2007

The role of the integrase c-terminus in replication and targeted integration of the yeast retrotransposon Ty5

Troy Leon Brady

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd
Part of the Genetics and Genomics Commons, and the Molecular Biology Commons

Recommended Citation

Brady, Troy Leon, "The role of the integrase c-terminus in replication and targeted integration of the yeast retrotransposon Ty5" (2007). Retrospective Theses and Dissertations. 15971.
https://lib.dr.iastate.edu/rtd/15971

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
The role of the integrase c-terminus in replication and targeted integration of the yeast retrotransposon Ty5

by

Troy Leon Brady

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

DOCTOR OF PHILOSOPHY

Major: Molecular, Cellular, and Developmental Biology

Program of Study Committee:
Daniel F. Voytas, Major Professor
Amy Andreotti
Kristen Johansen
Alan Myers
Thomas Petersen

Iowa State University
Ames, Iowa
2007

Copyright © Troy Leon Brady, 2007. All rights reserved.
# TABLE OF CONTENTS

**ABSTRACT**

**CHAPTER 1. GENERAL INTRODUCTION**

**CHAPTER 2. A CONSERVED MOTIF IN THE INTEGRASE C-TERMINI OF TY1/COPIA RETROTRANSPOSONS IS NECESSARY FOR STABILITY OF INTEGRASE AND REVERSE TRANSCRIPTASE**

- Abstract 18
- Introduction 19
- Results 20
- Discussion 28
- Experimental Procedures 32
- References 35

**CHAPTER 3. PHOSPHORYLATION REGULATES INTEGRATION OF THE YEAST TY5 RETROTRANSPOSON INTO HETEROCHROMATIN**

- Abstract 40
- Introduction 41
- Results 42
- Discussion 52
- Experimental Procedures 56
- References 59

**CHAPTER 4. RETROTRANSPOSON TARGET SITE SELECTION BY IMITATION OF A CELLULAR PROTEIN**

- Abstract 63
- Introduction 64
- Results 66
Discussion
Experimental Procedures
References

CHAPTER 5. GENERAL CONCLUSIONS
References

ACKNOWLEDGEMENTS
ABSTRACT

Mobile genetic elements have the inherent potential to shape the genomes of their hosts. The retroelements, which include the retrotransposons and retroviruses, are a particularly abundant class of mobile elements. The site of retroelement insertion in the host genome is largely determined by the element-encoded protein integrase (IN). Research presented here addresses three aspects of IN from the retrotransposon Ty5: the function of a conserved domain (GKGY) in the IN C-terminus, the role of IN post-translational modification in targeted integration, and the origin of the Ty5 targeting mechanism.

Sequence analysis of a conserved domain (GKGY) found in the C-termini of all retrotransposons of the Ty1/copia family revealed two variant forms of this motif: the ‘classic’ motif found in most Ty1/copia elements including Ty5, and a variant form characteristic of the retrotransposon Ty1 and closely related orthologues. Point mutations in the GKGY domains of both Ty1 and Ty5 resulted in greatly decreased transposition and low or undetectable levels of IN and reverse transcriptase (RT). Gag levels in all mutants were comparable to wild type, although less efficient processing of Gag and Pol was observed in Ty1. Lower levels of IN and RT were not due solely to low levels of polyprotein, as analysis of Ty1 protease mutants revealed only a modest decrease in polyprotein compared to wild type. These data collectively indicate that after processing of the polyprotein by protease, the GKGY motif has a role in forming a stable complex between IN and RT and/or in mediating their proper folding.

Ty5 IN encodes a short targeting domain (TD) at the very C-terminus that interacts with the heterochromatin protein Sir4. The TD/Sir4 interaction is responsible for Ty5’s preference to integrate into heterochromatin. Ty5 TD is phosphorylated at serine 1095, and using surface plasmon resonance, the TD/Sir4 interaction was shown to be dependent on the phospho-serine at position 1095. Mutations preventing phosphorylation of this residue disrupt the ability of TD to target Ty5 integration in vivo. Exposure of yeast cells to environmental stresses decreased TD phosphorylation, indicating an active role for the host cell in determining Ty5 target specificity in response to environmental signals. Together,
these results are the first to show direct involvement of the host cell in regulating target site selection of a transposable element.

The hypothesis was tested that the TD/Sir4 interaction imitates an interaction between a host protein and Sir4. Initially deletion analyses were used to define the N- and C-terminal boundaries of Sir4 that interact with IN. Multiple Sir4 missense mutants were also generated that failed to interact with IN. Using an assay that measures Ty5 targeting to sites to which the mutant Sir4 proteins are tethered, most mutants did not target Ty5 integration. With the IN-interacting domain of Sir4 characterized, we tested the hypothesis that IN imitates a host protein by identifying Sir4-interactors through a genome-wide yeast two hybrid screen. The interaction profiles between Sir4 and candidate proteins were compared to the Sir4/IN interaction profile. Esc1 was found to have the same Sir4 N- and C-terminal interaction boundaries as IN and failed to interact with the majority of Sir4 mutants. Esc1 through its interaction with Sir4 partitions DNA to daughter cells. The interaction of Sir4 with a small 34 amino acid region of Esc1 (aa 1440-1473) correlated with the ability to reliably partition DNA. A conserved 13 amino acid domain was found within this region, which along with a smaller 9 amino acid fragment, were shown to interact with Sir4 and to nucleate silent chromatin, a feature shared by IN and the TD. The functional equivalence of the TD and the 13 amino acid domain of Esc1 was demonstrated by exchanging these respective motifs to create chimeric IN and Esc1 proteins that maintained the ability to target integration and partition DNA, respectively. As with TD, mass spectrometry analysis of Esc1 (aa 1440-1473) revealed that this short domain is a phosphoprotein. These studies indicate that Ty5 imitates a conserved region of the host protein Esc1 in order to interact with Sir4 and target integration to heterochromatic regions. These data also suggest that like the IN/Sir4 interaction, the interaction between Sir4 and Esc1 and the \textit{in vivo} ability to partition DNA depends upon a small conserved region in the Esc1 C-terminus that is phosphorylated.
CHAPTER 1. GENERAL INTRODUCTION

Transposable elements arguably represent the most successful gene family. Genomes of nearly all organisms in every kingdom, including bacteria, plants, fungi, animals and protists encode numerous copies of these elements. Their structure and life cycles are diverse, yet they share the common feature of reproducing through mobilization of DNA that encodes proteins essential for additional rounds of replication.

DNA transposons and retroelements
Transposable elements can be divided into two general classes: DNA transposons and retroelements. DNA transposons are found in both prokaryotes and eukaryotes, whereas retroelements are most abundant in the eukaryotic kingdoms.

Transposons represent perhaps the simplest group of mobile elements. Their life cycle involves mobilization through a cut-and-paste mechanism; the DNA encoding the transposon is excised and inserted at a new genomic location. Transposons typically encode a single protein, transposase, which recognizes terminal inverted repeats (TIRs) flanking the transposase coding region. Transposases have a conserved catalytic domain capable of breaking the phosphodiester backbone of DNA to create double strand breaks that excise the transposon. Because the catalytic domain is capable of the reverse reaction, namely phosphodiester bond formation, transposase can also insert the transposon in a new location.

Retroelements perform an integration reaction very similar, if not identical, to that of the DNA transposons, but they do not excise themselves from the genome. Instead, the parental element is transcribed, and the resulting RNA serves a dual role: as a template for translation by host cell ribosomes and as a template for reverse transcription into DNA by the retroelement-encoded reverse transcriptase (RT). The resulting cDNA is then integrated into the host genome by integrase (IN), the retroelement equivalent of transposase (Figure 1). Because of these additional steps in replication, retroelement genomes are typically larger than those of the DNA transposons and encode more functions. Virtually all retroelements encode two genes, called gag and pol, which in some elements are separated by a shift in
Figure 1. Life cycle comparisons of DNA and retrotransposable elements. Life cycles of DNA transposons and retroelements are represented on the left and right sides, respectively. Replication of both classes of elements converges in a similar catalytic reaction at the point of DNA integration. DNA encoding the respective transposable element is shaded blue. Dashed lines represent the continuation of genomic DNA. IN, integrase; RT, reverse transcriptase; PR, protease.

... reading frame. gag shows little conservation between elements and encodes proteins that bind retroelement RNA and self-assemble into symmetrical particles. pol encodes three proteins: protease (PR), IN, and RT. PR is an aspartate protease whose active site is formed upon dimerization and which cleaves the retroelement polyproteins into individual subunits. Some retroelements, such as retroviruses, have additional genes that confer the ability to exit host cells and enter new cells, to alter the immune response, and to manipulate cellular defenses.

A more detailed view of the retroelement life cycle is given in Figure 2. This figure and the subject of this thesis concern retroelements that have long terminal repeats (LTRs) flanking an internal coding region. The LTR retroelement life cycle begins with transcription of the element, beginning and ending in the LTR. Some of the resulting mRNA is translated to produce Gag and Pol. After translation, PR processes both the Gag and Pol polyproteins into mature proteins, and Gag forms a viral particle (virus-like particle, or VLP, for retrotransposons). At this point, retroviruses are released from the host cell as infectious virions by budding from the cellular membrane. These virions, in turn, can infect other cells...
Figure 2. Life cycle of LTR retroelements. Replication of retroviruses and retroelements differ mainly in the ability of viruses to be released from the host cell and to infect new cells.

with which they come in contact. Retrotransposons lack the ability to exit their host, and so their VLPs remain in the cytoplasm. Reverse transcription is catalyzed by RT to produce cDNA, which IN binds to form a pre-integration complex (PIC). At some point during this process, or after infection of a new host in the case of the retroviruses, the retroelement particle dissociates, enabling the PIC to enter the nucleus and catalyze cDNA integration into the genome. Due to its central role in the integration reaction, IN is thought to determine the site of cDNA integration. Thus, knowledge of IN function is fundamental to understanding mechanisms of target site selection by these mobile elements.

**Domains and functions of integrase**

IN proteins from all LTR retroelements can be divided into three domains: an N- and C-terminal region separated by a catalytic core (Figure 3). The N terminal domain contains a zinc finger motif, which is thought to bind cDNA (Heuer and Brown, 1997) and requires zinc ions for correct folding and integrase multimerization in vitro (Heuer and Brown, 1997; Zheng et al., 1996). In some retroelements, however, this motif is not essential for DNA
binding or *in vitro* integration activity (Katz et al., 1996; Vincent et al., 1993) and its exact role in IN function is yet to be determined. The catalytic domain contains three highly conserved residues found in all retroelement integrases. These three residues (two aspartates followed by a glutamate 35 residues downstream) constitute the D,D(35)E motif. This motif is critical for catalyzing the integration reaction and is thought to coordinate binding of a divalent metal ion (Bujacz et al., 1996; Kulkosky et al., 1992). Mutations in these residues severely decrease integration efficiency and result in a catalytically inactive protein (Engelman and Craigie, 1992; Kulkosky et al., 1992; Leavitt et al., 1993). The catalytic core domain also has DNA binding properties and several residues, some of which are conserved, have been shown to be essential for this activity (Engelman et al., 1994; Esposito and Craigie, 1999). The C-terminal domain is the most divergent region of the protein (Johnson et al., 1986; Lutzke et al., 1994). It also displays DNA binding properties and, in retroviruses, is structurally similar to Src-homology three (SH3) domain protein folds (Eijkelenboom et al., 1995; Kishan and Agrawal, 2005; Lodi et al., 1995), which are known to support interaction with a variety of interfaces, including proteins, DNA, and RNA (Kishan and Agrawal, 2005). The exact function of this SH3-like domain in integration is not understood.

![Zinc Finger Catalytic Domain C-terminal Domain](image)

**Figure 3.** Comparison of integrases from a retrovirus and two retrotransposons of the Ty1/copia lineage. Conserved residues are marked with the corresponding amino acid position in the protein. NLS, nuclear localization signal; HIV, Human immunodeficiency virus; TD, targeting domain.

LTR retroelements can be divided into three groups: the Retroviruses, Ty1/copia, and Ty3/gypsy groups. The C-terminus of IN is considerably larger in retrotransposons of the Ty1/copia lineage than those of the Ty3/gypsy family and retroviruses (Figure 3). The majority of this extended C-terminus is highly divergent, even among close homologues.
Nonetheless, three common domains in the C-termini of IN have been reported. Many retrotransposons of the Ty3/gypsy group and some retroviruses, including Moloney murine leukemia virus, encode a cluster of conserved residues termed the GPY/F module. Also present in some Ty3/gypsy retrotransposons is a region resembling a chromodomain (Malik and Eickbush, 1999). The precise functions of these domains are not currently understood, although it is speculated that the chromodomain-like regions may mediate target site selection in these elements (Sandmeyer, 2003). For the Ty1/copia elements, the IN C-terminal regions are rich in serine and proline residues and encode a conserved motif directly adjacent to the core domain (Figure 3). This area is the only region with significant conservation in the C-terminus and measures some 60-80 residues in length. The exact function of this motif, named the GKGY domain (Peterson-Burch and Voytas, 2002), is not well understood but is investigated in this thesis. The GKGY motif is conserved in nearly all Ty1/copia elements from organisms as diverse as flies, plants, and yeast, suggesting it plays a critical role in the life cycle of these elements.

In addition to carrying out the integration reaction in all retroelements, IN has recently been shown to play other significant roles. RT requires functional IN in many retroelement model systems (Kirchner and Sandmeyer, 1996; Lai et al., 2001; Liu et al., 1999; Nymark-McMahon et al., 2002; Wilhelm and Wilhelm, 2005; Wu et al., 1999), and these two proteins from HIV have been shown to physically interact (Tasara et al., 2000; Zhu et al., 2004). IN also localizes the PIC to the nucleus for Ty1 and Ty3 (Kenna et al., 1998; Lin et al., 2001; Moore et al., 1998) and is thought to play an important role in nuclear localization of the PICs of HIV (Armon-Omer et al., 2004; Bouyac-Bertoia et al., 2001; Gallay et al., 1997). IN has also been shown to be the primary factor responsible for determining where retroelements insert their cDNA cargo (Harper et al., 2003; Lewinski et al., 2006). In an effort to learn more about these mobile elements, we have undertaken the study of a simple model system for retroelement target site selection, namely, the yeast retrotransposon Ty5. The focus of the research presented in this doctoral thesis is the C-terminal region of Ty5 IN and the investigation of domains needed for carrying out transposition and targeted integration.
**Target site selection of retroelements**

The importance of understanding how retroelements choose integration sites has been underscored by gene therapy trials of patients with severe combined immunodeficiency disorder due to mutations in their IL2γ genes. These patients were treated with retroviral vectors that carried the wild type IL2γ gene. While nine out of 10 patients showed significant improvement, three developed leukemias, two of which resulted from integration events near the LMO-2 proto-oncogene (Check, 2005; Hacein-Bey-Abina et al., 2003a; Hacein-Bey-Abina et al., 2003b). Genome-wide mapping of thousands of integration sites of retroviral vectors has revealed integration patterns and preferred sites for many retroviruses, including HIV, ASLV (avian sarcoma-leukosis virus), and MLV (murine leukemia virus) (Mitchell et al., 2004; Narezkina et al., 2004; Schroder et al., 2002; Wu et al., 2003). How these viruses determine where to insert their DNA must be understood if viral-mediated gene delivery and anti-viral drug development are to be realized. Due to the abundance of retroelements in human genomes and their potential to alter genome organization, obtaining a fundamental knowledge of retroelement integration mechanisms may also contribute to our understanding of how genomes are changed when mobilization occurs.

**Target site selection of Ty5**

The most detailed understanding of retroelement targeting mechanisms to date comes from studies of the yeast retrotansposon Ty5. *Saccharomyces cerevisiae* contains five retrotransposons designated Ty1 – Ty5. Of the five, only Ty5 does not target insertions upstream of genes transcribed by RNA polymerase III. Rather, Ty5 prefers to integrate into heterochromatic regions such as telomeres and silent mating loci. These regions share several common proteins involved in silencing DNA, including the presence of silent information regulator proteins, or Sir proteins. Ty5 IN interacts with the C-terminus of Sir4, and this interaction is essential for targeted integration (Xie et al., 2001; Zhu et al., 1999).

Research exploring regions of IN important for targeting has identified a six amino acid motif near the very end of the C-terminus. Termed the targeting domain (TD), this
motif contains four amino acids that are essential for wild type levels of targeting and transposition (Figure 4). The presence of two serines in this motif raised the possibility of post-translational modifications at these residues which, in turn, could play a role in the TD/Sir4 interaction. The reduction of overall transposition that is concomitant with reduced targeted integration (Gai and Voytas, 1998; Xie et al., 2001; Zhu et al., 1999) suggests that, in addition to interacting with Sir4, TD may also function as a regulator of IN activity to coordinate localization of the PIC with activation of IN.

How Ty5 acquired the ability to target integration is not known. *S. cerevisiae* has a relatively gene-dense genome with few gene duplications, making the likelihood of integration events causing fatal gene disruptions much greater than in higher eukaryotes. Because retrotransposons cannot leave their host cell like retroviruses, it is to their advantage to choose benign integration sites that do not affect the fitness of their host. Work presented in this dissertation attempts to address the origin of the Ty5 targeting mechanism and the role for post-translational modification in the function of TD. The results of this research will

---

**Figure 4.** The Ty5 integration complex targets insertions by interacting with Sir4. Sir proteins 2, 3, and 4 are recruited to the *HMR* locus by ORC, Rap1, and Abf1. Ty5 IN interacts with Sir4 to localize insertions to heterochromatin. The targeting domain (TD) of IN is needed for efficient targeting and integration. Essential residues of the domain are underlined.
shed light on the host-element relationship that exists in yeast and the molecular mechanisms of Ty5 targeting.

**Functions of the IN-interacting protein Sir4**

Sir4, the binding partner of IN, is a large protein known to interact with many host factors (reviewed in Gasser and Cockell, 2001). Because of these numerous binding partners, Sir4 is involved in many cellular processes apart from targeting Ty5 integration. These include the formation of heterochromatin at telomeres and the silent mating loci (*HMR* and *HML*) by association with the heterochromatin proteins Sir2 and Sir3 (Hecht et al., 1996; Strahl-Bolsinger et al., 1997), tethering telomeres to the nuclear periphery, and partitioning of DNA to daughter cells during mitosis through its interaction with nuclear periphery proteins (Andrulis et al., 2002; Ansari and Gartenberg, 1997). The very C-terminal region of Sir4 encodes a coiled coil domain that has been compared to the coiled coil repeats found in lamin proteins (Diffley and Stillman, 1989). *S. cerevisiae* does not have lamin proteins nor are there any known Sir4 homologues outside of closely related yeast species. For these reasons, it is thought that Sir4 serves an analogous function to that of lamins found in other organisms. The remainder of the Sir4 C-terminus contains another protein-interaction domain termed PAD (partitioning and anchoring domain, aa950-1250) involved in DNA partitioning during mitosis (Ansari and Gartenberg, 1997). The PAD domain interacts with Esc1, a large protein of 1658 amino acids which was identified by its ability to nucleate heterochromatin when tethered to DNA, presumably through its interaction with Sir4. Esc1 and Sir4 partially colocalize to the nuclear periphery and are required for partitioning of unstable plasmid DNA, properties that likely reflect their role in partitioning of chromosomes (Andrulis et al., 2002).

A Sir4 fragment of aa 971-1358 and residues between aa 971-976 have been identified as being important for interaction with IN (Zhu et al., 2003); however, the exact region or specific residues of Sir4 that are needed to support targeting have not been fully characterized. Research presented in this dissertation reports a thorough analysis of Sir4 requirements for interaction with IN and tests how Sir4 molecules that fail to interact with IN affect targeted integration of Ty5.
**Ty5 targeted integration is based on PIC tethering**

The target specificity of Ty5 is thought to result from tethering of the PIC, via the IN/Sir4 interaction, to regions of heterochromatin that are rich in Sir4. IN/Sir4 interaction data from yeast two hybrid assays and heterochromatin nucleation assays also support a tethering model that is dependent on TD (Xie et al., 2001). Indeed, previous work indicates that tethering is the principal determinant of targeting for Ty5. Although needed for interaction with Sir4, TD can be replaced with other peptide motifs without affecting IN activity. By substituting TD with small exogenous motifs from human proteins, Ty5 integration can be directed to a target plasmid to which the binding partner of the exogenous motif is tethered. This targeting is dependent upon the presence of the binding partner, indicating that tethering of Ty5 to the plasmid via interaction with these exogenous proteins is sufficient for targeted integration (Zhu et al., 2003).

**Integration of other retroelements supports a model of tethered integration**

The tethered integration model of Ty5 may apply to other retroelements as well. The retrotransposon Ty3 requires RNA polymerase III factors for targeted integration *in vitro*, suggesting an interaction between the Ty3 PIC and a component of the pol III transcription apparatus (Kirchner et al., 1995). This model is also attractive for explaining HIV target site selection. DNA-binding proteins such as BAF, Ini-1, HMGa1, Ku, and LEDGF enhance integration activity or bind to integration complexes (Cherepanov et al., 2003; Coffin, 1997; Kalpana et al., 1994). Of these host factors, lens epithelium-derived growth factor (LEDGF/p75) is the most favored for influencing HIV target site selection. LEDGF is a general transcription co-activator influencing the expression of many genes in humans (Ge et al., 1998). It binds the core domain of HIV IN through a C-terminal integrase-binding domain. LEDGF is needed for localization of IN to chromatin during infection, protects IN from proteasomal degradation, and may facilitate nuclear entry of IN (Llano et al., 2004a; Llano et al., 2004b; Vanegas et al., 2005). Knockdown of LEDGF levels in cell lines leads to an altered integration pattern (Ciuffi et al., 2005) and severe depletion of this host factor significantly reduces HIV infectivity (Llano et al., 2006). Interestingly, the LEDGF-mediated co-localization of IN and chromatin appears to be crucial for efficient HIV
infection, as expression of chromatin-binding deficient LEDGF in place of wild type protein fails to restore HIV infection (Llano et al., 2006).

Other host factors that do not directly interact with IN also play significant roles in retroviral infection. In human cells, reduction of emerin, a nuclear periphery protein that binds the nuclear lamina (Bengtsson and Wilson, 2004), or the chromatin-associated protein BAF (barrier to autointegration factor) inhibit HIV integration. BAF and emerin are needed for the association of HIV cDNA with chromatin, and this association is essential for wild type infectivity of the virus. Although BAF alone can bind chromatin and integrase, a cooperative interaction with emerin seems to be needed for HIV integration (Jacque and Stevenson, 2006). Together, these results suggest that multiple host factors, many of which interact with IN, bind DNA, and are localized at the nuclear periphery, are involved in integration of HIV.

![Diagram of protein environments during targeted integration of Ty5 and HIV](image)

Figure 5. Comparison of protein environments during targeted integration of Ty5 and HIV. Nuclear periphery proteins involved in integration of Ty5 and HIV are shown in addition to their respective binding partners. Continuous arrow indicate protein interactions required for targeted integration whereas discontinuous arrows indicate hypothetical relationships between proteins.

**Subnuclear localization: a common theme for models of tethered integration?**

The above observations draw interesting parallels between Ty5 and well-studied retroelements such as HIV. Ty5 interacts with a chromatin binding protein that is localized at the nuclear periphery. Tethering of this protein to a plasmid creates a new hotspot for Ty5 integration events, suggesting that both a Sir4/IN interaction and a Sir4/chromatin tethering
activity are needed for targeted integration. This closely mirrors the interactions of HIV IN and LEDGF (Figure 5). Also of note is the parallel between emerin and BAF and the Sir4/Esc1 interaction. In both cases, a chromatin-binding protein interacts with a protein partner at the nuclear periphery. Both emerin and BAF are needed for HIV integration, whereas only a requirement for Sir4 has been described for Ty5. In both cases, integration appears to occur at the nuclear periphery with the potential for other proteins in this region to influence integration. This issue is addressed in part by research presented here by documenting the interaction requirements for Sir4 and IN and how those requirements relate to targeted integration of Ty5.

Dissertation organization
Chapter II details the analysis of a conserved domain (GKGY) found in the C-termini of all retrotransposons of the Ty1/copia family (Peterson-Burch and Voytas, 2002). I performed all of the work described in this study.

Chapter III reports investigations into the relationship between Ty5 and its host, specifically the role of post-translational modification in targeted integration. This work was a collaborative effort on the part of Jiquan Gao, a current graduate student, two former graduate students, Junbiao Dai and Weiwu Xie, and myself. Using surface-plasmon resonance, I showed that phosphorylation of S1095, and not the adjacent serine at 1094, allows for \textit{in vitro} interaction of a TD-containing peptide with the Sir4 C-terminus. This demonstrated a direct interaction between Sir4 and TD and showed that site-specific phosphorylation of TD is necessary for optimal interaction with Sir4. This manuscript has been accepted by \textit{Molecular Cell}, pending revisions.

Chapter IV relates investigations leading to a better understanding of the targeting mechanism of Ty5 IN. This work was a collaborative effort between Peter Fuerst, a former graduate student, Clarice Schmidt, our lab manager, Robert Dick, a former undergraduate researcher, and myself. I directed the research efforts of Dick and Schmidt and designed and carried out all in-depth comparisons and analyses contrasting IN and Esc1.
Chapter V includes general conclusions from chapters II-IV as well as observations and future directions not mentioned in these previous chapters.

References


CHAPTER 2. A CONSERVED MOTIF IN THE INTEGRASE C-TERMINI OF TY1/COPIA RETROTRANSPOSONS IS NECESSARY FOR STABILITY OF INTEGRASE AND REVERSE TRANSCRIPTASE

Troy L. Brady and Daniel F. Voytas

1Department of Genetics, Development and Cell Biology, 1035A Roy J. Carver Co-Laboratory, Iowa State University, Ames, IA 50011-365

Abstract

Retrotransposons are found in nearly every eukaryotic organism, often as significant contributors to the genetic make-up of their hosts. The integrase (IN) protein plays a key role in the retroelement life cycle by mediating the nuclear import of cDNA and its integration into the host genome. The C-termini of integrases among all retroelements are highly variable, and the function of this domain is not completely understood. Alignments of integrase sequences representing the diversity of the Ty1/copia retrotransposon family previously revealed a conserved motif in the C-terminus, termed GKGY, approximately 65 residues downstream of the catalytic domain. Here, we identify two subclasses of GKGY motifs, designated as the Ty1 and classic groups. While differentially conserved, both motifs are predicted to be structurally similar, suggesting they perform similar functions. Mutational analysis of this domain resulted in decreased transposition of both Ty1 and Ty5, members of the Ty1 and classic groups, respectively. Levels of IN and RT were significantly reduced in these mutants while overall Gag levels remained close to WT. IN/RT instability in Ty1 could not be attributed solely to mutation-induced instability of the polyprotein, as polyprotein levels were maintained at decreased, but stable and detectable levels. The decrease in IN/RT levels in the presence of near-normal Gag levels suggests the GKGY domain may be important for IN/RT stability and/or proteolytic processing.
**Introduction**

Retrotransposons are non-infectious entities that replicate in a manner similar to retroviruses, and as such, they encode proteins homologous to retroviral integrase (IN), reverse transcriptase (RT) and Gag. IN plays a key role in the life cycle of all retroelements. It is comprised of three domains: the N-terminal zinc binding domain, whose exact function is not understood; a catalytic core domain which carries out processing of cDNA ends and host DNA cleavage to facilitate insertion of the newly synthesized cDNA into the host genome; and a highly divergent C-terminal domain.

Despite its divergence, the C-termini of various IN proteins are known to carry out essential functions. Some contain a nuclear localization signal essential for nuclear import of integrase and cDNA (Kenna et al., 1998; Lin et al., 2001; Moore et al., 1998), though cDNA nuclear import has also been reported to occur independent of IN import (Dvorin et al., 2002). IN dimerization and DNA binding domains have been mapped to the C-terminus (Andrake and Skalka, 1995; Jenkins et al., 1996; Khan et al., 1991; Lutzke et al., 1994), and selection of integration target sites is mediated by the IN C-terminus of the yeast retrotransposon Ty5 (Xie et al., 2001).

Data from IN mutants in various retroelements increasingly show that IN plays a significant role in reverse transcription. Deletions and point mutations in the C-terminus of the yeast retrotransposon Ty3 IN result in significant reductions of cDNA (Kirchner and Sandmeyer, 1996; Nymark-McMahon and Sandmeyer, 1999). Failure of reverse transcription was found to occur early in the replication process, resulting in low amounts of minus strand, strong stop DNA (Nymark-McMahon et al., 2002). Ty1, another yeast retrotransposon, also exhibits changes in reverse transcription when parts of IN are mutated (Wilhelm and Wilhelm, 2005). Furthermore, a C-terminal fragment of IN must be fused to the N-terminus of RT to produce active recombinant enzyme in *E. coli* (Wilhelm et al., 2000), indicating RT dependence on IN for activity *in vitro*.

In HIV, mutations in all three domains of IN have been observed to disrupt reverse transcription, some also resulting in slightly altered processing of Gag (Wu et al., 1999). Numerous C-terminal mutations in IN cause reverse transcription defects (Lu et al., 2005),
and a cysteine to serine substitution in the core domain inhibits production of early reverse transcription products (Zhu et al., 2004). Immunoprecipitations and pulldowns of IN demonstrate a direct interaction between HIV and MLV IN and RT (Hu et al., 1986; Tasara et al., 2000; Zhu et al., 2004). Deleting the 34 C-terminal residues or the entire IN protein of MLV results in severely reduced infectivity due to low levels of cDNA. Wild type IN supplied in trans rescues this phenotype, demonstrating a direct role for IN in maintenance of normal cDNA levels (Lai et al., 2001).

The many roles of IN C-termini suggest that specific regions may be conserved between elements. A conserved motif named GPF/Y motif has been described in many Ty3/gypsy retrotransposons as well as in some vertebrate retroviruses (Malik and Eickbush, 1999). The function of this domain and why it is only present in a portion of Ty3/gypsy elements is currently not understood. Similarly, a motif named the GKGY domain was found in the IN C-termini of all Ty1/copia elements examined to date (Peterson-Burch and Voytas, 2002). The function of the GKGY motif, named after four highly conserved residues, is unknown; however, its conserved nature suggests it plays a significant role in IN function and transposition of Ty1/copia retrotransposons.

Here we describe a functional analysis of the GKGY motif in Ty1 and Ty5, two yeast retrotransposons. The motif encoded by Ty5 is similar to the motif found in most Ty1/copia retrotransposons, whereas the motif in Ty1 and close homologues differs somewhat in amino acid conservation but is predicted to share a similar secondary structure. In either element, GKGY mutations resulted in a dramatic reduction in transposition activity, although phenotypes of particular mutations differed between elements. Transposition deficient mutants of both Ty1 and Ty5 exhibited reduced levels of IN and RT as well as signs of altered PR activity. These observations indicate involvement of the GKGY motif in maintaining normal levels of IN and RT as well as a possible role in proteolytic processing.
Results

Mutational analysis of the Ty5 GKGY motif.

To investigate the importance of the GKGY motif in the retrotransposon life cycle, alanine was substituted for each of the two conserved glycines in the Ty5 motif (G716A and G743A, Fig. 1A). The mutant Ty5 elements carry a *his3Δ1* marker gene, such that transposition or cDNA recombination (collectively referred to as mobilization) can be measured by quantifying the number of His\(^{+}\) cells that result after induction of Ty5 transcription by growth on galactose (Zou et al., 1996). Overall, mobilization was severely reduced in both mutants as measured by both qualitative (patch) and quantitative assays (Fig. 1B). Because Ty5 cDNA recombination can account for 40-50% of all mobilization events (Ke and Voytas, 1999; Zhu et al., 1999), the severe mutant phenotypes suggested that both integration and recombination was inhibited. This was further supported by performing mobilization assays with a *rad52Δ* strain that is deficient in homologous recombination and thus supports only

![Figure 1. Transposition frequency of Ty5 mutants. (A) Genomic organization of Ty5 showing the single reading frame encoding all Ty5 proteins. The GKGY motif is expanded with G,K,G and Y residues boxed and numbered. (B) Ty5 transposition patch assay with transposition frequencies and fold differences indicated beneath.](image-url)
integration (data not shown, Ke and Voytas, 1999). A decrease in the frequency of integration could result from mutation-induced instability or inactivity of IN. A reduction in recombination, however, could only be explained by either inhibition of recombination or reduced cDNA levels. The former would require some unknown IN-dependent mechanism for regulating homologous recombination of Ty5 cDNA whereas the latter implies a reduction in RT activity, protein level or cDNA stability.

To determine if the Ty5 GKGY mutations induced premature turnover of either IN or RT, we examined their protein levels by western blot analysis. To facilitate these studies, mutants were tagged in either IN or both RT and Gag with a RGSHIS epitope as described previously (Irwin and Voytas, 2001). Integrase levels in the mutants were reduced compared to wild type (Fig 2A), indicating the mutations may have disrupted stability of the protein. To our surprise, levels of RT were also significantly reduced in both mutants (Fig 2B). The levels of IN and RT for both mutants correlated with the fold reduction in transposition

Figure 2. IN and RT levels are reduced in Ty5 GKGY mutants. Uninduced cultures were grown in the presence of 2% glucose. All other cultures grown in 2% galactose. (A) Western blot of IN proteins from both WT and mutant elements shows a slight decrease in IN levels of GKGY mutants compared to WT. Total protein loads are shown below by the amido black stained membrane. (B) Western blot of RT and Gag in WT and mutant elements. RT and Gag species are designated with an arrow (→). Here, Gag also serves as an internal loading control.
(compare Fig 1B with Fig 2A and 2B); the G743A mutant displayed a greater reduction in protein levels and had less transposition activity. The decreased levels of RT could explain the reduced recombination rates observed in these mutants, as lower levels of reverse transcriptase would presumably result in the production of less cDNA.

In contrast to IN and RT, levels of Gag-p37 appeared relatively unchanged and Gag-p27 was only modestly reduced (Fig 2B). Ty5 encodes a single polyprotein that is processed by Ty5 protease to generate mature forms of IN, RT and Gag (both p27 and p37) (Irwin and Voytas, 2001). The presence of near wild type levels of Gag and reduced levels of IN and RT suggest that the mutations mediate their effects after polyprotein processing. Furthermore, the extent of polyprotein processing indicates that PR is at most modestly affected by the GKKGY mutations. In summary, we conclude that wild type levels of Ty5 RT and IN are dependent on a functional GKKGY motif.

**Ty1 encodes a GKKGY variant.**

To investigate whether the GKKGY motif functions similarly in other elements, we introduced corresponding mutations into the retrotransposon Ty1’s motif, based on previously published alignments (Peterson-Burch and Voytas, 2002). However, similar alanine substitutions in Ty1 (G255A and G276A) showed little effect on transposition compared to Ty5 as measured by qualitative patch assays (data not shown). This could be due to the more robust nature of Ty1 transposition (100 fold greater than Ty5), complementation from endogenously expressed Ty1 elements, or because Ty1 transposes in a GKKGY-independent manner. Closer examination of the Ty1 motif, however, revealed a significant difference in amino acid conservation in the vicinity of the second G in the motif when compared to other representative Ty1/copia elements (Fig 3, upper alignment). This raised the possibility that Ty1 has a different motif that does not perform the same function as its Ty5 counterpart and could explain the different responses to mutations in this region between these two retrotransposons.

To determine whether the observed amino acid sequence variation in GKKGY was unique to Ty1, we aligned the motifs of Ty1 relatives from different yeast species (Fig 3, lower alignment). Several conserved residues were observed in the region between the
conserved K and Y that were not found in the remaining Ty1/copia elements. While Ty1 and its homologues encoded a GY pair in this region, its position was significantly different from the other Ty1/copia elements (compare upper and lower alignments, Fig 3). The amino acid sequence similarity among the Ty1 homologues ends at the GKGY motif and does not extend into the IN N-terminus (data not shown), suggesting that this variant motif is functionally constrained. Secondary structure predictions were performed on the amino acid sequence alignments of the two groups using the protein structure prediction program Jpred (http://www-jpred@compbio.dundee.ac.uk) (Fig 3). Both motifs were predicted to be composed of multiple, similarly-spaced beta sheets, supporting the hypothesis that, while different in amino acid composition, these motifs perform similar functions in both groups of elements. We therefore designated the motifs carried by Ty1 and the remaining Ty1/copia elements as the Ty1 and classic GKGY motifs, respectively.

**Mutational analysis of the Ty1 GKGY motif.**

If the Ty1 GKGY motif functions in a manner similar to that of Ty5’s, then we would expect mutations disrupting function to result in phenotypes similar to that seen in the Ty5 mutants. We therefore proceeded to make additional Ty1 mutations, including Y288A, F254D, G287W, and a complete GKGY deletion (residues 254-291) (Fig 4A). The deletion and the latter two non-conservative amino acid substitutions were predicted to have the most severe
phenotypes. As a negative control, a Ty1 element was used that does not make IN and RT due to a frameshift mutation in PR (Ty1-fs). Mutations G287W and Y288A resulted in an approximately 70 to 80 fold reduction in transposition, respectively, whereas the F254D substitution and the deletion decreased transposition by over 1000 fold (Table 1). IN and RT levels were monitored in these mutants and, consistent with previous Ty5 data, both IN and RT levels were shown to be affected. As observed for Ty5 mutants, the greatest reduction in protein correlated with the greatest transposition defect (Fig 4B). These results support a universal role among members of the Ty1/copia family for the GKGY motif in both IN and RT stability.

Figure 4. IN and RT levels are reduced and Gag processing perturbed in Ty1 GKGY mutants. (A) Genomic organization of Ty1 showing the two overlapping reading frames represented by rectangles above or below a center line. The asterisk denotes the location of the frameshift mutation in the Ty1 negative control (Ty1-fs). The GKGY motif is shown expanded with G,K,G and Y residues boxed and numbered. Numbering is from the N terminus of IN. (B) Western blots of IN, RT and Gag proteins in WT and mutant Ty1 elements. Equal amounts of the same cell extracts were loaded for each individual analysis. IN, RT and Gag proteins indicated with arrows. Labels at the top of each lane pertain to each of the three blots.
To further define critical residues in the Ty1 GKGY motif, mutants with additional single or multiple alanine substitutions in the motif were tested for transposition activity. Because transposition of a marked Ty1 element can be influenced by expression of endogenous Tys, transposition assays were performed in a spt3Δ strain, which does not

Table 1. Transposition frequencies of Ty1 GKGY mutants in the SPT3 strain YPH499.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Transp freq</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ty1 WT</td>
<td>$8.5 \times 10^{-2} \pm 7.67$</td>
<td>1.0</td>
</tr>
<tr>
<td>F254D</td>
<td>$2.13 \times 10^{-5} \pm 2.27$</td>
<td>4010</td>
</tr>
<tr>
<td>Y288A</td>
<td>$2.26 \times 10^{-3} \pm 1.71$</td>
<td>37.7</td>
</tr>
<tr>
<td>G287W</td>
<td>$1.56 \times 10^{-3} \pm 0.51$</td>
<td>54.48</td>
</tr>
<tr>
<td>GKGYA</td>
<td>$8.8 \times 10^{-5} \pm 1.66$</td>
<td>968.46</td>
</tr>
<tr>
<td>Ty1-fs</td>
<td>$1.64 \times 10^{-6} \pm 0.77$</td>
<td>&gt;10000</td>
</tr>
</tbody>
</table>

Table 2. Transposition frequencies of additional Ty1 GKGY mutants in spt3Δ cells.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Substitution</th>
<th>Transp freq</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ty1 GKGY</td>
<td>NA</td>
<td>$1.71 \times 10^{-2} \pm 0.55$</td>
<td>1.0</td>
</tr>
<tr>
<td>AKAY</td>
<td>G255A</td>
<td>$1.67 \times 10^{-3} \pm 0.65$</td>
<td>10.6</td>
</tr>
<tr>
<td>GAGY</td>
<td>K268A</td>
<td>$3.07 \times 10^{-4} \pm 1.07$</td>
<td>57.5</td>
</tr>
<tr>
<td>GKAY</td>
<td>G276A</td>
<td>$7.67 \times 10^{-3} \pm 4.79$</td>
<td>2.2</td>
</tr>
<tr>
<td>GNGA</td>
<td>Y288A</td>
<td>$5.72 \times 10^{-6} \pm 2.24$</td>
<td>3080.8</td>
</tr>
<tr>
<td>AKAY</td>
<td>G255A, G276A</td>
<td>$1.24 \times 10^{-5} \pm 0.68$</td>
<td>1375.4</td>
</tr>
<tr>
<td>AKGA</td>
<td>G255A, Y288A</td>
<td>$1.47 \times 10^{-7} \pm 0.87$</td>
<td>&gt;10000</td>
</tr>
<tr>
<td>AAGY</td>
<td>G255A, K268A</td>
<td>$7.33 \times 10^{-5} \pm 10.4$</td>
<td>232.8</td>
</tr>
<tr>
<td>AAAY</td>
<td>G255A, K268A, G276A</td>
<td>$3.52 \times 10^{-5} \pm 6.05$</td>
<td>484.6</td>
</tr>
<tr>
<td>AAAA</td>
<td>G255A, K268A, G276A, Y288A</td>
<td>$1.47 \times 10^{-7} \pm 0.37$</td>
<td>&gt;10000</td>
</tr>
<tr>
<td>AAKGY</td>
<td>F254A, G255A</td>
<td>$1.78 \times 10^{-7} \pm 2.21$</td>
<td>&gt;10000</td>
</tr>
<tr>
<td>YGKGY</td>
<td>F254Y</td>
<td>$1.76 \times 10^{-3} \pm 1.57$</td>
<td>9.7</td>
</tr>
</tbody>
</table>
express native Ty1 elements (Boeke et al., 1986). Single mutations resulted in 2 to 7000 fold reductions in transposition, whereas transposition in elements with multiple mutations was nearly undetectable (Table 2). Mutations in Y288 and F254 reduced transposition most dramatically, and one combination of mutations (G255A and G276A) exhibited a synergistic effect. The F254Y substitution caused only a 6 fold reduction in transposition whereas the previously described F254D and a double F254A G255A mutants had severe transposition defects. This suggests that the hydrophobic nature of the residue at position 254 is critical.

Polyprotein levels in Ty1 GKGY mutants.

Ty5 encodes Gag and Pol on a single open reading frame (ORF). Because GKGY mutations had severe effects on IN and RT levels and only a modest effects on Gag, we concluded that the mutations mediated their effects after processing of the polyprotein. In contrast to Ty5, Ty1 expresses Gag and Pol on separate ORFs (Fig 4A). The first reading frame encodes Gag (p49), and a Gag-Pol polyprotein is synthesized by ribosomal frameshifting that occurs in the overlap between the two reading frames (Farabaugh, 1995). During virus-like particle maturation, the Gag-p49 species is cleaved by Ty1 PR to produce the predominant Gag-p45 species known as capsid (CA) (Merkulov et al., 2001). If the Ty1 polyprotein is made unstable due to GKGY mutations, little or no PR should be made, and this would result in the absence or lowered levels of CA. Thus, CA levels in GKGY mutants should be similar to what we observed for our negative control, which has a frameshift mutation in Pol just downstream of the naturally occurring Gag-Pol frameshift. The frameshift mutant produced only Gag-p49 and low levels of a Gag-PR fusion that has no apparent PR activity (Fig 4B, lane 4). The faint CA band for this negative control is produced by endogenous Ty1 elements (Fig 4B, compare lanes 1 and 4). Although processing of Gag-p49 was slightly impaired in the GKGY mutants, CA levels were not reduced to that of the negative control, indicating that PR was at least partially active in all of the mutants (Fig 4B, compare lanes 3, 5, 6, and 7 with lane 4). This indicates that active PR was produced in the Ty1 GKGY mutants, consistent with the Ty5 data indicating that GKGY mutations had their predominant effect on IN and RT stability.
To further evaluate the effect of GKGY mutations on Ty1 polyprotein stability, we assessed polyprotein levels in an inactive PR mutant background that expresses a single 199 kDa product (Monokian et al., 1994). Polyprotein levels were reduced but present in all mutants, showing that the polyproteins are not completely degraded due to instability (Fig 5). These results indicate that any instability of the polyprotein cannot account for the dramatic reduction of IN and RT levels observed in the GKGY mutants. Thus, in Ty1 and Ty5 mutants, instability and degradation of IN and RT appears to result primarily after processing of the polyprotein.

![Western blot](image)

**Figure 5.** Polyprotein in Ty1 GKGY mutants. Western blot of cell extracts from yeast expressing a Ty1 PR mutant and selected GKGY mutants. The filter was probed with antibody specific for IN. Polyprotein levels are slightly reduced in cells expressing mutant elements compared to wild type. IN and the polyprotein (Pptn) bands are indicated with arrows. The membrane was stained after immunoblotting to confirm equal amounts of protein in each lane.

**Discussion**

Retrotransposon and retroviral integrases are responsible for processing and integration of cDNA into the host genome, and these activities are carried out by the highly conserved catalytic domain of IN. Retrotransposon and retroviral integrases differ primarily in that the former have extended and highly divergent C-termini. Analyses of the abundant retrotransposon sequences made available by the various genome sequencing projects revealed that despite the heterogeneity of retrotransposon IN C-termini, they do have some
conserved sequence features. These include the GPY/F motif in some Ty3/gypsy element integrases and the GKGY motif in Ty1/copia integrases. Here we demonstrate that the GKGY motif of Ty1/copia integrases is essential for the replication of these ubiquitous mobile elements, particularly the stability of IN and RT.

Structure of the GKGY motif.
The GKGY motif abuts the IN catalytic domain and mirrors its characteristically high degree of amino acid sequence conservation. In fact, sequence divergence among GKGY motifs from various Ty1/copia elements (those that encode the ‘classic’ motif) is comparable to the levels of divergence seen for the catalytic domains of IN and RT. A notable exception is the related GKGY motifs of Ty1 and its close relatives. This is consistent with the observation that these elements comprise a unique, highly divergent lineage of Ty1/copia retrotransposons (Peterson-Burch and Voytas, 2002). Nonetheless, despite sequence divergence between the Ty1 GKGY and the classic motifs, secondary structure predictions suggest that they are highly similar, being comprised of several beta sheets. This predicted conservation in function is substantiated by our functional analyses indicating that mutations in the Ty1 motif have similar effects on transposition and protein stability.

Interestingly, the GKGY motif lies in the same area as the SH3-like (Src homology 3) motif described in the C termini of HIV, SIV, and RSV (Chen et al., 2000a; Chen et al., 2000b; Eijkelenboom et al., 1995; Lodi et al., 1995; Yang et al., 2000) and exhibits properties consistent with an SH3-like domain. Cellular SH3 domains are small (~60 aa long), composed of a compact beta-barrel generally containing 5 anti-parallel β-sheets and are usually involved in low affinity protein-protein interactions (Kishan and Agrawal, 2005; Mayer, 2001). The conserved region surrounding the GKGY motif is approximately 60 residues long and the structure of this region is predicted to be composed of four to six beta-sheets, consistent with C-terminal structures solved for RSV, SIV, and HIV which show five to six beta sheets. Conserved residues of the RSV C-terminus were found to be important for maintaining structure of the SH3-like domain, possibly explaining the conservation of GKGY residues (Yang et al., 2000). While the exact function of this SH3-like domain is not known in retroviruses, the HIV and RSV IN C-termini exhibit non-specific DNA binding
properties and are capable of dimerization (Esposito and Craigie, 1999; Yang et al., 2000). Structural analysis of the GKGY motif may shed light on whether or not this domain shares similarity to the SH3 domains found in related retroviral integrases.

**Phenotype of GKGY mutants.**

Mutations in the GKGY motif have severe consequences for transposition, and in many cases, completely abolish Ty mobilization. Consistent with this phenotype, Western blot analyses revealed reduced levels of IN in the mutants, and the extent of the reduction correlated with the severity of the transposition phenotype. Previous work has shown that Ty1 and Ty5 IN catalytic mutants still allow for Ty mobilization through cDNA recombination (Sharon et al., 1994; Zhu et al., 1999). However, the severity of the transposition defects in the GKGY mutants suggested that RT was also being affected and that Ty cDNA was not available for recombination. One the most surprising findings of our study was that mutations in the GKGY motif of IN also caused greatly reduced levels of RT.

A simple explanation for the effect of GKGY mutations on IN and RT levels is that they lead to increased protein turnover. Mutations affecting IN stability, however, would not be expected to promote degradation of the RT unless degradation occurred before processing of the polyprotein. In the case of Ty5, both Gag and Pol are encoded by a single polyprotein, and mutations causing polyprotein turnover should reduce levels of all Ty5 proteins, including Gag and PR. Gag levels in the G716A and G743A mutants, however, were comparable to wild type and high molecular weight unprocessed forms of the polyprotein were not observed.

In contrast to Ty5, Ty1 encodes Gag and Pol on separate reading frames. Most of the Ty1 mRNAs are translated to produce Gag-p49, but occasional ribosomal frameshifting results in the synthesis of a Gag-Pol fusion protein. Gag-p49 and the Gag-Pol fusion protein are processed to produce Gag-p45 by Ty1 PR. If GKGY mutants affected polyprotein stability, they should also affect Gag processing; however, at most a modest processing defect was observed. When a protease catalytic mutation was cloned into the GKGY mutants, polyprotein levels remained relatively stable compared to wild type, indicating that the observed reduction of mature IN and RT levels in Ty1 GKGY mutants was not simply
due to unstable polyprotein. Collectively, these data support the hypothesis that degradation of IN and RT occurs after processing of the polyprotein precursor.

**Potential functions of GKGY.**

The affect of IN GKGY mutations on RT protein stability suggests that the two proteins communicate with each other either directly or indirectly. Reverse transcription is known to require regions of IN in other retroelements. C-terminal deletions of Ty3 IN result in dramatic reductions in RT activity without significantly affecting RT protein levels (Kirchner and Sandmeyer, 1996). Likewise, charged-to-alanine scanning mutagenesis of Ty3 IN detected cDNA reductions in mutants with reduced IN levels (Nymark-McMahon and Sandmeyer, 1999). These mutants were defective for early cDNA intermediates, a similar phenotype reported for a replication compromised HIV due to a C130S mutation in IN (Zhu et al., 2004). Interestingly, this defect in Ty3 mutants could not be trans-complemented by co-expression of a wild type IN, suggesting Ty3 IN exerts its effect on RT activity in cis (Nymark-McMahon et al., 2002). Ty1 IN also appears to influence reverse transcription, as active RT can be purified from E.coli only if it is fused to a fragment of the IN C-terminus (Wilhelm et al., 2000). The exact role of Ty1 IN in reverse transcription is not clear, as the deletion of different regions of IN has opposite effects on the in vitro activity of RT purified from cells expressing these mutant Ty1 elements (Wilhelm and Wilhelm, 2005).

Retroviruses display a similar relationship between IN and RT. MLV IN is required for cDNA production, as IN deletions of 34 C-terminal residues or the entire IN protein significantly reduce infectivity due to a lack of cDNA (Lai et al., 2001). HIV IN and RT have been shown to physically interact by pulldown and immunoprecipitation assays (Tasara et al., 2000; Zhu et al., 2004). IN deletions removing the SH3-like domain or all of HIV IN were found to reduce levels of both IN and RT (Engelman et al., 1995). HIV and MLV IN mutants, usually defective for reverse transcription or Gag processing, partially restored cDNA synthesis and infection in complementation assays using exogenously expressed IN (Lai et al., 2001; Lu et al., 2005; Tsurutani et al., 2000; Wu et al., 1999). Likewise, the reverse transcription defects of chimeric HIV-1 constructs containing IN from HIV-2 could be complemented by expressing HIV-1 IN in trans (Liu et al., 1999). Naturally selected
mutations in these chimeras that augmented cDNA synthesis were located in both IN and RT, strongly indicating a role for IN in the reverse transcription process (Padow et al., 2003).

These studies show a common need for IN in reverse transcription in multiple retroelements. These reports found mutations affecting cDNA production to reside throughout IN. This suggests that although the majority of these mutations fall within the C-terminus, multiple regions of IN may function to influence reverse transcription in vivo. The exact mechanism by which IN and RT communicate and how IN influences RT activity, however, remains obscure. It is possible that the GKGY motif, which may adopt an SH3-like structure, defines an RT-interaction domain that is needed for a stable and active RT or RT-IN complex. The data presented here, however, are also consistent with mutations leading to fatal misfolding of IN. This perturbation in structure may then be communicated to RT or prevent RT processing, resulting in a concomitant turnover of both proteins. Future work will differentiate between these possibilities to determine whether this motif is important only for structural purposes or whether it is a modulator of RT activity in vivo. The conserved nature of the motif suggests an interaction with other evolutionarily constrained factor(s). It is interesting that this motif, although located in IN, appears to be intimately involved with RT and is found solely within the Ty1/copia family, a lineage defined using phylogenetic analyses of RT sequences (Peterson-Burch and Voytas, 2002). Understanding the role of this conserved domain in the lifecycle of these elements will lead to a greater understanding of retroelement transposition and evolution.

Experimental procedures

Yeast strains
YPH499 was used for all transposition and protein analyses of Ty5 and Ty1 with the exception of a BY4742 spt3Δ strain, which was used to quantify transposition of Ty1 alanine mutants without the influence of endogenous Ty1 elements. BY4742 spt3Δ was also used for analysis of polyprotein levels in Ty1 PR mutants.
Transposition assays and culture conditions

For monitoring protein levels, cells were cultured at 30°C overnight in SC media with dextrose (2%) followed by dilution to OD600 ~0.2 in SC media with raffinose (2%). Cells were then grown overnight at 30°C. Transposition was induced by diluting raffinose-grown cultures into SC media with galactose (2%) to a final OD of ~0.2 and growing 24 hours at room temperature before harvesting. For quantitative measurements of transposition, cultures were patched onto SC + dextrose plates, grown 2 days at 30°C, then replica-plated to SC + galactose media and incubated for 3 days at room temperature. Cells were scraped and resuspended in water. Serial dilutions were plated onto media lacking uracil or both uracil and histidine to obtain relative rates of transposition (Zhu et al., 1999).

Plasmids

GKGY mutations were introduced by four primer PCR followed by ligation of BspEI/BstBI or BstEII/AgeI fragments into subclones containing either the Ty5 SphI fragment (pTB19) or the XhoI/NheI fragment of pGTy1H3-HIS3AI (pTB42), respectively. Full length Ty5 elements were made by ligating the SphI subcloned fragment back into pYZ293 (a pRS315 version of pDR14 (Gao et al., 2002) to produce pTB29 (Ty5-GKAY) and pTB30 (Ty5-AKGY). For detecting protein levels, an XhoI/NotI fragment containing the entire Ty5 reading frame was transferred to the 2 micron vector pRS426 to generate plasmids pTB55 & 56 (AKGY), pTB58 & 57 (GKAY), and pTB59 & 60 (WT) with HIS-tagged RT and IN, respectively. An N-terminal HIS tag was added to the Gag reading frame of each RT-tagged element by cloning the XhoI/BstAPI fragment from pXG21 (Gao et al., 2002) into each respective plasmid to produce dual-tagged Ty5 elements on pTB95 (AKGY), pTB96 (GKAY), and pTB97 (WT). Full length Ty1 constructs were made by moving the subcloned XhoI/NheI fragments into pGTy1H3-his3AI to produce pTB52 (G255A), pTB53 (K268A), pTB54 (G276A), pTB68 (F254D), pTB69 (Ty1-fs), pTB70 (Y288A), pTB71 (G287W), pTB72 (254-291Δ), pTB98 (G255,K268,G276A), pTB99 (G255,G276A), pTB100 (F254Y), pTB101 (G255,Y288A), pTB102 (G255,K268A), pTB103 (F254,G255A), pTB111 (G255,K268,G276,Y288A). Protease-deficient Ty1 constructs were made by cloning the XhoI/NruI fragment pJEFF1105 PR-2 (Monokian et al., 1994) into the respective Ty1
subclone, followed by insertion into the full length element as described above to generate pTB104 (WT), pTB105 (F254D), pTB106 (Y288A), pTB107 (G287W), and pTB108 (254-291Δ). All plasmids constructed using PCR were sequenced for verification.

**Protein analysis**

Cultured cells were centrifuged, washed once with ice cold water, centrifuged again, and the resulting pellet was frozen. Ty5 proteins were prepared from 10 to 15 mL cultures using cell lysates generated by the glass bead method (Ausubel, 1987). Cell lysates were centrifuged at 4C for 30 minutes at 16,000xg. The pellet was resuspended in breaking buffer with 8M urea by breaking up the pellet and vortexing at 4C for five minutes. The resulting suspension was again centrifuged at 4C for 30 minutes. Relative amounts of total protein of the resulting supernatants were determined by BCA analysis and resuspended in appropriate amounts of SDS loading buffer (Ausubel, 1987) and heated at 95C for 5 minutes. Between 60-100 µg of protein were used for SDS-polyacrylamide gel electrophoresis and transferred at 30V overnight to a PVDF membrane for immunoblotting. Ty1 proteins were prepared in a similar fashion using the glass bead method, but urea extraction of the insoluble pellet was not necessary. Supernatants of cell lysates after centrifugation at 4C were directly used in BCA assays to estimate the total amount of soluble protein. After mixing with SDS loading buffer, 5-10 µg of total protein was used for analysis. After blocking with TBST + 3% milk, Ty5 proteins were detected using anti-penta HIS antibody (Qiagen, 1:3000 in TBST + 3% milk) and HRP-conjugated anti-mouse (1:3000 in TBST + 3% milk). PVDF membranes for Ty1 Gag, IN, and RT detection were blocked in TBST + 5% milk prior to immunodetection. Antibodies for Ty1 Gag detection were from sera Y589 (1:10,000 in TBST) with HRP-conjugated anti-rabbit (1:7000 in TBST) as secondary antibodies. For detection of Ty1 IN & RT, antibodies B2 & B8 were used, respectively, at a 1:7000 dilution in TBST. HRP-conjugated goat-anti-rabbit (Santa Cruz, 1:7000 in TBST) was used as the secondary antibody. For polyprotein analyses, the anti-IN antibody 8B11 was used (1:1000, TBST + 2% milk) after blocking with TBST + 10% milk. HRP-conjugated anti-mouse (Amersham, 1:2000 in TBST + 2% milk) was used as the secondary antibody. Protein detection was
performed using SuperSignal West Pico Chemiluminescent Substrate (Pierce) or ECL Western Blotting Detection Reagents (Amersham).

**Protein Alignments**

Retrotransposons representing various lineages of the Ty1/copia family were used for a comparative alignment of the GKGY domain (Peterson-Burch and Voytas, 2002). Ty1 homologues were obtained by BLAST searches using Ty1 reverse transcriptase sequence as an electronic probe. Protein sequences immediately surrounding the GKGY motif (as shown in the manuscript) were aligned using ClustalX and shaded in black and white using the program Boxshade version 3.21. These same alignments were submitted to the JPRED server for structure prediction (Cuff et al., 1998).

**Acknowledgements**

We would like to thank Jeff Boeke, currently at John Hopkins University, for kindly providing both wild type and PR mutant Ty1-his3AI elements as well as anti-Ty1 IN monoclonal antibody (8B11). We also thank David Garfinkel, currently at the National Cancer Institute-Frederick for kindly providing anti-Ty1 IN and RT antibodies (B2 and B8, respectively) and to Joan Curcio for generously providing anti-Ty1 Gag antibody (Y598).

**References**


Merkulov, G. V., Lawler, J. F., Jr., Eby, Y., and Boeke, J. D. (2001). Ty1 proteolytic cleavage sites are required for transposition: all sites are not created equal. J Virol 75, 638-644.


CHAPTER 3: PHOSPHORYLATION REGULATES INTEGRATION OF THE YEAST TY5 RETROTRANSPOSON INTO HETEROCHROMATIN

A manuscript accepted for publication in Molecular Cell, pending revisions.

Junbiao Dai$^{1,3}$, Weiwu Xie$^{1,2,3}$, Troy L. Brady$^1$, Jiquan Gao$^1$ and Daniel F. Voytas$^1$

$^1$Department of Genetics, Development and Cell Biology, 1035A Roy J. Carver Laboratory, Iowa State University, Ames, IA 50011-365
$^2$Present address: Biological Science Division, 117 Tucker Hall, University of Missouri-Columbia, Columbia, MO 65211
$^3$These authors contributed equally to this work.

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.
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 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.

Figure 1. Post-translational modification of the Ty5 targeting domain. (A) In vitro interaction between Sir4C and TD. Sir4C was expressed and labeled with 35S-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 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 UAS6 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.
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 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 His<sub>6</sub>-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<sub>6</sub> 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

![Image of Figure 2](image-url)

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 His<sub>6</sub>-tagged TD. The addition of His<sub>6</sub> 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 His<sub>6</sub>-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 HPO<sub>3</sub>, 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<sup>*</sup> and b<sup>**</sup> are produced by the loss of HPO<sub>3</sub> and one or two H<sub>2</sub>O molecules from b ions. b<sup>3</sup> and y<sup>3</sup> ions are produced by the loss of one H<sub>2</sub>O 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)
Table 1. MS/MS analysis of targeting domain phosphorylation.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Phosphorylated?</th>
<th>Site of Phosphorylation</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDSSPPNTS$H_6$</td>
<td>Yes</td>
<td>S1095</td>
</tr>
<tr>
<td>LDLSPPN$H_6$</td>
<td>Yes</td>
<td>S1095</td>
</tr>
<tr>
<td>$H_6$LDSSPPNTS</td>
<td>Yes</td>
<td>S1095</td>
</tr>
<tr>
<td>LDSAPPNTS$H_6$</td>
<td>No$^a$</td>
<td>—</td>
</tr>
</tbody>
</table>

$^a$When S1095 is changed to alanine, there is a trace amount of phosphorylation on S1094 that is absent in all other constructs.

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 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.

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>
<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).

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

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Chr. No.</th>
<th>Coordinate</th>
<th>Location</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>T²</td>
<td>Within PAF16 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 RAD55</td>
</tr>
<tr>
<td></td>
<td>XI/III</td>
<td>663,321/6,051</td>
<td>T²</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 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 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 ALG6 and YSP3</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>236,320</td>
<td>D</td>
<td>Within SED4</td>
</tr>
<tr>
<td></td>
<td>II, 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 ECM12</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>320,566</td>
<td>D</td>
<td>Between VAM6 and RXT3</td>
</tr>
<tr>
<td></td>
<td>VII</td>
<td>387,546</td>
<td>D</td>
<td>Within PYC1</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>765,652</td>
<td>D</td>
<td>Between CTH1 and GIR2</td>
</tr>
<tr>
<td></td>
<td>XII</td>
<td>434,128</td>
<td>D</td>
<td>Between YLR146W-A and SMD3</td>
</tr>
<tr>
<td></td>
<td>VIII</td>
<td>267,970</td>
<td>D</td>
<td>Within LRP1</td>
</tr>
<tr>
<td></td>
<td>VII</td>
<td>853,495</td>
<td>D</td>
<td>Between PBP1 and OKP1</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>419,780</td>
<td>D</td>
<td>Between YBR085C-A and IST2</td>
</tr>
<tr>
<td></td>
<td>IX</td>
<td>144,456</td>
<td>D</td>
<td>Within NUP159</td>
</tr>
</tbody>
</table>

¹T = telomeric and within 1500 bp of autonomously replicating sequence (ARS) in the telomeric X repeat; D = dispersed.
²Telomeric but more than 1500 bp from the X ARS.
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.

**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). 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

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.
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). Principle 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 PfuMI 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, 100mM 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 ZipTipc₁₈ (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 fragmented 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 rad52D 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 IV: RETROTRANSPOSON TARGET SITE SELECTION BY IMITATION OF A CELLULAR PROTEIN

Troy L. Brady, Peter Fuerst, Robert Dick, Clarice Schmidt, and Daniel F. Voytas

1Department of Genetics, Development and Cell Biology, 1035A Roy J. Carver Co-Laboratory, Iowa State University, Ames, IA 50011-365

Abstract

Target specificity of the retrotransposon Ty5 is mediated by a protein-protein interaction between Ty5 integrase and the host protein Sir4. We explored the hypothesis that Ty5’s targeting mechanism arose by acquisition of a Sir4 interaction domain from a cellular protein. Initially, deletion and mutation analyses of the Sir4 C-terminus (aa 950-1358) were carried out to establish criteria for the interaction between Sir4 and IN. These criteria were then used to screen for hypothetical cellular donors of Ty5 integrase’s targeting determinants. Esc1, which acts with Sir4 to partition DNA between mother and daughter cells, met our criteria: Esc1 interacts with the same domain of Sir4 as IN (aa 971-1082) and 75% of mutations that disrupt interaction between Sir4 and IN also completely disrupt or weaken Esc1/Sir4 interactions. Not surprisingly, the ability of these sir4 mutants to interact with Esc1 correlated with their capacity to partition DNA to daughter cells. Interestingly, Ty5 targeting was not abolished in all sir4 mutants tested. Targeted integration correlated strongly with the mutant protein’s ability to interact with Esc1, suggesting a role for Esc1 in target site selection in the absence of a Sir4/IN interaction. Tethering chromatin to the nuclear periphery did not appear to be important for integration specificity, as targeting was not affected in cells expressing wild type Sir4 but lacking either Esc1, Ku70, or both proteins. A small region of Esc1 (aa 1440-1473) previously reported to interact with Sir4 contains a sequence similar to the targeting domain of Ty5 IN. Like the Ty5 targeting domain, this motif is phosphorylated, and alanine substitutions in this motif disrupt two hybrid
interactions with Sir4. We conclude that Ty5 targets integration to heterochromatin by imitating the Esc1/Sir4 interaction.

**Introduction**

Transposable elements have inherent potential to alter genome organization. While some transposable elements do not have obvious insertion site preferences, many others frequently integrate into distinct regions of the host’s genome. Because mobile elements are dependent upon their host for survival, it would seem advantageous to both entities for insertion site preference to be regulated either by the host, the element, or both. Understanding how integration biases are determined will likely reveal new relationships between the host and mobile elements, increase understanding of how genomes are shaped by mobile DNA, and provide opportunity to harness mobile elements for genome modification such as therapeutic gene delivery.

In the baker’s yeast *Saccharomyces cerevisiae*, all retrotransposons (Ty1 – Ty5) display distinct target site preferences. Ty1 - Ty4 are located upstream of genes transcribed by RNA polymerase III, such as tDNA and rDNA. Ty5 integrates in regions of heterochromatin at telomeres and silent mating type loci. While the precise mechanism underlying Ty targeting to tDNA and rDNA remains elusive, targeting to telomeres and the silent mating loci results from a direct protein-protein interaction between Ty5 integrase (IN) and Sir4, a host protein integral to heterochromatin function. Essential for this interaction is a six amino acid motif in IN termed the targeting domain (TD). TD was recently shown to be a site of phosphorylation, and substitutions in TD that prevent phosphorylation abrogate interaction between IN and Sir4 leading to random integration of Ty5. The host, therefore, determines integration site specificity of Ty5 by post-translationally modifying IN to promote interaction with Sir4.

Sir4 is a large protein that serves as a molecular scaffold at the nuclear periphery. TD interacts with the Sir4 C-terminus (aa 951-1358), a region of the protein with similarity to nuclear lamins in that it encodes a coiled coil domain that contains lamin-like heptad repeats (Diffley and Stillman, 1989). This region can interact with numerous proteins, including Sir3
and the coiled coil domain of other Sir4 molecules (Chang et al., 2003; Murphy et al., 2003; Tsukamoto et al., 1997). Upstream of the coiled coil domain is a region of Sir4 involved in DNA partitioning and anchoring (PAD, 950-1262). When the PAD domain is ectopically tethered to otherwise unstable plasmids, it facilitates their transmission to daughter cells during mitosis (Ansari and Gartenberg, 1997). The PAD domain carries out its activity by interacting with Esc1, a protein important in tethering heterochromatin to the nuclear periphery (Andrulis et al., 2002). Due to the localization of Sir4 at chromosome ends, PAD-mediated tethering is also thought to be involved in chromosome segregation during mitosis. Chromosomal tethering occurs via redundant pathways involving Esc1 and yKu, a DNA repair protein complex that binds chromosome ends and interacts with Sir4 (Bertuch and Lundblad, 2003; Hediger et al., 2002; Taddei and Gasser, 2004; Tsukamoto et al., 1997).

Localization of DNA to the nuclear periphery is also thought to reinforce silencing by bringing heterochromatic regions into close proximity with silencing factors, such as Sir2 and Sir3, which localize to the nuclear periphery and also interact with Sir4. While tethering may function to strengthen silencing in these regions, it is not essential for maintenance of silent chromatin (Gartenberg et al., 2004; Taddei et al., 2004).

The relationship between Ty5 IN and Sir4 bears some similarities to interactions between retroviruses and host proteins localizing to chromatin and the nuclear periphery. Because IN appears to be the primary retroelement factor determining integration site choice (Lewinski et al., 2006; Zhu et al., 2003), host proteins that interact with IN have the potential to influence the site of integration. LEDGF/p75, a co-transcription factor, is one such protein that interacts with HIV IN and significantly impacts target site selection of this retrovirus. While HIV normally prefers to integrate in transcriptionally active regions, reducing LEDGF levels in vivo significantly lowers HIV integration and target site bias (Ciuffi and Bushman, 2006; Llano et al., 2006; Vandekerckhove et al., 2006). Furthermore, by tethering LEDGF to ectopic sites in vitro, it was demonstrated that integration could be directed to sites where LEDGF was located (Ciuffi et al., 2006).

HIV has also been shown to require the integral inner-nuclear envelope protein emerin for efficient cDNA integration (Jacque and Stevenson, 2006). Emerin, like other LEM (LAP2, Emerin, and MAN1 domain) containing proteins, binds barrier-to-
autointegration factor (BAF), an essential protein associated with chromatin structure and nuclear assembly (Lee et al., 2001; Segura-Totten et al., 2002; Zheng et al., 2000). Through its anchoring in the nuclear membrane and its interaction with lamin A and nesprin proteins, emerin is thought to mediate chromatin localization to the nuclear periphery (Bengtsson and Wilson, 2004). While BAF is known to reduce the autointegration of cDNA by HIV IN, the exact role of emerin in HIV integration is unclear. The lack of cDNA integration in cells deficient for these proteins suggests an interesting link between proteins involved in chromatin structure and localization and retroviral integration.

As the requirements for mobile element integration are revealed, an increasingly close relationship between these elements and host integration determinants is becoming apparent. Based on these observations, we hypothesized that Ty5 IN may be imitating or perhaps may have even commandeered a Sir4-interaction domain from some other cellular factor. Such a domain would have similar requirements for interaction with Sir4 as those of IN. To determine if a host protein containing such a region could be found, we characterized requirements for the IN/Sir4 interaction through deletion and mutagenesis studies. We then identified Sir4-interacting host proteins by a yeast two-hybrid screen and compared the interaction requirements of these proteins with those of IN. We demonstrate here a near equivalency in the Sir4-interaction profiles of IN and Esc1 and conclude that Ty5 targets integration to heterochromatin by imitating the host factor Esc1.

**Results**

**Defining the region of Sir4 that interacts with IN**

We previously demonstrated that Ty5 IN interacts with the C-terminus of Sir4 (residues 951-1358) (Xie et al., 2001), and using two hybrid assays, we showed that N-terminal truncations beyond residue 971 severely impair interaction with IN (Figure 1A) (Zhu et al., 2003). To define the C-terminal boundary of the IN-interacting domain, various Sir4 C-terminal truncations were tested in two hybrid assays (Figure 1B). Deletions removing up to 276 residues still interacted with IN (Figure 1A), indicating that the coiled coil domain of Sir4 (residues 1271-1346) is not needed. All Sir4 truncations were expressed at equivalent levels,
except for C-terminal truncations extending beyond 1082 (Figure 1B). Smaller C-terminal truncations may be capable of interacting with IN, but this could not be determined by yeast two hybrid assays due to protein instability.

Two residues in the IN interacting domain (W974 and R975) were previously shown to abrogate Sir4/IN interactions when mutated to alanine (Zhu et al., 2003). To identify additional residues, alanine point substitutions between 976 to 990 were performed and tested in yeast two hybrid assays for IN interaction (Figure 1C). This identified four additional residues (W978, L982, I985, V987). Together with the previously described mutations, six of the seven substitutions replace strongly hydrophobic residues, suggesting an interaction with IN mediated by hydrophobic residues or a need for these residues in maintaining proper structure of the interacting region.

The relatively large number of residues identified by alanine scanning (7 out of 19 tested) suggested that multiple residues throughout the 971-1082 region are important for IN interaction. To identify additional residues, a reverse two-hybrid screen was performed. Randomly mutagenized constructs encoding LexA-Sir4 (951-1358) were tested for loss of
interaction with IN (data not shown). Western blots were performed to eliminate those mutants that did not express Sir4 efficiently. Forty-one mutants that passed this screen were sequenced, and ninety percent (37 of 41) contained at least one mutation between residues 971 and 1082, correlating well with the deletion data. Of the 41 mutants, 4 had mutations in residues identified as important in the alanine scanning mutagenesis, suggesting that the screen did not reach saturation. The multiplicity of mutations isolated throughout Sir4 lead us to conclude that this minimal region of 971-1082 contains multiple residues that are needed to support interaction with IN. Six of the 41 mutants had single nucleotide changes that resulted in an amino acid replacement. These mutants along with three double mutants were characterized further. All of these mutations are located within the 971 to 1082 region of Sir4 required for interaction with IN with the exception of two mutations, one in each of the double mutants L984P, K1123R and T957L/K1037E (Figure 2A). The region between 954 and 1114 is conserved among Sir4 proteins from related yeast species, and all mutations except K1123 reside within this domain of conservation.

Figure 2. Conserved Sir4 residues are important for interaction with IN. (A) The location of Sir4 mutations are indicated on an alignment of Sir4 homologues from different species of *Saccharomyces*. The IN-interaction boundaries of Sir4 (971-1082) are designated by the area within the dashed rectangular box. The locations of alanine substitution mutants are marked by a circle (O) and reverse-two hybrid mutants by a triangle (▼). (B) Immunoblots of LexA-Sir4 proteins. LexA-Sir4 mutants are stably expressed compared to wild type.
Sir4-interacting proteins and their relationship to IN

Because of the conservation of the region of Sir4 that interacts with IN, we reasoned that this domain carries out a critical function and likely accomplishes this by interacting with other host proteins. To identify such factors, a yeast two hybrid screen was carried out using LexA-Sir4C (951-1358) as bait. Because Sir4 interacts with Sir2, Sir3, and itself, all candidate two hybrid interactions were tested in strains lacking these proteins to screen out interactions that may be bridged by one of these proteins (data not shown). The screen identified five proteins: Sum1, Chd1, Nma2, Esc1 (a known Sir4-interacting protein (Andrulis et al., 2002)) and Sir4 (which is known to homodimerize). Sum1 is a transcriptional repressor of middle sporulation-specific genes and is involved in telomere maintenance (Askree et al., 2004; Laurensen and Rine, 1991; Xie et al., 1999). Chd1 remodels nucleosomes and regulates transcription elongation (Simic et al., 2003; Tran et al., 2000) and Nma2 (Nicotinic acid mononucleotide adenylyltransferase) is involved in the NAD(+) salvage pathway (Emanuelli et al., 1999) and strengthens silent chromatin when overexpressed (Anderson et al., 2002). Esc1 is a recently described protein known to interact with Sir4 to tether telomeres and partition plasmids to daughter cells (Andrulis et al., 2002).

We hypothesized that IN may mimic one of the identified cellular proteins with respect to its interaction with Sir4. We therefore tested these four proteins against various Sir4 deletions and the W974A point mutant to determine whether their interactions resembled GAD-IN. In contrast to IN, Sum1 required the entire Sir4 fragment for interaction (951-1358), as it failed to interact with Sir4 deletions lacking the coiled coil domain (950-1250) or the N terminal region (1082-1358) (Figure 3). Nma2 and Chd1 both required the coiled coil of Sir4 but interacted with 1082-1358. Previous work has shown that Esc1 interacts with the region of Sir4 that includes residues 950-1150 (Andrulis et al., 2002), and our data supports this observation, demonstrating that the interaction does not require the coiled coil domain but does require the N terminal region for interaction. In assessing allele specificity, the Nma2/Sir4 interaction was not disrupted by the W974A mutation whereas interactions with Sum1, Chd1, and Esc1 were not observed with this mutant (Figure 3). Based on these results, we concluded that Esc1 shares the most similarity with IN in terms of its Sir4 interaction.
Figure 3. Interaction profiles of Sir4-interacting proteins. Yeast two hybrid assays measuring interaction of candidate proteins with various LexA-Sir4 constructs. (Top) Sir4 deletions used are shown in graphic form with labeled interaction domains. The interaction of these constructs with IN is shown to the right. (Bottom) Interaction profiles of Sum1, Nma2, Chd1, and Esc1 with Sir4 truncations or a Sir4 mutant are shown in the right and left panels, respectively.

**Esc1 interacts with the same region of Sir4 as IN**

To further characterize the similarity between the IN/Sir4 and Esc1/Sir4 interactions, Esc1/Sir4 two hybrid assays were conducted with the various Sir4 truncations and point mutants generated and tested with IN. Two fragments of Esc1 were tested: a GAD-Esc1 C-terminal fragment (1361-1658) isolated in our two hybrid screen and a 34 amino acid region of Esc1 reported previously to interact with Sir4 (Andrulis et al., 2002). Esc1 1361-1658 failed to interact with N-terminal truncations past 971 and C-terminal truncations preceding 1080 (Figure 4). The smaller Esc1 fragment (1440-1473) had only slightly different requirements, failing to interact with N-terminal truncations past 961 and C-terminal truncations preceding 1082 (Figure 4). This suggests that this small fragment of Esc1 is the primary region that interacts with Sir4.
We next tested Esc1 interactions with our collection of LexA-Sir4 mutants. Ten of the sixteen mutants failed to interact with the full length Esc1 C-terminus (Figure 5C, data not shown). Of the six that interacted, three (R975G, T957L/K1037E, and V987A) exhibited slow growth, indicating a weakened interaction. This effect was more pronounced on media containing higher concentrations of 3AT (data not shown). The shorter Esc1 fragment, 1440-1473, failed to interact with R975G and T957L/K1037E, suggesting that these mutations disrupt Sir4-interactions within the 1440-1473 region (Figure 5D). Mutant K1064M interacted as strongly as wild type Sir4 with either Esc1 fragment, and we conclude that this residue is specific for interaction with IN but is not needed for interaction with Esc1. Mutant K1050M/R1075G showed a slightly weaker interaction with Esc1 1440-1473, as evidenced by slow growth compared to wild type. Conclusions could not be drawn regarding the interaction of Esc1 1440-1473 with the alanine scanning mutants, because they were made in a shorter Sir4 fragment (971-1358), which failed to support an interaction at levels above the negative control. Collectively, the data show that Esc1 interacts with Sir4 in a fashion remarkably similar to that of IN and that the 34 amino acid Esc1 fragment encodes a strong Sir4-interacting region that behaves like the larger 1361-1658 C-terminus in its interactions with the C-terminus of Sir4.
Figure 5. Esc1 and IN interaction profiles with sir4 mutants are similar. Yeast two hybrid assays measuring the interaction of two Esc1 fragments with various Sir4 mutants. The majority of Sir4 mutants that fail to interact with IN also do not support interaction with either fragment of the Esc1 C-terminus. Slow growth of colonies expressing mutants R975G and T957L.K1037E suggests that interaction with the 34 amino acid fragment of Esc1 is important for establishing a stable interaction with Sir4.

**Correlation of mitotic stability and interaction with Esc1 in sir4 mutants**

The region of Sir4 that interacts with IN and Esc1 has been show to be involved in partitioning DNA to daughter cells during mitosis (Ansari and Gartenberg, 1997; Longtine et al., 1992). This domain, referred to as PAD (partitioning and anchoring domain) spans residues 950-1262. Partitioning activity of PAD was demonstrated in assays in which a LexA-Sir4 PAD fusion was tethered to a plasmid without a centromere or 2 micron origin. Interaction of the Sir4 PAD with Esc1 allows the plasmid to be partitioned, and mitotic stability is measured by retention of a plasmid marker after cell division. While the full PAD is required for optimal partitioning, a smaller region (950-1150) retains partial partitioning capability; however, removal of an additional 100 residues from the N- or C-terminus of this smaller fragment reduces partitioning to background levels (Ansari and Gartenberg, 1997).

We tested the sir4 mutants identified in the reverse two hybrid screen for their ability to confer mitotic stability to an otherwise unstable plasmid (Ansari and Gartenberg, 1997; Longtine et al., 1992). The two mutants that strongly interacted with Esc1 1361-1658 (K1064N and K1050M/R1075G), partitioned at levels indistinguishable from wild type. Similarly, all five of the mutants that failed or showed weakened interactions with Esc1 in two hybrid assays showed significantly reduced levels of plasmid partitioning. The data obtained with F1076L and L984P/K1123R were inconclusive due to high standard
deviations. In general, the ability of mutants to engage in yeast two hybrid interactions correlates well with their ability to efficiently partition plasmids.

**Sir4 mutants and targeted integration**

We next measured the ability of various Sir4 mutants to direct Ty5 integration to a target plasmid that has LexA-Sir4 proteins tethered to LexA operators. In this assay (Zhu et al., 2003), a plasmid-borne Ty5 under control of the *GAL1* promoter is induced for transposition

<table>
<thead>
<tr>
<th>Table 1. DNA partitioning of <em>sir4</em> mutants</th>
<th>Table 2. Targeting efficiencies of <em>sir4</em> mutants</th>
</tr>
</thead>
<tbody>
<tr>
<td>LexA-Sir4 (951-1358)</td>
<td>Mitotic Stability</td>
</tr>
<tr>
<td>Wild type</td>
<td>38.4 ± 7.4</td>
</tr>
<tr>
<td>LexA</td>
<td>8.6 ± 2.6</td>
</tr>
<tr>
<td>R975G</td>
<td>10.8 ± 4.4</td>
</tr>
<tr>
<td>S1031R</td>
<td>7.2 ± 3.4</td>
</tr>
<tr>
<td>S1047P</td>
<td>7.8 ± 5.8</td>
</tr>
<tr>
<td>K1064N</td>
<td>35.3 ± 14.7</td>
</tr>
<tr>
<td>F1076L</td>
<td>24.7 ± 18.2</td>
</tr>
<tr>
<td>V1077G</td>
<td>7.0 ± 1.8</td>
</tr>
<tr>
<td>T957L, K1037E</td>
<td>11.6 ± 3.6</td>
</tr>
<tr>
<td>L984P, K1123R</td>
<td>18.4 ± 18.4</td>
</tr>
<tr>
<td>K1050M, R1075G</td>
<td>27.8 ± 6.5</td>
</tr>
<tr>
<td>Wild type (971-1358)</td>
<td>9.37 ± 2.9</td>
</tr>
<tr>
<td>R975A</td>
<td>6.51 ± 0.5</td>
</tr>
<tr>
<td>L979A</td>
<td>9.2 ± 2.8</td>
</tr>
<tr>
<td>V987A</td>
<td>7.29 ± 3.0</td>
</tr>
<tr>
<td>L982A</td>
<td>7.92 ± 1.1</td>
</tr>
<tr>
<td>LexA</td>
<td>-</td>
</tr>
<tr>
<td>W974A</td>
<td>0.4 ± 0.2</td>
</tr>
<tr>
<td>W978A</td>
<td>0.26 ± 0.1</td>
</tr>
<tr>
<td>I985A</td>
<td>0.37 ± 0.2</td>
</tr>
</tbody>
</table>

* Esc1 1361-1658

by growth on media containing galactose. A subset of Ty5 insertions will occur on the target plasmid, and the ratio of target plasmids with and without a Ty5 provides a measure of targeting efficiency. Of the 16 mutants that failed to interact with Ty5 IN in two hybrid assays, eight showed a decrease in targeting efficiency greater than three-fold. Surprisingly, several of the mutants still targeted efficiently, some at levels equal to wild type (Table 2).
To better assess target specificity, the site of Ty5 integration was determined by DNA sequencing. Between seven and eleven independent transposition events were analyzed for each of several sir4 mutants (Figure 3). Even though these sir4 mutants failed to interact with IN, all but one of the insertion patterns resembled wild type with 90% of the insertions occurring within a window ±1000 base pairs flanking the LexA operators. This pattern held regardless of the mutant’s targeting efficiency. These results suggest that regions of IN or Sir4 other than those tested in the two hybrid assay interact and assist in mediating targeted integration. Alternatively, the IN/Sir4 interaction was not abolished by the mutation, but weakened below the threshold of detection by two hybrid assays.

The S1047P allele exhibited the greatest decrease in targeting efficiency and also showed the only significantly altered integration pattern. Of 11 insertions characterized, six occurred near the LexA operators, one occurred on the plasmid backbone, and the remaining four clustered near the plasmid’s autonomously replicating sequence (ARS). Adjacent to the ARS is an Abf1 binding site. Because Abf1 recruits Sir4 to telomeres, we tested the requirement of endogenous Sir4 for this alternative targeting pattern. Of seven insertions mapped in a sir4 deletion strain, two occurred near the lexA binding sites and the remainder appeared to have inserted randomly throughout the plasmid and were not associated with the ARS. This suggests that ARS-targeting is dependent on endogenous Sir4 and does not represent a secondary targeting mechanism that was unmasked in S1047P. This observation correlates with previous findings that show Ty5 preferentially integrates near X repeats, telomeric sequences that encode binding sites for the origin recognition complex and Abf1 (Zou et al., 1996). The effect of endogenous Sir4 on targeting efficiencies was tested for the remaining Sir4 mutants. Only mutant F1076L exhibited a significant decrease in targeting efficiency. F1076L displays the second lowest targeting efficiency, and it appears endogenous Sir4 supports the limited targeted transposition observed in this mutant.

**Ty5 transposition is not affected by loss of Esc1**

Of the sixteen sir4 mutants that failed to interact with IN, half still had targeting efficiencies comparable to wild type. A strong correlation was noted between the targeting efficiencies of these sir4 mutants and their ability to interact with Esc1 in yeast two-hybrid assays. In
fact, seven of eight mutants that targeted efficiently also interacted with Esc1 1361-1658, the exception being L982A (Table 2). This suggested a role for Esc1 in Ty5 targeting in the absence of a Sir4/IN interaction, perhaps serving as a co-factor for targeted integration.

To examine whether Esc1 was involved in Ty5 transposition, overall transposition frequencies and targeting efficiencies were monitored in esc1 mutants. Transposition frequencies were conducted in such a way so as to distinguish between cDNA recombination and integration, and targeting was measured to a tethered wild type LexA-Sir4 fusion (Ke and Voytas, 1999; Zhu et al., 2003). No significant differences between the esc1 mutant and wild type were observed for transposition, cDNA recombination or target specificity (Table3, data not shown). One function of Esc1 is to tether telomeres to the nuclear periphery, and we
reasoned that Esc1 may mediate its effect on localizing Ty5 targets to this specific nuclear compartment. Ty5 transposition frequencies were therefore also measured in strains lacking the redundant telomere tethering factor, Ku70, as well as in esc1, ku70 double mutants. Neither the single nor double mutants showed much difference in overall transposition frequencies, relative levels of cDNA recombination or target specificity.

Table 3. Ty5 targeting in the absence of nuclear periphery proteins

<table>
<thead>
<tr>
<th>LexA-Sir4</th>
<th>Fold decrease from WT Sir4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WT</td>
<td>sir4Δ</td>
</tr>
<tr>
<td>951-1358 wild type</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>R975G</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>K1064N</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>T957L, K1037E</td>
<td>1.0</td>
<td>1.1</td>
</tr>
<tr>
<td>K1050M, R1075G</td>
<td>0.9</td>
<td>1.5</td>
</tr>
<tr>
<td>971-1358 wild type</td>
<td>1</td>
<td>N.D.</td>
</tr>
<tr>
<td>R975A</td>
<td>1.4</td>
<td>N.D.</td>
</tr>
<tr>
<td>L979A</td>
<td>1.0</td>
<td>N.D.</td>
</tr>
</tbody>
</table>

While no difference in target specificity was observed with tethered wild type Sir4, we next tested sir4 mutants capable of targeting integration for their ability to target integration in a strain lacking Esc1. None of the mutants tested showed a significant change in targeting in the absence of Esc1, suggesting that Esc1 itself does not mediate targeting in the absence of a Sir4/IN interaction, and only the ability to interact with Esc1 correlates with the ability to target integration. Targeting of these mutants also did not change when either Sir4, Ku70, or both Esc1 and Sir4 or Ku70 and Esc1 were absent, showing that the ability of some sir4 mutants to mediate targeting is not dependent on these proteins.

Similarities between the C-termini of Ty5 IN and Esc1

In addition to similarities in their Sir4-interaction profiles, the C-termini of IN and Esc1 share common features at the amino acid sequence level. First, the PAD domain of Esc1 (1395-1551) is serine and proline rich (11.9% and 9.3%, respectively), a feature shared with
the Ty5 IN C-terminus (934-1131, 11.6% serine and 7.1% proline) and the integrase C-termini of other retrotransposons of the Ty1/copia family (Peterson-Burch and Voytas, 2002). This amino acid sequence composition is well above the average for proteins in the UniProtKB database (254609 sequence entries), which average 6.82% serine and 4.82% proline. In addition, we noticed that the 1440-1473 fragment of Esc1 encodes a sequence motif (1448-1453, LPSDPP) that has three of the four residues in Ty5 TD that are required for targeted integration. An alignment of Esc1 homologues from five different yeast species showed that this motif is located in a block of 13 highly conserved residues, suggesting this region plays an important role in Esc1 function (Figure 7A).

Ty5 TD alone can interact with Sir4 in two hybrid assays. To determine the minimal fragment of Esc1 1440-1473 capable of interacting with Sir4, we tested Gal4 binding domain
(GBD) fusions of the 13 amino acid conserved domain, nine amino acids that included the TD-like motif (LPSDPPSDK), and two alanine point mutants (S1450A and D1451A) of the 34 amino acid fragment for interaction with GAD-Sir4. Of all the proteins tested, only GBD-Esc1 (S1450A) failed to interact with Sir4. All proteins were expressed well, including the S1450A mutant (Figure 7B and C). This indicated that within this conserved domain, Esc1 encodes a TD-like motif capable of interacting with Sir4 and that the serine at position 1450 is critical for interaction with Sir4.

We tested the ability of these GBD-Esc1 fragments and point mutants to nucleate heterochromatin in a manner similar to TD. When tethered adjacent to an HMR silencer with mutations in cis-acting sequences important for silencer function, TD can nucleate heterochromatin and silence expression of a TRP1 reporter gene integrated at the HMR locus of a reporter strain (Chien et al., 1993; Xie et al., 2001). The extent to which cells grow on media lacking tryptophan reflects the ability of the expressed GBD fusion protein to nucleate factors of heterochromatin (Figure 8A). Consistent with the yeast two hybrid assays, the

![Diagram](image)

**Figure 8.** Minimal fragments of Esc1 and IN nucleate heterochromatin. (A) A representation of the heterochromatin nucleation assay with GBD-TD represented at the upstream activation sequence (UAS) binding site of GBD. The silencer is indicated on the left with wild type elements in upper case and defective elements in lower case (A, ARS consensus sequence; E, Rap1 binding site; B, Abf1 binding site). (B) Heterochromatin nucleation assays in cells with or without UAS binding sites and different combinations of defective silencer elements. Cells were serially diluted prior to spotting onto non-selective and selective media.
S1450A mutant was the only Esc1 construct unable to establish heterochromatin, presumably due to its lack of interaction with Sir4 (Figure 8B). Collectively, we conclude that the conserved region of Esc1 1443-1455 and the TD-like motif it encodes can interact with Sir4 and nucleate silent chromatin, closely resembling the heterochromatin-nucleating ability of TD.

**Esc1 1440-1473 is a phosphoprotein**

The addition of a phosphate to the second serine of the Ty5 IN targeting domain is necessary for efficient IN/Sir4 interaction and targeted integration. Due to the many similarities between Esc1 and IN, we questioned whether Esc1 1440-1473, like TD of IN, was a phosphoprotein. This fragment of Esc1 was expressed as a dual-tagged fusion protein in yeast and subjected to mass spectrometry after purification. Analysis by Maldi-TOF showed a mono and di-phosphorylated form of the peptide in addition to the unmodified form (Figure 9), indicating that this region of Esc1 can be phosphorylated *in vivo* and that Esc1 1440-1473, like TD, is post-translationally modified.

![Figure 9. Esc1 1440-1473 is phosphorylated. Purified Esc1 protein was subjected to Maldi-TOF mass spectrometry analysis. Circled peaks represent, from left to right, unphosphorylated, mono, and di-phosphorylated Esc1 peptides.](image-url)
Functional equivalence of IN and Esc1 domains

Given the similar interaction requirements of IN and Esc1 with Sir4 and the existence of small Sir4 interacting motifs in both proteins, we hypothesized that these domains are functionally equivalent and therefore able to functionally substitute for one another. The surprising modularity of the TD was shown previously when small peptide ligands were used to replace TD and redirect Ty5 integration to plasmids displaying a tethered ligand binding partner (Zhu et al., 2003). The high degree of conservation of Esc1 residues 1443-1455, however, suggested it may not share the modular properties observed for TD.

To test whether Ty5 TD and the conserved region of Esc1 were functionally equivalent, a domain swap was performed by replacing a TD-containing region of Ty5 with the conserved region of Esc1 (1443–1455) or the TD-like motif in this region (LPSDPP). S1451A mutations were introduced into both Ty5-Esc1 chimeras and tested along with the wild type versions for targeted integration using our tethered targeting assay. A target plasmid lacking LexA operators was used as a negative control. As expected, both chimeras with wild type Esc1 sequence targeted as well as wild type Ty5. Similar results for both S1450A mutant chimeras, however, were unexpected (Figure 10). These results indicate that while this mutation disrupts interaction with Sir4 in yeast two hybrid assays, it has no effect when placed in the context of the Ty5 IN protein.

<table>
<thead>
<tr>
<th>Ty5 Construct</th>
<th>AA Sequence</th>
<th>Targeting Efficiency</th>
<th>Fold difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>...PPSI0SPPPNTSNF...</td>
<td>13.00 ± 0.63</td>
<td>1</td>
</tr>
<tr>
<td>Neg Control</td>
<td>...PPSI0SPPPNTSNF...</td>
<td>0.96 ± 0.22</td>
<td>0.07</td>
</tr>
<tr>
<td>Ty5 s-1</td>
<td>...PPSI0SPPPNTSNF...</td>
<td>12.13 ± 0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>Ty5 s-1 (S1450A)</td>
<td>...PPSI0SPPPNTSNF...</td>
<td>14.71 ± 1.89</td>
<td>1.13</td>
</tr>
<tr>
<td>Ty5 s-2</td>
<td>...PPSIPADPNTSNF...</td>
<td>12.5 ± 4.2</td>
<td>0.96</td>
</tr>
<tr>
<td>Ty5 s-2 (S1450A)</td>
<td>...PPSIPADPNTSNF...</td>
<td>8.03 ± 2.48</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Figure 10. Targeting efficiencies of chimeric Ty5 elements. Ty5 elements containing wild type or mutant Esc1 sequence in place of TD were tested for the ability to target integration using the tethered targeting assay described previously.

The reciprocal swap was also performed, replacing Esc1 1443-1455 with TD-containing sequence from Ty5 or an equivalently sized region of GST predicted to have a
similar structure to the Esc1 motif. A C-terminal polyhistidine epitope was added to monitor protein levels. These Esc1 chimeras were introduced into an esc1Δ strain on single copy plasmids and tested for their ability to partition DNA as previously described. Replacement of the conserved region of Esc1 with GST abolished all partitioning activity, indicating this Esc1 domain as essential for DNA partitioning. Despite TD being the only Ty5 sequence with similarity to this highly conserved Esc1 domain, the TD-containing Esc1 chimera

**Figure 11.** DNA partitioning efficiencies of chimeric Esc1 proteins that contain Ty5 TD or GST sequences. The exchanged sequences are highlighted, with the wild type IN and Esc1 proteins shown above. The presence of absence of LexA operators on the marker plasmid is indicated (+/-).

retained low, but significant, levels of DNA partitioning. The ability of these domains to functionally replace one another confirms their equivalency and supports the hypothesis of Ty5 targeting by an IN-Sir4 interaction that mimics a conserved mechanism for partitioning DNA during mitosis.

**Discussion**

Retroelement replication requires extensive use and manipulation of host cell machinery. For the yeast retrotransposon Ty5, integration into regions of heterochromatin is a selective process mediated by the interaction of IN with Sir4. We hypothesized that Ty5 IN did not evolve a unique Sir4-interacting domain, but instead mimics an existing Sir4 interacting protein. Such mimicry predicts that the two proteins would interact with Sir4 in a similar fashion.
Characterization of the Sir4/IN interaction

To measure how similar other host factors were to IN, we first established criteria for the interaction of IN with Sir4. Using yeast two hybrid assays to measure interaction between these proteins, we generated Sir4 N- and C-terminal deletions and mapped the interaction boundaries of Sir4 to between residues 971 and 1082. We also generated Sir4 mutants that failed to interact with IN using a combination of directed alanine scanning and random mutagenesis approaches. Although we cannot rule out the possibility of mutation-induced misfolding of these Sir4 mutants, all were expressed at levels comparable to wild type.

Sir4 mutants vary in their ability to partition DNA and target integration

Sir4 interacts with many proteins to facilitate cellular processes, such as DNA partitioning, heterochromatin formation, and targeted integration of Ty5. To investigate whether any of our sir4 mutants failed to function in any of these roles, we used two previously described assays: a mitotic stability assay for measuring DNA partitioning (Ansari and Gartenberg, 1997) and a tethered targeting assay to measure Ty5 target specificity (Zhu et al., 2003). For the former assay, Sir4 is tethered at LexA operators on an unstable plasmid, and partitioning to daughter cells occurs in an Esc1-dependent manner. Five of nine sir4 mutants showed background levels of partitioning, two showed highly variable partitioning, and two consistently partitioned at near wild type levels.

The tethered targeting assay measures the percentage of transposition events that occur on a target plasmid. This plasmid contains four tandem LexA operators to which LexA-Sir4 fusions bind and direct Ty5 integration to these sites. Although none of the sir4 mutants we generated interacted with IN by yeast two hybrid assays, half of them still retained the ability to target integration at levels close to wild type. A strong correlation was noticed between the ability of a mutant sir4 protein to target integration and its ability to interact with Esc1. Of the eight LexA-sir4 mutants that still sustained targeted integration, seven of them interacted with Esc1 by yeast two hybrid assays. We questioned whether Esc1 could be involved in mediating integration. To distinguish between a need for Esc1 versus other factors involved in DNA localization also present at the nuclear periphery, we measured levels of targeting in strains lacking Esc1, Sir4, and Ku70, a DNA repair and
chromatin-anchoring protein essential for localization of telomeres to the nuclear periphery (Laroche et al., 1998). Strains lacking a combination of Esc1/Sir4 or Esc1/Ku70 were also tested. Targeting in all of these backgrounds remained unchanged, indicating none of these nuclear periphery proteins are needed for targeting or integration in this assay. The correlation of targeting and the ability to interact with Esc1 could stem from the ability of Sir4 to recruit an uncharacterized Esc1-like protein that is involved in integration. Alternatively, 2 micron plasmid localization to the nuclear periphery may remain unchanged in these genetic backgrounds, negating the need for chromatin anchoring factors.

Because transposition involves a full length Ty5 element, it is possible that regions of IN that were not included in the GAD-fusion protein used for yeast two hybrid assays could also mediate sufficient interaction for targeted integration. The correlation of targeted integration and the ability to interact with Esc1, however, suggests other factors may be involved. For example, a complex of proteins may be involved in Ty5 targeting with Sir4 being the nucleating protein for this complex. Interaction with Sir4 may be crucial for bringing Ty5 in close proximity with other proteins located at sites of Sir4 localization. Targeting was unaffected in the deletion strains we tested, but the possibility remains that other factors, likely localized at the nuclear periphery, could somehow compensate for a weakened Sir4/IN interaction.

**Comparison of Esc1 and IN**

To find host factors that IN may imitate in its interaction with Sir4, we isolated Sir4-interacting proteins in a genome-wide yeast two hybrid screen (using Sir4 951-1358 as bait) and screened them for an interaction profile similar to that of IN. This was accomplished using Sir4 truncations that lacked the PAD or coiled coil domains or that contained a point mutation known to disrupt Sir4/IN interaction by yeast two hybrid assays. Of the four candidate proteins tested, only Esc1 proved to depend on the PAD domain for interaction and failed to interact with the *sir4* point mutant, a profile matching that of IN. The C-terminus of Esc1 is known to interact with the PAD domain but does not require the coiled coil of Sir4 (Andrulis et al., 2002). This same study reported that 100 amino acid deletions beyond 950-1150 disrupted interaction with Esc1. We reasoned that the actual interaction boundaries may
be smaller and more closely resemble those of IN. Using yeast two hybrid assays to measure
binding of Esc1 and our collection of Sir4 N- and C-terminal truncations, this was shown to
be true with the interaction boundaries of the Esc1 C-terminus found to be nearly identical to
those of IN (971 and 1080 versus 971 and 1082, respectively). We also asked whether Esc1
could interact with the same Sir4 mutants that failed to interact with IN. Of 16 mutants
tested, ten also failed to interact with Esc1 (1368-1658). Of the six that did interact, three
appeared to have a weakened interaction with Sir4. These data indicate that the C-termini of
both Esc1 and IN share many of the same requirements for interaction with Sir4.

To locate the IN-like region within Esc1, we further characterized the interaction
between Esc1 and Sir4. To determine whether a smaller region of Esc1 was primarily
responsible for interacting with Sir4, we tested a small region of 34 amino acids previously
reported to interact with Sir4 (Andrulis et al., 2002) using our collection of Sir4 deletions and
various mutants. This fragment interacted with Sir4 in a near-identical fashion to that
observed with the larger Esc1 C-terminal fragment, exhibiting N and C-terminal interaction
boundaries of 961 and 1082. It also failed to interact with seven of the nine Sir4 mutants
tested, interacting strongly with the same two mutants as the larger C-terminal fragment. In
addition, a correlation between the ability of a Sir4 mutant to interact with this small Esc1
fragment and its ability to reliably partition DNA was observed. This indicates that the
ability of Sir4 to interact with this small region of Esc1 determines the efficiency of DNA
partitioning during mitosis. Together, these data suggest that Esc1 1440-1473 is a major
Sir4-interacting region of the Esc1 C-terminus and encodes Sir4-interaction properties
similar to that of IN.

A minimal Sir4-interacting region of Esc1 nucleates heterochromatin

The targeting domain of IN consists of six amino acids, four of which are essential for
targeting (L_SS_P). Aligning homologues of Esc1 from five different yeast species showed
Esc1 (1440-1473) to contain a highly conserved block of 13 residues, inside of which resides
a sequence very similar to that of TD (1448-1453, LPSDPP). This entire region of 13
conserved residues as well as a smaller region of nine amino acids (containing the TD-like
sequence) was sufficient for interaction with Sir4 by yeast two hybrid assays. Alanine
substitutions into two residues of the TD-like sequence revealed a requirement for S1450, but not for the adjacent residue, D1451. This may reflect a difference in the molecular interactions of TD and this TD-like region of Esc1 with their partner, Sir4. Mutation of either serine in TD abrogates interaction with Sir4 whereas the negatively charged aspartate of Esc1 (D1451), which could correspond to TD’s phosphoserine, does not appear to be required. Consistent with yeast two hybrid data, only the S1450A mutant failed to recruit silencing factors in a manner similar to that of TD when bound adjacent to a mutated silencer. These data demonstrate that the ability of Esc1 1443-1455 to interact with Sir4 and nucleate heterochromatin mirrors that of IN and its Sir4-interacting domain TD. Analysis of two alanine substitutions in Esc1 also suggests that, while these domains may serve to mediate interaction with Sir4, they may be specialized for the biological function of the respective proteins.

**Functional equivalence of IN and Esc1 domains**

To determine if the IN and Esc1 domains that interact with Sir4 encode equivalent functions in vivo, we exchanged them by replacing TD (LDSSPP) with the 13 amino acid motif from Esc1 and vice versa. Targeting of the chimeric Ty5 elements was found to be indistinguishable from wild type, indicating that substitution of Esc1 residues in place of TD efficiently targets Ty5 integration. Although the S1450A mutation abolishes yeast two hybrid interaction of Esc1 with LexA-Sir4, it has no detectable effect when placed in the context of the IN protein. The chimeric Esc1-TD protein still functioned to partition plasmids above background levels while substituting sequence from GST instead of TD reduced partitioning to background levels. The inability of Esc1-GST to partition plasmids supports data indicating this region as being critical to DNA partitioning and shows conclusively that IN and Esc1 share functionally equivalent domains, regions that are essential to the function of both proteins and that are crucial in mediating interaction with Sir4.

The equivalence of these two domains raises many important questions regarding both Esc1 and IN. Phosphorylation of IN has been shown to be required for proper interaction with Sir4 and targeted integration. Is the Esc1-Sir4 interaction regulated in a
similar manner? Is this conserved domain of Esc1 the only region regulating this interaction? The kinase responsible for regulation of Ty5 targeting is not yet known. Does the same kinase act on both proteins? The need for regulating interaction between Esc1 and Sir4 is likely needed during cell division. What is the timing of this regulation and how may it relate to the integration of a mobile element? Is Ty5 transposition optimal during a particular phase of the cell cycle? If so, how would IN be held in check until the time is right for integration? Gaining answers to these questions is the subject of ongoing research. Of particular interest will be the identification of the kinase that phosphorylates TD. This kinase, along with the parallels between Esc1 and IN in recognizing Sir4, will make it possible to further probe this interesting relationship between the Ty5 retrotransposon and its host.

Experimental Procedures

Yeast Strains and Media
Complete supplement mixtures from BIO101 Systems were used for yeast growth under selective conditions. YPH499 or isogenic strains with gene deletions were used in the tethered targeting and mitotic stability assays. Esc1, Ku70, and Sir4 deletions were performed by the one step gene knockout method (Ausubel, 1987), using plasmids pFA6a-hphNT1 and pFA6a-KANMX to amplify fragments containing genes for Hygromycin B and G418 resistance, respectively. Deletions were confirmed by phenotypic growth and PCR analysis. The yeast two hybrid reporter strain L40 was used for measuring protein-protein interactions with LexA-Sir4 by spotting serial dilutions of an overnight culture on media lacking histidine and supplemented with 1mM 3AT to reduce background. Strain PJ69 (James et al., 1996) was used to measure interaction of GBD-Esc1 proteins with GAD-Sir4 in a similar manner. Strains YSB1, 2, 35, and 41 for monitoring nucleation of heterochromatin are described elsewhere (Chien et al., 1993). Strain JTY142 (MATa his3-A200, ura3-52, trp1Δ63, leu2Δ1, can1, prb1-Δ1.6R, pep4::HIS3) was used for expression and purification of GST-Esc1-6HIS proteins. All cultures were grown at 30C unless noted otherwise.
Plasmid Construction

LexA-Sir4 constructs were made by PCR amplification using as template a LexA-Sir4 (950-1358) plasmid described previously (Ansari and Gartenberg, 1997). PCR products were digested with EcoRI/BglII and ligated into the EcoRI/BamHI sites of pBTM116 (Moretti et al., 1994). A primer-encoded six-histidine tag was added to the C terminus of these 1440-1473 constructs. GAD-Esc1 () were made by annealing complementary primers and ligating into the BamHI/XhoI and Sall/XhoI sites, respectively, of pGAD-C1. GBD-Esc1 plasmids were generated by digesting GAD-Esc1 plasmids (made by PCR amplification or primer annealing followed by ligation into EcoRI/BamHI for 1440-1473 constructs or BamHI/XhoI and Sall/XhoI sites of pGAD-C1 for 1443-1455 and 1448-1456 constructs, respectively) with EcoRI/EcoRV and ligating into EcoRI/MscI of pGBD-U (James et al., 1996). GAD-Sir4 (950-1358, pJB67) was made by moving the Sir4-containing EcoRI/PstI fragment from pYZ127 into the same sites of pYZ277, a LEU2-marked version of pBTM116. To make pYZ277, a TRP1-containing PstI/PvuII fragment of pGBD was cloned into the same sites of pGAD-C1. Chimeric Ty5 and Esc1 plasmids were made using four primer PCR to introduce the desired sequences and cloned into the BspEI/PflMI sites of Ty5 (pTB60). pTB60, a derivative of pDR14, was created by cloning the BspEI/NotI fragment of pWW32 to tag IN with an RGS6HIS sequence to create pTB32. The XhoI/NotI fragment of pTB32 was then moved into pRS426 to create a 2um based, IN-tagged Ty5 to facilitate analysis of IN protein levels. Chimeric his-tagged Esc1 constructs were made by cloning 4 primer PCR products into ApaI sites of pDZ45, a centromeric plasmid containing the entire Esc1 reading frame and reported to restore portioning to esc1Δ cells (Andrulis et al., 2002). Addition of a 15 histidine tag was added to the very C terminus to monitor protein levels. LEU2-marked target plasmids for the tethered targeting assay were made by ligating a 4 kb BglI/NsiI, LEU2-containing fragment from pRS425 into a 10.3 kb, BglI/Nsil-gapped pYZ316 and 317 (REF) to create pCS434 and pCS430 containing 4 and 0 tandem LexA operators, respectively. To create a suitable plasmid for generating a LexA-Sir4 mutant library by recombination in yeast, pCS439 was made by substituting silent mutations into pYZ127 (LexA-Sir4 950-1358) at Sir4 residues K971 and R1331 to introduce Sacl and PpuMI sites, respectively. All plasmids constructed using PCR were verified by sequencing.
Yeast Two-Hybrid screens and interactions

Isolation of Sir4 mutants was performed by construction of a mutant LexA-Sir4 library using mutagenic PCR conditions described previously (Cadwell and Joyce, 1994). Briefly, PCR amplified Sir4-6HIS was used as template for amplification with oligos dvo3608 and 3609. Polymersase from Eppendorf (Triple Master, 5 units) was used in a 100 µL reaction containing final concentrations of 7 mM MgCl$_2$, 0.5 MnCl$_2$, with dATP and dGTP at 0.2 mM and dCTP and dTTP at 1mM. After a two minute denaturation step at 94 C, 20 cycles were performed under the following parameters: 94C 20sec, 55C 20sec, 68C 90sec, and completed with a ten min incubation at 68C. Mutagenized Sir4 fragments were recombined into a gapped pCS439 plasmid (SacI/PpuMI) and plated on media selecting only for the presence of the bait and prey plasmids. Transformants were stamped after two days growth at 30C to selective media lacking histidine and scored for growth defects after two to three days incubation. Candidates were streaked onto SC-LT + dextrose, inoculated into 96 well plates, and retested for loss of interaction with Sir4 by replicating to selective media. Colonies unable to interact were subjected to colony PCR using primers annealing to the LexA plasmid to detect the presence of a Sir4 insert. Reactions giving no PCR product were repeated to minimize false negatives. The remaining pool of candidates was then screened by immunoblot analysis to detect nonsense mutations or unstable Sir4 proteins. Plasmids were rescued from yeast by the glass bead method (Ausubel, 1987) and sequenced.

Measurement of Plasmid Segregation

Mitotic Stability was measured as described previously (Ansari and Gartenberg, 1997), with the exception of cell growth in 96 well plates at 30C until cultures reached late log density (36-48 hours). Segregating properties of chimeric Esc1 proteins were performed in like manner, but with the addition of a LEU2 marked Esc1-containing plasmid encoding either wild type HIS-tagged Esc1 (pTB227) or TD (pTB223) or GST (pTB225)-containing derivatives of pTB227. Cells were assayed in the same manner with the exception of being grown in SC –TLU and plated to SC-TL and SC-TLU + dextrose.
**Tethered targeting and transposition**

Targeted integration of Ty5 was measured using a modified version of the published protocol (Zhu et al., 2003). Briefly, cells containing a Ty5 (URA3), LexA-Sir4 (TRP1), and target plasmid (pCS430 or 434, LEU2) were grown overnight in 96 well plates in selective media containing glucose. Cells were resuspended and approximately 80 µL were spotted onto selective agar media containing 2% Galactose and 0.5% Raffinose. Cells were allowed to grow three days at room temperature (22C). Plates were replica-plated to SC-LH + dextrose to select for the target plasmids and transposition events and grown for two to three days at 30C. Resulting colonies were then scraped, resuspended in 5 mL of liquid SC-LH + dextrose and refreshed overnight. The following day cells were pelleted, washed once with water, and plasmids recovered using the glass bead method (Ausubel, 1987). 2µL of plasmid-containing supernatant was used to transform 100 µL of the hisB deficient E.coli strain eDW335 and cells were allowed to recover in 1.5 mL SOC media for 45-50 minutes at 37C with shaking. 500 µL of LB + Chloramphenicol (Chlr, 100 µg/mL) was then added and cells were grown an additional three hours. Cultures were transferred to 2 mL microcentrifuge tubes, pelleted one minute at 8,000xg, washed with one mL of water, and pelleted again. The resulting pellet was resuspended in water and plated to LB + Chlr (20µg/mL) with a ten-fold greater volume being plated to M9 + Chlr. Plates were incubated one to three days before counting.

**Heterochromatin nucleation studies**

Yeast strains YSB1, 2, 35, and 41 were transformed with plasmids encoding GBD-Esc1 fragments and single colonies were grown overnight in SC-U + dextrose to saturation. Serial ten-fold dilutions were replicated to SC-U or SC-TU media and grown 2 to 3 days at 30C to monitor respective levels of silencing at the HMR locus.

**Protein analyses**

Cultures were grown in selective media to mid log phase (OD1-2), pelleted, washed with water, and the pellet frozen immediately at -20C. Cells were prepared for analysis by resuspending the pellet in 2x sodium dodecyl sulfate (SDS) sample buffer (Ausubel, 1987) to equal ~6x10^7 cells/5µL and boiling 5 minutes using a thermocycler. Between 3-6x10^7 cells
worth of lystate was separated by denaturing polyacrylamide gel electrophoresis. Protein levels were analyzed by immunoblotting with antibodies recognizing a penta-histidine epitope (Qiagen), GBD (Santa Cruz Biotechnology), or LexA (Upstate) at 3,000, 1,000, and 10,000 fold dilutions in TBST + 3%, 3%, and 5% milk, respectively. HRP-conjugated secondary antibodies recognizing IgG antibodies from mouse or rabbit were used at 5,000 and 10,000 fold dilutions in respective solutions of TBST + milk. Proteins were detected using ECL Western Blotting Detection Reagents (Amersham). Mass spectrometry was performed at the Roy J. Carver Co-Lab Proteomics Facility on proteins purified from cell lysates generated using the glass bead method (Ausubel, 1987). GST-Esc1-6HIS expression was induced by addition of galactose and cultures were allowed to reach late mid to late late phase before harvesting. Esc1 proteins were purified by nickel chelate chromatography following by a second purification using agarose beads with covalently linked glutathione. The resulting eluate was digested with thrombin protease to remove GST and purified again by nickel chelate chromatography. The resulting eluate was subjected to mass spectral analysis

**Protein Alignments**

Homologues of Sir4 and Esc1 were obtained by tBLASTn searching using the Sir4 PAD domain (aa950-1250) and Esc1 C terminus (aa1395-1551). Protein sequences were aligned using ClustalX and shaded in black and white using the program Boxshade version 3.21.

**Acknowledgements**

We would like to thank Philip James for the generous gift of GAD yeast two hybrid library DNA (James et al., 1996), Marc Gartenberg for the gift of pAA0 and pAA6 for measuring mitotic stability (Ansari and Gartenberg, 1997), Rolf Sternglanz for the gift of the Esc1-containing plasmid, pDZ45 (Andrulis et al., 2002) and yeast strains YSB1, 2, 35, and 41 for use in heterochromatin nucleation assays (Chien et al., 1993), and David Amberg for kindly providing the yeast strain JTY142. We also thank Siquan Luo at the Iowa State University Proteomics Facility for carrying out the MS analysis.
References


GENERAL CONCLUSIONS

Integrate (IN) is essential for the replication of retroelements and functions primarily as a molecular machine that breaks and joins DNA. In order for IN to carry out this critical activity, it must also mediate other processes, such as nuclear entry and chromosomal localization of integration complexes. The yeast retrotransposon Ty5 serves as a simple model system to learn about fundamental activities of IN that can then be applied to help understand more complex mobile elements such as the retroviruses.

The GKGY domain functions to stabilize IN and RT

The family of mobile elements to which Ty5 belongs carries a conserved motif at the beginning of their IN C-termini that is referred to as the GKGY motif. Here we characterized this motif and explored its function through mutational studies of Ty5 and Ty1. A variant of this motif was found to reside in Ty1 and its close homologues. While the Ty1 motif significantly differed from the motif found in Ty5 and other Ty1/copia elements, both motifs are predicted to have similar secondary structures and both have comparable mutant phenotypes, suggesting they carry out similar functions. While the sensitivity of each element to mutation in this domain varied, the collective results were the same: decreased transposition was always accompanied by reduced levels of IN and RT. Protein instability appeared to be specific to IN and RT, as overall levels of Gag remained unchanged in the mutants. In addition, polyprotein levels of Ty1 mutants did not reflect the same reduction seen with mature IN and RT protein levels. We conclude that the GKGY domain is required for the stability of these two proteins and may be required for efficient processing by PR. Future work should determine whether this domain acts in cis or trans, whether RT requires IN for reverse transcription, and if IN and RT of Ty1 and Ty5 directly interact. Should IN/RT interaction be confirmed, it would be important to know whether the GKGY motif is needed for such protein-protein interactions.
**Phosphorylation of Ty5 mediates targeted integration**

The site of integration is of critical importance to both the element and the host. Integration of Ty5 into heterochromatic regions is mediated by the interaction of IN with the host protein Sir4. A targeting domain (TD) at the very end of IN is necessary for this interaction and for targeting Ty5 insertions. Here we investigated the molecular determinants of the TD/Sir4 interaction and their relevance for targeted integration. The targeting domain was found to be phosphorylated, a modification necessary for interaction with Sir4 in vitro and targeted integration in vivo. Mutations preventing phosphorylation of TD abrogated its ability to nucleate heterochromatin when fused to GBD. Furthermore, when such mutations were introduced into transposition-competent Ty5 elements, they resulted in non-targeted insertions that landed in coding regions more than half of the time. The host plays an active role in regulating phosphorylation of TD, as growth of the host strain under stress conditions, specifically starvation for amino acids and carbon, reduced phosphorylation of TD. Future work will focus on determining the kinase responsible for modifying TD and exploring how kinase activity affects Ty5 transposition and targeting.

**Ty5 targets integration by imitating the host factor Esc1**

How mobile elements gain the ability to target insertions is likely to reveal new insights into mechanisms of targeted integration and transposition. Here we investigated the hypothesis that Ty5 mimics a cellular factor in order to interact with Sir4. We found that IN and Esc1 interact with Sir4 in a very similar fashion. We also found a conserved region of Esc1 that interacts with Sir4 that is involved in DNA partitioning and can mediate Ty5 targeting when substituted for TD. Similarly, TD can partially maintain Esc1 function when substituted for this conserved Esc1 domain. These two domains, therefore, are functionally equivalent, and we conclude that IN, by use of the TD, mimics a conserved region of Esc1 that is essential for Esc1 function in vivo. Future work will ask whether what is currently known about the TD/Sir4 interaction can be applied to the Esc1/Sir4 interaction to understand how it stably partitions DNA during mitosis. Also under current investigation is whether IN and Esc1 must compete for interaction with Sir4 and whether the ability of Sir4 to interact with Esc1 influences targeted integration.
**Biology of Ty5 integrase: insights into the timing and mechanics of Ty5 integration and replication**

The data described in this thesis collectively provide several insights into the timing and mechanics of Ty5 integration and replication. Preliminary studies involving the GKGY domain of IN indicate a role in IN/RT stability, potentially influencing reverse transcription. The similarity between this motif and the SH3-like domains of retroviruses is intriguing. Should the GKGY motif prove to be structurally similar to SH3 domains, a functional study of this domain in retrotransposons may shed light on the enigmatic function of the IN C-terminus in retroviruses. The high level of conservation found in the GKGY motif of Ty1/copia elements suggests that an understanding of this domain would apply to all elements of this family. Because IN seems to be required in vivo for optimal RT activity in multiple systems, understanding the relationship between IN and RT during replication has broad applications in the development of antiviral therapies. Also of interest is why this domain is unique to the Ty1/copia family and whether this motif served as a basis for the evolution of this unique and pervasive lineage of retroelements.

Targeting of Ty5 to “safe havens” within the genome is shown in this thesis to be actively controlled by the host. TD phosphorylation, a step that is required for the TD/Sir4 interaction and targeted integration, is maintained by cellular machinery but can be perturbed under certain conditions. How regulation of TD phosphorylation occurs is not yet known. Host control over Ty5 targeting may be an indirect consequence of cellular stresses. It is possible, however, that either the host or Ty5 itself, under certain stress conditions, actively inhibits the ability to target insertions with the hope of inducing genomic modifications that result in greater cellular fitness under the given stress condition. How Ty5 transcription, translation, and integration are affected by these stresses is under current investigation. Of prominent interest are the kinase(s) involved in the regulation of targeting and how changes in kinase activity affect Ty5 transposition.

The exact function of TD remains to be determined. While shown to interact with Sir4, in vitro interaction data suggests TD is not the sole determinant of the interaction between IN and Sir4 (Xie, 2003). TD can be replaced by exogenous motifs to constitute a chimeric IN with activity equal to that of a WT Ty5 (Zhu et al., 2003). While interacting
with Sir4, TD may, in fact, act as a regulatory element for IN activity or as a means to access DNA in heterochromatin. Other regions of IN may interact with Sir4 to dock IN at silent DNA, but phosphorylation of TD may be required for a more intimate interaction with Sir4 that leads to heterochromatin remodeling or to an active integration complex. Replacing TD with motifs that conserve IN structure may relieve IN from the regulation imposed by TD, explaining the ability of these swaps to target integration at levels indistinguishable from wild type. This model is not without precedence, as the retrotransposon Tf1 in Schizosaccharomyces pombe encodes a chromodomain-like motif that appears to inhibit IN activity: removing this motif from Tf1 leads to enhanced activity and less substrate specificity in vitro (Hizi and Levin, 2005).

The mechanism by which Ty5 gains access to DNA in regions of heterochromatin is also of interest. Integration may occur following chromatin remodeling by IN itself or by host factors recruited to integration sites by Ty5. Alternatively, integration may be linked to certain points in the cell cycle when heterochromatic regions are dismantled, such as during DNA replication. Data presented here provides potential insights into this second model by describing the similarities in the interaction between Sir4, a component of heterochromatin, and IN and Esc1. Esc1 interacts with Sir4 to tether chromatin to the nuclear periphery, and IN interacts with Sir4, presumably at the nuclear periphery, by mimicking a Sir4-interacting region of Esc1. Although the biology of the Esc1/Sir4 interaction has not yet been thoroughly studied, the interaction of these molecules is likely conditional and influenced by the cell cycle. We found that a 34 amino acid region of Esc1 containing the conserved TD-like domain is phosphorylated, supporting a role for phosphorylation in the regulation of Sir4/Esc1 interaction. Determining which residues are phosphorylated and whether these modifications are needed for interaction with Sir4 is of primary importance. As with the TD, knowing which kinase carries out these modifications and how this kinase is regulated in relation to Esc1 function will also shed light on how tightly IN and Esc1 are linked in terms of their interaction with Sir4. Should the Esc1/Sir4 interaction be regulated, it will be interesting to compare the dynamics and effects of that regulation on the IN/Sir4 interaction and Ty5 integration.
Lastly, recent studies have uncovered what appears to be a correlation between integration and the localization of other integration complexes to chromatin and the nuclear periphery. LEDGF and BAF, both of which are DNA binding proteins, are required for efficient HIV infection. Emerin, a protein-binding partner of BAF, has also been reported to be required for integration and is localized at the nuclear periphery (Jacque and Stevenson, 2006; Llano et al., 2006). It is interesting to compare these findings with Ty5, which interacts with a DNA-binding protein that is also localized at the nuclear periphery (Sir4) through its imitation of another protein at the nuclear periphery, Esc1. Very little is known about Esc1 and the proteins it interacts with, but due to the similarities between IN and Esc1, it will be interesting to determine if any such Esc1-interactors influence integration activity or target site selection.

References


ACKNOWLEDGEMENTS

I would like to thank my professor, Daniel Voytas, for his patience and perseverance in teaching and guiding me to think critically yet creatively and for his mentorship in science, generosity and friendship. His influence has become part of who I and will always remain with me. I gladly acknowledge many good friends for their contribution to my training and education as well as their good company. Their presence helped make the long hours and frustrations bearable and even, at times, enjoyable. I particularly thank Junbiao Dai for his encouragement and mentorship toward me as well as his enthusiasm for science. I also thank Jiquan Gao for the many conversations about our favorite retrotransposons and for his ability approach problems with a critical and objective view. I would also like to thank lab members who have encouraged and helped me to perform this research, namely Weiwu Xie, Xiang Gao, Peter Fuerst, Yi Hou, Ericka, and particularly thank Robert Dick and Clarice Schmidt without whom this research would not have been possible. I thank my committee members for demanding more of me that I would improve and strengthen my weakest areas. Finally, I sincerely thank my dear wife, Sherrie, and my two sons, Taylor and Cole, for their patience and unconditional support and my God for the invaluable opportunity to have these good people in my life.