine, 1998). A recent study showing that the phosphorylation of Ty5 integrase by host cells controlled the Ty5 integration site choice corroborates the hypothesis that host cells might have a mechanism to actively control the integration of retrotransposons into “safe havens”.
McClintock hypothesized that transposable elements contribute to genetic variability and restructure the host genome in response to environmental stresses (McClintock, 1984). In this model, the mobility of transposable elements could be suppressed due to the potential deleterious effect of insertion under normal growth conditions. In response to stressful environmental change, the transposable elements could be activated, increasing the mutation rate, generating genetic variability, and reshaping the host genome to provide the diversity needed for a population to survive the stress.

Although this hypothesis was raised based on the observation of maize DNA transposons activation by stress, other studies of the activation of retrotransposons in various species under diverse conditions and the mutation of host factors regulated by stress corroborate McClintock’s hypothesis. For example, Tnt1, a retrotransposon in tobacco, and Tos17, a retrotransposon in rice, are more active in cell culture than in normal plant growth (Hirochika, 1993; Hirochika et al., 1996). Tnt1 hypertransposition is also triggered by pathogens or wounds. The promoter of TLC1.1, a retrotransposon in Solanum chilense, is activated by salicylic acid (SA), abscisic acid (ABA), methyl jasmonate (MeJA), hydrogen peroxide (H2O2) and the synthetic auxin 2,4-D. Heat shock, copper sulfate, and oxidative stress can activate the retrotransposon MAGGY resident in the plant pathogenic fungus Magnaporthe grisea. L1, a non-LTR retrotransposon in humans, is activated by the genotoxic stress by benzo(a)pyrene, a ubiquitous environmental carcinogen (Wessler, 1996).

Extensive study of the transposition of the yeast LTR retrotransposons Ty1, Ty3, and Ty5 largely strengthens the understanding of transposition activation by stress. Ty1, an LTR retrotransposon in S. cerevisiae, can be activated by many different stress conditions at the transcriptional and post-transcriptional levels. Exposure of yeast cells to DNA damaging agents, including UV light, γ irradiation, and chemicals such as 4-NQO, MMS and EMS, can activate Ty1 mobility by increasing Ty1 mRNA levels (Bradshaw and McEntee, 1989; Sacerdot et al., 2005; Staleva Staleva and Venkov, 2001). Yeast growth below 30 degrees Celsius can induce Ty1 transposition 100-fold higher than cultivation at 30 degrees (Paquin
One mutant surviving from lethal heat treatment is constitutively resistant to lethal heat shock and other stresses such as UV light and ethanol. This mutant has a Ty1 insertion upstream of the adenylate cyclase gene $CDC35$. This Ty1 insertion rendered lower adenylate cyclase activity, contributing to the multistress resistance over the wild type (Iida, 1988). Elevated Ty1 transposition due to increased mRNA levels was also observed under severe adenine starvation (Todeschini et al., 2005). A recent study showed that Ty1 insertion drives the transcription of adjacent coding and noncoding sequences using a promoter in the Ty1 5' LTR promoter that responds to adenine starvation. This series of studies demonstrate that Ty1 transpositional activation and simultaneous Ty1-driven transcription of adjacent coding and non-coding sequence results upon adenine starvation (Servant et al., 2008).

When starved for nitrogen, yeast cells undergo filamentous or invasive growth, allowing them to forage for nutrients in distant environments. Nitrogen starvation also stimulates Ty1 retrotransposition through activation of the filamentous growth kinase/Kss1 MAPK cascade. The Ste12 and Tec1 transcription factors, which are activated by the Kss1 MAPK cascade, can act on FRE (filamentous responsive elements) located within the Ty1 coding sequence to induce Ty1 transcription and thereby transposition. Since invasive/filamentous growth is a response to environmental stress, it was proposed that the activation of Ty1 transposition by this pathway may help survival of yeast under stress conditions by creating adaptive mutants (Morillon et al., 2000). In yeast, the invasive/filamentous signal transduction pathway shares several components with the pheromone response pathway. These two pathways are distinguished by their distinct MAPKs: Fus3 acts in the pheromone response pathway, and Kss1 acts in the invasive/filamentous pathway. Different from the active role of Kss1, Fus3 seems to play a role in controlling Ty1 dormancy by destabilizing Ty1 virus-like particle–associated proteins. In $fus3$ deletion strains, Ty1 integrase, reverse transcriptase, TyA protein, and cDNA are all increased, while Ty1 RNA levels, protein synthesis, and proteolytic processing are not altered relative to $FUS3$ strains. The suppression of Ty1 transposition in naive cells also requires basal levels of Fus3 activation. The enhanced activation of Fus3 when yeast cells are treated with the opposite mating pheromone inhibits Ty1 transposition (Xu and Boeke, 1991).
In contrast to their repressive role on Ty1 transposition, pheromones can activate Ty3 and Ty5 mobility. Pheromone responsive elements (PREs) located within the 5’ LTRs of Ty3 and Ty5 are responsible for transcriptional activation by the pheromone cascade (Bilanchone et al., 1993; Ke et al., 1997). Interestingly, Ty3 does not transpose when yeast cells are arrested in G1 by treatment with the mating pheromone $\alpha$-factor. The inhibition by pheromone is cell-cycle specific, because far1 mutants, which abolish cell-cycle arrest in G1, can abolish the Ty3 transposition blockage by pheromones (Menees and Sandmeyer, 1994). Furthermore, the block in Ty3 transposition can be reversed when yeast cells are released from G1 arrest. In addition, Ty3 can efficiently transpose into cells of the opposite mating type (Kinsey and Sandmeyer, 1995). Taken together, the data indicate that natural transposition of Ty3 is regulated temporally to occur during mating. Moreover, Ty1, Ty3, and Ty5 transposition are cell-type regulated, since diploid cells can repress their transposition due to the binding of an a1/a2 repressor complex, which is specific to diploid cells, to their transcriptional regulatory sequences (Bilanchone et al., 1993; Company and Errede, 1988; Errede et al., 1985; Fulton et al., 1988).

All of the activated transpositions of retrotransposons mentioned above were shown to result from increased transcription of retrotransposons, which was mediated via stress-responsive cis-regulatory elements in the retrotransposons. Hypertransposition of retrotransposons can also result from the increased cDNA levels that are induced under stress. For instance, Ty1 retrotransposition can be activated substantially in parallel with telomere erosion, and partially deactivated in survivors (Scholes et al., 2003). Experimental results suggest that Ty1 transposition is stimulated by substantially increased Ty1 cDNA levels under telomere stress.

Retrotransposons can also reshape the host genome through recombination between multiple copies at different loci induced by stress. When $HTA1-HTB1$, one of two gene pairs encoding histone 2A and 2B in $S.\,cerevisiae$, was deleted, two Ty1 elements flanking $HTA2-HTB2$ recombined together to form a new small circular chromosome. This chromosome amplified the copy number of $HTA2-HTB2$ to compensate for the change in histone dosage (Libuda and Winston, 2006). Similarly, Ty1 pairs also can be induced to mediate
interchromosomal translocations under reduced levels of the replicative α DNA polymerase (Lemoine et al., 2005). During long-term cultivation in glucose-limited media, Ty sequences were the major source of sequence alterations and chromosome rearrangements associated with adaptive changes (Adams and Oeller, 1986). Strikingly, some Ty associated breakpoints were close to the genes whose expression could confer selective advantage to glucose-deprived yeast populations (Dunham et al., 2002). Elegant work done in the Petes’ lab has shown that the chromosomal sites where Ty1 elements reside become double strand break hot spots and cause various Ty1 mediated chromosome rearrangements, including translocations and duplications upon treatment with UV irradiation and in cells with compromised levels of DNA polymerase (Lemoine et al., 2005).

In summary, environmental changes can regulate retrotransposition at both transcriptional and post-transcriptional levels, particularly in yeast. This regulated mobility is thought to enhance host survival in the face of stress. Furthermore, recent work on the yeast retrotransposon Ty5 showed posttranslational modification as one novel mechanism to regulate the integration pattern of Ty5 in response to stress.

YEAST RETROTANSPOSON TY5 AS A MODEL TO STUDY THE RELATIONSHIPS AMONG RETROELEMENTS, HOST, AND ENVIRONMENTAL STRESS

The genome of the yeast retrotransposon Ty5 encodes a single ORF that can be processed into Gag, integrase, reverse transcriptase, RnaseH, and protease. During the Ty5 life cycle, integrase binds cDNA synthesized by reverse transcriptase to form a preintegration complex (PIC). Integrase then directs PIC to transport cDNA into the nucleus and mediate its integration into the chromosome.

One of the most interesting characteristics of Ty5 transposition is that Ty5 preferentially integrates into Sir2/3/4–mediated silent chromatin, including telomeres and silent mating loci (HML and HMR): about 90% of de novo Ty5 transposition events occur in
a 3Kb window centered on an autonomously replicating consensus sequence (ACS) in the subtelomeric X repeat or HMR silencer (Zou et al., 1996a, 1996b; Zou and Voytas, 1997).

Ty5’s integration preference conforms to the safe haven hypothesis, namely that retrotransposons selectively integrate into gene-free regions such as heterochromatin to minimize the deleterious effect of insertions. Interestingly, besides Ty5, other yeast LTR retrotransposons such as Ty1 and Ty3 in S. cerevisiae and Tf1 in S. pombe also support this hypothesis. Ty1 and Ty3 preferentially integrate respectively 750bp upstream and 1–3bp upstream of PolIII transcribed genes, whereas Tf1 shows a strong bias toward integrating upstream of PolII genes. This hypothesis makes sense, considering that the yeast genome is highly compact with coding sequences and the random insertion of retrotransposons will be highly mutagenic for yeast cells.

To date, the tethering model provides the best explanation for the mechanism of yeast retrotransposon target specificity. In this model, retrotransposon integrase interacts with chromatin factors bound at the target region on the chromosome and the subsequent interaction directs integration to the target region. The study of Ty5 target specificity provides the most concrete support for this tethering model. A small motif (LDSSPP) at the C-terminus of Ty5 integrase, named the targeting domain (TD), has been shown to be responsible for determining Ty5 integration site choice. In addition, single amino acid substitutions within TD randomize Ty5 integration patterns (Gai and Voytas, 1998; Xie et al., 2001). TD directly interacts with a protein component of heterochromatin, silent information regulator 4 (Sir4) (Xie et al., 2001; Zhu et al., 2003), and the TD/Sir4 interaction is necessary and sufficient for target specificity (Zhu et al., 2003).

Sir4, a heterochromatic protein required for transcriptional repression in silent chromatin, serves as a scaffold protein to interact with many proteins including Esc1, a nuclear periphery protein that has the ability to partition DNA (Andrulis et al., 2002). Esc1 interaction with Sir4 is one of two redundant pathways anchoring silent chromatin to the nuclear periphery (Gartenberg et al., 2004). It has been shown that Sir complex–mediated silent chromatin associated with Esc1 and Ku70 comprises one specialized subnuclear
compartment localized at the nuclear periphery. Recent work discovered that Esc1 has a small domain resembling Ty5 TD that interacts with Sir4. A swapping experiment suggested that the Esc1 TD-like domain and Ty5 TD domain are functionally equivalent in that both direct Ty5 integration and partition plasmids when expressed in either Esc1 or Ty5 IN. Therefore, it is speculated that Ty5 may have acquired the targeting determinant by imitating the host factor Esc1 (Brady et al., 2008).

Given the functional equivalence between Ty5 TD and the Esc1 TD-like domain, it is likely that host cells might actively control Ty5 integration by imposing regulation on TD function. It is also highly possible that a particular growth condition or developmental stage will regulate TD function to change Ty5 integration specificity and make Ty5 act as an insertional mutagen to generate greater genetic variation in response to certain conditions. After the Ty5 preintegration complex is tethered at silent chromatin by recognizing Sir4, the subsequent Ty5 integration into the highly compacted regions may impose another means to regulate the maintenance of silent chromatin and the subnuclear compartment through cooperation with Esc1 or other Sir4-binding proteins. These possibilities are in part addressed in the studies described in this dissertation. The relevant results we hope will advance understanding of the relationships among retrotransposons, host cells, and environmental changes.

**DISSERTATION ORGANIZATION**

Chapter II was published in *Molecular Cell*. The paper was previewed by *Molecular Cell*, 27, 180-181, (Stress Management: How Cells Take Control of Their Transposons) and highlighted by the Faculty of 1000 as a must-read research article. In this paper, we show that phosphorylation at serine 1095 of Ty5 TD is essential for Ty5 targeted integration into silent chromatin, and TD S1095 phosphorylation was reduced under nutrient deprivation. The results suggest that host cells may regulate Ty5 integration specificity by phosphorylating TD in response to environmental stress. This study supports McClintock’s hypothesis that transposable elements can restructure the host genome in response to stress. The work in this chapter was carried out collaboratively by three former graduate students, Junbaio Dai,
Weiwu Xie, and Troy Brady, and me. Junbaio Dai and Weiwu Xie found that TD is phosphorylated at S1095 and this specific phosphorylation is essential for Ty5 targeting specificity. Using surface plasmon resonance (SPR) spectroscopy, Troy Brady confirmed that S1095 phosphorylation at TD is required for direct interaction with Sir4. Using an antibody that specifically recognizes S1095 phosphorylation, I showed that nutrient deprivation can regulate the phosphorylation level of TD. I also showed a change in Ty5 integration specificity (not included in the publication).

Chapter III is a manuscript prepared for submission to *Molecular and Cellular Biology*. The paper presents evidence that the Ty5 targeting domain and the Esc1 TD-like domain trigger Sir4 turnover. This turnover depends on the ubiquitination pathway involving Ris1, an ubiquitin ligase specific for sumo conjugates, and Ubc4, an ubiquitin-conjugating enzyme. The work in this study was done in collaboration with former graduate student Peter Feurst, who initiated the study and found that TD overexpression disrupted the silencing at heterochromatin and caused Sir4 turnover in a Ris1-dependent manner. I showed that Ris1 has a RING finger dependent auto-ubiquitination activity *in vitro* and cooperates with Ubc4 to degrade Sir4 through the proteasome pathway. In addition, I showed that the Esc1 TD-like domain disrupts silencing in a Ris1- and Ubc4-dependent manner, similar to the action of Ty5 TD. This study demonstrated one novel mechanism by which Ty5 can regulate heterochromatin through the ubiquitination pathway and supports the hypothesis that Ty5 imitates host factors to acquire recognition specificity and the capacity to regulate chromatin.

Chapter IV is a project I initiated. In this work, I show that DNA damage can reduce the integration efficiency of Ty5.

Chapter V is a manuscript prepared for submission to *Genetics*. The paper for the first time comprehensively presents data for the global distribution of Ty5 integration sites and provides a new method for analyzing the relationship between retrotransposon integration patterns and chromatin features.
REFERENCES


CHAPTER 2. PHOSPHORYLATION REGULATES INTEGRATION OF THE YEAST TY5 RETROTRANSPOSON INTO HETEROCHROMATIN

A paper published Molecular Cell

Junbiao Dai\textsuperscript{2,4}, Weiwu Xie\textsuperscript{2,3,4}, Troy L. Brady\textsuperscript{2}, Jiquan Gao\textsuperscript{2} and Daniel F. Voytas\textsuperscript{2}

ABSTRACT

The yeast Ty5 retrotransposon preferentially integrates into heterochromatin at the telomeres and silent mating loci. Target specificity is mediated by a small domain of Ty5 integrase (the targeting domain, TD), which interacts with the heterochromatin protein Sir4 and tethers the integration complex to target sites. Here we demonstrate that the TD is phosphorylated and that phosphorylation is required for interaction with Sir4. The yeast cell, therefore, through post-translational modification, controls Ty5’s mutagenic potential: when TD is phosphorylated, insertions occur in gene-poor heterochromatin, thereby minimizing deleterious consequences of transposition; however, in the absence of phosphorylation, Ty5 integrates throughout the genome, frequently causing mutations. TD phosphorylation is reduced when yeast cells are exposed to stress, specifically starvation for amino acids, nitrogen or fermentable carbon. This suggests that Ty5 target specificity changes in response to nutrient availability and is consistent with McClintock’s hypothesis that mobile elements restructure host genomes as an adaptive response to environmental challenge.

\textsuperscript{1} Reprinted with permission from Mol Cell, 2007, 27:289-299
\textsuperscript{2} Department of Genetics, Development and Cell Biology, 1035A Roy J. Carver Co-Laboratory, Iowa State University, Ames, IA 50011-365
\textsuperscript{3} Present address: Biological Science Division, 117 Tucker Hall, University of Missouri-Columbia, Columbia, MO 65211
\textsuperscript{4} These authors contributed equally to this work.
INTRODUCTION

Transposable elements are a powerful force that shape eukaryotic genomes (Craig et al., 2002). Insertions of transposable elements into new genomic loci can alter coding sequences, effect gene expression or initiate chromosomal rearrangements. A critical determinant of the impact of transposable elements on genomes is the site of element integration. Consequently, factors that affect target site choice influence both host genome integrity and evolution.

Eukaryotic genome sequencing efforts have revealed that transposable elements are not randomly distributed on chromosomes (Hua-Van et al., 2005). A variety of forces contribute to non-random distribution patterns, including selection against integration in gene-rich euchromatin or variation across chromosomes in rates of recombination resulting in differential accumulation or loss of mobile elements (Charlesworth and Langley, 1989). Increasingly, the presence of transposable elements in certain chromosomal regions is thought to be a consequence of targeted integration (Bushman, 2003). Target site choice for some elements, such as the R2 retrotransposons of insects, is mediated by recognition of specific DNA sequences (Yang et al., 1999). Other elements, like the budding and fission yeast retrotransposons, recognize specific types of chromatin or DNA-bound protein complexes (Bowen et al., 2003; Chalker and Sandmeyer, 1992; Devine and Boeke, 1996; Zou and Voytas, 1997). Transposable element targeting mechanisms, therefore, likely play a significant role in determining how mobile elements shape the genome.

An important model for understanding mechanisms of target site choice is the Saccharomyces retrotransposon Ty5. Like retroviruses and other retrotransposons, Ty5 encodes a reverse transcriptase, which copies element mRNA into cDNA, and an integrase (IN), which inserts the cDNA into the host genome. Ty5’s preferred target site is heterochromatin: about 90% of de novo Ty5 transposition events occur in heterochromatin at the telomeres or silent mating loci (HML and HMR) (Zou et al., 1996a; Zou et al., 1996b; Zou and Voytas, 1997). Ty5 IN determines target site choice through a small motif at the C-terminus (the targeting domain, TD), and single amino acid substitutions within TD
randomize Ty5 integration patterns (Gai and Voytas, 1998; Xie et al., 2001). TD interacts with a protein component of heterochromatin, namely silent information regulator 4 (Sir4) (Xie et al., 2001; Zhu et al., 2003). This interaction tethers the integration complex to target sites, and the TD/Sir4C interaction is necessary and sufficient for target specificity (Zhu et al., 2003).

Ty5’s target preference conforms to the ‘safe haven hypothesis’, which posits that transposition is favored in gene poor regions such as heterochromatin to minimize the deleterious effects of transposition (Boeke and Devine, 1998; Craig, 1997). Ty5’s target specificity, therefore is presumed to be an adaptation to life in the compact yeast genome. In this view, the yeast cell contributes passively to target site choice, and Ty5 uses Sir4 simply as a molecular marker for heterochromatin. In contrast to this passive view, we demonstrate here that the yeast cell actively participates in deciding where Ty5 integrates. The targeting domain of Ty5 integrase is phosphorylated by a host kinase, and phosphorylation is necessary for productive interactions with Sir4. In the absence of phosphorylation, Ty5 integrates throughout the genome, and therefore the host cell – through post-translational modification – determines how Ty5 affects the integrity of its genome.

RESULTS

The Ty5 TD/Sir4C interaction is blocked by phosphatase treatment

We define the Ty5 targeting domain (TD) as a span of nine amino acid residues at the C-terminus of Ty5 integrase (1092-LDSSPPNTS) (Xie et al., 2001). Genetic analyses determined that mutations in four residues (in bold), including two serines, result in loss of targeting to heterochromatin. When TD is expressed as a Gal4 DNA binding domain fusion protein (GBD-TD) and purified from yeast cells, the fusion protein interacts in vitro with the C-terminus of Sir4 (Sir4C) (Figure 1A). In contrast, when TD is fused to glutathione S-transferase (GST) and expressed and purified from E. coli, no binding to Sir4C is observed (Figure 1A). It is possible that the different protein scaffolds (i.e. GBD vs. GST) affect the ability of TD to interact with Sir4, although
Figure 1. Post-translational modification of the Ty5 targeting domain

(A) *In vitro* interaction between Sir4C and TD. Sir4C was expressed and labeled with $^{35}$S-methionine by coupled transcription and translation. GBD-TD (LDSSPPNTS) and GBD-td (LDLSPPNTS) were expressed in yeast and immunoaffinity purified with anti-GBD agarose beads. GST-TD and GST-td were expressed in *E. coli* and purified with glutathione agarose beads. As indicated by the top row of lanes, Sir4C interacts with TD purified from yeast but not from *E. coli*. The interaction between Sir4C and TD is abrogated when GBD-TD immunopurified from yeast is treated with l protein phosphatase before mixing with Sir4C. For each experiment, the bottom row of lanes indicates the amount of input protein used to test *in vitro* interactions. For GST and the GST fusion proteins, a Coomassie stained gel of the input proteins is shown. For GBD and the GBD-TD fusions, the amount of input protein was assessed by immunoblotting using an anti-GBD antibody.

(B) Amino acid substitutions in TD suggest serine phosphorylation in a tethered silencing assay. Silencing is established when GBD-TD is tethered to a silencing-impaired *HMR* locus with UAS$_s$ sites. Serial, 10-fold dilutions of cells were plated onto non-selective (SC-Ura) or selective (SC-Ura-Trp) media to measure silencing of the *TRP1* reporter gene at *HMR*. Amino acid sequences of TD carried by the fusion proteins are shown on the left of the figure. Mutated residues are in bold. At the bottom of the figure is an immunoblot performed with an anti-GBD antibody to show that expression of each fusion protein was equivalent.

(C) Target specificity of various Ty5 mutants as measured by the plasmid-based targeting assay (Gai and Voytas, 1998). The Y-axis indicates the percentage of target plasmids with an *HMR* locus that acquired a Ty5 insertion.
TD/Sir4C two-hybrid interactions were observed when TD was expressed in yeast as a fusion to the Gal4 activation domain (GAD) (Xie et al., 2001). Alternatively, TD purified from yeast cells could be post-translationally modified, allowing for its interaction with Sir4C. To test this latter hypothesis, and more specifically a role for phosphorylation, GBD-TD was purified from yeast and treated with λ phosphatase before mixing with Sir4C. Binding of phosphatase-treated GBD-TD to Sir4C was reduced to background levels (Figure 1A).

Amino acid substitutions suggest serine phosphorylation of TD

The diminished TD/Sir4C interaction upon phosphatase treatment suggested that one or more of the serines within TD was phosphorylated. We reasoned that the TD/Sir4C interaction might not be compromised if the serines are substituted with threonine or negatively charged amino acids that mimic phosphorylation (e.g. glutamic acid). To test the effect of such TD mutations on Sir4C interactions, an in vivo tethered silencing assay was employed. This assay measures the ability of GBD-TD to establish transcriptional silencing at a weakened HMR-E transcriptional silencer with Gal4 binding sites (UASg) (Chien et al., 1993; Xie et al., 2001). The effectiveness of GBD-TD in silencing is measured by the transcriptional status of a TRP1 reporter gene at HMR. GBD-TD silences TRP1 in this assay, presumably because it interacts with Sir4 and thereby establishes heterochromatin. When yeast cells were plated on control and selective media, wild type TD silenced TRP1 at least 100 times more effectively than the control (GBD only) and the S1094L mutant – the original mutant identified that abrogates TD/Sir4C interactions (Xie et al., 2001) (Figure 1B). The S1095E mutant was as effective in establishing silencing as the wild type, whereas neither S1094E nor the S1094E, S1095E double mutant silenced the reporter. The threonine substitutions at positions 1094 and 1095 had an intermediate phenotype, indicating that they cannot completely substitute for the native serines. Furthermore, substitution of residue 1095 with cysteine, which is structurally similar to serine but lacks a hydroxyl group that can be phosphorylated, failed to promote silencing.
To examine the effect of TD mutations on Ty5 target specificity, several of the mutants, including S1094E, S1095E and the double mutant, were moved into full-length Ty5 elements and tested in our plasmid-based targeted integration assay (Gai and Voytas, 1998) (Figure 1C). This assay monitors integration of Ty5 to a plasmid-borne *HMR* locus. We have previously shown that for wild type Ty5, approximately 8% of transposition events occur on the plasmid, whereas targeting-defective mutants (e.g. S1094L) transpose to the plasmid at frequencies ranging from 0.5 to 3% (Gai and Voytas, 1998; Xie et al., 2001). Consistent with the data from the tethered silencing assay, wild type Ty5 and the S1095E mutant targeted to the plasmid at frequencies approximating 6% (Figure 1C). In contrast, the S1094E mutation decreased target specificity approximately five fold. Similar decreases in targeting efficiencies were observed for the double mutant (S1094E, S1095E) as well as for mutants with S1095C and S1095A substitutions. The tethered silencing and targeted integration results indicate that both S1094 and S1095 are required for the TD/Sir4C interaction, and that productive interactions can only be maintained by substituting S1095 with a negatively charged glutamic acid residue, which mimics phosphorylation. This supports the hypothesis that TD is phosphorylated and implicates S1095 as the site of modification.

*TD is phosphorylated in vivo and phosphorylation is required for the TD/Sir4C interaction*

Mass spectrometry was used to test directly whether TD is phosphorylated and to determine the precise site of modification. Because native IN is insoluble and lacks cleavage sites necessary to generate peptides for analysis, we assessed the phosphorylation state of a TD fusion protein that was expressed and purified from yeast cells. The fusion protein has a six-histidine tag (His<sub>6</sub>) at the N- or C-terminus of TD to facilitate purification by nickel affinity gel. The His<sub>6</sub>-tagged TDs were tested to ensure that the tags did not compromise function. This was accomplished by fusing the tagged domains to GAD and testing for interactions with Sir4C by the two-hybrid assay. As shown in Figure 2A, the His<sub>6</sub>-tagged TDs interacted with Sir4C at levels significantly above the mutant TD fusion proteins (GAD-td and GAD-tdH<sub>6</sub>). The His<sub>6</sub>-tagged TD
Figure 2. TD is phosphorylated in vivo and phosphorylation is required for the interaction with Sir4C. (A) Two-hybrid assays measure interactions between Sir4C and His6-tagged TD. The addition of His6 at either the N- or C-terminus of TD does not interfere with TD’s ability to interact with Sir4C. Non selective medium is SC-Leu-Trp; selective medium is SC-Leu-Trp-His + 1mM 3-AT. (B) Deconvoluted mass spectrum of His6-tagged TD. Values above each peak are the observed mass of the most abundant isotopic species. The observed mass difference between unphosphorylated TD and monophosphorylated TD closely corresponds to the calculated monoisotopic mass of HPO₃, which is 79.996 Da. (C) Ion fragmentations from tandem MS of the ion fragment corresponding to monophosphorylated TD (m/z 1819.69 in (B)). More than 70% of the observed peaks could be assigned when S1095 was considered to be the phosphorylation site. b* and b** are produced by the loss of HPO₃ and one or two H₂O molecules from b ions. b₅ and y₅ ions are produced by the loss of one H₂O from b ions and y ions, respectively. (D) Surface plasmon resonance demonstrates an interaction between phosphorylated TD and Sir4C. A significant association is only detected when the second serine within TD is phosphorylated (pS1095) was then fused to GST, expressed in yeast and purified with a nickel affinity gel. The eluted protein was subjected to a second round of affinity purification using glutathione agarose. The nine-amino acid peptide along with either the N- or C-terminal His6-tag was cleaved from GST by factor Xa and again nickel affinity purified and concentrated. The purified 15-amino acid peptides were then analyzed by mass spectrometry.
The mass spectrum revealed both the 15-amino acid peptide and a species 80 Da greater in molecular mass, consistent with a single phosphorylation site within TD (Figure 2B). Tandem MS was performed with peptides with both N- and C-terminal His$_6$ tags to determine the site of phosphorylation. A representative spectrum is given in Figure 2C, and the MS/MS data are summarized in Table 1. Consistent with the genetic data, the MS/MS analyses revealed that S1095 was phosphorylated. No phosphorylation was detected at either S1094 or the downstream serine (T1099) or threonine (S1100). We also analyzed an S1095A TD variant and found a very weak but detectable signal, indicating that S1094 was phosphorylated in this mutant (data not shown). We believe that if S1094 is phosphorylated \textit{in vivo}, the levels are very low, or perhaps phosphorylation only occurs when S1095 (the preferred site) is mutated. Interestingly, S1095 is unambiguously phosphorylated in the S1094L mutant (data not shown). The S1094L substitution greatly reduces Ty5 target specificity and abrogates interactions with Sir4C (Gai and Voytas, 1998; Xie et al., 2001). The MS data indicate that these phenotypes are not due to lack of TD phosphorylation. Rather, the mutation may secondarily perturb the interaction with Sir4C or alternatively, reduce the extent to which S1095 is phosphorylated.

\begin{table}
\centering
\caption{MS/MS analysis of targeting domain phosphorylation.}
\begin{tabular}{lll}
\hline
Construct & Phosphorylated? & Site of Phosphorylation \\
\hline
LDSSPPNTS$_{H_6}$ & Yes & S1095 \\
LDLSSPPNTS$_{H_6}$ & Yes & S1095 \\
H$_6$LDSSPPNTS & Yes & S1095 \\
LDSAPPNTS$_{H_6}$ & No$^a$ & — \\
\hline
\end{tabular}
\end{table}

$^a$When S1095 is changed to alanine, there is a trace amount of phosphorylation on S1094 that is absent in all other constructs.
To test directly whether S1095 phosphorylation is required for the interaction with Sir4C, surface plasmon resonance (SPR) spectroscopy was used to monitor associations between GST-Sir4C and unphosphorylated or phosphorylated TD peptides. Three, twelve amino acid peptides were synthesized that span residues 1089 – 1100. One peptide was unphosphorylated and two had phosphoserines at either S1094 (pS1094) or S1095 (pS1095), making it possible to test the importance of both phosphorylation and the position of the phosphate group in the interaction with Sir4C. pS1094 and the unphosphorylated peptide exhibited very weak binding affinity, evidenced by the observation that binding of pS1094 to GST-Sir4C never reached saturation (Figure 2D). Phosphorylation of S1095, however, increased the affinity of TD for Sir4C over 12-fold. Binding of pS1095 to GST-Sir4C required high concentrations of peptide, indicating a very weak affinity. We speculate that high-affinity binding requires either a protein scaffold such as GBD or GST or additional residues in the IN C-terminus that extend beyond the 12 amino acid residues tested. Consistent with the latter hypothesis, two hybrid interactions with Sir4C are at least 100-fold stronger when a larger portion of the integrase C-terminus is used instead of TD (data not shown). Despite the weak affinity of the peptides, the data demonstrate that phosphorylation is important for TD’s interaction with Sir4C in vitro, and that the phosphate-enhanced interaction is specific for modifications on S1095.

*Phosphorylation is required for high efficiency transposition and targeted integration*

We previously observed that S1094L mutations decreased transposition frequency approximately five-fold (Gai and Voytas, 1998). Similarly, strains lacking Sir4 also show a comparable decrease in transposition (Zhu et al., 1999). Quantitative transposition assays were performed to test the effects of S1095 mutations on transposition efficiency (Table 2). Transposition frequencies of the wild type and the S1095E mutant were statistically indistinguishable. Both the S1095A and S1095C mutants, however, were reduced approximately five-fold. The S1095E mutant interacts with Sir4, but not S1095A or S1095C (Figure 1B), suggesting that interaction with Sir4 is required for high efficiency transposition.
Table 2. The effect of S1095 mutations on Ty5 transposition frequencies.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Transposition frequency (X 10⁻⁵)</th>
<th>Fold reduction in transposition frequency</th>
<th>p-value (compared to wild type)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>4.88 ± 0.77</td>
<td>Not applicable</td>
<td>Not applicable</td>
</tr>
<tr>
<td>S1095E</td>
<td>3.91 ± 0.94</td>
<td>1.25</td>
<td>0.162</td>
</tr>
<tr>
<td>S1095A</td>
<td>1.04±0.41</td>
<td>4.68</td>
<td>0.001</td>
</tr>
<tr>
<td>S1095C</td>
<td>0.87±0.26</td>
<td>5.64</td>
<td>0.001</td>
</tr>
</tbody>
</table>

In addition to affecting transposition frequency, the targeted transposition data presented in Figure 1C suggest that Ty5 mutants that lack S1095 phosphorylation do not target to heterochromatin. This data was obtained using our plasmid-based targeting assay, which previously has proven a reliable measure of chromosomal integration specificity (Gai and Voytas, 1998). To determine unequivocally the impact of phosphorylation on chromosomal target specificity, we mapped the genomic locations of Ty5 insertions generated by elements bearing the S1095A, S1095C or S1095E mutations (Table 3).

In previous work, we defined targeted insertions as those occurring within a 3 kb window centered on the autonomously replicating sequence (ARS) of the telomeric X repeat or the HMR silencers; over 90% of wild type Ty5 insertions occur in these windows (Zou et al., 1996a; Zou et al., 1996b; Zou and Voytas, 1997). Consistent with a requirement for S1095 phosphorylation in targeting, seven of eleven insertions generated by a Ty5 element bearing the S1095E mutation met the definition of being targeted to heterochromatin; two others were telomeric and just outside the 3 kb window. In contrast, none of the seven chromosomal insertions generated by the S1095A mutant were targeted. Similarly, targeting was not observed for any of the ten insertions generated by the S1095C mutant. Insertions generated by these latter two mutants were not associated with any obvious chromosomal feature, and nine of seventeen insertions occurred within genes, indicating that in the absence of TD phosphorylation, Ty5 is an effective mutagen. Collectively, the data demonstrate that
either a negatively charged amino acid or the addition of a phosphate group on S1095 is necessary for Ty5 to identify heterochromatic target sites on yeast chromosomes.

Table 3. Chromosomal insertion sites of \textit{de novo} transposition events generated by Ty5 S1095 mutants.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Chr. No.</th>
<th>Coordinate</th>
<th>Location(^1)</th>
<th>Nearby chromosomal features</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1095E</td>
<td>I</td>
<td>228,809</td>
<td>T</td>
<td>Within YAR073W and 633 bp from ARS in X repeat</td>
</tr>
<tr>
<td></td>
<td>XI</td>
<td>2,392</td>
<td>T(^2)</td>
<td>Within \textit{PAU16} and 1,632 bp from ARS in X repeat</td>
</tr>
<tr>
<td></td>
<td>XV</td>
<td>264</td>
<td>T</td>
<td>Within X repeat and 535 bp from the X ARS</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>599,223</td>
<td>D</td>
<td>Within \textit{RAD55}</td>
</tr>
<tr>
<td></td>
<td>XI/III</td>
<td>663,321/6,051</td>
<td>T(^2)</td>
<td>Between YKR106W and YKRWomega1 and 2,259 bp from the X ARS/between YCLWomega2 and YCL073C and 4,991 bp from the X ARS</td>
</tr>
<tr>
<td></td>
<td>XII</td>
<td>11,851</td>
<td>T</td>
<td>Within X repeat and 186 bp from the X ARS</td>
</tr>
<tr>
<td></td>
<td>VI</td>
<td>269,960</td>
<td>T</td>
<td>Within X repeat and 193 bp from the X ARS</td>
</tr>
<tr>
<td></td>
<td>IX</td>
<td>439,367</td>
<td>T</td>
<td>Within X element core sequence 254 bp from X ARS</td>
</tr>
<tr>
<td></td>
<td>XIII</td>
<td>510,682</td>
<td>D</td>
<td>Between \textit{RPL15B} and YMR122C</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>1,430</td>
<td>T</td>
<td>Within YCLWTy5-1 and 257 bp from the X ARS 370</td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>570,082</td>
<td>T</td>
<td>Adjacent to X repeat and 440 bp from the X ARS</td>
</tr>
<tr>
<td>S1095A</td>
<td>IV</td>
<td>1,410,803</td>
<td>D</td>
<td>Within YDR476C</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>1,125,173</td>
<td>D</td>
<td>Between YDR326C and \textit{SKP1}</td>
</tr>
<tr>
<td></td>
<td>XII</td>
<td>152,025</td>
<td>D</td>
<td>Within YLR001C</td>
</tr>
<tr>
<td></td>
<td>IX</td>
<td>39,729</td>
<td>D</td>
<td>Within YIL161W</td>
</tr>
<tr>
<td></td>
<td>XV</td>
<td>331,259</td>
<td>D</td>
<td>Between \textit{ALG6} and \textit{YSP3}</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>236,320</td>
<td>D</td>
<td>Within \textit{SED4}</td>
</tr>
<tr>
<td></td>
<td>XII, VII, XIII, or XV</td>
<td>Unknown</td>
<td>D</td>
<td>Within a Ty1 LTR</td>
</tr>
<tr>
<td>S1095C</td>
<td>IV</td>
<td>1,472,174</td>
<td>D</td>
<td>Within YDR514C</td>
</tr>
<tr>
<td></td>
<td>VIII</td>
<td>149,348</td>
<td>D</td>
<td>Within \textit{ECM12}</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>320,566</td>
<td>D</td>
<td>Between \textit{VAM6} and \textit{RXR3}</td>
</tr>
<tr>
<td></td>
<td>VII</td>
<td>387,546</td>
<td>D</td>
<td>Within \textit{PYC1}</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>765,652</td>
<td>D</td>
<td>Between \textit{CTH1} and \textit{GIR2}</td>
</tr>
<tr>
<td></td>
<td>XII</td>
<td>434,128</td>
<td>D</td>
<td>Between YLR146W-A and \textit{SMD3}</td>
</tr>
<tr>
<td></td>
<td>VIII</td>
<td>267,970</td>
<td>D</td>
<td>Within \textit{LRP1}</td>
</tr>
<tr>
<td></td>
<td>VII</td>
<td>853,495</td>
<td>D</td>
<td>Between \textit{PBP1} and \textit{OKP1}</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>419,780</td>
<td>D</td>
<td>Between YBR085C-A and \textit{IST2}</td>
</tr>
<tr>
<td></td>
<td>IX</td>
<td>144,456</td>
<td>D</td>
<td>Within \textit{NUP159}</td>
</tr>
</tbody>
</table>

\(^1\)T = telomeric and within 1500 bp of autonomously replicating sequence (ARS) in the telomeric X repeat; D = dispersed.

\(^2\)Telomeric but more than 1500 bp from the X ARS.
TD phosphorylation is regulated by environmental stress

The requirement for phosphorylation for targeted integration suggests that the host regulates Ty5 target site choice. To seek evidence for regulated TD phosphorylation, we generated an antibody that specifically recognizes the phosphorylated form of TD. The antibody reacts with GST-TD expressed in yeast, but very poorly with GST-TD produced in *E. coli* (Figure 3A).

Figure 3. The phosphorylation state of TD is regulated by stress. (A) Specificity of the phospho-TD antibody. The phospho-TD antibody reacts strongly with GST-TD expressed and purified from yeast, whereas little or no reaction is observed with GST-TD purified from *E. coli*. (B) TD phosphorylation levels decrease upon starvation for amino acids, nitrogen or carbon. GST-TD phosphorylation was monitored using an antibody specific for phospho-TD after exposure of yeast cells to a variety of stress and growth conditions (upper panel) (see also Experimental Procedures for a more detailed description of the stress regimes). An anti-GST antibody was used to assess overall levels of GST-TD protein expression (lower panel). Numbers above each lane indicate the hours after induction of GST-TD expression that the cells were harvested. (C) Proteins isolated from 3 hour time points for the no stress control and the nutrient deprivation treatments were normalized with respect to GST using the anti-GST antibody. As a control for the specificity of the phospho-TD antibody, 50 ng of GST-TD protein that had been previously isolated from yeast or *E. coli* was included on the filter.
We felt that assessments of TD phosphorylation by the antibody should reflect the phosphorylation status of the native IN protein, since all mutations tested that affect phosphorylation of TD fusion proteins also affect Ty5 target specificity (Figure 2, Tables 1 and 3).

TD phosphorylation levels were measured with the phospho-specific antibody after subjecting yeast cells to a variety of growth and stress regimes, including heat shock, DNA damage, osmotic shock, oxidative stress, cell cycle arrest, exposure to heavy metals and starvation for amino acids, carbon and nitrogen. To calibrate the extent of TD phosphorylation, overall GST-TD protein levels were determined using an anti-GST antibody. Most stress regimes did not affect TD phosphorylation, although significant decreases were observed upon nutrient deprivation, specifically starvation for amino acids, carbon and nitrogen (Figure 3B, 3C). The actual effect of carbon limitation may be even greater than that observed, since it was necessary to add galactose to the starved cells to induce expression of the GST-TD protein. Numerous signal transduction pathways in Saccharomyces mediate the metabolic and transcriptional response to changes in nutrient availability (Schneper et al., 2004). Principles among these are the protein kinase A pathway (important for carbon signaling) and the TOR pathway (important for nitrogen signaling). Although the specific regulators of TD phosphorylation remain to be identified, the data demonstrate that the host cell regulates TD phosphorylation in response to environmental challenge, specifically starvation.

DISCUSSION

The ‘safe-haven’ hypothesis suggests that transposable elements can better persist in host genomes if they integrate into gene poor regions, thereby offsetting deleterious genetic consequences of random integration (Boeke and Devine, 1998; Craig, 1997). An underlying tenet of the hypothesis is that transposable elements are genomic parasites, and selective pressures act on mobile elements to adopt strategies to identify genomic safe havens. In contrast, host influences on target site choice are generally viewed as passive. For example, the association of retroviral insertions with transcription units has long been viewed as a
consequence of integration complexes being restricted from access to DNA by chromatin (Bushman, 2003). The data presented here, however indicate that the host can play a critical role in determining target site choice. Under normal circumstances, integrase is phosphorylated by the host, and Ty5 transposes to its heterochromatic safe haven, thereby minimizing deleterious consequences of integration. However, in the absence of phosphorylation, Ty5 becomes a potent endogenous mutagen that integrates throughout the genome. Because the Ty5 encoded targeting domain and the host kinase are both required for targeted integration, the element and host cooperatively participate in how Ty5 influences genome integrity.

Ty5-encoded targeting determinants

The Ty5 targeting domain serves as molecular glue to tether the integration apparatus to sites of Sir4: Ty5 normally recognizes Sir4 at the telomeres and silent mating loci (Zou et al., 1996a); however, the rDNA becomes a target when Sir4 moves to the nucleolus during aging (Zhu et al., 1999), and integration hotspots can be created by tethering Sir4 to ectopic DNA sites (Zhu et al., 2003). Ty5’s targeting domain is surprisingly modular. Target specificity can be altered by swapping TD with peptides that recognize different protein partners, resulting in integration at sites to which the new protein partners are tethered (Zhu et al., 2003). Remarkably, the precision and efficiency with which the engineered Ty5 elements recognize their new targets is comparable to that of the wild type element.

The targeting domain lies within the IN C-terminus, which among retrotransposons is poorly conserved relative to the N-terminal zinc finger and catalytic core (Malik and Eickbush, 1999; Peterson-Burch and Voytas, 2002). Both the high degree of sequence divergence and the modularity of Ty5’s targeting domain suggest that the IN C-terminus is a rapidly evolving platform that interacts with chromatin and perhaps other host factors. In support of this, the IN C-termini of some retrotransposons encode chromodomains (Koonin et al., 1995; Malik and Eickbush, 1999). In cellular proteins, chromodomains interact with modified histone residues, suggesting that these retrotransposons target sites with specific histone modifications (Jacobs and Khorasanizadeh, 2002; Nielsen et al., 2002). We
anticipate that the observed sequence divergence among retrotransposon IN C-termini reflects different targeting solutions reached by mobile elements in order to adapt to specific host cell environments.

Host control of target specificity

The host imposes control on Ty5 target specificity by regulating the phosphorylation status of S1095. Lack of phosphorylation has two consequences: it leads to an approximately five-fold decrease in transposition efficiency, and it randomizes integration. We believe the decrease in transposition is due to integrase’s inability to interact with Sir4: approximately five-fold decreases in transposition are observed in sir4D strains and for all TD mutations that abrogate TD/Sir4C interactions (Gai and Voytas, 1998; Zhu et al., 1999) (see also Table 2). In contrast, the S1095E mutation, which interacts with Sir4 in tethered silencing assays, has wild type levels of transposition. Environmental or growth conditions that lower TD phosphorylation levels would be predicted to decrease transposition efficiency and thereby minimize the deleterious consequences of random integration, and so the requirement for a Sir4 interaction for maximal transposition efficiency may serve to keep Ty5 mutagenesis in check. Interestingly, the human transcription factor LEDGF/p75, which directs HIV integration to target sites, is required for HIV integration, suggesting that the regulation of integration efficiency by targeting determinants may be widespread (Ciuffi et al., 2005; Llano et al., 2006).

It is possible that other sites within or adjacent to the TD are also modified by phosphorylation and contribute to the regulation of target specificity. For example, we observed phosphorylation of S1094. This residue, however, was only modified in a fusion protein with a S1095 mutation, and constructs with glutamic acid substitutions, which mimic phosphorylation, only interacted with Sir4 if the substitution was at position 1095. Nonetheless, these data do not exclude a negative regulatory role for S1094 phosphorylation in target specificity. Whereas S1095 phosphorylation is necessary for productive interactions with Sir4, it is not sufficient: S1095 is
phosphorylated in constructs with an S1094L mutation, yet such constructs do not recognize Sir4.

Phosphorylation plays a role in regulating the related yeast retrotransposon, Ty1. For example, Ty1 is activated by the MAPK cascade that regulates filamentous growth (Conte and Curcio, 2000; Morillon et al., 2000), and the DNA damage response triggered upon telomere erosion (Scholes et al., 2003). How the kinases in these pathways specifically act in Ty1 replication remains to be determined, and since Ty1-encoded targeting determinants have not been identified, no evidence exists as to whether Ty1 integration specificity is regulated by phosphorylation. Phosphorylation of Ty5 proteins may also regulate other steps in the Ty5 life cycle, and we anticipate that post-translational modification of retroelement proteins is likely to be an important means by which the host and mobile element communicate to negotiate a shared genome.

The status of S1095 phosphorylation can have significant consequences for the host if phosphorylation levels decrease under certain conditions resulting in random integration. We surveyed a variety of cellular stresses and growth conditions and observed a significant decrease in TD phosphorylation when yeast cells were starved for amino acids, carbon or nitrogen (Figure 3B). Our experiments used a highly expressed TD fusion protein, which we recognize is less than ideal for monitoring sensitive regulatory events such as changes in phosphorylation levels; biologically relevant changes in phosphorylation of native integrase may also occur under other treatments. The finding that nutrient status regulates TD phosphorylation will necessitate a change in the approach we use to monitor transposition. Expression of our Ty5 constructs is controlled by carbon source (i.e. galactose), confounding interpretation of experiments that use existing reagents to test the effect of nutrient availability on target specificity.

Multiple signaling pathways are used by Saccharomyces to respond to environmental nutrient status. Principle among these are the protein kinase A and TOR pathways, which are involved in sensing carbon and nitrogen availability, respectively (Schneper et al., 2004). Both pathways are highly interconnected, and under starvation
conditions, both activate a common cellular response (called the environmental stress response) by transmitting signals to the transcription factors Msn2 and Msn4. These transcription factors are principally responsible for the stereotypical changes in gene expression triggered by diverse environmental stresses (Causton et al., 2001; Gasch et al., 2000). We anticipate that the kinase or phosphatase that regulates TD phosphorylation is part of this interconnected nutrient signaling network or acts downstream of a point of integration. Future work will utilize various signaling mutants to determine the specific cellular regulators that control Ty5 target specificity.

The regulation of TD phosphorylation by stress is consistent with Barbara McClintock’s prediction that mobile elements provide the cell with a pre-wired mechanism to reorganize the genome in response to environmental challenge (McClintock, 1984). Whereas various stress treatments activate the expression of diverse mobile genetic elements, to the best of our knowledge, this study is the first to suggest that stress can change the mutagenic potential of a transposable element by altering target specificity. Host control of target specificity implies that Ty5 is not a mere genomic parasite, but rather a mutualist, and that by abdicating control of target specificity to the yeast cell, Ty5 provides its host with a mechanism to regulate genome integrity.

EXPERIMENTAL PROCEDURES

Plasmids and strains

The tethered silencing assay was carried out in strain YSB35 (Aeb::3×UASg) (gift of D. Shore, University of Geneva) (Chien et al., 1993). The yeast two-hybrid assay used strain L40 (MATa his3-200 trp1-901 leu2-3,112 ade2 lys2-801am LYS2::(lexAop)4-HIS3 URA3::(lexAop)8-lacZ GAL4) (gift of R. Sternglanz, SUNY Stony Brook) (Chien et al., 1991). Strain JTY142 (MATa his3-D200 ura3-52 trp1D63 leu2D1 prb1-D1.6R can1 pep4::HIS3) and YPH499 (MATa ura3-52 lys2-801 ade2-101 trp1D63 his1D200 leu2D1) were used for all protein expression experiments (gift of D. Amberg, Upstate Medical University and A. Myers, Iowa State University, respectively).

For the tethered silencing assay, GBD-TD and its mutant derivatives were constructed
by ligating double-stranded oligonucleotides into EcoRI/BglII-digested pGBDU plasmid (James et al., 1996). The same approach and plasmid pGAD were used to make the His6-tagged TD-GAD fusion constructs for two-hybrid assays. The GST fusion constructs were generated in pEG(KG) (Mitchell et al., 1993) by inserting a DNA fragment obtained either by PCR amplification and cloning or by annealing two complementary oligonucleotides. PCR-based mutagenesis (Ausubel et al., 1987) was used to introduce the TD mutations into a full-length Ty5 element on pNK254 (Ke and Voytas, 1997). This was accomplished by PCR-amplifying pNK254 with two pairs of primers: 1) a forward primer and mutagenic primer; 2) a reverse primer and a second, complementary mutagenic primer. The amplification products were used as a template for a second PCR reaction with the forward primer and the reverse primers. Amplification products were then digested with BspEI and PflMI and used to replace the corresponding fragment in pNK254. All plasmids were sequenced subsequent to recovery from E. coli to verify the presence of the mutation and to ensure no unwanted mutations were incorporated. Sequences of all oligonucleotides used in this study are available upon request.

Monitoring TD phosphorylation in vitro

GST-TD fusion proteins were purified from E. coli using glutathione agarose and methods supplied by the manufacturer (Sigma). GBD-TD fusion proteins were immunoprecipitated from yeast cell extracts by methods previously reported (Xie et al., 2001). Assays measuring the interaction between TD-fusion proteins and Sir4 were performed as previously described (Xie et al., 2001). To treat the immunoprecipitated GBD fusion proteins with λ phosphatase (New England Biolabs), the final wash was performed with λ phosphatase buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 0.1 mM EGTA, 2 mM DTT, 0.01% Brij 35, and 2 mM MnCl₂) instead of PBS. The beads were incubated at 30°C with 0.5 μl (200 units) enzyme for 30 minutes and washed with PBS before mixing with the labeled Sir4C as described (Xie et al., 2001).

For mass spectrometry, GST-TD fusion proteins expressed in yeast were first purified using HIS-Select™ Nickel Affinity Gel (Sigma) under native conditions. Eluates were combined and purified again with glutathione-agarose (Sigma). To isolate TD, fusion proteins were incubated at room temperature with 2 μl (20 units) factor Xa (New England
Biolabs) for six hours. The digestion mixture was diluted in equilibration buffer (50 mM sodium phosphate, pH 8.0, with 0.3 M sodium chloride and without imidazole) and purified again using HIS-Select™ Nickel Affinity Gel (Sigma). MALDI-TOF MS/MS analyses were performed using a QSTAR XL quadrupole TOF mass spectrometer (AB/MDS SCIEX) equipped with an oMALDI ion source. The mass spectrometer was operated in the positive ion mode, and mass spectra were acquired over m/z 500 to 4000. All samples were desalted and concentrated with a 10 μl ZipTip®c18 (Millipore) following the instructions provided by the manufacturer. Peptides were eluted with 1 μl of 50% acetonitrile/0.1% trifluoroacetic acid saturated with α-cyano-4-hydroxycinnamic acid and deposited onto the MALDI target plate. After every regular MS acquisition, MS/MS acquisition was performed against the most intensive ions. The molecular ions were selected by information dependent acquiring in the quadrupole analyzer and fragmentated in the collision cell.

Real-time protein-protein interactions were examined using a BIAcore T100 instrument (BIAcore). An anti-GST antibody was immobilized on a CM5 sensor chip using a GST capture kit (BIAcore). GST-Sir4C was purified from E. coli with glutathione-agarose (Sigma), following the protocol supplied by the manufacturer, and eluted in 50mM HEPES pH 7.6, 350mM NaCl, 1mM DTT, 20mM glutathione, 25% glycerol. The fusion protein was diluted to 10 μg/ml in HBS-EP buffer (BIAcore) and injected with a constant (10 μl/min) flow rate at 25°C for 20 min. Synthetic TD or phospho-TD (2.5 mM) (Sigma-Genosys) was injected at a flow rate of 30 μl/min for 60 seconds at 25°C respectively. Due to the fast off-rate of all peptides tested, the same bound GST-Sir4C protein was used to compare affinity of all three peptides by allowing sufficient wash time between addition of new peptides to re-establish the baseline. The chemical binding surface was regenerated between replicates by washing with regeneration solution (10 mM glycine-HCl, pH 2.2) for 2 min at 20 μl/min. Sensorgrams were subjected to global analysis using BIAcore T100 evaluation software 1.0. Two independent binding comparisons were performed and the data averaged.

**Monitoring in vivo levels of TD phosphorylation, TD/Sir4C interaction and transposition**

The tethered silencing, yeast two-hybrid, plasmid targeting and transposition assays were performed as previously described (Gai and Voytas, 1998; Xie et al., 2001). For the plasmid targeting assay, we used the target plasmid pXW72 with a TRP1 marker rather than
pXW78 with a LEU2 marker. Transposition data presented in Table 2 was performed in a rad52Δ derivative of YPH499. Chromosomal Ty5 insertions and flanking sequences were recovered by inverse PCR as previously described (Zhu et al., 1999).

To monitor regulation of TD phosphorylation, yeast strain YPH499 with plasmid pJB149 (encoding a galactose-inducible GST-TD fusion protein) was grown overnight in SC-Ura plus glucose medium. The overnight culture was centrifuged and resuspended to O.D. 0.2 in SC-Ura plus raffinose medium and allowed to grow to O.D. 0.5. Aliquots of the culture (30 ml) were subjected to stresses using protocols similar to those previously described (Gasch et al., 2000). Briefly, stress agents were added to the culture to yield the following final concentrations: 0.7M NaCl; 1 M sorbitol; 0.3 mM H2O2; 500 mM CuSO4; 0.5 mM alpha factor; 10 mg/ml nocodazole; 150 mM hydroxyurea; 0.02% MMS. Heat shock was applied by placing the cultures at 37 C. For nutrient deprivation, yeast cells were resuspended in SC-Ura without raffinose (carbon starvation) or minimal medium without amino acids (YNB-AA-AS, 2% raffinose; amino acid starvation) or minimal medium without amino acids and with limiting concentrations of ammonium sulfate (0.025% ammonium sulfate, YNB-AA-AS, 2% raffinose; nitrogen starvation). The cultures were allowed to grow for 2 or 3 hrs and then expression of GST-TD was induced by addition of galactose to 2%. Aliquots of cells were harvested each hour after induction for 4 hrs. Control cells were treated as above, except that after centrifugation, cells were resuspended in SC-Ura plus 2% raffinose. Antibodies specific to TD with S1095 phosphorylation were prepared commercially (PhosphoSolutions). Immunoblotting was carried out in Tris-buffered saline (TBS) with 0.1% Tween 20, 0.1% Triton X-100 and with 3% BSA at room temperature. Signals were detected using Amersham ECL Western Blotting Detection Reagent (Amersham).

ACKNOWLEDGEMENTS

We thank William Lewis and Siquan Luo at the Iowa State University Proteomics Facility for carrying out the MS and MS/MS analysis. This work was supported by National Institutes of Health Grant GM061657.
REFERENCES


Devine, S. E., and Boeke, J. D. (1996). Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. Genes Dev 10, 620-633.


CHAPTER 3. THE RIS1 UBIQUITIN LIGASE MEDIATES SIR4 TRUNOVER
UPON INTERACTION WITH THE TARGETING DOMAIN OF THE YEAST
RETROTRANSPOSON TY5

A manuscript to be submitted to *Molecular Cellular Biology*

Jiquan Gao¹, Peter Feurst², and Daniel F. Voytas³

ABSTRACT

Heterochromatin is a structurally condensed region bound by many heterochromatic proteins including Sir2/3/4 complex in *Saccharomyces cerevisiae*. Many studies suggested that post-translational modifications of Sir proteins might regulate the heterochromatic silencing such as phosphorylation, sumoylation, and ubiquitination. Sir4 served as the scaffold protein for its interaction with many proteins involving ubiquitin pathway such as Ubp10, Ubp3 and Ris1. However, the ubiquitin mediated regulation of heterochromatin was not yet discovered. In this study, we demonstrated that Ris1, a putative ubiquitin ligase of sumo-conjugates, regulate Sir4 stability and silencing status when the targeting domain of the yeast retrotransposon Ty5 is overproduced. We further showed the evidence that Ris1 RING finger domain has RING finger dependent autoubiquitination activity in vitro and TD mediated Sir4 turnover requires the intact RING finger domain of Ris1, Ubc4 and functional proteasome pathway. In addition, a small domain functionally equivalent to TD of Esc1, a nuclear periphery protein interacting with Sir4 can also break the silencing in Ris1, Ubc4 dependent manner when overproduced. Therefore, we for the first time showed that heterochromatic silencing could be regulated via Sir4 ubiquitination by Ris1. Meantime, considering the similar effect of Ty5 TD domain and Esc1 TD like domain on silencing, we speculated that the regulation of Sir4 stability and heterochromatin could occur at multiple cellular events such as cell cycle progression, the integration of Ty5.

¹ Primary researcher and author
² Second author, the former graduate student who initiated this project
³ Professor and corresponding author, Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN 55455
INTRODUCTION

The eukaryotic genome is organized into transcriptionally active regions known as euchromatin and transcriptionally repressed regions called heterochromatin or silent chromatin. Heterochromatin is densely compact and structurally inaccessible to many enzymes such as RNA polymerase, Dam DNA methyltransferase, restriction enzymes and DNA repair proteins. Heterochromatin plays many critical roles such as maintaining chromosome stability, regulating gene expression and repressing recombination. In *Saccharomyces cerevisiae*, three genomic regions are known to be transcriptionally silenced: telomeres, the silent mating type loci *HML* and *HMR*, and the rDNA locus (Huang, 2002; Rusche et al., 2003). The cis-acting DNA sequences and relevant proteins that establish and maintain silencing at these regions have been identified. Telomeres and the *HM* loci share a very similar silencing mechanism that is mediated by the multiprotein, nucleosome-binding SIR complex. The SIR complex is composed of Sir2, Sir3, and Sir4 (Gasser and Cockell, 2001). Sir2 is a histone deacetylase that acts on histone H3 and H4 tails allowing Sir3 to tightly bind. Sir4 proteins interact with Sir2 and Sir3 and also many other proteins involved in the maintenance of silencing. Sir2 is also required for silencing at the rDNA locus, but Sir3 and Sir4 are dispensable for silencing at the rDNA. This suggests a mechanistic difference in silencing at the telomeres, *HM* loci and rDNA.

Although heterochromatin is inaccessible to many enzymes and can persist through mitotic and meiotic cell divisions (Grewal, 2000), recent studies have shown dynamic regulation of silencing at the telomeres and *HM* loci. Mating pheromone, starvation and heat shock can induce Sir3 hyperphosphorylation via MAP kinase pathway and therefore, strengthen the silencing at telomeres (Stone and Pillus, 1996). Constitutive Sir3 hyperphosphorylation via the TOR signaling pathway is correlated with reduced subtelomeric silencing, whereas hyperphosphorylation by the pheromone response pathway increases telomeric silencing. In both cases, silencing at the *HM* loci is unaffected. Reduced silencing at telomeres by TOR signaling derepresses transcription of subtelomeric genes and subsequently enhances cell wall synthesis and stress resistance (Ai et al., 2002). In another study, commitment to rapid cell growth results in Sir3 phosphorylation by the Slt2 kinase.
When the yeast cells are in state of growth quiescence, Slt2 kinase is not activated and unphosphorylated Sir3p accumulates. As a result, telomeric silencing is decreased whereas silencing at the HM loci and rDNA loci is enhanced. The redistribution of silencing functions between the telomeres, HM, and rDNA loci is also associated with extended lifespan (Ray et al., 2003). Interestingly, Sir3 redistribution from telomeres and consequential loss of telomeric silencing are induced by double strand DNA breaks (Martin et al., 1999; McAinsh et al., 1999; Mills et al., 1999). Sir4 relocation is also observed in response to DNA damage, and the relocation of both Sir3 and Sir4 are dependent on the Mec1/Rad9 pathway and are thought to facilitate DNA repair at the site of double strand breaks (Martin et al., 1999).

Whereas changes in phosphorylation status and the redistribution of Sir proteins contribute to silent chromatin dynamics, Sir protein turnover is also likely involved. A recent study suggests that the ubiquitin machinery affects silent chromatin due to interactions between Sir4 and Ubp3, Ubp10 and Ris1. The deletion of Ubp3, a deubiquitinase, results in improved silencing at telomeres and HM loci (Moazed and Johnson, 1996). The direct protein target of Ubp3, however, is not known. Ubp10, another deubiquitnase, removes ubiquitin from histone H2B. This activity is thought to maintain proper levels of histone modification to allow optimal Sir protein binding. Thus, loss of Ubp10 disrupts telomeric silencing (Gardner et al., 2005b). Ris1 is a large multidomain protein first identified for its role in HM silencing (Zhang and Buchman, 1997). More recently, Ris1 has been suggested to be a RING-type ubiquitin ligase that mediates the ubiquitin-dependent proteolysis of SUMO conjugates in vivo (Uzunova et al., 2007). In vitro ubiquitin ligase activity of Ris1, however, was not reported. Moreover, it is not known whether Ris1 can ubiquitinate Sir4 or any other silencing protein and whether ubiquitination plays a role in regulating silencing. The San1 ubiquitin ligase is known to ubiquitinlate Sir4 (Dasgupta et al., 2004), and preferentially acts on Sir4 proteins produced by the temperature sensitive Sir4-9 allele. The exquisite specificity of San1 for Sir4-9 and other aberrant nuclear proteins is thought to be one means by which yeast cell controls the quality of nuclear proteins (Gardner et al., 2005a).

Silent chromatin is the favorite integration target site for the yeast retrotransposon Ty5 (Zou et al., 1996; Zou and Voytas, 1997). Ninety percent of Ty5 integrations occur in
silent chromatin at the telomeres and \textit{HM} loci. Targeting depends on an interaction between Sir4 and the targeting domain (TD) located at the very C-terminus of Ty5 integrase. TD is sufficient and necessary to mediate targeted integration via its interaction with Sir4. In a previous study, overproduction of TD was found to break telomeric silencing (Xie et al., 2001b), but the mechanism of TD anti-silencing activity was not understood. In this study, we demonstrate that TD overproduction breaks silencing by promoting Sir4 turnover, which is dependent on Ris1 and the proteasomal protein degradation pathway. We also are the first to report \textit{in vitro} ubiquitin ligase activity of the Ris1 RING finger domain, which is required to mediate the Sir4 turnover in the presence of TD. This is the first report that Ris1, through ubiquitination, regulates Sir4 degradation and silencing status. We speculate that Ty5, with help from Ris1p, accesses DNA within heterochromatin to complete the integration reaction.

\section*{RESULTS}

\textit{Requirements for Ty5 TD to break transcriptional silencing}

Over-expressing components of yeast heterochromatin frequently disrupts transcriptional silencing, presumably because silencing complexes are perturbed by altered stoichiometry of protein components or because critical factors are titrated away. As an example, transcriptional silencing is disrupted by over-expressing Sir4p or Sir4p interacting proteins such as Ris1p, Dot4p or portions of Sir3p (Gotta et al., 1998; Kahana and Gottschling, 1999; Zhang and Buchman, 1997). We have previously shown that the Ty5 targeting domain (TD), when fused to the Gal4 DNA binding domain (GBD), disrupts silencing of telomeric \textit{URA3} and \textit{ADE2} genes, as evidenced by the \textit{URA}+ 5FOA sensitive and white colony color phenotype conferred by GBD-TD expression (Fig. 1A) (Xie et al., 2001a). In these experiments, only the six amino acid TD and three adjacent amino acid residues are fused to GBD (\textbf{LDSSPPNTS}). A mutation in the second serine of TD (S1094L, designated td) abrogates interactions with Sir4p and fails to break silencing.

Here we show that the anti-silencing effect of TD is not limited to telomeres, as GBD-TD can break transcriptional silencing of a \textit{TRP1} reporter located at \textit{HMR} (Fig. 1B). This is evidenced by the ability of cells expressing the TD fusion to grow on media lacking
Fig 1. The overexpression of Ty5 TD breaks the silencing at both telomere and HM loci, but not rDNA. The six amino acid targeting domain, and three adjacent amino acid residues, were fused to GBD (LDSSPPNTS; TD is underlined). (A) Expression of TD fused to GBD breaks the telomeric silencing of an URA3 gene as seen by the ability to grow on media lacking uracil (see arrow SC-L-U). Expression of GBD or a biologically inactive TD mutant (td; LDL*SPPNTS) fusion does not effect telomeric silencing of the URA3 reporter gene. (B) GBD-TD expression breaks silencing at the HM loci as seen by the ability of a strain with TRP1 inserted into HMR to grow on media lacking tryptophan. (C) Expression of GBD-TD does not break silencing of a MET15 reporter gene integrated into the rDNA. In the absence of MET15 hydrogen sulfide ions build up in yeast cells and precipitate lead nitrate incorporated into the media, resulting in a brown pigment. No difference in the silencing of the MET15 gene was observed when comparing strains expressing GBD-TD and GBD-td.

tryptophan, in contrast to the TD mutant. GBD-TD expression, however, has no influence on the silencing of a MET15 reporter gene integrated at the rDNA locus, as assayed by the brown pigment phenotype resulting from the precipitation of lead nitrate incorporated into the media (Fig. 1C). In summary, GBD-TD expression can cause the loss of silencing at telomeres and HM loci, but not the rDNA. Further, the anti-silencing activity depends on the ability of GBD-TD to interact with Sir4.

**GBD-TD breaks silencing by lowering levels of Sir4p**

Since TD interacts with Sir4, it is possible that overexpression of GBD-TD breaks silencing by titrating Sir4 from telomeres and HM loci, thereby disrupting the SIR complex
Fig 2. The silencing defect caused by TD overexpression results from the downregulation of Sir4 protein. (A) Expression of GBD-TD has no effect on Sir2 level as seen by western blot analysis (arrows) compared with expression of GBD-td. (B) Sir3 level is not affected by TD overexpression either. A c-Myc tagged SIR3 allele was used, as effective commercial antibodies were not available. (C) Sir4 levels are lowered in the strain expressing GBD-TD (arrow). (D) Expression of GBD-TD does not repress the transcription of Sir4 promoter driven LacZ reporter gene compared with expression of GBD-td. (E) Sir4 expressed from a plasmid-born promoter is lowered by expression of GBD-TD. Western blot analysis indicates that expression of GBD-TD decreases Sir4 levels expressed from this construct compared to GBD-td.

or even affecting the stability of Sir proteins. To understand TD anti-silencing, we assessed the steady state levels of Sir2, Sir3 and Sir4, all of which are required for transcriptional silencing at the telomeres and HM loci. In western blots using an anti-Sir2 antibody, Sir2 levels were not affected in cells expressing GBD-TD versus GBD-td (Fig. 2A). To assess the Sir3 levels, we constructed an N-terminal myc-tagged version of the protein, whose ectopic expression on a CEN plasmid complemented the sir3 delta mutation in silencing and mating assays (data not shown). We found Sir3 protein levels were not affected by expressing TD relative to td (Fig. 2B). In contrast, Sir4 protein levels were significantly reduced by overexpression of GBD-TD compared with GBD-td (Fig. 2C).
Since GBD has the ability to bind the DNA, it is possible that GBD-TD reduces Sir4 expression at the transcriptional level. To rule out this possibility, we evaluated expression of a LacZ reporter gene under the control of the SIR4 promoter. No significant difference in expression of LacZ from the SIR4 promoter was observed in the strain expressing GBD-TD (5.3±2.1 units of ONPG hydrolyzed) compared with GBD-td (3.4±1.2 units of ONPG hydrolyzed) (Fig. 2D). As an additional control for this experiment, a plasmid containing the SIR4 gene was introduced into a sir4 delta strain and found to be sensitive to GBD-TD (Fig. 2E). Furthermore, northern blot analysis also showed no difference in steady state SIR4 mRNA levels in strains expression GBD-TD compared with GBD-td (data not shown). These data indicate that GBD-TD overexpression disrupts silencing by causing Sir4 turnover.

**TD-mediated Sir4 turnover requires Ris1, Sir2 and Sir3**

It was reported that the San1 ubiquitin ligase specifically ubiquitinates Sir4 proteins expressed from the temperature sensitive allele sir4-9, leading to degradation via the proteasome pathway. In another report, the half-life of wild type Sir4 was shown to be regulated by San1. Although there is inconsistency between these reports with respect to San1’s action on Sir4, San1 is still a very good candidate as a regulator of TD-induced Sir4 turnover. To our surprise, however, TD-induced Sir4 turnover is not affected in san1 delta strains (Fig 3A).

In heterochromatin, Sir4 plays a critical structural role by linking Sir2 and Sir3 proteins with telomere binding protein yKu70 and silencer-binding proteins such as Rap1. In addition, Sir4 interacts with many other proteins important for silencing such as Zds1, Zds2, Esc1, Ubp3, Dot4 and Ris1. Among the Sir4 interactors, Ubp3 and Ubp10/Dot4 are deubiquitinases, regulating protein degradation by the ubiquitin-proteasomal pathway. However, deletion of ubp3 or dot4 has no effect on TD-mediated Sir4 turnover (data not shown). To identify factors important for TD’s effect on Sir4, we screened strains with deletions in genes encoding Sir4 interacting proteins (SIR2, SIR3, RAP1, YKU70, DOT4, UBP3, RIS1, ZDS1, ZDS2 and ESC1). These strains were transformed with GBD-TD or GBD-td, and Sir4 was measured by Western blot analysis.
Of the genes tested, *SIR2*, *SIR3* and *RIS1* were required for TD mediated turnover (Fig. 3B, 3D). Because loss of *SIR2* or *SIR3* leads to expression of the silent mating loci, we evaluated TD mediated Sir4 degradation in *sir2Δ hmrΔ*, MATα and *sir3Δ hmrΔ*, MATα strains. Expression of GBD-TD in these strains had no effect on Sir4 levels, indicating that the loss of TD mediated Sir4 turnover in a *sir2Δ* or *sir3Δ* background is not explained by derepression of the silent mating loci (Fig. 3B and data not shown). Furthermore, Sir2 enzymatic activity did not appear to be important for Sir4 turnover, because turnover was observed in a strain with a deletion of *NPT1* (data not shown), which synthesizes the cofactor required for *SIR2*-mediated silencing (Smith et al., 2000).

**Fig. 3** GBD-TD mediated Sir4 degradation is dependent on *RIS1*, *SIR2* and *SIR3*. Strains with deletions of genes encoding Sir4 interacting genes were transformed with GBD-TD or GBD-td. Western blot analysis was used to monitor Sir4 levels. (A) GBD-TD induced Sir4 turnover does not require San1 (see Results for a complete list of deletions that had no effect). (B) Strains with *SIR2*, *SIR3* deleted no longer support Sir4 turnover associated with GBD-TD expression. Expression of the silent mating loci in *sir* knockout strains does not account for the lack of Sir4 turnover. Deletion of the non-MAT mating locus in a *sir2Δ* strain does not restore Sir4 degradation in a strain expressing GBD-TD. (C) *RIS1* is required for GBD-TD mediated Sir4 degradation. (D) *RIS1* is required for GBD-TD mediated Sir4 anti-silencing. A *ris1Δ* strain with a telomeric *URA3* was transformed with GBD-TD or GBD-td. The *ris1Δ* strain expressing GBD-TD had an approximately ten-fold growth defect compared to the same strain expressing GBD-td (SC -L). GBD-TD no longer breaks telomeric silencing of a telomeric *URA3* marker gene if *RIS1* is deleted. This is evidenced by the inability of the *ris1Δ* GBD-TD strain to grow on media lacking uracil (SC -L-U). Over expression of Sir4 in this strain was used as a positive control of anti-silencing.
The mechanism of Sir2, Sir3 requirement for TD caused Sir4 turnover is unclear. Since sir2 or sir3 deletion will disrupt the silencing and change the structural environment in which Sir4 normally exists, we speculate that TD induced Sir4 turnover depends on the native silencing scenario.

The Ris1 RING finger is essential for ubiquitin ligase activity, TD-mediated Sir4 turnover, and anti-silencing

Ris1 is a member of ubiquitin ligase of SUMO conjugates and has previously been shown to be required for efficient mating type switching (Uzunova et al., 2007; Zhang and Buchman, 1997). Unlike deletion of SIR2 or SIR3, deletion of RIS1 does not have an effect on transcriptional silencing at the telomeres and silent mating loci. If TD-mediated anti-silencing is the result of lowered Sir4 levels, and lowered protein levels depend on Ris1, then telomeric silencing should not be effected by GBD-TD expression in a ris1 strain. To test this, a ris1 delta strain with telomeric marker genes was constructed and transformed with plasmids expressing GBD-TD and GBD-td. The strain was spotted onto media lacking uracil, and telomeric silencing of a URA3 marker gene was not broken by expression of GBD-TD (Fig. 3D). The lack of effect of GBD-TD expression on URA3 expression in the ris1 deletion strain is consistent with the lack of effect of GBD-TD on Sir4 protein levels.

Ris1 contains a C3HC4 type RING finger motif characteristic of ubiquitin ligases (Uzunova et al., 2007; Zhang and Buchman, 1997). Alignment of the Ris1 RING finger amino acid sequence with the amino acid sequences of other RING finger domains shows high levels of conservation to yeast ubiquitin ligases (Fig. 4A). The presence of a RING finger and the requirement of Ris1 for TD-mediated Sir4 turnover suggested that Ris1 leads to Sir4 turnover through ubiquitin-mediated degradation. Attempts by us and others, however, failed to directly show ubiquitin ligase activity of Ris1 due to very poor expression and insolubility of recombinant Ris1. Here, we expressed and purified the Ris1 RING finger motif (AA 1290-1445) from yeast using a GST tag. The GST-Ris1 RING finger fusion showed in vitro autoubiquitination activity (Fig. 4B), although weaker than the GST-San1 control, whose activity was demonstrated in a previous study (Gardner et al., 2005a).
Fig 4. The RING finger motif of Ris1 has RING finger dependent ubiquitin ligase activity and is required for TD mediated Sir4 degradation and anti-silencing. (A) The sequence alignment between Ris1 and other RING type ubiquitin ligases reveals that Ris1 contains highly conserved RING finger domain. (B) Ris1 RING finger domain has *in vitro* ubiquitinylation activity. Ris1R (AA 1290-1445), Ris1R point mutants and San1 were expressed as GST fusions. After incubating purified GST fusion proteins with ubiquitin and necessary cofactors, reaction mixture were run on a polyacrylamide gel, transferred to a membrane and probed with antibodies to GST or ubiquitin. GST antibodies indicate that proteins of expected sizes were made. San1p and Ris1R were ubiquitinated as visualized by the presence of polyubiquitinylated products, while Ris1R is less active than San1. Negative controls lacking either the Uba or Ubc reagent were negative. (C) Point mutations in the RING finger domain of Ris1 abrogated ubiquitin ligase activity. (D) The Ris1 RING finger motif is required for GBD-TD anti-silencing and Sir4 degradation. The absence of GBD-TD mediated anti-silencing, in a *ris1* deletion background, can be complemented by expression of *RIS1* from a plasmid (*RIS1* TD-L-T-U). Telomeric silencing is compromised in a *ris1Δ* strain transformed with GBD-TD and a plasmid born *RIS1*. *RIS1* alleles with mutations in the RING finger motif are unable to fully restore GBD-TD-mediated anti-silencing, as seen by the decreased anti-silencing associated with GBD-TD expression (ring4 TD and ring2 TD SC-L-U). The *RIS1* ringΔ4 allele has four conserved residues in the RING finger motif replaced with alanine, while the *RIS1* ringΔ2 allele has two conserved residues replaced with alanine. Western blot analysis of Sir4 expression in these strains indicates that the ring finger mutants no longer facilitate GBD-TD mediated Sir4 turnover.

Moreover, the autoubiquitination activity of Ris1 requires an intact RING finger, since two serine substitution mutants at conserved cysteines (AA 1348 and 1385) eliminated activity
(Fig. 4C). Two mutant alleles of *ris1* were constructed to test the requirement of the RING finger motif. Alanine substitutions were introduced in conserved cysteine residues of the RING finger, and wild type and mutant genes were introduced into a *ris1* delta strain with a telomeric *URA3* reporter gene. The wild type gene was able to complement the *RIS1* deletion, as evidenced by the restoration of GBD-TD mediated anti-silencing. Anti-silencing is reduced more than 100-fold in the RING finger mutants (Fig. 4D). Western blot analysis confirmed that the lack of TD mediated anti-silencing in strains expressing *RIS1* alleles with mutant RING fingers was due to the failure of GBD-TD to decrease Sir4 levels (Fig. 4D).

*TD*-mediated Sir4 turnover occurs through the proteasome pathway and requires Ubc4, an ubiquitin conjugating enzyme associated with Ris1

It has previously been shown that Ubc4 and its homologue, Ubc5, coordinate with Ris1 to regulate the degradation of SUMO conjugated proteins (Uzunova et al., 2007). Furthermore, Ubc4 physically binds Ris1, as demonstrated by pull-down experiments. We therefore investigated whether TD-mediated anti-silencing also requires Ubc4. In our silencing assay, we found that Ubc4 is indeed required for TD-mediated anti-silencing at both telomeres and *HM* loci (Fig. 5A).

Polyubiquitinated proteins usually are degraded via the proteasomal pathway (Ciechanover, 1998). In order to test whether TD triggers Sir4 degradation via the proteasome, we added the proteasome inhibitor MG-132 to yeast cells expressing GBD-TD or GBD-td. To enhance the efficacy of MG-132, we deleted the multidrug transporter *PDR5* (Lee and Goldberg, 1998). We found that MG-132 prevented Sir4 degradation in the presence of GBD-TD. This indicates that Ris1-mediated Sir4 turnover occurs through the proteasomal pathway (Fig. 5B). In summary, the data collectively suggest that Ris1, an ubiquitin ligase, mediates Sir4 ubiquitination upon interaction with TD, and this targets Sir4 degradation through the proteasomal pathway.
Fig5. GBD-TD mediated antisilencing requires Ubc4, an ubiquitin conjugating enzyme associated with Ris1 and GBD-TD mediated Sir4 turnover occurs through proteasomal pathway.

The overexpression of the TD-like domain of Esc1 also breaks silencing in a Ris1- and Ubc4-dependent manner

The nuclear protein Esc1 anchors silent chromatin to the nuclear periphery through interactions with Sir4 (Gartenberg et al., 2004; Taddei et al., 2004). The absence of Esc1 has modest effects on telomeric silencing, suggesting that Esc1 also contributes to the maintenance of telomeric heterochromatin. Recently, we showed that Esc1 contains a short motif (AA 1443-1455), namely TD-like domain, at its C-terminus that shares sequence similarity with Ty5 integrase’s TD (Brady et al., 2008). Further, we showed functional equivalence to TD by demonstrating that the Esc1 domain interacts with the same region of Sir4 as does TD. Interestingly, the Ty5 TD domain can efficiently substitute for the Esc1 TD-like domain to mediate plasmid partitioning during mitosis, and the Esc1 TD-like domain can target Ty5 integration when it is used to replace Ty5’s native TD.
Since overexpression of Ty5 TD breaks silencing, we were curious to know whether the Esc1 TD-like domain has similar activity. In our experiment, the Esc1 TD-like domain was overexpressed as a GBD fusion protein and found to break silencing at telomeres and \( HM \) loci. Further, like the Ty5 TD, the anti-silencing activity of the Esc1 domain required Ris1 and Ubc4 (Fig. 6). This result further confirms the functional equivalency between Ty5 TD and the Esc1 TD-like domain and implies that Esc1 regulates silencing through its TD-like domain in addition to promoting silencing and anchoring silent chromatin to the nuclear periphery.

DISCUSSION

\textit{Ris1 ubiquitin ligase and Sir4}

In our initial work and in previous studies, the characterization of Ris1’s ubiquitin ligase activity was hindered by the fact that recombinant, full-length Ris1 is poorly expressed and largely insoluble. We observed this to be the case in both yeast and \textit{E. coli}. To counteract this problem, we expressed just the Ris1 C-terminal fragment from AA 1290-1445, which contains the RING finger motif. A similar approach proved successful in demonstrating autoubiquitination activity of the TEB4 RING domain (Hassink et al., 2005).
Like the TEB4 experiments, we found that the GST-RING finger fusion protein prepared from yeast autoubiquitinated. To our surprise, however, the same GST fusion made in *E. coli* failed to autoubiquitinate (data now shown). This difference suggests that the activity of Ris1 ubiquitin ligase may be subject to posttranslational modification. In addition, a larger C-terminal fragment of Ris1 bearing AA 894-1619 and prepared from both yeast and *E. coli* failed to autoubiquitinate in vitro (data now shown). To explain the difference in activity of the two fragments, we speculate that Ris1 has a built-in mechanism to autoinhibit its ubiquitin ligase activity. A built-in mechanism to regulate Ris1 activity with respect to antisilencing was proposed in the first report of Ris1 based on several observations: 1) overexpression of an N-terminal Ris1 fragment in wild type strains breaks silencing, but overexpression of full-length Ris1 does not; 2) antisilencing caused by overexpression of the N-terminal Ris1 fragment requires the C-terminal catalytic domain; 3) the N-terminal Ris1 fragment interacts with Sir4 strongly in yeast two hybrid assays, but the full-length Ris1 only weakly interacts with Sir4; 4) a GAD fusion with the Ris1 N-terminal fragment derepresses silencing 10-fold more strongly than a GAD fusion of full-length Ris1; a GAD fusion with the Ris1 C-terminal fragment shows only weak derepression. Collectively, all of the above observations support the view that Ris1 has a built-in mechanism to regulate its ubiquitin ligase and antisilencing activities.

Recently, it was suggested that Ris1 is the ubiquitin ligase for SUMO-conjugated proteins, based on several experimental observations: 1) The N-terminus of Ris1 interacts with SUMO through its SUMO interaction motif; 2) in the absence of Ris1, high molecular weight sumo conjugates (HMW-SC) accumulate. Accumulation of HMW-SCs also was observed in an *ubc4, ubc5* double mutant and upon treatment of cells with MG132; 3) Ris1 interacts with Ubc4 (Uzunova et al., 2007). Furthermore, Sir4 is sumolyated in yeast, and Ris1 interacts with nuclear proteins such as Ebp2, Cdc3 and Cdc11 in a SUMO-dependent manner (Denison et al., 2005). Finally, both Ris1 and Sir4 are nuclear proteins. All the data, including data reported in this manuscript, suggest that Ris1 ubiquitinates Sir4 in the presence of TD and then targets Sir4 degradation via the proteasomal pathway. We failed to detect Sir4 ubiquitination when we immunoprecipitated myc-Sir4 from strains expressing TD (data not shown). In our immunoprecipitation experiment, however, myc-Sir4 was
overexpressed, and the failure to detect Sir4 ubiquitination could result from changes in silent chromatin structure due to overexpression. Our model is that TD enhances the sumolyation of Sir4, which is then sensed by Ris1, leading to Sir4 ubiquitination and subsequent degradation.

Recently, Slx5/Slx8 complex, another protein complex in sumolyation dependent ubiquitin ligase physically associated with nuclear pore in *Saccharomyces cerevisiae*, selectively ubiquitinated and degraded the yet unknown protein bound at collapsed fork or resected breaks to facilitate the collapsed replication fork recovery (Nagai et al., 2008). Although the substrates of Slx5/Slx8 has not yet been identified, clear parallels can be drawn between Slx5/Slx8 and Ris1 in terms of mediating ubiquitination and proteasomal dependent degradation through sensing the change in sumolyation of substrate proteins.

*Ris1 and yeast cell growth*

Slx5/Slx8 has been shown to interact with Sir2 (Darst et al., 2008). Mutations in components of the Slx5/Slx8 complex are defective in telomeric silencing, and cells exhibit a growth defect. This growth defect is thought to result from the accumulation of excess sumolyated proteins in cells, since mutants in both *slx5* and *ulp2*, a desumolyation proteinase, cause the accumulation of sumolyated conjugates. Mutations in *RIS1* do not disrupt silencing nor do they cause a growth defect. Interestingly, expression of TD in a *ris1* mutant causes a growth defect compared to a *ris1* mutant expressing td (Fig.3D). One explanation is that TD stimulates Sir4 sumolyation, and in the absence of Ris1, excess sumolyated Sir4 accumulates in cells to cause the growth defect. Consistent with this hypothesis, deletion of *SIR2*, *SIR3* or *SIR4* from a *ris1* delta strain abrogates the growth defect (data now shown). This result suggests that like Slx5/Slx8, Ris1 may promote cell viability by preventing excessive accumulation of sumolyated conjugate.

Esc1 is a protein that anchors Sir4 to the nuclear periphery and contributes to telomeric silencing. We reported a small motif of Esc1 that is functionally equivalent to Ty5 TD in terms of its ability to interact with Sir4 and direct Ty5 integration. When
overexpressed, the Esc1 TD breaks silencing at telomeres and \textit{HM} loci. Like Ty5 TD, the Esc1 TD also requires Ris1 and Ubc4 to break silencing. Esc1 was recently shown to regulate the accumulation of SUMO conjugates (Lewis et al., 2007). Esc1 mutants partially suppress the growth defect caused by deletion of ulp2, which results in the accumulation of SUMO conjugates in the nucleus. The ability of esc1 mutants to suppress ulp2 deletions is independent on the status of telomeric silencing and is due to mislocalization of ulp1, another desumolyating enzyme that normally associates with nuclear pore complex. It is also possible that in the \textit{esc1} mutant, the accumulation of sumo conjugates is reduced, although no obvious change in HMW-SCs was observed in esc1, ulp1 mutant and esc1, ulp2 mutant in comparison to ulp1 and ulp2 single mutants. Esc1 could regulate sumolyation of low abundance proteins such as Sir4, and therefore a slight reduction in levels could be obscured by other highly abundant SUMO conjugates. It is likely that Esc1 contributes to the sumolyation of Sir4 and enhanced sumolyated targets Sir4 for degradation by Ris1.

\textit{The role of TD in Ris1-mediated Sir4 turnover}

It is important to note that Ris1 interacts with Sir4 via its N-terminal domain, a characteristic of interactions between ubiquitin ligases and their targets. However, full-length Ris1 only weakly interacts with Sir4. It is thought that the Sir4-interacting activity of Ris1 is normally masked, and events such as DNA damage or mating type switching activate the interaction between Ris1 and Sir4. In our study, Ris1-mediated Sir4 degradation depends on the interaction between TD and Sir4; td does not interact with Sir4 and does not cause Sir4 degradation. We speculate that TD interacts with Sir4 and increases the sumolyation of Sir4 or induces a conformational change in Sir4 or Ris1 that consequently stimulates Sir4 degradation by Ris1.

We previously reported that Ty5 with a td mutant has a 20-fold decrease in the efficiency of targeted integration (Gai and Voytas, 1998). It is possible that Ty5 integration into silent chromatin is facilitated by Ris1-mediated Sir4 turnover. If Ris1 is required for Ty5 integration, this may explain the mechanism by which Ty5 integrates into highly compact heterochromatin. However, we saw neither a defect in Ty5 integration efficiency nor a
change in integration preference in ris1 delta strains (data not shown). One explanation is that some unknown host factor plays a redundant role in facilitating Ty5 integration. HIV integrase interacts with the chromatin tethering factor LEDGF/p75. Like the interaction between Ty5 integrase and Sir4, the HIV integrase-LEDGF/p75 interaction explains HIV integration site preference. Moreover, the integrase-LEDGF/p75 interaction stabilizes integrase and protects integrase from degradation via the ubiquitin-proteasome pathway. Our study and the HIV studies both suggest ubiquitination plays an important role during the integration of retroelements.

Esc1 is well known to be a nuclear anchor which links Sir4 mediated silent chromatin to nuclear periphery (Gartenberg et al., 2004; Taddei et al., 2004). Therefore, Esc1 promote the Sir4 mediated silencing. How do we explain the contradictory roles of Esc1 in silencing? The previous study suggested that the anchorage of silent chromatin into nuclear periphery is dynamic and cell cycle regulated (Gartenberg et al., 2004). Esc1 mediated anchorage of silent chromatin into nuclear periphery peaks in S-phase, while yKu70/yKu80 pathway predominates in G1 phase. Ris1 is required for mating type switch, which requires the change of silent chromatin. In our study, our result that Esc1 had a motif to break silencing in Ris1 dependent manner suggest that it is likely that Ris1 helps Esc1 regulate silent chromatin to facilitate mating type switch. It is also imaginable that Ris1 helps Esc1 regulate silent chromatin during cell cycle and in response to stress conditions.

In addition, our previous study suggested that Esc1, host cellular factor, is imitated by Ty5 retrotransposon to acquire the integration bias into silent chromatin via the interaction of TD domain and Sir4. Current study demonstrated that Esc1 TD like domain, like Ty5 domain, regulated the silent chromatin in Ris1 and Ubc4 dependent manner. This similarity consolidated our hypothesis that Ty5 imitated Esc1 to obtain the domain required for integration preference into silent chromatin.
EXPERIMENTAL PROCEDURES

Plasmids and Strains

Plasmid construction. GAD-TD: GAD-TD was constructed by ligating an EcoRI/BglII fragment containing TD from pWW45 into pGAD-C3 (James et al., 1996).

SIR4 promoter construct: The SIR4 promoter was PCR amplified from Saccharomyces cerevisiae genomic DNA using oligonucleotides 2424 and 2425, and ligated into the SacI/PstI sites of pRS424, yielding pPF225 (Sikorski and Hieter, 1989). LacZ was PCR amplified from pRS404 using oligonucleotides 2426 and 2427 and ligated into the PstI/XhoI sites of pPF225 yielding pPF226. SIR4 construct: A SacI/ClaI fragment containing SIR4 under its native promoter was cut out of eSZ269 and ligated into pRS414 (Xie et al., 2001a).

RIS1 construct: RIS1 was amplified from S. cerevisiae genomic DNA using primers 2308 and 2309. The PCR product was digested with NotI and Acc65I and ligated into the same sites in pRS414 to yield pPF231. A c-Myc tag was added to the C-terminus of RIS1 by amplifying the RIS1 C-terminus with primers 2639 and 2641 and ligating this fragment into the BglII and SpeI sites in pPF231 to yield pPF233. RIS1 RING finger mutations: Two rounds of PCR were used to mutate the RING finger motif of RIS1. The first reaction used oligonucleotides 2677 or 2679 and 2676 alongside 2680 or 2678 and 2639. The PCR products (2677-2676 and 2678-2639 or 2679-2676 and 2680-2639) were purified and used as template for a second round using oligonucleotides 2639 and 2676. PCR products were purified, digested and ligated into the BglII/SpeI sites of pPF234 yielding pPF235 (ringΔ4; C1330A, C1303A, C1320A and C1322A) and pPF236 (ringΔ2; C1382A and C1385A).

Expression of Ris1 was not detected by immunoblot analysis when strains were transformed with low copy number plasmids. However, equivalent levels of Ris1 and Ris1 RING finger mutant alleles were observed when expressed from a high copy number plasmid constructed by ligating RIS1 and adjacent sequences into the NotI and Acc65I sites of the high copy vector pRS424. C-Myc tagged SIR3 allele: A c-Myc tagged SIR3 allele was constructed by PCR amplifying the C-terminus of SIR3 with oligonucleotides 2296 and 2297. The PCR product was purified and digested with XhoI and ligated into pPF186, which was constructed by ligating the SIR3 containing NotI/SalI fragment from pYZ1 into pRS316 (Zhu et al.,
1999). Wild type function of the c-Myc tagged *SIR3* allele was verified by complementing the telomeric silencing and mating defects of a *sir3Δ* strain with the c-Myc *SIR3* allele.

**GST-*RIS1* expression constructs:** The plasmid pJQ58 (GST-*RIS1* R) was made by using primers 4268 and 4269 to amplify the RING finger of *RIS1*, corresponding to amino acids 1290-1445, from *S. cerevisiae* genomic DNA. The PCR product was digested with BamHI and SalI and ligated into the same sites of pGEX-4T1. The *Ris1C1348S* and *Ris1C1385S* alleles were constructed by mutagenic PCR. Primers 1310 and 1311 or 1312 and 1313 were used to introduce the C1348S or C1385S mutations in a first round of PCR. Amplification products were used as template in a second round of PCR using oligonucleotides 4268 and 4269.

**SAN1** construct: The *SAN1* gene was PCR amplified using oligonucleotides 3637 and 3638 and ligated into the BamHI and SalI sites of pGEX-4T1 to yield pJQ5.

**Strain Construction. YPF331:** YPF331 was made by mating UCC3505 to BY4742 *ris1Δ* and sporulating the resulting diploids. Tetrads were isolated and tested for the desired genotype. *Sir* knockouts: *Sir* knockouts were made by amplifying selectable markers from the pRS series of plasmids with *SIR2*, *SIR3* or *SIR4* flanking sequences using the following oligonucleotides: *SIR4*, 1202 and 1203 (kan-mx); *SIR2*, 2294 and 2295; *SIR3*, 2240 and 2241 and *SIR4* 2011 and 2012 (pRS auxotrophic markers). Purified PCR product was transformed into desired strains and recombinant cells were selected by auxotrophy or G418 resistance. *SIR* knockouts were confirmed by Western blot analysis, PCR analysis and the inability to mate followed by the restoration of mating by transformation of an appropriate *SIR* plasmid. Other deletion strains used in this study were generated in the yeast gene deletion project (Winzeler et al., 1999).

**Western blot analysis.**

Strains were transformed with appropriate plasmids and grown overnight in 1 ml selective media and diluted with 1 ml YPD in the morning. After 3 hours growth cell density was measured and equal volumes (about 600 μl) were centrifuged, washed and resuspended in 40 μl sample buffer. Cells were boiled for five minutes and 15 μl of suspended solution was loaded. Standard protocols were used to transfer and visualize the immunoblot (Ausubel et al., 1987). Three independent transformants were tested for each protein assayed. All
antibodies used in this study were acquired from Santa Cruz Biotechnology.

**Northern blot analysis.**

Total yeast RNA was isolated from strains expressing GBD-TD and GBD-td according to manufacturer’s instructions (Ambion). 20 μg RNA was loaded and run on an 8% agarose gel and transferred to a nylon membrane. The membrane was prehybridized in Church buffer for 30 minutes at 65°C and hybridized overnight in the same buffer. Sir4 mRNA was detected using a probe labeled with \([\gamma^{-32P}]\) CTP to the C-terminus of Sir4p, amplified with oligonucleotides 1194 and 1195. The membrane was washed twice sequentially in 2x SSC 0.1% SDS, 0.5xSSC 0.1% SDS and 0.1x SSC 0.1% SDS for fifteen minutes.

**Transcription analysis.**

PPF226 was transformed into YPF237. LacZ levels were determined in triplicate by assaying ONPG hydrolysis as previously described (Current Protocols in Molecular Biology: Chapter 13; basic protocol 6.2) (Ausubel et al., 1987).

**In vitro ubiquitinylation assay.**

The *in vitro* autoubiquitinylation assay was performed by testing the ability of purified protein to attach ubiquitin to itself as previously described (Lorick et al., 1999). Each 20 μl reaction containing 500 ng human recombinant UbcH5a, 100 ng ubiquitin activating enzyme, 2.5 μg ubiquitin, 1 μl ubiquitinylation buffer (50 mM Tri.Cl, pH 7.5, 2.5 mM MgCl₂, 0.5 mM DTT and 2 mM ATP), and 5 μl of purified GST-Ris1 R, was incubated at room temperature for 2 hours. Equal volume of SDS loading buffer was added to each reaction and boiled for 5 minutes. Half of the reaction was run on 8% SDS-PAGE gels, and probed with antibodies to GST (1:3000) and ubiquitin (1:400). Human recombinant UbcH5a, ubiquitin activating enzyme, ubiquitin were purchased from Boston Biochem. The GST-SAN1 plasmid was generously provided by Dr. David Auble.

**Yeast culture and dilution dot assay.**
Yeast strains were grown in SC media supplemented with required amino acids. Strains were transformed using a lithium acetate heat shock protocol (Current Protocols in Molecular Biology: Chapter 13) (Ausubel et al., 1987). Yeast strains were grown to an O.D. 600 of 0.6. 10 fold serial dilutions were made in SC media lacking amino acids (SC). 5 μl of each dilution was spotted onto an appropriate culture plate and incubated 1-2 days at 30° C.

ACKNOWLEDGEMENTS

We thank Dan Gottschling, Guido Cuperus, and David Auble for providing some of the strains used in this study. We appreciate Zhiguo Zhang for generously offering us anti-Sir4 antibody used in western blot analysis.

This study is supported by NIH 420-22-44.

REFERENCES


CHAPTER 4. DNA DAMAGE BLOCKS INTEGRATION OF THE YEAST RETROTRANSPOSON TY5

A manuscript to be submitted to Mobile DNA
Jiquan Gao¹ and Daniel Voytas²

ABSTRACT

Integration of cDNA into the host genome is a critical step in the retroelement life cycle. It has been observed that some environmental conditions influence the efficiency of integration, and some host factors regulate integration. Here we report that integration of the Saccharomyces cerevisiae retrotransposon Ty5 is blocked by DNA damage produced by treatment with hydroxyurea or methyl methanesulfonate (MMS). Ty5 cDNA recombination with the parental plasmid, however, was not affected by treatment with either drug. The integration blockage was not due to cell cycle arrest, since treatment with the mating pheromone alpha factor, which arrests the cell cycle in G1, had no effect on Ty5 integration. Furthermore, Ty5 integrase, reverse transcriptase, and Gag protein levels were unaffected by treatment with either hydroxyurea or MMS. Since Ty5 cDNA recombination with the parental plasmid occurred at a normal frequency, we predict Ty5 cDNA levels are not affected by DNA damage. Moreover, we showed that Sml1, a downstream substrate of the DNA damage response pathway, was required for Ty5 integration, whereas Dun1 and Chk1 were dispensable. Strains with deletions in mec1 and rad53, key regulatory kinases involved in the DNA damage response, showed enhanced Ty5 cDNA recombination in combination with a sml1 deletion. It is known that a sml1 deletion releases repression of ribonucleotide-disphosphate reductase, resulting in higher cellular dNTP pools. We speculate that dNTP pool levels or a yet unknown function of Sml1 regulates Ty5 cDNA integration.

¹ Primary researcher and author, Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA 50011
² Professor and corresponding author, Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN 55455
INTRODUCTION

Retrotransposons comprise a class of transposable element that moves from one genomic locus to another through a process that involves an RNA intermediate (Craig, 2002). Retrotransposons consist of non-LTR retrotransposons and LTR retrotransposons, depending on whether the genome of retrotransposons is flanked by long terminal repeats (LTRs). In *Saccharomyces cerevisiae*, there are five distinct retrotransposon families, designated Ty1, Ty2, Ty3, Ty4 and Ty5 (Boeke and Devine, 1998).

The structure of the retrotransposon genome is similar to retroviruses, except retrotransposons do not have an *env* gene. Both retrotransposons and retroviruses encode Gag, integrase (IN), reverse transcriptase (RT), protease and Rnase H. Like retroviruses, LTR retrotransposons undergo four major steps during their life cycle: transcription, particle formation, reverse transcription, and integration. LTR retrotransposons are first transcribed into mRNA by the host RNA Pol II transcription machinery. The mRNA is exported into the cytoplasm where it is translated into proteins, which assemble with mRNA to form virus-like particles (VLPs). Within the VLP, the mRNA is reverse transcribed to a linear, double-stranded cDNA by RT. cDNA associates with IN to form a pre-integration complex (PIC), which will be imported into the nucleus and mediates cDNA integration into the chromosome. Recent studies have shown that cDNA levels are the limiting factor for the transposition efficiency of retrotransposons, whereas integration is the most important step in terms of the impact on genome integrity of the host cell. Therefore, understanding of integration mechanisms and integration site preference can help understand and control the genetic impact of both retrotransposons and retroviruses.

cDNA integration is catalyzed by IN, which binds both LTRs at the end of cDNA. IN creates a nick at the target DNA and inserts cDNA at the break site, leaving a single strand gap at both ends. The gaps are filled by host DNA repair machinery, which generates the target site duplication (TSD), a hallmark of integration. How the host DNA repair machinery completes the integration reaction is not clearly defined. Recent studies on Ty1 and HIV integration suggest that components of the DNA damage response system might contribute to
DNA repair at the gap created by integrase. For instance, one study indicated that although Mec1 and Tel1 – the yeast homologues of ATM and ATR kinases that sense DNA damage and trigger the DNA damage response – are dispensable for Ty1 integration. However, Ku proteins, involved in DNA double strand break repair by non-homologous end-joining (NHEJ), are important for integration. The role of yKu in Ty1 integration is supported by data indicating that neither Ty1 transcription nor VLP formation is affected in the absence of Ku, and Ku is associated with Ty1 cDNA and Ty1 integrase in the Ty1 preintegration complex (Downs and Jackson, 1999). Consistent with the role of yKu in Ty1 integration, Ku proteins are also required for HIV integration (Daniel et al., 2003). The results together suggest that NHEJ might repair the DNA gap created by cDNA integration. In support of this model, other components of NHEJ in humans, including DNA-PKcs, XRCC4 and ligase IV, are required for efficient transduction by both HIV-1-based and MLV-based vectors (Daniel et al., 2003; Smith and Daniel, 2006). Further study showed that ATM and ATR are also important for HIV integration, and the ATM/ATR inhibitor caffeine and caffeine-related methylxanthines inhibit HIV infection (Daniel et al., 2003; Lau et al., 2005; Nunnari et al., 2005). However, one recent study suggests NHEJ is dispensable in retroviral integration, making the role of NHEJ controversial (Ariumi et al., 2005). Experimental differences, such as the retroviral vector used, the transduced reporter genes, and the promiscuous activity of inhibitors and genetic techniques used to down-regulate the expression of NHEJ genes, may explain the discrepancy of conclusions drawn from different studies (Smith and Daniel, 2006).

It has been proposed that many proteins related to DNA repair also contribute to the regulation of the mobility of retrotansposons and retroviruses. For instance, mutations in XPB, XPD, RAD18 and RAD52 increase the efficiency of retroviral transduction (Lau et al., 2004; Lloyd et al., 2006; Yoder et al., 2006). Mutations in the yeast homologues of these four genes, SSL2, RAD3, RAD18 and RAD52, respectively, also increase the transposition efficiency of native Ty1 elements (Curcio et al., 2007; Lee et al., 2000; Rattray et al., 2000). The role of these four genes was suggested to act at the stage of preintegration, since their mutations increase the level of retroviral cDNA and Ty1 cDNA, often a limiting factor for retromobility.
Cell cycle arrest and DNA integrity are closely related, because it is imperative to leave adequate time during cell division for DNA repair. Cell cycle progression also determines chromatin status, which is recognized by retroelements upon integration. Chromatin status also impacts the host cell’s physiological state relevant to DNA synthesis and repair. It is highly plausible, therefore, that retroelement integration is also regulated by and adapted to cell cycle progression. For instance, integration of both Ty1 and Ty3 are blocked when cells are arrested at G1 phase by treatment with the mating pheromone α-factor (Menees and Sandmeyer, 1994; Xu and Boeke, 1991). This integration deficiency is due to reduced cDNA levels, since transcription and VLP assembly are intact. The restriction of retroviral infection also occurs when host cells are arrested in their cell cycle. The infection of murine leukemia virus (MLV) is blocked during G1 and also shows restriction in S and G2 phases (Richter et al., 1984; Roe et al., 1993). The infection of RSV is restricted when the host cell is arrested at G1 phase (Hsu and Taylor, 1982). In both cases, the cDNA is normally made but integration is blocked. The blockage of MLV infection during G1, S and G2 phase results from the requirement of nuclear entry of the MLV preintegration complex, which only occurs before the cell completes M phase. For Rous sarcoma virus (RSV), it is not known how integration is restricted. On the other hand, cell arrest at G2/M phase may promote the integration of retroelements, due to some characteristic of the cellular environment, such as an increase in the pool of deoxynucleotides or the presence of a newly replicated DNA. For example, Ty1 cDNA levels and retromobility are increased when the yeast cell cycle is arrested at G2/M upon treatment with nocodozole, or when the yeast cell cycle is delayed at G2/M by est2 mutations (Scholes et al., 2003). Data derived from human immunodeficiency virus-1 (HIV-1) infection support that the cell cycle at G2/M promotes HIV-1 infection by increasing cDNA levels (Groschel and Bushman, 2005).

Ty5 preferentially integrates into heterochromatin at the telomeres and silent mating loci (HM loci) (Zhu et al., 1999; Zou and Voytas, 1997). Heterochromatin is thought to define a special subnuclear compartment that is highly compact and anchored to the nuclear periphery through the interaction between Sir4, a heterochromatic protein, and Esc1, a nuclear periphery protein or the yKu70/80 complex (Gartenberg et al., 2004). The anchorage of heterochromatin to the nuclear periphery is regulated by cell cycle and DNA damage. In
this study, we explored the possibility that cell cycle arrest and DNA damage treatment might also regulate Ty5 transposition efficiency and affect Ty5 integration site preference. We found that Ty5 integration is independent of cell cycle, since Ty5 can still transpose in host cells arrested in G1 phase by mating the pheromone α-factor. DNA damage induced by hydroxyurea and MMS, however, can inhibit Ty5 integration. DNA damage does not affect Ty5 cDNA levels, because Ty5 cDNA recombination occurs normally and the stability of Ty5-encoded proteins, including Gag, integrase, and reverse transcriptase, is unaffected. Furthermore, we showed that Sml1, the repressor of ribonucleotide-disphosphate reductase, is required for efficient Ty5 integration, whereas loss of Mec1 and Rad53 enhances the recombination of Ty5 cDNA with the parental plasmid in the absence of Sml1. It is known that a sml1 deletion releases repression of ribonucleotide-disphosphate reductase, resulting in higher cellular dNTP pools. We speculate that dNTP pool levels or a yet unknown function of Sml1 regulates Ty5 cDNA integration.

RESULTS

DNA damage can block Ty5 integration, but not recombination.

It has been suggested that the integration of retrotransposons and retroviruses is regulated by the cell cycle and DNA repair pathways. To test the effect of these signaling pathways on Ty5 integration, we used the well-characterized Ty5 transposition assay based on pGalTy5-His3AI (Zou and Voytas, 1997) (Fig. 1). The assay uses a GAL1 promoter-driven Ty5 donor element with a His3AI marker inserted between the end of reverse transcriptase and the 3’ LTR. Upon galactose induction, transcription of Ty5-His3AI is activated. Subsequently, the artificial intron (AI) is spliced out from the Ty5-His3 transcript. Hence, reverse transcribed cDNA will reconstitute a Ty5 element and a functional HIS3 gene. Once the newly-synthesized cDNA either integrates into the host genome or recombines with the parental plasmid, His+ cells are formed. Cells with plasmid recombination events, however, cannot grow on SC-H+5FOA media, due to the presence of the URA3 gene on the plasmid (i.e. selection for HIS3 retains the plasmid).
Fig. 1. A genetically marked derivative of Ty5 for assaying transposition. The Ty5 donor element is constructed on a *URA3*-based vector under the control of the *GAL1* promoter. Upon galactose induction, Ty5 cDNA is reverse transcribed using mRNA from the donor element, which reconstitutes a functional *HIS3* gene. The cDNA can either integrate into the host genome or recombine with the donor element to give rise to His⁺ cells. Cells in which Ty5 has recombined with the donor element cannot grow on SC-H+5FOA media, due to the presence of the *URA3* gene. Overall Ty5 transposition rate = total His⁺ cells / total cells. Ty5 integration rate = His⁺ and 5-FOA⁻ cells / total cell amount. Ty5 recombination rate = His⁺ and 5-FOA⁺ cells / total cells.

The efficiency of Ty5 mobility will be calculated as the ratio of the total His⁺ cells versus the total cell population. Integration efficiency can be measured as the ratio of the subpopulation of 5FOA resistant and His⁺ cells versus the total cell number. Recombination efficiency is the ratio of 5FOA sensitive and His⁺ cells versus the total cell amount. One advantage of this assay is that we can specifically monitor the transposition events in the pretreated cell population. The second advantage is that the galactose-inducible system gives rise to a relatively constant amount of Ty5 transcript under diverse treatments.

We treated early log phase yeast cells containing pGalTy5-*His3AI* and growing on raffinose-based growth medium with the mating pheromone α-factor or hydroxyurea (HU) or MMS. The data from flow cytometric analysis suggested that the yeast cells were arrested at
G1, S and S phase by α-faror, HU and MMS, respectively (Fig. 2). After a 3 hour treatment with these drugs, we added galactose to 2% to induce Ty5 transposition while constantly treating the yeast cells with the corresponding drugs at the same dosage. The result from the flow cytometric analysis confirmed that the yeast cells were constantly arrested during the induction of Ty5.

In contrast to the great repression of Ty1 and Ty3 transposition by mating pheromone, the transposition of Ty5 only showed two-fold reduction by mating pheromone. The integration efficiency of Ty5 was reduced proportionally. Our results suggest that Ty5, unlike Ty1 and Ty3, can still efficiently integrate in yeast cells arrested at G1 phase. Interestingly, we found that when the yeast cells were arrested in S phase by HU or MMS, Ty5 integration was largely inhibited, but the overall mobility of Ty5 was comparable to cells arrested in G1.

Fig. 2. DNA damage can block Ty5 integration, but not recombination. Yeast cells bearing pNK254 were grown on SC-U+raffinose to early log phase and treated with α-factor (10ug/ml), hydroxyurea (150mM), or MMS (0.02%) for 3 hours. Galactose was added to a final concentration of 2% to induce Ty5 transposition for 3 hours. A dilution of cells was plated on YEPD media to count the total cell number, and the appropriate amount of cells was plated on SC-H media to give rise to 100-200 colonies. After 3-4 days of growth, His+ cells were replica stamped onto SC-H+5FOA plates to select for Ty5 integrants. The overall Ty5 transposition rate, integration rate, and recombination rate were calculated and plotted. Below the graph, data is presented for the flow cytometric analysis. The green graph indicates pretreated cell, the blue graph indicates after 3hr treatment, but before gal induction, and the red graph indicates after 3 hours gal induction.
phase. During Ty5 transposition, cDNA levels are the limiting factor that determines transposition efficiency. In our assay, both integration and plasmid recombination require cDNA as a substrate. For Ty5, unlike Ty1, integration and recombination do not compete for the same cDNA pools. The observed high efficiency of recombination suggests that cDNA is formed normally in the presence of HU and MMS, similar to what was observed with a-factor treatment.

Ty5 protein levels and targeting specificity were not affected by DNA damage

The Ty5 genome encodes a single polyprotein, Gag-Pol, which is processed into functional units including Gag37, Gag27, integrase, and reverse transcriptase. We checked protein levels of these functional proteins when cells were subjected to DNA damage caused by HU and MMS.

![Amino black staining](image)

Fig. 3. Ty5 protein levels were not affected by DNA damage. The RGS\textsubscript{H6}-tagged Ty5 was grown using the same conditions for the transposition assay except induction on galactose was for 24 hours. Integrase, reverse transcriptase, and Gag protein levels were checked using an anti-\textsubscript{His\textsubscript{6}} antibody.
Previous studies demonstrated that a single RGSHis₆ tag can be used to label the Ty5 proteins without affecting transposition and integration. Using the antibody against the RGSHis₆ tag, we found that all of the major Ty5 functional proteins including Gag37, Gag27, integrase and reverse transcriptase are expressed at comparable levels both under DNA damage and in wild type conditions.

In addition to the effect on cell cycle progression, HU and MMS trigger the DNA damage response signaling in different ways. HU pauses DNA replication fork progression by reversibly inhibiting ribonucleotide reductase (RNR), preventing the reduction of ribonucleotides to deoxyribonucleotides. MMS mainly induces double strand breaks in the host genome.

![Diagram](image)

**A**

*Inverse PCR assay to identify Ty5 integration site*

**B**

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Tested integration events</th>
<th>Integrations into euchromatin</th>
<th>Integration into heterochromatin</th>
<th>Targeting into heterochromatin</th>
</tr>
</thead>
<tbody>
<tr>
<td>No stress</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>95%*</td>
</tr>
<tr>
<td>HU</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>100%</td>
</tr>
<tr>
<td>MMS</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>80%</td>
</tr>
</tbody>
</table>

Fig.4. Ty5 still targets into heterochromatin under HU and MMS. (A) Inverse PCR assay to identify Ty5 integration site. (B) The randomly tested integration of Ty5 under HU and MMS still occurs at heterochromatin.
In *Saccharomyces cerevisiae*, MMS induces the translocation of the heterochromatic protein Sir4 and Ku70 from telomeres to sites of double strand breaks. This translocation presumably helps the repair of double strand breaks. Since Sir4 is essential and sufficient to direct Ty5 targeted integration, we explored the possibility that Ty5 targeted integration into heterochromatin is also changed in parallel with the translocation of Sir4 induced by DNA damage. To our surprise, most Ty5 integration still occurred at heterochromatin under both HU and MMS treatment. This result suggests that Ty5 integration preference is not affected by DNA damage.

*SML1 is required for efficient Ty5 integration*

In *S. cerevisiae*, HU and MMS trigger DNA damage response signaling pathway via the Mec1/Rad53 pathway (Sclafani and Holzen, 2007). A previous study suggested that the Ty5 integrase protein is highly serine and threonine rich and phosphorylated at multiple sites, one of which is serine 1095 within the Ty5 targeting domain. Phosphorylation at this site is responsible for Ty5 target specificity. We posited that the activation of this pathway by HU and MMS might posttranslationally modify and inactivate the Ty5 integrase protein and cause integration defects.

![Graph](image_url)  
*Fig. 4. Ty5 transposition in diverse DNA damage signaling mutants. Sml1, a repressor of ribonucleotide-diphosphate reductase, is required for efficient Ty5 integration, whereas Dun1 and Chk1 are dispensable. Interestingly, mec1 and rad53 deletions enhance Ty5 recombination in combination with an sml1 deletion. Data was generated from three independent colonies for each genetic strain carrying pNK254.*
To test the requirement of DNA signaling pathway for Ty5 integration deficiency, we deleted *MEC1* and *RAD53* in a background with an *sml1* deletion, which suppresses the lethality of *mec1* and *rad53* single mutants. We found that the *sml1* single mutant alone caused a Ty5 integration defect, but it did not affect recombination frequency (Fig.4). It is of note that Sml1 is also degraded in the presence of HU and MMS via the Mec1/Rad52/Dun1 pathway (Zhao and Rothstein, 2002). However, *mec1*, *sml1* and *rad53*, *sml1* double mutants have no added effect on Ty5 integration, but largely increase the Ty5 recombination frequency. In contrast, the deletion of Chk1, another kinase and DNA checkpoint effector, has no effect on Ty5 integration. Similarly, Dun1, the kinase that phosphorylates Sml1 to mediate its degradation has no effect on Ty5 integration. These results are consistent with previous observation that the deficiency of Ty5 integration depends on the availability of Sml1.

Finally, we also observed that the requirement of Sml1 also pertains to Ty5 elements with targeting defects. It has been shown that three Ty5 elements with substitutions in S1094L, S1095C, or S1095A randomly insert into the host genome due to

![Graph](image-url)  
*Fig. 5. Transposition of targeting defective Ty5 elements in DNA damage signaling mutants. Ty5 integrase mutants S1094L, S1095C, and S1095A are targeting defective. These mutants require Sml1 for efficient integration.*
the inability to recognize Sir4. Although these mutant Ty5 elements integrate at diverse loci in the genome, they all require Sml1 for efficient integration (Fig.5). The mec1 deletion mutant also enhances the recombination frequency of the three targeting defective mutants in combination with an sml1 deletion.

DISSCUSSION

Ty5 integration is independent of the cell cycle

Previous studies on Ty1 and Ty3 transposition showed that both retrotransposons do not transpose in host cells arrested in G1 phase (Menees and Sandmeyer, 1994; Xu and Boeke, 1991). The basis for the Ty1 and Ty3 transposition defect is repressed cDNA synthesis. The experiments with Ty1 and Ty3, together with this Ty5 study, all used elements expressed from the GAL1 promoter, and induction of Ty transposition was typically done after the host cells had been effectively arrested at G1 phase by the mating pheromone α-factor. In contrast to Ty1 and Ty3, Ty5 integration is only modestly restricted in the yeast cells arrested in G1.

Cell cycle arrest also does not affect the recombination frequency between Ty5 cDNA and the parental plasmid. This indicates that the formation of Ty5 cDNA, unlike Ty1 and Ty3, is normal in G1 arrested cells. The mechanism by which the cell cycle mediates its opposite effects on Ty5 vs. Ty1 and Ty3 transposition is not known. Three possible features may help explain the different effects of the cell cycle on these retrotransposons: (1) Ty5 reverse transcriptase may have an intrinsic feature that is resistant to effects of the cell cycle; (2) Although Ty5 transcription is driven from the same GAL1 promoter as Ty1 and Ty3, Ty5 transcription is much lower, and this may help Ty5 escape the effects of the cell cycle better than Ty1 and Ty3; (3) Ty1 and Ty3 integrate to sites of Pol III transcription, whereas Ty5 integrates into silent chromatin (Boeke and Devine, 1998). Although none of the above features provide an obvious explanation for the differences in cell cycle regulation, the first is the most compelling. Perhaps Ty1 and Ty3 reverse transcriptase is post-translationally modified by signaling pathways responsive to cell cycle. Mass spectrometry of reverse
transcriptase to identify such modifications may offer one experimental approach to explore this hypothesis.

*Ty5 integration is sensitive to DNA damage*

Ty5 integration is not affected by the mating pheromone a-factor, but is repressed by HU and MMS, which cause two different types of DNA damage, namely replication fork stalling and DNA double strand breaks, respectively (Alvino et al., 2007; Choy and Kron, 2002; Ui et al., 2005). Both forms of damage result in arrest of the cell cycle at S phase. When yeast cells are treated with HU and MMS, the Mec1/Rad53/Dun1 signaling pathway is activated. This leads to the phosphorylation and consequent degradation of Sml1, which, in turn, releases the repression of ribonucleotide reductase and results in the increased cellular dNTP pool levels needed to facilitate DNA repair (Zhao and Rothstein, 2002). We found that an *sml1* single deletion indeed represses Ty5 integration, and Mec1, Rad53 and Dun1 have no obvious effect on Ty5 integration. The Sml1 requirement for Ty5 integration applies to three Ty5 elements that are defective in targeting. Ty1 integration is also independent of Mec1.

To date, the only reported effect of an *sml1* deletion is an increased level in the dNTP pools (Zhao et al., 1998). This occurs because Sml1 negatively regulates ribonucleotide reductase. To test whether the levels of dNTP pools directly affect Ty5 integration, we intend to overexpress *RNR1*, the gene coding for ribonucleotide reductase. This experiment should help determine if Sml1 acts through RNRI or has some independent, yet undescribed function. If the overexpression of RNR1 fails to regulate Ty5 integration, it is likely that the repression activity of sml1 on Ty5 integration could result from some yet unknown function. We plan to do genetic studies to map the region of Sm1 responsible for this repression on Ty5 integration. It will be also interesting to know whether the sml1 deletion and the level of dNTP pools repress the integration of Ty1, Ty3 and the retroviruses.
EXPERIMENTAL PROCEDURES

Plasmids and strains

Plasmids. The pGALTy5HisAI plasmid used in this study is pNK254 (Zou and Voytas, 1997). This plasmid has been demonstrated for Ty5 integration preference and widely used in diverse Ty5 studies. pWW32 encodes IN RGS-His6 tagged version of Ty5 derived from pNK254. Similarly, RGS-His6 epitope is used to tag RT, Gag 37 and Gag27 proteins of Ty5 to generate pIP37, derivative of pNK254 (Haase and Reed, 2002; Irwin and Voytas, 2001).

Strains. The yeast strain yPH499 was used as wild type strain in this study. All of the deletion strains derived from yPH499 were generated by one-step gene knockout method, using plamids pFA6a-KANMX and pFA6a-HphNT1 to amplify fragments containing genes for G418 and Hygromycin B resistance, respectively. The deletions were confirmed by phenotypic growth and PCR analysis.

Cell cycle arrested Ty5 transposition assay.

The yeast strain containing pNK254, grown in SC-U+Raffinose liquid medium up to early log phase (OD600=0.6), was treated by 10ug/ml alpha-factor, 150mM HU and 0.02% MMS for 3 hours. The yeast cells was collected, resuspended into SC-U+Galactose with addition of fresh drugs including 10ug/ml alpha factor, 150mM HU and 0.02% MMS and allowed to grow for additional 3 hours to induce Ty5 transposition. The appropriated portion of yeasts before galactose induction and at the end of galactose induction was subject to flow cytometric analysis described by S. Haase (Haase and Reed, 2002). The appropriate dilution of cells was plated on YEPD media to count the total cell number, and the proper amount of cells was plated on SC-H media to give rise to 100-200 colonies. After 3-4 days of growth, His+ cells is replica stamped onto SC-H+5FOA plate to select for Ty5 integrants. The overall Ty5 transposition rate, integration rate, and recombination rate were calculated accordingly.

Western analysis and inverse PCR analysis were performed according to the method described before (Irwin and Voytas, 2001; Zou and Voytas, 1997).
ACKNOLEGEMENT

We thank Dr. Shawn Rigby for helping us with yeast flow cytometric analysis.

REFERENCES


CHAPTER 5. GENOME-WIDE ANALYSIS OF INTEGRATION SITES OF THE YEAST RETROTRANSPOSON TY5

A manuscript to be submitted to Genetics
Jiquan Gao1, Joshua Baller2, Nirav Malani3, Frederic Bushman3 and Daniel Voytas2

ABSTRACT

Integration of reverse transcribed cDNA into the host genome is one of the critical steps in the life cycle of retroelements. During integration, many yeast LTR retrotransposons, including Ty1, Ty3, and Ty5 from Saccharomyces cerevisiae and Tfl from Schizosaccharomyces pombe, show strong integration preferences according to randomly surveyed integration events. However, a global view of the distribution of yeast retrotransposon integration sites is lacking. In this study, we applied linker-mediated PCR followed by high throughput pyrosequencing to identify the genome-wide integration pattern of the yeast retrotransposon Ty5. Our further bioinformatic analysis revealed that Ty5 integration strongly favors silent chromatin at the telomeres and HM loci, in accordance with previous studies. In addition, the global analysis of Ty5 integration revealed several integration hotspots in rDNA and euchromatin regions. Interestingly, mapped Ty5 integrations in all the above regions showed strong biases for intergenic regions. The analysis of Ty5 integration patterns in diploid cells showed the same bias to integrate outside of coding sequences. Therefore, our findings support the “safe haven” hypothesis, namely that retrotransposons actively select gene-poor regions as integration sites to avert the potential deleterious effect of integration on the host. Furthermore, the experimental and biocomputational methods described in this study for the first time demonstrate the comprehensive, genome-wide mapping of retrotransposon insertion sites with single base pair resolution, an approach that will be of great value for analyzing and understanding integration site preferences of other retrotransposons.

1 Primary researcher and author, Iowa State University, Ames, IA 50011
2 University of Minnesota, Minneapolis, MN 55108
3 University of Pennsylvania, Philadelphia, PA 19104
INTRODUCTION

Retrotransposons make up a large portion of eukaryotic genomes as exemplified by the non LTR retrotransposon L1 and Alu elements, which constitute 30% of the human genome. During the life cycle of retrotransposons, integration of reverse transcribed cDNA into the host genome is one of the essential steps required to establish these genomic parasites and enable them to be vertically inherited through transmission to progeny. However, random integration could be potentially deleterious to the host by causing insertional mutations, introducing regulatory elements to genes adjacent to integration sites, and repressing the expression of adjacent genes. Thus, the retrotransposon and host cell have co-evolved multiple strategies to regulate the integration of retrotransposons that mutually benefit the survival of both.

One strategy adopted by retrotransposons and the host cell is to direct integration into gene-poor regions of the host genome, namely “safe havens”, to avert catastrophic insertion events (Boeke and Devine, 1998). There are now spectacular examples that support the “safe haven” hypothesis. For instance, Ty1 and Ty3, LTR retrotransposons in *Saccharomyces cerevisiae*, preferentially integrate upstream of PolIII transcribed genes, whereas Ty5 favors silent chromatin that includes telomeres and *HM* loci as integration sites. The Tf1 LTR retrotransposon of *S. pombe* preferentially jumps upstream of PolII genes. These observed integration preferences are concluded from small numbers of randomly selected insertions that have been surveyed. Thus, the small-scale integration surveys might mask some potential integration hotspot due to the limited size of the integration dataset. Furthermore, the small-scale survey likely fails to show the subtle changes in integration pattern that may occur under some particular conditions such as environmental stress. Therefore, it is increasingly desired to develop an approach to analyze the genome-wide integration pattern of retrotransposons.

Recently, microarray-based methods have successfully been applied to identify the integration sites of transposable elements, including Ty1 retrotransposons in *S. cerevisiae*, artificially introduced bacterial transposons in *S. cerevisiae* and IAP retrotransposons in mice
(Gabriel et al., 2006; Takabatake et al., 2008; Wheelan et al., 2006). However, several limitations of the microarray technique inhibit this method from wide application: (1) low resolution of 280bp. (2) exclusion of repeat sequences, including telomeres, rDNA, and preexisting transposable elements; (3) limited to a single sample; (4) empirical determination on the cutoff of hybridization; and (5) only a 80% positive rate for identifying new insertions. Accordingly, a novel method to accurately map the genome-wide integration pattern of retrotransposons is in high demand.

More recently, high throughput sequencing of PCR fragments representing virus-host junction DNA fragments has been applied to map the global integration sites of retroviruses including human immunodeficiency virus-1 (HIV-1), murine leukemia virus (MLV) and avian Sarcoma Leukosis Virus (ASLV) (Bushman et al., 2008; Uren et al., 2009; Wang et al., 2007; Wu et al., 2003). This method has proven strongly effective and allows: (1) the ability to map integration at 1bp resolution; (2) high throughput conferred by massively parallel sequencing; (3) capacity to analyze multiple samples with different treatments in a single sequencing run by barcoding samples individually; (4) stringent criteria applied to ensure positive results. However, the described methods that analyze data generated from pyrosequencing are constrained to those sequences that uniquely map onto the reference genome. All integration sites occurring in repeated sequences such as telomeres, rDNA, and preexisting elements are not included in the current method (Ciuffi et al., 2009; Wang et al., 2007).

The yeast retrotransposon Ty5 strongly favors integration at telomeres and \textit{HM} loci, both of which contain highly repetitive sequences (Zhu et al., 1999; Zou et al., 1996). Ty5 integration specificity is mediated through the interaction between Ty5 integrase and Sir4, a heterochromatic protein bound at both telomeres and \textit{HM} loci (Xie et al., 2001). Recent studies demonstrated that Ty5 integrase is phosphorylated at serine 1095, and this phosphorylation is essential for both target specificity and interaction with Sir4. Furthermore, nutrient starvation attenuates S1095 phosphorylation levels, and Ty5 integrates into euchromatin at modest frequencies according to integration sites surveyed under stress conditions (Dai et al., 2007). In addition, previous studies based on randomly selected
integration events revealed that chromosome III (accounting for 1/40 of the total genome) is a favored target of Ty5 integration with insertions occurring at a rate of 30%. Moreover, the right telomere of chromosome III seemed to be less favored since none of 19 Ty5 insertions at chromosome III occurred in right telomere, whereas the left telomere of chromosome III and the \( HML \) and \( HMR \) loci were hit at comparable frequencies (Zou and Voytas, 1997).

Finally, information as to whether Ty5 inserts into essential genes is unknown, since any Ty5 insertion into essential genes would have killed the haploid cells used in previous studies, thus making it impossible to recover insertion events within essential genes. Also, although it has long been known that Ty5 integrates into euchromatin at low frequencies, it is not known if there is a pattern underlying euchromatic insertions. Comprehensive understanding of Ty5 integration and regulation by stress requires an effective method to analyze the genome-wide pattern of Ty5 integration. In this study, we applied linker-mediated PCR followed by pyrosequencing to exhaustively characterize the genome-wide integration pattern of the yeast retrotransposon Ty5.

RESULTS

A global view on Ty5 integration patterns

Linker mediated PCR (LM-PCR) has proven to have high sensitivity and specificity for cloning virus-host junction DNA fragments and been successfully applied in combination with pyrosequencing to identify the global insertion sites of retroviruses including HIV, MLV and ASLV (Wang et al., 2007; Wu et al., 2003). We adopted this strategy for Ty5 integration site mapping. The schematic workflow of LM-PCR is shown in Fig1. To observe the global integration pattern of Ty5, we used a well-characterized Ty5 transposition assay to create a Ty5 integrant library composed of about 400,000 independent integrants derived from sixteen independent Ty5 transposition assays. Each Ty5 transposition assay gave rise to a pool of roughly 25,000 Ty5 integrants. Genomic DNA was prepared from the pools and
Fig 1. The schematic workflow of linker mediated PCR (LM-PCR) to clone the junction sequence adjacent to Ty5 integration sites. TaqI, one 4bp restriction enzyme, was selected as an example to illustrate the workflow of LM-PCR: the genomic DNA carrying Ty5 insertion was subjected to TaqI digestion, ligation with linker DNA, AseI digestion and nested PCR to generate amplicons carrying junction sequence adjacent to Ty5 integration site flanked by 454 sequencing primers.

subjected to the workflow involving LM-PCR to generate individually barcoded PCR product libraries (Fig 1).

Using this strategy, we generated a total of sixteen individually barcoded PCR product libraries representing eight pools each from haploid and diploid cells. The sixteen PCR product libraries were mixed together in equal proportions and subjected to pyrosequencing using the GS FLX platform. To minimize the bias introduced by restriction enzymes and to maximize the representation of distinct Ty5 integration sites, we chose two sets of restriction enzyme pairs (MspI and HinplI; AciI and TaqI) to digest the genomic DNA. Each enzyme cuts at a particular four base pair recognition site at approximately 300-1000bp intervals in the yeast genome.
A collection of over 300,000 raw sequence reads was generated through pyrosequencing. We used the INSIPID software platform to trim the raw sequences and to recover high quality matches. Typically, sequences are required to have (1) a perfect match to the terminus of the Ty5 LTR, (2) a 98% match to genomic DNA, and (3) a match to host cell DNA beginning within 3 bp of the end of the vector DNA. Trimmed sequences passing these quality controls are stored in separate databases. Table 1 shows the number of sequences remaining in each pool after each step. INSIPID divided the data into two sets, those that mapped to a unique sequence in the genome and those that mapped to multiple sequences in the genome. The multi hit sequences have significant positional overlap. In many cases the same position may be identified by many different sequences.

Table 1. Summary of the sequences processed by INSIPID

<table>
<thead>
<tr>
<th>Pool #</th>
<th>Sequences</th>
<th>Aligned</th>
<th>Unique Hits</th>
<th>Multi Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haploid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20024</td>
<td>12529</td>
<td>1609</td>
<td>5128</td>
</tr>
<tr>
<td>2</td>
<td>20343</td>
<td>12919</td>
<td>1723</td>
<td>4551</td>
</tr>
<tr>
<td>3</td>
<td>20608</td>
<td>13792</td>
<td>1478</td>
<td>5618</td>
</tr>
<tr>
<td>4</td>
<td>21366</td>
<td>13784</td>
<td>1650</td>
<td>5337</td>
</tr>
<tr>
<td>5</td>
<td>20260</td>
<td>12831</td>
<td>1619</td>
<td>4973</td>
</tr>
<tr>
<td>6</td>
<td>19291</td>
<td>12070</td>
<td>1561</td>
<td>4856</td>
</tr>
<tr>
<td>7</td>
<td>19836</td>
<td>12748</td>
<td>2077</td>
<td>5200</td>
</tr>
<tr>
<td>8</td>
<td>20374</td>
<td>12953</td>
<td>2058</td>
<td>4961</td>
</tr>
<tr>
<td>Totals:</td>
<td>162102</td>
<td>103626</td>
<td>13775</td>
<td>40624</td>
</tr>
</tbody>
</table>

| Diploid|           |         |             |            |
| 9      | 16948     | 10757   | 1014        | 5676       |
| 10     | 18445     | 11807   | 629         | 5728       |
| 11     | 17471     | 11254   | 633         | 6019       |
| 12     | 18709     | 11174   | 700         | 5305       |
| 13     | 17685     | 11300   | 687         | 5593       |
| 14     | 19955     | 12141   | 718         | 6100       |
| 15     | 19317     | 12164   | 1098        | 6080       |
| 16     | 17599     | 10508   | 654         | 5067       |
| Totals:| 146129    | 91105   | 6133        | 45568      |
Table 1 shows the number of unique positions covered by the multi hit and single hit data sets, both within pools and after combining the pools. The multi hit data made up three-quarters of the positional values. While this value is potentially inflated by the number of positions a single multi hit sequence can map to, ignoring the multi hit data could introduce significant positional bias to our analysis. According to the criteria, we mapped approximately 20,000 sequences to unique locations in the yeast genome. A total of approximately 85,000 sequences were mapped to multiple loci with the same quality. The large number of sequence aligned at multiple loci is consistent with the known observation of Ty5 targeting to regions with highly repetitive sequences.

Each pool represents an independent set of insertions and therefore the presence of insertions at the same site in multiple pools provides additional information as to the frequency of insertions in an area. Similarly, two insertions in opposite orientations are distinguishable events. With 8 pools and 2 orientations each position can have from 0 to 16 distinguishable insertions. By this measurement, approximately 12,000 sites were hit more than once.

We mapped the unique Ty5 insertions onto the yeast chromosomes as indicated in Fig2. For the insertions that mapped to multiple loci, we are uncertain about their accurate position on the genome. Because these data are important in understanding Ty5 targeting, we assigned a weight value to each location and mapped them onto the genome with green bars. Visual inspection of Ty5 integration patterns confirms that Ty5 shows a marked preference for heterochromatic regions in the telomeres and HM loci. In order to facilitate further analysis, we defined these regions as regions of preferential insertion as spanning from the ends of the chromosomes to the most distant window for which insertion density is less than 50 insertions per kb. In addition to the regions of preferential insertion at the telomeres, we identified two collections of insertions requiring further analysis. One such collection of insertions is at the rDNA locus on chromosome XII. Previous work has shown that the rDNA is a site of frequent insertions, especially in the sir4-42 genetic background (Zhu et al., 1999). The other collection is found at scattered hotspots throughout the genome. These collections provide information about the fine scale targeting of Ty5 (see below).
Fig 2. The global distribution of Ty5 integration sites on 16 chromosomes. Each chromosome is represented by a horizontal line. The Ty5 insertions derived from haploid cells are indicated above the line, whereas the Ty5 insertions derived from diploid cells are indicated below the line. The black bar stands for unique Ty5 insertions sites, and the green bar stands for Multi Hit insertions.

*Ty5 integration strongly favors the telomeres and silent HM loci*

Previously, we demonstrated that Sir4, a host factor bound at silent chromatin such as the telomeres and HM loci, predominantly mediates Ty5 integration specificity. Among the sixteen chromosomes, chromosome III is unique because it contains the silent mating loci HML and HMR, in proximity to the end of left and right telomeres, respectively. Consistent with Sir4 binding at both telomeres and HM loci, we observed that Ty5 integrations span the region from the HM loci to the end of chromosome III. Previously, we observed that 90% of
Ty5 integration occurs at a 3 Kb window centering on the autonomously replicating consensus sequence (ACS) within the telomeric X repeats and HM silencers (Zou and Voytas, 1997). Accordingly, this 3 Kb window was designated the targeting window. However, our global survey reveals that Ty5 integration can occur far beyond the previously defined window.

As typical examples, Ty5 integration pattern at the right end of chromosome III and chromosome IV are shown in Fig3. HMR is around 22.6 Kb distant from the end of the right telomere on chromosome III. This region encodes several characterized genes such as GIT, PAU3, ADH7, RDS1, AAD3 and unverified gens such as YCR99C, YCR100C, YCR101C and YCR102C.

Fig3. Ty5 insertions in right telomeres of chromosomes III and IV. The black bars represents for Ty5 unique insertions, whereas the green bars stands for Ty5 multi-hit insertions.
We found that Ty5 integration is non-uniformly enriched in the region from \textit{HMR} to the end of the telomere. Ty5 interestingly tends to avoid inserting into coding sequences. Further, Ty5 insertions are clustered upstream of coding sequence regardless of the orientation of the coding sequence. The integration pattern generated from haploid and diploid strains looks nearly same allowing us to rule out the possibility that failure to insert into genes results from selection after Ty5 transposition. This avoidance of coding sequence suggests that Ty5, in cooperation with host cell, may actively integrate into gene-poor regions.

\textit{Ty5 integration associates with rDNA}

In addition to the telomeric regions, Ty5 integration also clusters at the rDNA locus (Fig4). The rDNA is transcriptionally repressed, but typically binds Sir2 and not Sir4 (Huang, 2002). Insertions within the rDNA locus are dense. This density is accentuated by a simplified representation of the locus in the published DNA sequence. While two copies of the rDNA sequence are present in the published \textit{S. cerevisiae} genome, the estimated count is closer to 120. A previous study showed that 3\% of Ty5 insertions occur in the rDNA array and the ratio of Ty5 targeting to rDNA was increased up to 26\% in \textit{sir4-42} strains (Zhu et al., 1999). The \textit{sir4-42} allele produces a truncated Sir4 protein and causes the relocation of Sir3 and Sir4 from telomeres and \textit{HM} loci to the rDNA.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{rDNA_locus.png}
\caption{Ty5 insertions at the rDNA locus. The black bars represents for Ty5 unique insertions, whereas the green bars stands for Ty5 multi-hit insertions.}
\end{figure}
In \textit{sir4-d4}, Ty5 integrations predominantly occur at two locations within the rDNA -- one near the 5S rRNA gene and the other near the junction of the 5.8S and 25S genes. Similarly, of three wild-type rDNA insertions generated in a small-scale study, two were near the 5S and 25S hotspots. Intriguingly, we also found two clusters of Ty5 integration near 5S rRNA genes and the junction of the 5.8S and 25S genes. The combined results suggest that Ty5 insertion in the rDNA is nonrandom and rDNA may serve as backup “safe haven” to bear Ty5 insertions when silent chromatin is altered.

\textit{Ty5 integrates into particular regions in euchromatin}

Very few Ty5 insertions are found outside of silent chromatin, with the exception of a few small hotspots. We arbitrarily defined a hotspot as a continuous region, not adjacent to the silent chromatin, with an insertion density per kb greater than 5 (Table 3).
We identified 78 hotspots in euchromatin according to this criterion. Three typical euchromatic hotspots for Ty5 integration are shown in Fig. 5. Our analysis shows that Ty5 hotspots in euchromatin seldom occurred in verified ORFs and instead opted for dubious ORFs or regulatory binding regions upstream of verified ORFs. This same trend is also found in diploid cells, though to a less significant extent, likely because the diploid dataset had fewer insertions. The mechanism for the creation of Ty5 hotspots in euchromatin is not clear. However, the enrichment of Ty5 insertions in intergenic sequences supports the “safe haven” hypothesis that Ty5 tends to jump into gene-poor regions.

DISCUSSION

Ty5 target specificity

Previously, we showed that Sir4 is a host factor required for Ty5 integration into silent chromatin (Zhu et al., 1999). Furthermore, using a tethering assay, we demonstrated that Sir4 is necessary and sufficient for directing Ty5 integration (Zhu et al., 2003). According to results from ChIP-Chip experiments and other genetic experiments, Sir4 associates with telomeric regions, ranging from 600bp to 19Kb inward from chromosome ends (Lieb et al., 2001). This variable Sir4 binding region is somewhat consistent with the domain of transcriptional silencing at telomeres. Telomeric silencing was mapped to a region 3.5Kb distant from the chromosome end and became weaker around 6Kb from the chromosome end. The domain of silencing extended to 11-12Kb in telomeres containing a 6.7Kb Y’ subtelomeric repeat (Renauld et al., 1993). Although we cannot accurately report the association of Ty5 integration with Sir4 (or Sir3 or Sir2) occupancy due to the lack of a detailed landscape of Sir4 binding, the overlapping region between Ty5 integration and general Sir4 occupancy, in combination with the data showing the role of Sir4 in directing Ty5 targeting, suggests that Ty5 integration is closely related with and dependent on Sir4’s chromosomal location.

In addition to the telomeres, HMR and HML are bound by the Sir complex. At HML, Sir2 and Sir4 bind a nearly continuous region of 15kb from telomere 3L to YCL064C,
whereas Sir3 is localized to a region of 7.1kb from YCL069W to YCL064C (Lieb et al., 2001). Consistently, we observed Ty5 integration in a wide and continuous region from telomere 3L to YCL064C, the region occupied by Sir4. Similarly, Sir4 was also shown to bind the region between \textit{HMR} and right telomere of chromosome III. The ability of Sir4 to bind a broad region near the telomere and between \textit{HM} loci and the telomere explains the presence of Ty5 integration at this region. Our analysis of the \textit{HM} loci and telomeres indicates that the preferred window for Ty5 integration is much greater than the one we previously reported, namely the 3 kb window centered on the ACS in the telomeric X repeats and the silencers of the \textit{HM} loci. This definition was based on the clustering of randomly chosen Ty5 integration events and the newly described targeting window provides a more accurate description of Ty5’s preferred target sites.

It should be noted that Ty5 insertions are clustered in the intergenic regions at the telomeres and near the \textit{HM} loci. Due to the absence of fine scale Sir4 occupancy data in this region, we are unable to evaluate at a detailed level the association between Ty5 integration and Sir4 binding. Nonetheless, the Ty5 integration pattern supports the safe heaven hypothesis, namely that retroelements preferentially insert in gene-poor regions to minimize potential harmful effects on host genome integrity.

Very interestingly, the bias for intergenic regions is also observed when Ty5 jumps into the rDNA locus and the occasional euchromatic hotspot. The mechanism for Ty5 targeting to the rDNA locus and euchromatic hotspots is unclear. However, the results from CHIP-CHIP experiments show that 15% of Sir4, 12% of Sir2 and 33% of Sir3 are distributed at regions other than the telomeres and \textit{HM} loci (Lieb et al., 2001). As indicated by previous studies, Sir2 is associated with the rDNA locus constitutively and mediates transcriptional repression at the rDNA. However, the role for Sir4 and Sir3 at internal regions of the genome is unclear, since transcriptional levels of genes associated with Sir4 and Sir3 in internal regions are not changed in Sir4 and Sir3 mutants. Future analysis of the association between Ty5 integration sites and Sir4/Sir3 occupancy in euchromatic regions will likely be important in understanding the mechanism for Ty5 integration at euchromatic sites.
Extensive and systematic research has made *S. cerevisiae* an ideal model for comprehensively understanding chromosome biology, and exhaustive datasets are available that include a nucleosome atlas, global histone modifications, pre-existing Ty insertion sites, replication origins and data on the timing of DNA replication (Kim et al., 1998; Lee et al., 2007; Roh et al., 2004; Wyrick et al., 2001). Analyses of Ty5 integration relative to the above chromatin features are ongoing. The results generated will contribute to the understanding of Ty5 integration site preference, and the technical approach to exhaustively map Ty5 insertions sites described here can be used to further dissect aspects of Ty5 targeting and its regulation. Finally, the described experimental approaches provide a template for understanding the integration of other retrotransposons, including Ty1 and Ty3 and artificial DNA transposons in yeast.

**EXPERIMENTAL PROCEDURES**

*Plasmids and Strains.*

8 independent haploid cell yPH499 and 8 independent diploid cell yPH501 carrying pNK254 were used to generate Ty5 integrant pools according to the published methods.

*Linker Mediated PCR.*

The linker mediate PCR procedure was applied from Angela Ciuffi with some adjustment. The genomic DNA isolated from each Ty5 integrant pool will be treated by two sets of restriction enzyme, AciI plus TaqI and MspI plus HinfII, respectivley. The digested DNA will be ligated to the linker DNA annealed from Dvo4621 (/Phos/CGGTCCCTTAAGCGGAG/3AmM/) and Dvo4622 (GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC). The ligation product generated from the set of AciI plus TaqI will be subject to AseI digestion, whereas the ligation product derived from the set of MspI plus HinfII will be digested by EcoRI. The first PCR reaction will be accomplished using Ty5 LTR specific primer Dvo495 (CCATAGTTTCTGTGTACAAGAGT) and linker specific primer Dvo4632 (GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC). The second nested PCR reaction will be accomplished using Dvo4665 (GCCTTGCCAGCCCGCTCAG...
AGGGCTCCGCTTAAGGGAC and barcoded Ty5 LTR primer Dvo
(GCCTCCCTCGGCCATCAG-barcode-ACATATAAGCTAGATCGTAATTCACT).
The PCR product with 100bp-500bp length will be gel purified, quantitated using PicoGreen DNA Quantification Kit. The appropriate amount of DNA samples from 16 barcoded pools will be mixed together and sequence using Roche 454 GLX sequencer.

Mapping of Cloned Sequences on the Yeast Genome.

The processing and mapping of raw sequences from 454 sequencing were accomplished using INSIPID method described by Bushman group. Integration sites were judged to be authentic if the sequences began within 3 bp of HIV LTR ends and had a >98% sequence match when aligned to the genome of Saccharomyces cerevisiae using BLAST. The INSIPID will generate two sets of integration sites, first set of integration sites are derived from the sequence which was aligned to unique best position on genome (unique hit insertions) and the second set of integration sites are derived from the sequences which were aligned to multiple positions on the genome (multiple hits insertions).

REFERENCES


CHAPTER 6. GENERAL CONCLUSIONS

When the maize geneticist, Barbara McClintock, discovered transposons in the 1940s, many scientists deemed transposons as useless or “junk” DNA for decades thereafter. This view is partially persuasive, since transposable elements do not generate useful products that benefit host growth under normal conditions. Moreover, transposable elements, including transposons and retrotransposons, parasitize the host cell, and show full dependence on the host for energy, building materials and the cellular machinery needed to complete their “life” cycle and sustain survival from generation to generation. In addition, the host cell developed diverse transcriptional, post-transcriptional and post-translational mechanisms to limit the mobility of transposable elements and thereby minimize the harmful effect of transposition.

The view of transposable elements as junk DNA has changed substantially after scientists increasingly collected evidence that transposable elements help the host cell restructure its genome as an adaptive response to stress conditions, a hypothesis that McClintock originally put forward (McClintock, 1984). Therefore, the study of the interaction between transposable elements and the host cell and the mechanism by which transposable elements and the host cell have co-evolved to respond to environmental stresses has been greatly intriguing.

McClintock’s hypothesis and Ty5

Several retrotransposons including Ty5 show strong integration preference into gene-poor regions. The best understanding for the mechanism of targeted integration was attained by the study of Ty5. Our previous work demonstrated that integration of Ty5 into heterochromatin is mediated by the interaction between Ty5 integrase and the heterochromatic protein Sir4 (Xie et al., 2001). The targeting determinant of Ty5 integrase was mapped to a small domain at the integrase C-terminus (called the Targeting Domain or TD). TD contains several serine/threonine residues, and here we investigated the phosphorylation of serine 1095 in TD and its relevance for targeted integration. We found that phosphorylation at serine 1095 is essential for the interaction of Ty5 integrase and Sir4
and Ty5 target specificity. Mutants that lack phosphorylation at serine 1095 change Ty5 insertion preference from heterochromatin to euchromatin, and as a result, half of the insertions land within coding sequences. Furthermore, the phosphorylation of serine 1095 within TD is subject to regulation by stress conditions such as nutrient deprivation. The results from this study substantially support McClintock’s hypothesis that transposable elements can reshape the host genome in response to stress.

Future work will focus on determining the kinase responsible for phosphorylation of TD at serine 1095. Discovery of the relevant signaling pathway will elucidate the mechanism of phosphorylation control in more detail. In addition, phosphorylation might be a general mechanism for controlling the integration of retrotransposons, since the integrase of the yeast retrotransposon Ty1 is also serine/threonine rich and phosphorylated in vivo. Future work on characterizing the role of these phosphorylations and regulation will contribute to the understanding of the interaction between the host cell and retrotransposons.

The Ty5 targeting domain regulates heterochromatin

In *Saccharomyces cerevisiae*, heterochromatin is tethered to the nuclear periphery via two redundant pathways, Sir4/yKu70 and Sir4/Esc1 (Gartenberg et al., 2004; Taddei et al., 2005). Heterochromatin is densely compact and normally bound by the large complex of Sir2/3/4 proteins, making it inaccessible to many enzymes. How does Ty5 integrase gain access to this specialized compartment and overcome many steric barriers to complete cDNA integration?

In this study, we showed that the Ty5 integrase TD has the ability to mediate Sir4 turnover. This turnover is dependent on the ubiquitin ligase Ris1, the ubiquitin conjugating enzyme Ubc4, and the proteasome pathway. We are the first to demonstrate that the Ris1 RING finger domain has auto-ubiquitination activity, which requires the intact form of the RING finger domain. Furthermore, we observed that the TD-like domain of Esc1, a nuclear periphery protein that recognizes Sir4 in a manner similar to TD, has the same ability as TD to regulate silencing in a Ris1 and Ubc4 dependent manner. These results suggest that Ty5,
together with Esc1, might regulate silencing and Sir4 stability under certain conditions, such as during Ty5 integration, at specific stages in the cell cycle, during mating type switching or under stress conditions. Future work will evaluate the regulation of Sir4 by Ris1 under these conditions and should contribute to a better understanding of heterochromatin dynamics. Furthermore, the study of the possible regulation of Sir4 by Esc1 will add to our knowledge about how heterochromatin is anchored to the nuclear periphery.

**Ty5 integration requires Sml1**

Speculating that Ty5 integration might be regulated by the cell cycle and DNA damage, we further assessed Ty5 transposition under these conditions to understand the interrelationship between Ty5, the host cell and the environment. We found that Ty5 integration is not inhibited in yeast cells arrested in G1 by the mating pheromone α-factor, although it is known that transposition of Ty1 and Ty3 is repressed by alpha factor treatment. In addition, we demonstrated that Ty5 cDNA integration is inhibited by the presence of two types of DNA damage, namely replication fork pausing caused by HU and double strand breaks created by MMS. Under both types of DNA damage, Ty5 cDNA recombination (and thereby cDNA synthesis) is not affected. By performing genetic analyses, we confirmed that Sml1, which regulates the level of cellular dNTP pools under DNA damage, is required for Ty5 integration. The mechanism by which Sml1 regulates Ty5 integration is under investigation.

**Ty5 genome-wide integration patterns**

The preference for Ty5 to integrate into silent chromatin conforms to the safe haven hypothesis that retrotransposons preferentially insert into gene-poor regions to minimize their harmful effect on host genome integrity. We used high throughput pyrosequencing to identify the genome-wide distribution of Ty5 insertion sites. The fine scale of Ty5 integration patterns generated from haploid and diploid cells confirmed the preference of Ty5 to integrate into silent chromatin and identified a bias for the occasional euchromatic insertion to occur in intergenic regions.
The described experimental and biocomputational methods will provide a new tool to study the regulation of Ty5 target specificity by stress and in various genetic backgrounds. Additionally, we are using this method to study the integration specificity of another yeast retrotransposon, Ty1. All together, the results from the global distribution of both Ty5 and Ty1 integration sites will help us to understand the relationships among retrotransposons, the host cell and environmental conditions.

REFERENCES


ACKNOWLEDGEMENTS

I would like to thank my mentor, Dr. Dan Voytas for his guidance and patience in teaching me with scientific thoughtfulness and for his kindness, generosity and friendship throughout my time as his student. He is always ready to help me with any kinds of problems which I came with during my study at his lab. His guidance and influence will benefit not only my scientific research, but my career and personal development in future.

I gladly acknowledge many lab members who have encouraged me and helped me to perform this research, namely Troy Brady, Junbiao Dai, Peter Fuerst, Yi Hou, Fengli Fu, Xiang Gao, Joshua Baller.

I would also like to thank my committee members for helping me improve my weakness.

Finally, I sincerely thank my wife, Xiaohong Li and my son, Richard Haotian Gao, for their patience and unconditional support. Without their support and company, I could not finish this study.