The mechanism of integration preference to heterochromatin of yeast retrotransposon Ty5

Weiwu Xie
Iowa State University

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

Recommended Citation
The mechanism of integration preference to heterochromatin of yeast retrotransposon Ty5

by

Weiwu Xie

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

DOCTOR OF PHILOSOPHY

Major: Genetics (Computational Molecular Biology)

Program of Study Committee:
Daniel F. Voytas, Major Professor
John Mayfield
Alan Myers
David Oliver
Jo Anne Powell-Coffman

Iowa State University
Ames, Iowa
2003
Graduate College
Iowa State University

This is to certify that the doctoral dissertation of

Weiwu Xie

Has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program
TABLE OF CONTENTS

ABSTRACT vii

CHAPTER I. GENERAL INTRODUCTION 1

Transposable elements 1
Integrase and integration 3
Ty elements and targeted integration 6
Heterochromatin and silencing 9
Ty5 targeting 15
Dissertation organization 19

References 20

Figure 1 29
Figure 2 30
Figure 3 31
Figure 4 32
Figure 5 33
Figure 6 34
Figure 7 35
Figure 8 36
Figure 9 37
CHAPTER II. TARGETING OF THE YEAST TY5 RETROTRANSPOSON TO SILENT CHROMATIN IS MEDIATED BY INTERACTIONS BETWEEN INTEGRASE AND SIR4P

Abstract

Introduction

Materials and Methods

Mutagenesis of Ty5 elements

Tethered silencing

Two-hybrid assays

Protein analyses

Results

Defining the Ty5 targeting domain

The targeting domain is located in the integrase C-terminus

Tethering the targeting domain to DNA nucleates silent chromatin

Overexpression of the targeting domain disrupts telomeric silencing

Sir4p interacts with Ty5 integrase

Discussion

The integrase C-terminus and the targeting domain

The targeting domain and silent chromatin

The targeting domain and Sir4p

Concluding remarks
CHAPTER III. KINASES REGULATE TARGETING OF THE YEAST RETROTRANSPOSON TY5 TO HETEROCHROMATIN

Introduction 80

Results 82

In vitro binding of TD to Sir4p is blocked by protein phosphatase. 82

Amino acid substitutions suggest serine phosphorylation of TD. 83

Kinases modify the targeting domain. 87

Confirming the kinase mutants 90

Discussion 92

The targeting domain is phosphorylated in vivo. 92
ABSTRACT

Ty5 is a retrotransposon in \textit{Saccharomyces cerevisiae}. In wild type yeast strains, over 90\% of Ty5 insertions occur at the telomeres or \textit{HM} loci. These regions are bound in silent chromatin, which is analogous to heterochromatin of higher eukaryotes. Our laboratory previously described a targeting mutant of Ty5 that has an amino acid substitution near the C-terminus of integrase. Targeting to silent chromatin is reduced more than 20-fold in the mutant. We further defined the Ty5 targeting domain (TD) by saturation mutagenesis. All of the targeting mutations mapped to a stretch of six amino acids (LDSSPP). We tethered TD to a crippled \textit{HMR-E} silencer by fusing it to the GAL4 DNA binding domain (GBD-TD). The GBD-TD fusion could nucleate the assembly of silent chromatin, as measured by the transcriptional status of an adjacent marker gene. This silencing was dependent upon the silent chromatin components Sir2p, Sir3p, and Sir4p. When over-expressed, GBD-TD was found to disrupt silencing at the telomeres. These results support our model that Ty5 target specificity results from a protein-protein interaction, which recruits the Ty5 pre-integration complex to silent chromatin. We further determined that TD interacts with Sir4p, a structural component of silent chromatin. This interaction between TD and Sir4p C-terminus was demonstrated by two-hybrid assays and \textit{in vitro} affinity binding. We noticed that TD is able to bind Sir4p \textit{in vitro} when it is expressed and purified from yeast cells but not from \textit{E. coli} cells. In addition, protein phosphatase treatment of TD disrupts this binding. When we replaced the serines of TD with threonine or glutamate, TD is partially functional, and
mutant Ty5 elements are able to target effectively. We inferred that TD is phosphorylated by kinases and that TD function depends on this modification. A collection of kinase knockout strains was used to identify the kinases that modify TD. Dunlp, the DNA damage checkpoint kinase, is the most likely kinase involved in Ty5 targeting and TD modification. This finding sheds light on how Ty5 targeting is regulated by the host and further defines the close relationship between Ty5 and its host.
CHAPTER I. GENERAL INTRODUCTION

Transposable elements

A gene is typically a DNA sequence, which stores a unit of genetic information and guides synthesis of RNA and then protein molecules. Genes are usually arranged linearly in huge DNA molecules called chromosomes. Most genes normally have a fixed position in a chromosome relative to other genes; only in rare events can they be inverted or translocated. However, there are certain genes or DNA sequences able to move from one position to another, and they are called transposable elements (TEs).

TEs include DNA transposons and retrotransposons. A typical DNA transposon consists of a transposase gene flanked by inverted repeats at both ends (Fig. 1). Transposase excises the element from its original position and inserts it into a new site. This event leaves behind 4-12 bp direct repeats. This well-known “cut and paste” mechanism does not typically increase the copy number of the element in the genome. On the contrary, a retrotransposon utilizes a “copy and paste” mechanism to multiply itself through transposition. Retrotransposons are first transcribed into mRNA. This RNA is translated to produce proteins, including a reverse transcriptase (RT). RT can use the retrotransposon mRNA as a template to make cDNA. The cDNA molecules are then inserted into a new position in the genome.

Retrotransposable elements are divided into two big groups, the LTR (long terminal repeat) and non-LTR elements. An intact non-LTR element encodes one or two open reading frames
(ORFs) without flanking repeats (Fig. 1). An RNA Pol II promoter, fortuitously located in the upstream of an element, drives transcription. One of the translated proteins functions as both reverse transcriptase (RT) and restriction-like endonuclease (EN). This protein binds to the poly(A) mRNA of the element and forms a RNA-protein complex (RNP). The RNP then moves to the nucleus and finds a T-rich track. EN makes a nick at the tract, and the poly(A) end of the element mRNA pairs to a short sequence of the T-track. RT then starts synthesis of the cDNA from the nick using the mRNA template. The host DNA repair system finishes the job by linking the ends and completing synthesis of the other strand of the element (Kazazian and Goodier 2002). Non-LTR retrotransposons are also called LINEs, meaning “long interspersed nuclear elements”. SINEs (short interspersed nuclear elements) resemble LINEs, but do not encode any genes and are transcribed by RNA Pol III. SINES are believed to replicate by exploiting the replication machinery of LINEs (Weiner 2002).

LTR retrotransposons are closely related to retroviruses. They share a similar genomic organization (Fig. 1) and life cycle (Fig. 2). Between the LTRs are encoded proteins like RT, integrase (IN) and Gag. Retroviral genomes additionally encode an envelope gene (env). LTRs usually contain a Pol II promoter, and the whole retroelement genome is transcribed into a single mRNA molecule. RT and IN (and sometimes Gag) are translated into a single polyprotein. A retroelement-encoded protease usually cleaves the polyprotein into individual peptides (e.g. RT and IN). Multiple molecules of Gag form a capsid in the cytoplasm. Within the capsid are packaged two copies of the mRNA along with RT, IN and a host tRNA, which is used as a primer for reverse transcription. Retrovirus particles may leave the cell after Env proteins assemble on the cell surface. However, for retrotransposons, the
particles are trapped in the cell and reverse transcription happens within the particle in the cytoplasm. After the double-strand cDNA is synthesized, the particle dissembles and forms a pre-integration complex (PIC) near the nuclear periphery. This PIC primarily consists of the cDNA and multimeric IN. The PIC passes through the nuclear membrane pores or waits for nuclear membrane degeneration during mitosis, and then the PIC inserts the cDNA to the genome.

**Integrase and integration**

The integrase of retrotransposable elements and the transposase of DNA transposons are functionally equivalent (Haren et al. 1999). During transposition, the enzymes dimerize (or possibly tetramerize in the case of integrase) and join the two ends of the element DNA to form a synaptic complex (the PIC for example). The enzymes then catalyze two steps: first they nick the element DNA ends to generate free 3’OH, then they use these two 3’OH to attack two phosphodiester bonds on different strands of the target DNA. The two phosphodiester bonds are typically separated by 4-12bp. The first step of the reaction is called end-processing, and the second step is called strand transfer or joining. The donor DNA of the retroelement is the cDNA synthesized by RT. Integrase works like an endonuclease to cut off two nucleotides at the 3’ ends of the LTRs, leaving a 2-nucleotide 5’ overhang. The 3’OH then serves as nucleophile for the next step of DNA strand transfer in which the 3’ end of each strand of the donor DNA is linked to the 5’ end of each strand of the recipient DNA. In the final step, the host DNA repair system finishes the integration reaction
by filling in the gap and ligating the 5'-end of the donor with the 3'-end of the recipient (Fig. 3).

The catalytic (core) domain of integrase is conserved structurally and functionally among all retroelements and with the transposases of most DNA transposons (Capy et al. 1996). Integrase and transposase contain a DDE amino acid signature (the second aspartate is typically separated by 35 aa from the glutamate; Fig. 4). In the known crystal structures of integrase catalytic domains, these 3 aa residues are close together within the catalytic pocket (Haren et al. 1999). Furthermore, in the recently resolved crystal structure of a Tn5 transposase-DNA complex, the DNA ends, which mimic the digested retroelement ends, are very close to the DDE residues (Rice and Baker 2001). This structural data clearly indicates a core role for these residues in catalysis. Divalent metal ions also have been shown to be captured by these acidic amino acids. Although only one ion is shown in the crystal structures, a reaction model has been proposed that utilizes two magnesium ions in the catalytic pocket. This model is based on many other enzymes that catalyze phosphoryltransfer reactions (Haren et al. 1999).

During transposition, both ends of an element should be processed simultaneously in order to insert the element into a single target site. To accomplish this, dimerization of the catalytic domain and synapsis of the donor DNA ends are necessary. In support of synapsis, deletions placed in the end of the 3' LTR affect processing of both ends of the cDNA (Murphy and Goff 1992). In the Tn5 crystal structure, the contacts between the DNA and protein and between the two subunits are extensive and interwoven. Therefore, a concerted strategy has
been proposed for both LTR elements and DNA transposons in which synapsis is a prerequisite condition (Rice and Baker 2001). Although the pairing of catalytic domains seems the only solution here, data regarding dimerization of integrase is contradictory. In vitro integration studies, in which a mutated DDE motif is complemented by N-terminal or C-terminal truncated IN (see below), did not support the dimer model (Engelman et al. 1993; van Gent et al. 1993). Thus, a tetramer or even octomer was suggested.

There is not any structural data yet to help describe how the strand-transfer step happens. Nevertheless, it is believed that the same complex nonspecifically binds to the target DNA and executes the second reaction in the same pair of active sites. Indeed, the structure of the Tn5 protein-DNA complex has revealed a highly positively charged groove between the active sites, which is big enough to fit the target DNA (Rice and Baker 2001).

Besides the core domain, retroelement integrases have two other conserved domains at the N- and C-termini. These domains are not conserved in transposases, even though transposases also have two domains in the corresponding positions (Fig. 4). The N-terminal domain of IN is a HHCC zinc finger. This domain is suggested to play an important role in LTR-end recognition and in synapsis, because mutations in any of the residues in the HHCC motif of HIV nearly or completely abolish end-processing and joining reactions (Engelman and Craigie 1992; Engelman et al. 1995). The C-terminal domain is about 40 amino acids after the catalytic domain. This domain is the least conserved among retroelement integrases. In HIV, the C-terminus contains a SH3 fold motif and is suggested to function in nonspecific binding to DNA (Brown 1997). Deletion mutations also showed that the C-terminus is
needed for multimerization. In contrast to the short C-terminal domain of the viral INs, the retrotransposons frequently have much longer IN C-termini (e.g. Ty1 and Ty5 in Fig. 4). The functional study of the retrotransposon Ty5 IN C-terminus is part of my thesis work.

**Ty elements and targeted integration**

Ty1-Ty5 are 5 families of retrotransposons in the yeast *Saccharomyces*. These elements belong to two lineages of retroelements. Ty3 is in the Ty3/gypsy group, in which reverse transcriptase is encoded upstream of the integrase. Others are in the Ty1/copia group, in which RT and IN are encoded in the opposite way (Fig. 1). Because *S. cerevisiae* is an ideal model organism for molecular biology, Ty1, Ty3, and Ty5, as representatives of diverse retrotransposons, are under extensive study. They are especially of interest because of their integration target specificity.

The idea for retroelement integration preference originally came from some observations with retroviruses. All retroviruses show preferences for some target sites over others. Certain individual sites have been observed with >1000 fold more insertions than other sites within a few Kb stretch of DNA (Shih et al. 1988; Kitamura et al. 1992; Withers-Ward et al. 1994). However, there is no significant evidence indicating any consensus DNA for the retroviral targets. In vitro integration of MLV and HIV showed preference of nucleosomes to naked DNA (Pryciak et al. 1992a; Pryciak et al. 1992b; Pryciak and Varmus 1992). The pattern of integration into nucleosomal DNA sites indicated that the sites of the major groove
facing outward were favored (Pruss et al. 1994). The data suggest the accessibility of the target DNA determines local target specificity. Supporting this hypothesis, proteins bound to DNA (obviously in a different way than histones in a nucleosome) block integration (Pryciak and Varmus 1992; Bushman 1994).

Not until the availability of the human genome sequence, has a study of the genome-wide distribution of HIV integration sites been possible. This study indicates that more than 60% of the insertions are associated with RNA Pol II-transcribed genes, and there was a clear preference for genes that are actively transcribed. Additionally, regional hotspots for integration were found that are enriched in active genes (Schroder et al. 2002). Although there is not a simple explanation for the data, it is suggested that the structure and organization of the human genome, which affects the accessibility of the virus PIC, should play an important role.

The Ty retrotransposons have a much stronger preference for choosing specific chromosomal regions for integration. Ty1, Ty3, and Ty5 are targeted to their preferred sites over 90% of the time (Kim et al. 1998). Ty1 and Ty5 choose target sites within well-defined windows. Ty1 targets are within 750 bp upstream of RNA Pol III genes, including tRNAs, U6 and 5S rRNA (Devine and Boeke 1996). Ty5 preferentially inserts into silent chromatin, which can span several Kbs at the telomeres and silent mating loci (Zou et al. 1996). On the contrary, Ty3 integration sites are often restricted to within a few bases of the transcription start sites of RNA Pol III genes (Chalker and Sandmeyer 1992).
Studies with Ty3 argue against the idea that the element chooses specific DNA sequences for integration. RNA Pol III promoter mutations, which affect the ability of transcription factors to bind at positions distant from target sites, also block Ty3 integration. Therefore, a mechanism based on interaction between Ty3 and the Pol III transcriptional machinery is reasonable to explain Ty3 targeting. Pol III transcription factors TFIIIB and TFIIIC are involved in the targeting (Kirchner et al. 1995). More recent studies show that the Brf and TBP subunits of TFIIIB direct Ty3 integration in vitro (Yieh et al. 2000). TFIIIC is not essential for the targeting, but it directs the orientation of TFIIIB for transcription initiation and hence Ty3 insertion (Yieh et al. 2002). Because of the lack of an effective assay to identify mutations in Ty3 that affect targeting, it is not yet known which Ty3-encoded factors are responsible for interactions with the transcription factors.

It has been suggested that the Pol III transcription machinery is also involved in Ty1 targeting, since promoter mutations have similar negative effects on Ty1 integration (Devine and Boeke 1996). However, because Ty1 integration occurs within a “window” instead of at specific sites, this implies that the transcription factors themselves do not directly guide integration. In support of this, in vitro experiments with purified transcription components failed to restore targeting (Devine and Boeke, unpublished data). Some other feature of chromatin within the window, which could be certain chromatin binding proteins or histone modifications, is most likely involved. Recently, many reports support that chromatin domains are extensively marked by different combinations of histone modifications (called histone codes), so this should not be ignored when considering Ty1 targeting.
Heterochromatin and silencing

An overview of silent chromatin is provided before talking about how Ty5 selects these chromatin domains during integration.

Silent chromatin is defined as chromatin domains that are not conducive to transcription. Silent chromatin is referred to as heterochromatin in higher organisms. In *Saccharomyces cerevisiae*, silent chromatin includes the telomeric regions, the silent mating loci $HMR$ and $HML$, and the rDNA repeats (Laurenson and Rine 1992; Huang 2002).

**The silent mating loci.** The silent mating loci are among the best-studied domains of silent chromatin. There are three mating loci located on chromosome III. The non-silent locus, $MAT$, expresses genes that determine the mating type of the cell (either $a$ or $\alpha$). The $HMR$ and $HML$ loci also encode genes that determine $a$ ($HMR$) and $\alpha$ ($HML$) mating type, but they are not transcribed. Mating type can switch when the $HM$ locus that has the opposite genes to $MAT$ acts as a donor for the gene conversion of $MAT$ by homologous recombination. Silencing is regulated at the $HM$ loci by four Sir (silent information regulator) proteins. Sir2p, Sir3p, and Sir4p are essential components of silent chromatin. Strains with mutations in these genes are viable but unable to mate. The other Sir gene, $SIR1$ partially affects silencing and mating (Haber 1998).

*Cis*-acting sequences are also needed for silencing at the $HM$ loci. The sequences involved flank $HML$ and $HMR$, and they are called silencers. The silencer on the left of a given $HM$
locus is designated as E, and the silencer on the right is designated I (Abraham et al. 1984)(Fig. 5). Analyses of expression of plasmid-borne \textit{HM} loci have shown that the E sites are essential for silencing, and deletion of I sites weakens silencing (Abraham et al. 1984; Feldman et al. 1984). However, in the genome, only deletion of \textit{HMR}-E is able to disrupt silencing, whereas deletion of \textit{HMR}-I does not have an effect (Brand et al. 1985). Combinations of both \textit{HML}-E and \textit{HML}-I deletions are needed to break silencing of the genomic \textit{HML} locus (Mahoney and Broach 1989).

Further dissection of the \textit{HMR}-E silencer indicated that it contains three protein-binding sequences, called A, E and B (Brand et al. 1987)(Fig. 5). Deletion of one binding site has little effect on silencing, but combining any two deletions causes transcriptional derepression. E and B bind the proteins Rap1p and Abf1p, respectively. Sequence A is an ARS (autonomously replicating sequence), that binds ORC (origin replication complex). ORC recruits Sir1p, and Sir1p recruits Sir4p. Sir4p is also recruited by Rap1p (Laurenson and Rine 1992; Huang 2002).

**Telomeric silencing.** With a few exceptions, silencing at the telomeres is very similar to silencing at the \textit{HM} loci. Sir1p does not participate in silencing at telomeres, but the other Sir proteins play similar roles as at \textit{HM} loci (Aparicio et al. 1991). Within the telomeric repeat (C\textsubscript{1,3}A\textsubscript{n}), multiple Rap1p binding sites are present (Buchman et al. 1988; Longtine et al. 1989). Also, in the sub-telomeric \textit{Y'} and \textit{X} repeat sequences, ORC and Abf1p binding sequences are found (Chan and Tye 1983)(Fig. 5). Sir4p has been shown to bind to Rap1p,
which is bound to the telomeric repeats, and this binding is independent of other Sir proteins (Luo et al. 2002).

**A model for silencing.** A simple model for silencing is the following (Fig. 6): the DNA binding proteins Rap1p, ORC, and Abf1p recognize the silencers, and then they recruit Sir1p (only at the *HM* loci) and Sir4p to initiate silencing. After that, Sir2p, Sir3p, and Sir4p bind to the nucleosomes and to each other. The binding starts with Sir4p at the initiation site, and binding extends outward to form the heterochromatin domain. Two-hybrid experiments have shown that Sir4p interacts directly with Rap1p, Sir2p, Sir3p, and itself; Sir3p and Sir4p interact with deacetylated histone H3 and H4 N-terminal tails (Luo et al. 2002). Sir2p is a histone deacetylase (Imai et al. 2000). This function suggests that Sir2p's role in maintaining and establishing silent chromatin is to deacetylate H3 and H4 and make possible the interaction of Sir3p and Sir4p with the nucleosomes. This recruitment model contains a positive feedback loop: binding of Sir3p and Sir4p to histones recruits Sir2p, and Sir2p deacetylates the next histones, facilitating Sir3p and Sir4p binding.

The recruitment model assumes that the reiterated binding of Sir proteins and histones makes a huge complex and condenses the chromatin. In this way, heterochromatin blocks the access of the transcription machinery and represses gene expression. This assumption is supported by the observation that heterochromatin is inaccessible to digestion by micrococcal nuclease (Nasmyth 1982). However, more recent results demonstrated that the pre-initiation complex for transcription (TBP plus Pol II) occupies the promoters of genes located in
heterochromatin (Sekinger and Gross 2001). Hence, the assumption that genes are not accessible to transcription factors is obviously oversimplified.

The recruitment model is generally satisfactory to explain the maintenance of transcriptional repression at silent loci. Experiments have shown the silencer sequences are dispensable when the proteins are recruited by other means. For example, a marker gene at \textit{HMR} is derepressed when any two or all three protein binding sites in \textit{HMR-E} are deleted. Repression is restored when Sir1p is tethered to the silencer (Chien et al. 1993). This implies that establishment of silencing is as simple as the model suggests. However, there are some observations pertaining to establishment that cannot be explained by the model. A \textit{SIR3} temperature-sensitive mutation abolishes silencing at non-permissive temperatures, whereas silencing is reestablished at permissive temperatures. However, the restoration can only happen after passing through S-phase of the cell cycle (Miller and Nasmyth 1984). DNA replication was suggested to be involved in establishing silencing, since S-phase when DNA synthesis takes place. Seventeen years later, this possibility was excluded by using extrachromosomal rings carrying the \textit{HMR} locus that cannot undergo DNA replication (Kirchmaier and Rine 2001; Li et al. 2001). The rings were able to resume silencing after S phase. The requirement for passage through S phase remains puzzling.

The other issue unexplained by the recruitment model, is the inheritance of silencing. There are two states for expression of a gene located in silent chromatin: on or off. In the overwhelming majority of the cases at the \textit{HM} loci, expression is off. Either transcriptional state is inherited through cell division. Switching between the two states happens at a very
low frequency. In *sir1* mutant strains, the phenotype of partial loss of silencing is actually a mixture of cells in both transcriptional states. With respect to *HML*, 80% of cells are derepressed and 20% are repressed in *sir1* strains (Pillus and Rine 1989). Both expression states are still inherited, although the switch rate is higher than in wild type strains. Obviously, epigenetic information has to be duplicated to the progeny cells through mitosis, and the mechanism by which this occurs is not yet part of the recruitment model.

**Other factors that influence silencing.** Many proteins other than the ones mentioned above are involved in silencing, and the list is still growing. Rif1p and Rif2p compete with Sir3p for Rap1p interactions and hence negatively regulate telomeric silencing (Hardy et al. 1992; Wotton and Shore 1997). Sum1p and the Sir2p homolog Hst1p represent an alternative way of silencing at *HM* loci, because they can suppress *sir* mutants (Sutton et al. 2001). Sir proteins have also been demonstrated to participate in DNA repair by non-homologous end-joining (NHEJ) (Tsukamoto et al. 1997; Boulton and Jackson 1998). When this happens, the Sir proteins mobilize from the telomeric reservoir to break sites, and this weakens telomeric silencing (McAinsh et al. 1999). Ku proteins, which are needed for NHEJ, are also associated with telomeres and positively regulate silencing. The redistribution of Ku and Sir proteins requires passing from G1 to S phase of the cell cycle and is dependent on the DNA damage checkpoint pathway (Martin et al. 1999; Mills et al. 1999). As a consequence, many proteins in the DNA damage pathway, such as Mec1p, Mrc1p, are indirectly involved in telomeric silencing.
As stated above, the hypoacetylation of histones in silent chromatin is required for gene repression, and hence proteins for histone acetylation (HATs) and deacetylation (HDACs) are entangled in silencing. Deletion of HDACs, namely *HDA1* and *RPD3*, and other genes in the Rpd3p HDAC complex (e.g. *SIN3*, *SDS3* and *PHO23*) increases repression at all silent regions (Rundlett et al. 1996; Vannier et al. 1996). Two HATs, Sas2p and Sas3p, play roles in repression (Reifsnnyder et al. 1996). Sas2p pairs with Sir2p to control the expression of telomeric regions (Kimura et al. 2002; Suka et al. 2002). Other histone modification factors such as Rad6p, Dot1p, Set1p, and Set2p affect silencing directly or indirectly by changing the histones locally or globally (Huang 2002). Interestingly, the chromatin assembly factor CAF-I and the DNA replication factor PCNA are also involved (Monson et al. 1997; Enomoto and Berman 1998; Shibahara and Stillman 1999; Zhang et al. 2000). This observation provides a possible picture of how epigenetic memory is transmitted after DNA and chromatin replication.

**Silencing at the rDNA.** Transcriptional silencing at the rDNA repeats only needs Sir2p and not the other Sir proteins. At the rDNA, there are two other factors, Net1p and Cdc14p, that form a complex with Sir2p (Shou et al. 1999; Straight et al. 1999). Net1p recruits Sir2p to the rDNA repeats. In a strain with a *sir4-42* mutation, Sir3p and Sir4p are located at the rDNA (Kennedy et al. 1997). Although the mechanism of rDNA silencing is yet to be understood, the basic framework will likely be similar to silencing at other sites.
**Ty5 targeting**

The intact Ty5 retrotransposon has a genome of 5.4 kb including two 251 bp LTRs. Ty5 has a single long ORF, which encodes Gag, IN and RT (Voytas and Boeke 2002). The polyprotein is processed by a Ty5-encoded protease (PR) activity. IN and RT are estimated to be 80 kD and 59 kD respectively, while Gag has two forms of 37 kD and 27 kD (Irwin and Voytas 2001).

In the S288c strain of *S. cerevisiae*, there is one non-functional copy of Ty5 located at the telomere of chromosome III as well as a few solo LTRs. Five of the solo LTRs are associated with telomeres and the other two with *HMR*. An active copy of Ty5 was retrieved from the closely related species *S. paradoxus*. Numerous Ty5 insertions are found in the *S. paradoxus* genome, and most are also located in silent chromatin (Zou et al. 1995).

The functional Ty5 was modified so that expression is driven by a galactose inducible promoter. A *HIS3* gene was inserted in reverse orientation after the stop codon of RT. An artificial intron (AI) was inserted in this *HIS3* gene in the correct orientation to be spliced from the Ty5 transcript (Fig. 7). This marker gene is advantageous for indicating reverse transcription and integration because it is not active until the AI is removed after Ty5 is transcribed, spliced, and reverse transcribed to cDNA. To gain stable expression, the cDNA must also be integrated. Fortunately, the marker’s expression is not completely blocked by silent chromatin (Gottschling et al. 1990), and *HIS3* only needs very low expression to confer a His+ phenotype. When Ty5 inserts into silent chromatin, the marker is still readily
selectable. Transposition assays show that the Ty5 transposes at a very low efficiency (10^{-6} to 10^{-4} per element per cell). The new transposition events demonstrate a strong preference to integrate in silent chromatin: all but one of 19 events on chromosome III were within silent chromatin (Zou et al. 1996) (Fig. 8).

About 10% of Ty5 integration events occur near HMR (Zou and Voytas 1997). As mentioned above, HMR-E contains three protein-binding sites (A, E and B) that are essential for silencing. Combinations of deletions in these sites were tested for Ty5 targeting. Correlated with silencing, the targeting to hmr loci with mutant silencers was reduced from a few to more than fifteen fold (Zou and Voytas 1997). The data suggested that the chromatin state is the key for targeting. To clarify this hypothesis further, strains with deletions in the core components of silencing, namely SIR2, SIR3, and SIR4 were tested (Zhu et al. 1999). A number of Ty5 insertions in the strains were analyzed. Ty5 lost target specificity in the sir3 and sir4 strains as expected. However, the sir2 mutant still had half of the Ty5 insertions targeted, although sir2 mutations disrupt silencing as effectively as sir3 and sir4 mutations. This observation implies that targeting does not depend on the silencing itself, but rather certain proteins that participate in silencing (e.g. Sir3p or Sir4p). This is supported by a further observation: the rDNA region, which has Sir2p as a constituent, is not a favored target for Ty5. Sir3p and Sir4p are not normally located in the rDNA, but they are in a sir4-42 strain, which expresses a C-terminal truncation of Sir4p. In sir4-42 strains, Ty5 integration to the rDNA (10% of the genome) increased from 3% to 26% (Zhu et al. 1999).
Interestingly, not only does Ty5 integration favor silent chromatin, but so does recombination of Ty5 cDNA to its genomic homologues (Ke and Voytas 1999). This was shown by first preventing integration by mutating the signature residues (DDE) of the integrase catalytic domain. Two strains with Ty5 inserted at silent chromatin and euchromatin, respectively, were used for the analysis. Ty5 cDNA was found to recombine with the substrate in silent chromatin at least three times more frequently than in euchromatin. The preference depended on the Ty5 targeting function: the mutation (M3, see below), that abolishes Ty5 targeting also abolishes the recombination preference. Therefore, we hypothesize that the Ty5 PIC containing IN and cDNA is recruited to silent chromatin, possibly by Sir3p or Sir4p. The consequence of recruitment is the bias in integration and cDNA recombination.

To investigate the Ty5-encoded factors required for targeting, a targeting assay was designed by Xiaowu Gai, a former student in our laboratory (Gai and Voytas 1998)(Fig. 9A). A special yeast strain is used in the assay that has the \textit{RAD52} gene deleted to prevent homologous recombination. Ty5 on a plasmid is transformed into the strain and cells with integration events are selected by the \textit{HIS3} marker (Fig. 7). Another plasmid -- the target plasmid -- is also present in the cell to indicate the targeting bias. This plasmid has the \textit{HMR} locus and an \textit{ADE2} marker. Silencing at \textit{HMR} makes this plasmid a favorite Ty5 target. The \textit{ADE2} marker is used to show the presence of the plasmid by the difference of the colony color. The strain is \textit{ade2}, and colonies are red because they accumulate an adenine precursor; the colonies with the plasmid are white. The plasmid is quickly lost without selection pressure. This is due to the antagonism between the two DNA replication origins -- the \textit{CEN} and the \textit{ARS} in \textit{HMR-E}. Therefore, when an integration event occurs on the target
plasmid, the selection for histidine will select the plasmid with the ADE2 marker. Consequently, the colonies will be white. On the contrary, a genomic integration event will lead to a red or sometimes sectored colony (in cases where the plasmid is not lost in some cells). For a wildtype Ty5, the ratio of the white to red colonies is consistently 8-10% (Fig. 9B).

To identify Ty5 mutants with altered targeting specificity, the Ty5 genome was randomly mutagenized and screened by the targeting assay. One targeting mutant, M3, was identified out of 3000 Ty5 clones. M3 decreased the frequency of white colonies to 0.5% (~20 fold) (Fig. 9B). The phenotype was confirmed by sequencing ten Ty5 insertions in the genome, none of which was associated with silent chromatin. The M3 mutation is located at the border of the IN C-terminus and the RT N-terminus, a region not conserved in other retrotransposons. The results suggest that Ty5 encodes the targeting function, possibly at the IN C-terminus.

At this point I started my thesis work and set out to address the following specific aims: first, to define the Ty5 determinants in charge of targeting; then to characterize the host factors that interact with these determinants; finally, to clarify the mechanism of Ty5 target specificity.
Dissertation organization

Chapter II of my dissertation is a paper published in *Molecular and Cellular Biology*. In this paper, the targeting domain (TD) of Ty5 integrase was defined and shown to be a functional unit. I mutagenized the Ty5 integrase-reverse transcriptase boundary region and screened for targeting mutants with the help of Xiaowu Gai (co-author). The mutations comprise of a short motif (6 aa). Xiaowu performed the tethered silencing assay to show TD interacts with silent chromatin. Yunxia Zhu and David Zappulla (from Rolf Sternglanz's lab, SUNY Stony Brook) showed TD interacts with the Sir4p C-terminus. We obtained from the Sternglanz lab the strains for tethered silencing and two-hybrid tests. I purified the GBD-TD fusion from yeast and showed it bound to Sir4p *in vitro*. In addition, I determined that TD is part of integrase and not RT. Finally, I carried out all additional experiments necessary for the publication of this manuscript.

In Chapter III, I demonstrated that TD is phosphorylated. *DUN1* and other candidate genes were identified that affect Ty5 targeting. In this work, Junbiao Dai characterized the kinases involved in TD phosphorylation by using the three assays (tethered silencing, two-hybrid and disrupting silencing); these assays were developed in my first paper (Chapter II). Troy Brady helped me test some TD mutants by two-hybrid assays.

In the last chapter, I provide general conclusions and point out some information that is not included in the papers and is worth mentioning regarding future work.
References


Devine, S.E. and J.D. Boeke. 1996. Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. Genes Dev 10: 620-33.


Reifsnyder, C., J. Lowell, A. Clarke, and L. Pillus. 1996. Yeast SAS silencing genes and human genes associated with AML and HIV-1 Tat interactions are homologous with


Figure 1. The genomes of transposons and retrotransposable elements. The boxes with arrowheads inside are to show the repeated sequences at the ends of the elements. A DNA transposon has inverted repeats. Retrotransposons and retroviruses have long terminal repeats (LTRs). Patterns of the boxes show the similarity among the genes. The endonuclease (EN) activity encoded by the transposase, or integrase (IN) is essential for all the elements. The reverse transcriptase (RT) gene is conserved for the retroelements. The protease (PR) activity is conserved for the LTR retroelements.
Figure 2. Life cycle of retrotransposons and retroviruses. A cell (the square) and its nucleus (the big circle) are shown. A circle with an octagon represents a virus-like particle. Steps specific to the retroviruses are in italics.
**Figure 3.** The integration reaction of a LTR element. The nucleophilic attacks by H$_2$O or by the exposed 3' hydroxy of the LTR ends are indicated. The phosphodiester bonds in the acceptor DNA strands are shown by the small letter P. In the third step, the mismatched nucleotides are removed and the gaps are filled by the host DNA repair system. This step results in the target site duplication (TSD). The conserved LTR end sequences are shown.
Figure 4. Conserved domains of the integrases of the retrovirus HIV and the retrotransposons Ty1 and Ty5. The conserved HHCC residues for the zinc finger domain and DDE signature for the catalytic domain (shaded gray) are shown. NLS indicates the location of the Ty1 nuclear localization signal. M3 is the targeting mutation of Ty5.
Figure 5. Structure of the telomeric and HM silencers. The protein binding sequences are shown for the silencers (open boxes). The silencing state is represented by the shadow of dark orange. Rap1p, repressor activator protein 1; Abf1p, ARS-binding factor 1; ORC, origin recognition complex; ARS, autonomously replicating sequence. [Redrawn from Lustig 1998, Curr Opin Genet Dev 8, 233-239.]
**Figure 6.** The model of yeast telomeric silencing. The silencing is initiated at a telomeric repeat, which recruits Rap1p then Sir4p. The nucleosomes in silent chromatin are condensed and hypoacetylated. Sir2p is recruited by Sir4p. It deacetylates H3 and H4 and thus facilitates Sir3p and Sir4p binding to the nucleosomes. The nucleosomes in non-silent chromatin are less condensed.
1. The *HIS3* marker gene is inactivated by an intron (AI) that cannot be spliced from the *HIS3* mRNA.

2. Ty transcription and splicing.

3. Reverse transcription and integration.

4. The *HIS3* gene is functional after reverse transcription of the spliced Ty transcript.

**Figure 7.** Transposition assay. A *GAL* promoter is used to replace part of the 5' LTR, which makes the element expression inducible on galactose media. Only after a new insertion of the element, can the cell grow on His- media. SD, splicing donor; SA, splicing acceptor.
Figure 8. Ty5 integrates preferentially into yeast heterochromatin. The red arrows indicate the locations of new integration events on chromosome III. The regions of left telomere, HML and HMR loci are enlarged to show the detailed locations of the insertions. The closed arrowheads are to show the native Ty5 LTRs.
Figure 9. The targeting assay. A. A diagram of the assay. Expression of an episomal Ty5 element is induced. The new cDNA can be integrated either to a chromosome or to the target plasmid. The fate of the integration is indicated by the color (white or pink) of a colony in a selective plate (His-). B. The results of the wildtype Ty5 and the M3 targeting mutant transposition. For the wildtype, the percentage of white colonies is about 8%, and for the mutant, it is 0.5%.
CHAPTER II. TARGETING OF THE YEAST TY5 RETROTRANSPOSON TO SILENT CHROMATIN IS MEDIATED BY INTERACTIONS BETWEEN INTEGRASE AND SIR4P

A paper published in Molecular and Cellular Biology

Weiwu Xie\textsuperscript{1}, Xiaowu Gai\textsuperscript{1}, Yunxia Zhu\textsuperscript{1}, David C. Zappulla\textsuperscript{2}, Rolf Sternglanz\textsuperscript{2}, and Daniel F. Voytas\textsuperscript{1}

Abstract

The Ty5 retrotransposons of \textit{Saccharomyces} integrate preferentially into regions of silent chromatin at the telomeres and silent mating loci (\textit{HMR} and \textit{HML}). We define a Ty5-encoded targeting domain that spans six amino acid residues near the C-terminus of integrase (LXSSXP). The targeting domain establishes silent chromatin when tethered to a weakened \textit{HMR-E} silencer, and it disrupts telomeric silencing when overexpressed. As determined by both yeast two-hybrid and \textit{in vitro} binding assays, the targeting domain interacts with the C-terminus of Sir4p – a structural component of silent chromatin. This interaction is abrogated by mutations in the targeting domain that disrupt integration to silent chromatin, suggesting that recognition of Sir4p by the targeting domain is the primary determinant in Ty5 target specificity.

\textsuperscript{\textcopyright} Reprinted with the permission of the ASM journals publisher

\textsuperscript{1} Department of Zoology & Genetics, 2208 Molecular Biology Building, Iowa State University, Ames, IA 50011-3260
\textsuperscript{2} Department of Biochemistry and Cell Biology, State University of New York, Stony Brook, NY 11794-5215
Introduction:

The long terminal repeat (LTR) retrotransposons are a large and ubiquitous class of mobile genetic elements. Like their cousins the retroviruses, they replicate by reverse transcribing an element mRNA and then integrating the cDNA product into their host’s chromosomes. LTR retrotransposons are typically abundant components of nuclear genomes, constituting a few percent of the *Saccharomyces cerevisiae* genome to over 50% of the genomes of some plants such as maize (30, 44). As the genome sequencing projects progress, it is apparent that most retrotransposons are not randomly distributed on chromosomes. In *Drosophila melanogaster* and *Arabidopsis thaliana*, for example, retrotransposons are highly enriched in the pericentromeric heterochromatin (18, 43). This non-random distribution may be the result of preferential integration to these sites. It has been suggested that the low gene density of heterochromatin may offer a safe haven for transposition, which ensures persistence of retrotransposons by avoiding the harmful consequences of mutations that might occur if integration were random (4). Because repetitive sequences can form heterochromatin in some species, the accumulation of retrotransposons in certain regions of the genome may, in turn, contribute to the formation of chromatin domains (17).

How is it that retrotransposons identify certain chromosomal regions during integration? One model suggests that the integration apparatus recognizes specific chromatin states or DNA-bound protein complexes (6). This tethers the integration machinery to target sites and results in the observed target site biases. This model is best supported by studies of the *S. cerevisiae* retrotransposons (Ty1-Ty5). Over 90% of native Ty1-Ty4 insertions are
located upstream of genes transcribed by RNA polymerase III (RNAP III) (30). These regions are often gene-poor, and like heterochromatin, may provide a safe haven for transposition within the streamlined *S. cerevisiae* genome (4). For Ty1 and Ty3, the association with sites of RNAP III transcription is due to targeted integration. Targeting requires assembly of the RNAP III transcription complex, and promoter mutations in target genes that prevent transcription complex assembly render them inefficient targets (7, 8, 12). *In vitro* targeted transposition assays have been developed for Ty3 in which binding of TFIIIB and TFIIIC to tRNA gene templates is sufficient for targeting (31). The critical factors within these complexes appear to be the TATA binding protein and Brf (also called TFIIIB70) (49). These data support the model that targeting results when the retrotransposon preintegration complex recognizes specific DNA-bound proteins.

In contrast to the other *S. cerevisiae* retrotransposons, native Ty5 elements are not located at sites of RNAP III transcription. Rather, like retrotransposons in many other organisms, Ty5 insertions are predominantly found within the heterochromatin-like domains of the *S. cerevisiae* genome, such as at the telomeres and silent mating loci (HMR and HML) (56). The chromatin at these sites is referred to as silent chromatin, because it represses transcription of genes located in these regions. Silent chromatin is made up of a large number of proteins that assemble at specific DNA sequences (reviewed in (34, 35). Several proteins or protein complexes bind the E and I silencers that flank HMR and HML, including the origin recognition complex (ORC), the transcription factor Abf1 and the repressor/activator protein Rap1p. These proteins also bind to sequences near the telomeres: ORC and Abf1 bind to the subtelomeric X repeat, and Rap1p binds to the telomeric repeat sequences (TG1-3). These DNA-bound proteins recruit additional components of silent
chromatin, including the well-studied Sir proteins. Sir2p is a histone deacetylase (20, 32, 46), and Sir3p and Sir4p are considered structural components of silent chromatin. All three proteins interact with each other, and they nucleate at the silencers and spread outward along the chromosome (16, 39).

Ty5 integrates preferentially into regions of silent chromatin. Over 95% of de novo Ty5 transposition events occur within a 3 kb window on either side of the HM silencers or the subtelomeric X repeat (53, 54). Silent chromatin is required for this target choice, because mutations in HMR-E that prevent its assembly abolish targeting to this locus (55). Targeting decreases by approximately 50% in sir2Δ strains, whereas it is virtually abolished in sir3Δ or sir4Δ strains (52). An allele of SIR4 (sir4-42) causes a dramatic change in the chromosomal distribution of the Sir complex (28). In sir4-42 strains, Sir3p and Sir4p move from the telomeres and silent mating loci to the rDNA (29). This change in Sir protein distribution is related to mother cell aging, and sir4-42 strains are long-lived (28). Ty5 target specificity changes with the chromosomal distribution of the Sir complex in sir4-42 strains, and over 25% of the insertions occur within the rDNA (52). These results suggest that the Sir complex, particularly Sir3p and Sir4p, determines Ty5 target choice.

Ty5-encoded proteins are also important for target site selection. A Ty5 missense mutation decreases targeting more than 20-fold and provides the first direct evidence that retroelements encode their own targeting determinants (13). In this paper, we further define Ty5-encoded factors required for targeting and describe a short targeting domain near the integrase C-terminus. This targeting domain interacts with silent chromatin, because when tethered to a defective HMR-E silencer, reporter genes at HMR are transcriptionally silenced in a Sir-dependent fashion. Overexpression of the targeting domain disrupts telomeric
silencing, likely by titrating away critical silencing components. We show that the targeting
domain interacts with Sir4p in both yeast two-hybrid and in vitro binding assays. Sir4p,
therefore, appears to be the primary host determinant mediating Ty5 target specificity, and
the interaction between the Ty5-encoded targeting domain and Sir4p appears to determine
target choice.

Materials and Methods:

Mutagenesis of Ty5 elements

A BspEI-PflMI fragment of Ty5 (Figure 1) was mutagenized by PCR, using two
different protocols to minimize mutation biases. The first used the nucleoside triphosphate
analogue dPTP, which pairs with both A and G and thereby increases the mutation spectrum
(50). A typical 20 μl reaction included 5 ng of template DNA (pXW27, a plasmid containing
the BspEI-PflMI fragment), 2.5 units of Taq polymerase, 2 μl of 10x buffer, 0.5 μl of the
universal and reverse primers (20 mM), 1.6 μl of 25 mM MgCl2, 4 μl of dNTPs (2.5 mM
each) and 2.5 μl of dPTP (400 μM). The PCR reaction was carried out for 5 cycles as
follows: 92°C for 1 minute, 50°C for 1.5 minutes and 72°C for 5 minutes. An aliquot of the
reaction (0.5 μl) was then used for PCR amplification without dPTP (3). The second
mutagenesis method used Mn²⁺ (rather than Mg²⁺) and biased amounts of dNTPs (45). A
typical 50 μl reaction included 5 ng of template DNA, 2.5 units of Taq polymerase, 5 μl of
10x buffer, 0.5 μl of each primer (20 mM), 14 μl of 25 mM MgCl2, 0.75 μl of 10 mM
MnCl2, 4 μl of dNTPs (2.5 mM each), 4 μl of dCTP (10 mM) and 4 μl of dTTP (10 mM).
The PCR reaction was carried out for 13 cycles as follows: 94°C for 30 seconds, 50°C for 45 seconds and 72°C for 3 minutes. An aliquot of the reaction (0.5 μl) was used as a template in a standard PCR amplification (3). The mutant Ty5 library was constructed by replacing the mutagenized BspEI-PflMI fragment with the corresponding fragment in a wild type Ty5 element on pNK254 (26).

Mutations were made by PCR-based site-directed mutagenesis to define the targeting domain. pNK254 was PCR-amplified using the reverse primer and a mutagenic primer. The amplification product was digested with EcoRI and then inserted into the EcoRI site of pWW37 -- a Ty5 subclone containing a Hpal-SacI fragment. The BspEI-PflMI fragment of the recombinant plasmid was used to replace the corresponding fragment in pNK254. The mutagenic primers were as following: DVO754 for mutant pWW39 (5'-GGA-ATT-CAA-TCG-AAT-CTC-CTC-CAT-CGG-TGG-ATT-CAT-C), DVO755 for mutant pWW40 (5'-GGA-ATT-CAA-TCG-AAT-CTC-CTC-CAT-CGT-TGG-CTT-CAT-CGC-C), DVO756 for mutant pWW41(5'-GGA-ATT-CAA-TCG-AAT-CTC-CTC-CAT-CGT-TGG-ATT-CAT-CG-CGG-CTC-CAA-ATA-C), DVO757 for mutant pWW42 (5'-GGA-ATT-CAA-TCG-AAT-CTC-CTC-CAT-CGT-TGG-ATT-CAT-CTA-TTA-AC).

The Ty5 mutants rut-3, rut-15, rut-31, rut-38, rut-41, and rut-46 were constructed by replacing an EcoRI-PflMI fragment of the wild type Ty5 element on pNK254 with the
corresponding fragment from mutants ut-3, ut-15, ut-31, ut-38, ut-41 and ut-46. Because of
the multiple EcoRI sites in Ty5, this was accomplished in two steps: first, an EcoRI-EcoRI
fragment from the original mutant element was inserted into the EcoRI site of plasmid
pWW37, which contains a HpaI-SacI fragment of Ty5; secondly, the BspE1-PflMI fragment
of the resulting plasmid was then used to replace the corresponding fragment of pNK254.
For all mutant elements tested in this study, targeting was measured using our plasmid-based
targeting assay described in detail in our previous study (13).

**Tethered silencing**

GBD fusion proteins were generated using pGBD plasmids, which contain the GAL4
DNA binding domain under the control of the *ADHI* promoter (23). The plasmid expressing
GBD-INC (pXW140) contains amino acids 879-1136 of Ty5 inserted between the BamHI
and PstI sites of pGBDU; the GBD-inc construct (pXW158) is identical, except that it carries
the S1094L mutation (13). The plasmid expressing GBD-TD (pXW205) was generated by
inserting into the EcoRI-BgIII sites of pGBDU a short DNA fragment created from two
complementary oligonucleotides DVO690 (5'-AAT-TCT-TGG-ATT-CAT-CGC-CTC-CAA-
ATA-CCT-CA) and DVO691 (5'-GAT-CTG-AGG-TAT-TTG-GAG-GCG-ATG-AAT-
CCA-AG); the GBD-td construct (pXW213) is identical, except that it has the S1094L
mutation. Versions of these plasmids (pWW48, GBD; pWW49, GBD-TD; pWW50, GBD-
td) were generated in which the *TRP1* marker gene was replaced with *HIS3*. This was
accomplished by replacing the EcoRV-XbaI fragment with a *HIS3*-containing *NruI*-XbaI
fragment. *LEU2*-based expression plasmids (pWW44, GBD; pWW45, GBD-TD; pWW46,
GBD-td) were generated by swapping the GAD-encoding *Sph*I fragment from pGAD (which has the *LEU2* marker (24)) with an *Sph*I fragment encoding the various GBD fusion proteins.

To test silencing of the *HMR* reporter gene, the above expression plasmids were transformed into strains with different *HMR*-E mutations (9). These strains include YSB1 (aeB no UASg), YSB2 (aeB::3X UASg) and YSB35 (Aeb::3X UASg). *sir* derivatives of YSB2 and YSB35 include RS1072 and RS112 (*sir1::URA3*), RS1042 and RS1132 (*sir2::URA3*), RS1061 and RS1133 (*sir3::URA3*) and RS1067 (*sir4::URA3*) (1). The strain used to assess telomeric silencing was UCC3505 (kind gift of D. Gottschling) (15).

Complementation of GBD-TD-induced loss of telomeric silencing was tested by introducing *SIR* genes (kind gift of J. Rine) on 2 µM plasmids (pRS424 (10)). The *SIR2* plasmid (pSZ270) carries a *Notl*-*Xhol* fragment from pRS315-*SIR2*, the *SIR3* plasmid (pSZ282) carries a *BamHI*-*Sall* fragment from pJR104, and the *SIR4* plasmid (pSZ269) carries a *SacII*-*ClaI* fragment from pRS316-*SIR4*. To measure silencing, an overnight culture was grown to saturation for each strain and adjusted to OD$_{600}$ 1. Ten-fold serial dilutions were made; 10 µl of each dilution were spotted onto both the test plate and the control plate. The plates were incubated at 30°C for two days.

**Two-hybrid assays**

Two hybrid assays were performed using yeast strain L40 (19), which has *HIS3* and *LacZ* reporter genes under the control of upstream LexA operators. The LexA-*SIR4C* construct was previously described (2) and includes the C-terminal region of Sir4p (aa 950 to 1358). Control strains expressed LexA from plasmid pBTM116 (4). GAD-INC fusions were constructed by inserting a *Xmal*-PstI fragment from pXW140 into pGAD (23); the GAD-inc
construct is identical, except that it carries the S1094L mutation. The control expressed GAD from pGAD. Strains with the relevant plasmids were inoculated into 2 ml of selective media and shaken at 30°C for 24 hours. Ten-fold serial dilutions of the cultures were made and 5 μl of each dilution were plated onto SC-Trp-Leu or SC-Trp-Leu + 5 mM 3-amino-1,2,4-triazole (3AT) media; plates were incubated at 30°C for three days.

**Protein analyses**

Immunoblot analyses of Ty5 and GBD proteins were conducted as previously described using antibodies specific to the RGS-H6 tag (Qiagen) or GBD (Santa Cruz Biotechnology) (21). Two of the epitope-tagged Ty5 elements were described in this previous report (pWW32 and pIP19). The element with the targeting domain replaced by the RGS-H6 epitope (pWW59) was constructed by PCR mutagenesis (3). The Ty5 fragment within pWW37 (see above) was amplified using the universal primer, the reverse primer, DVO1180 (5'-TG A-TGG-TGA-TGC-GAT-CCT-CTC-GAT-GGA-GGA-GAT-TCG-ATT-G) and DVO1181 (5'-CGC-ATC-ACC-ATC-ACC-ATC-ACA-ATA-CT-CAT-TTA-ACG-CGG-C). The BspEl-PfMI fragment contained within the amplification product was used to replace the corresponding fragment in the wild type Ty5 element carried on pSZ152 (53).

To measure *in vitro* interactions between the targeting domain and Sir4p, we first constructed a plasmid expressing the C-terminus of *SIR4* (aa 950-1358) by PCR-amplifying pSZ269 (see above) with primers DVO1137 (5'-GAA-GGA-TCC-AGA-GGA-TCG-CAT-CAC-CAT-CAC-CAT-CAC-AGA-AGA-GTG-TCG-CAT-AGT-G) and DVO1085 (5'-TGA-TCT-CGA-GTC-AAT-ACG-GTT-TTA-TCT-CC). The amplification product was digested with *Bam*HI and *Xho*I and inserted into pCITE-2a(+) (Novagen) to generate pWW56.
pWW56 DNA (0.5 µg) was used in a 50 µl coupled transcription/translation reaction (Promega) containing 20 µCi $^{35}$S-methionine. GBD fusion proteins were immuno-affinity purified as previously described (40) from 250 ml yeast cultures (OD$_{600}$ 0.8-1.2) with either plasmid pXW205, pXW213 or pGBDU. Cells were harvested, washed with ice-cold water and resuspended in 1.8 ml lysis buffer B (50 mM HEPES pH 7.5, 0.5 M NaCl, 10% glycerol, 0.1% IGPEL CA-630 (Sigma), 2 mM DTT, 2.5 mM benzamidine, 0.5 mM PMSF, 2 µg/ml leupeptin, 2 µg/ml bestatin and 2 µg/ml pepstatin). Cells were disrupted by the glass bead method (3) and the lysate was centrifuged at 12,000 rpm at 4°C for 30 min. Levels of fusion protein in the supernatant were assessed by immunoblot analysis using anti-GBD antibodies (Santa Cruz Biotechnology). The GBD fusion proteins were immunoaffinity purified from 300 µl of supernatant using 10 µl of an anti-GBD agarose-bead slurry (Santa Cruz Biotechnology). After a 2 hr incubation, the beads were collected by centrifugation (500Xg, 2 min), washed twice with 300 µl of wash buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 5 mM Mg(OAc)$_2$, protease inhibitors) and once with 1X PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na$_2$HPO$_4$, 1.4 mM KH$_2$PO$_4$). The 50-µl in vitro transcription/translation reaction containing the labeled Sir4p was then added to the washed anti-GBD agarose beads, and the mixture was incubated at room temperature for 30 minutes. The beads were collected by centrifugation (500Xg, 2 min) and washed three times with PBS. 10 µl of 2X SDS/sample buffer (3) was added to each tube; samples were heated (95°C for 10 min) and separated by SDS PAGE. The gel was dried and exposed to X-ray film overnight.
Results:

Defining the Ty5 targeting domain

We previously found that a single amino acid substitution at position 1094 in the Ty5 polyprotein (S1094L) dramatically decreased targeting to the telomeres and silent mating loci (13). This indicated that Ty5 plays an active role in selecting targets and pointed to a possible targeting domain around S1094. To further define Ty5-encoded targeting determinants, a 758 bp BspEI-PflMI restriction fragment encompassing S1094 was mutagenized by PCR (Figure 1A). The PCR products were used to replace the corresponding wild type fragment, and sequencing of several recombinants revealed that each carried 2-8 base pair changes. Over 2100 mutagenized elements were screened for targeting defects using our plasmid-based targeting assay, which we have previously shown is an effective measure of chromosomal integration patterns (13). In this assay, targeting is quantified as the percentage of integration events that occur at a plasmid-borne HMR locus. Eleven elements were identified that were impaired in targeting to varying degrees (Table 1). DNA sequencing revealed multiple nucleotide changes in the BspEI-PflMI fragments of these elements. Whereas some mutations were silent, the number of amino acid substitutions ranged from two (mutant ut41 and ut38) to eight (mutant ut46) (Table 1).

We made three observations regarding the targeting mutants: 1) one mutant (ut5) had a S1094L substitution identical to the mutation described in our previous study; 2) all of the remaining mutants had an amino acid substitution in the vicinity of S1094 (encompassing a span of six amino acid residues); 3) independent mutations in the same residue near S1094 had similar targeting defects (e.g. mutants ut23, ut31 and ut33). These observations
suggested that the mutations near S1094 were primarily responsible for the loss of target specificity. To test this hypothesis, seven new mutants were constructed that carried only one amino acid substitution near S1094 (Table 1). The S1094L mutation was not included because it was previously characterized (13). Targeting assays indicated that in all seven cases, the single mutations conferred a targeting defect nearly identical to the original mutants. This was the case even for conservative substitutions; for example, the L1092V mutation dramatically decreased targeting. Based on these data, we concluded that Ty5 encodes a targeting domain (TD) and that it may be limited to a short stretch of six amino acids (LDSSPP).

Because our mutagenesis recovered multiple substitutions in the same amino acid, this suggested that the mutagenesis of the PCR fragment was saturated. To confirm the boundaries of the targeting domain and to ensure that all critical amino acid residues near the LDSSPP motif had been identified, directed PCR mutagenesis was used to change residues near S1094 to alanine that were not identified as important for targeting. This included two residues (Asp1093 and Pro1096) that are located between other amino acids critical for targeting as well as residues upstream and downstream of the LDSSPP motif (Table 2). These substitutions had at most a modest effect on targeting (e.g. P1096, pWW41) or no effect at all (e.g. N1098, pWW42). It is interesting to note that each of the targeting mutants identified in this study had transposition frequencies from 2- to 6-fold lower than wild type (Table 1). In contrast, the mutants without altered target specificity transposed at near wild type levels (data not shown). This observation was also made in our original study wherein the S1094L mutation caused a 4-fold decrease in transposition (13). This indicates that targeting domain mutations also affect transposition efficiency.
The targeting domain is located in the integrase C-terminus

Ty5 encodes a single open reading frame that is processed by a Ty5-encoded protease into several proteins, including integrase (IN; 80 kD) and reverse transcriptase (RT; 59 kD) (21). Extrapolating molecular weights from the Ty5 amino acid sequence, we predict the protease cleavage site that separates IN from RT is within the vicinity of the TD, and therefore the TD could reside either within the C-terminus of IN or within the N-terminus of RT. To distinguish between these possibilities, the TD was replaced with an epitope tag (RGS-I_h). Like the other TD mutants, the epitope-tagged element transposed more than 5-fold lower than wild type (data not shown). Proteins were prepared from strains expressing the TD-tagged element (pWW59) as well as from control strains expressing elements with the same epitope at either the C-terminus of RT (pIP19) or within the middle of the IN coding region (pWW32) (Figure 1A). Immunoblot analysis indicated that the element with the tag located at the targeting domain expressed a protein with the same mobility on SDS-PAGE gels as IN (Figure 1B). Other tagged proteins revealed by immunoblotting represent various processing products or intermediates (21). All lanes contained equivalent amounts of total protein, yet levels of the TD-tagged IN were several-fold lower than the other tagged IN, suggesting that mutations in the TD may affect protein stability. Nonetheless, we could conclude from this experiment that the TD resides within IN, consistent with its role in target site selection.

Tethering the targeting domain to DNA nucleates silent chromatin

Our model for target specificity predicts that the TD interacts with silent chromatin to tether the integration apparatus to its target sites. If this is the case, then the converse may also be true: a targeting domain tethered to DNA may recruit silencing factors and establish
silent chromatin. To test this idea, an assay was utilized that evaluates a protein’s ability to establish transcriptional silencing (9). The test protein was first fused to the Gal4p DNA-binding domain (GBD), and the fusion protein was expressed in a yeast strain with a weakened HMR-E silencer that contains Gal4p binding sites (UASg). The effectiveness of the tethered proteins in nucleating silent chromatin was measured by the transcriptional status of a reporter gene at HMR (e.g. TRP1). Four Ty5-GBD fusion proteins were constructed (Figure 1C), one of which has 258 amino acids of the Ty5 integrase C-terminus (GBD-INC) and another which has only nine Ty5 amino acids (GBD-TD), six of which constitute the targeting domain. The remaining two fusion constructs differed only by the S1094L mutation (GBD-inc, GBD-td). Immunoblot analysis indicated that the wild type and mutant forms of the fusion proteins were expressed equivalently in yeast (Figure 1D).

The fusion constructs were introduced into yeast strains with two different HMR-E mutations: one lacked binding sites for ORC and Rap1p (aeB), and the other lacked binding sites for Rap1p and Abf1p (Aeb) (9). The ability of the fusion proteins to establish silencing was measured by spotting 10-fold serial dilutions of the various strains onto media lacking tryptophan. The GBD-INC fusion was found to repress transcription over 100-fold (Figure 2). Surprisingly, even the nine amino acid GBD-TD fusion protein silenced TRP1 more than 10-fold. In both cases, the transcriptional silencing required the presence of the UASg, and in agreement with previous work (1, 9), the fusion proteins were more effective at silencers with a wild type ORC binding site (Aeb). If the TD interacts with silent chromatin as predicted, then mutations that disrupt targeting should decrease its effectiveness in recruiting silencing factors (Figure 2C). Consistent with this hypothesis, both GBD-inc and GBD-td were unable to establish transcriptional silencing. All of the above observations were
confirmed in a strain with a *URA3* reporter gene at *HMR*, indicating that the silencing was not reporter gene dependent (data not shown).

Silent chromatin, by definition, requires the action of Sir2p, Sir3p and Sir4p. To determine whether the TD fusions establish silent chromatin or rather act in some other way to occlude the transcriptional machinery from the *TRP1* promoter, the TD fusions were introduced into various *sirA* strains (9) (Figure 3). Sir1p was not required for TD-mediated silencing, consistent with its primary role in recruiting components of silent chromatin to the *HM* loci (47). However, the structural components of silent chromatin, Sir2p, Sir3p and Sir4p, were all required, indicating that the Ty5 TD represses the reporter gene at *HMR* by establishing silent chromatin.

**Overexpression of the targeting domain disrupts telomeric silencing**

Transcriptional silencing is very sensitive to the expression level of some components of silent chromatin. Overexpression of Sir4p, for example, disrupts telomeric silencing, presumably by titrating away other components of silent chromatin or by disrupting complex formation (22, 38) (See also Figure 4B). We tested whether overexpression of GBD-TD could disrupt telomeric silencing by monitoring the expression of telomeric (and therefore normally silenced) *URA3* and *ADE2* genes (15). Overexpression of GBD-TD (Figure 1D) resulted in loss of telomeric silencing, as evidenced by the ability of cells to grow on media lacking uracil (Figure 4A). *ADE2* expression was also evident by the white colony phenotype rather than the pinkish color characteristic of *ADE2* repression and adenine precursor accumulation. As observed in the tethering experiments, expression of the fusion with the S1094L mutation did not affect telomeric silencing.
To test whether GBD-TD disrupts telomeric silencing by titrating away components of silent chromatin, the SIR genes were ectopically expressed by introducing them on high-copy 2μ plasmids (Figure 4B). Overexpression of Sir4p by this means has previously been shown to disrupt telomeric silencing (38), and we made the same observation regardless of whether or not GBD-TD was co-expressed. Overexpression of Sir2p did not restore telomeric silencing to GBD-TD-expressing strains; however, it did increase telomeric silencing in GBD or GBD-td strains. In contrast, overexpression of Sir3p overcame the GBD-TD-dependent loss of telomeric silencing. This suggests that Sir3p is titrated away (either directly or indirectly) by interacting with the TD. Alternatively, excess Sir3p could dominantly restore silencing by bypassing a factor being titrated away by TD. In addition, a growth defect was observed on non-selective media when Sir3p and GBD-TD were co-overexpressed (Figure 4B).

**Sir4p interacts with Ty5 integrase**

Several lines of evidence suggest that Sir3p and Sir4p are likely candidates for interacting with the Ty5 integrase to mediate target specificity: 1) targeting is largely abolished in sir3Δ and sir4Δ strains (52); 2) in sir4-42 strains, Ty5 integration specificity changes with the chromosomal localization of Sir3p and Sir4p; 3) as described above, loss of telomeric silencing due to overexpression of TD can be complemented by overexpression of Sir3p. To test whether Ty5 integrase interacts with Sir3p or Sir4p, two-hybrid assays were conducted. An interaction was detected between the C-terminus of Sir4p (SIR4C, aa 950-1358 expressed as a LexA fusion) and the Ty5 integrase C-terminus (expressed as a GAD fusion). This interaction strongly activated the HIS3 reporter gene and enabled growth of
yeast cells on selective media (Figure 5A). Consistent with the role of the targeting domain in silencing, this interaction required a wild type TD: the S1094L mutation greatly weakened the two-hybrid interaction. These data indicate that integrase C-terminus binds to Sir4p (either directly or indirectly) and that the TD is required for this interaction.

To confirm the two hybrid data, we tested whether the Sir4p C-terminus and the Ty5 targeting domain could interact *in vitro*. GBD, GBD-TD or GBD-td were expressed in yeast and immunoaffinity-purified using anti-GBD agarose beads. Beads with the bound GBD proteins were incubated with SIR4C that had been labeled with $^{35}$S-methionine. The beads were washed, and the proteins eluted and separated by SDS-PAGE. Sir4p bound to GBD-TD -- the fusion protein with the wild type targeting domain. Lower, background levels of binding were observed for GBD and GBD-td (Figure 5B). These *in vitro* data support the results obtained in the two-hybrid assays and collectively suggest that the biological activity of the targeting domain is mediated by interactions with Sir4p.

**Discussion:**

In a simple model to explain LTR retroelement target specificity, the interaction between the preintegration complex and DNA-bound proteins tethers the integration machinery to target sites and results in integration site biases (6). Support for this model comes from the study of the yeast Ty retrotransposons. These elements have different target preferences: Ty1 and Ty3 prefer sites of RNAP III transcription (8, 12), and Ty5 prefers silent chromatin (55). In both cases, DNA-bound protein complexes are required for targeting. HIV integrase interacts with a human homolog of the transcription factor SNF5 in two hybrid assays (24); however, there is no evidence that this interaction mediates target site
choice. In this study, we define a Ty5-encoded targeting domain (LXSSXP) and show that it interacts with Sir4p, a component of silent chromatin. To our knowledge, this interaction between a retroelement-encoded protein and a chromatin factor provides the first direct evidence for the targeting model.

**The integrase C-terminus and the targeting domain**

Retrotransposon and retroviral integrases consist of three distinct domains (25): 1) an N-terminal region with a zinc-binding motif that is required for integrase activity and likely binds cDNA; 2) a catalytic domain that executes the integration reaction; and 3) a C-terminal region, which for the retroviruses and some retrotransposons, is required for cDNA 3'‐end processing. The retrotransposon integrase C-termini are considerably larger than their retroviral counterparts; the Ty1 and Ty5 C-termini constitute more than half of integrase (e.g. HIV IN is 288 aa; Ty1 IN is 635 aa). Little is known about the function of the retrotransposon C-terminal extensions, with the exception that the very C-terminus of Ty1 integrase encodes a nuclear localization signal that is required for the preintegration complex to gain nuclear access (27, 41). For Ty1 and Ty5, the coding region of IN lies upstream of RT, and both are released from the polyprotein by proteolytic cleavage. In our earlier study of a Ty5 targeting mutant, we could not determine whether the targeting mutation was located in the C-terminus of IN or the N-terminus of RT (13). By replacing the targeting domain with an epitope tag and comparing its electrophoretic mobility to other tagged forms of RT and IN, we demonstrate here that the targeting domain resides within integrase. This indicates that integrase is responsible for target specificity and demonstrates a new function for the integrase C-terminus. For some members of the more distantly related Ty3/gypsy...
group retrotransposons (Metaviridae), the integrase C-terminus encodes a chromodomain, a motif implicated in targeting proteins to chromatin (36). The integrase C-terminus, therefore, may generally be used by retroelements for integration site selection.

Despite the large size of the integrase C-terminus, the Ty5 targeting domain identified through our mutant screen only spans 6 amino acid residues. This short domain is biologically active: when as few as nine amino acids encompassing the targeting domain are expressed as part of a fusion protein (e.g. GBD-TD), they nucleate silent chromatin or disrupt telomeric silencing. Fusion proteins expressing larger fragments of the IN C-terminus are at least 10-fold more effective in nucleating silent chromatin. This suggests that other regions of the IN C-terminus play a role in targeting. These regions may not have been identified by our mutant screen if they are required for transposition. Transposition defects were observed during characterization of several targeting mutants. For example, the transposition efficiency of mutant ut15 was more than 30-fold lower than wild type. This mutant had three missense mutations in the IN C-terminus in addition to the mutation in the targeting domain (Table 1). When the ut15 targeting domain was evaluated in isolation (see rut15, Table 1), transposition was only 5-fold lower than wild type. Additionally, we observed that most of the mutagenized elements (~60%) were unable to transpose or showed greatly reduced levels of transposition (unpublished data). Some of these likely carry stop codons or frameshift mutations that prevent synthesis of the downstream RT; however, the frequency of non-transposing elements was too high based on the extent of mutagenesis (~0.5%). This implies that regions of the C-terminus outside of the targeting domain are important for transposition.

In addition to loss of target specificity, a second phenotype shared by all of the TD mutants is an overall decrease in transposition (2- to 6-fold). For wild type Ty5, more than
94% of Ty5 integration events occur within regions of silent chromatin, whereas the remainder appear to be randomly distributed throughout the genome (54, 56). If the TD is required for integration, then TD mutations should alter transposition frequencies. However, if the TD is required only for targeted integration, then the transposition frequencies would be predicted to drop ~15-fold (94% divided by 6%). The intermediate effect of TD mutations on transposition suggests that the TD plays a role beyond tethering the integration apparatus to its preferred target sites. Possibilities include facilitating the integration reaction, facilitating reverse transcription (which is supported by recent data indicating that the C-terminus of integrase is required to produce functional Ty1 RT in vitro (48)), or facilitating nuclear localization (the Ty1 NLS is located in approximately the same region of the protein as the Ty5 TD (27, 41)). In strains expressing a modified Ty5 in which the TD was replaced by an epitope tag, we observed significantly lower levels of integrase protein (Figure 1). It is also possible that a wild type TD is required for protein stability.

The targeting domain and silent chromatin

A large number of proteins are involved in the assembly and maintenance of silent chromatin (reviewed in (34, 35). These include proteins such as Sir3p and Sir4p that carry out structural or scaffolding roles, and proteins like Sir1p and Rap1p that recruit and nucleate the structural proteins at the silencers. Tethering either class of these proteins to weakened HMR-E silencers effectively establishes silencing (9, 37). Likewise, the integrase C-terminus is very effective in nucleating silencing and causes a 100 to 1,000-fold decrease in the expression of reporter genes at HMR. The targeting domain does not simply recruit a protein complex that occludes the RNA polymerase II machinery from the reporter gene. Rather, it
establishes silent chromatin as defined by the requirement for Sir2p, Sir3p and Sir4p. TD-mediated silencing does not need Sir1p, which primarily acts at $HM$ loci and is not required for silencing at the telomeres or rDNA. At the $HM$ loci, Sir1p interacts with ORC (47), and it is interesting to note that the TD is more effective when the ORC binding site is intact. This has also been observed with other nucleators of silent chromatin (1, 9). Although a number of $S. cerevisiae$ proteins have motifs that match the TD consensus (LXSSXP), none are known components of silent chromatin.

We have previously shown that in contrast to $sir3A$ and $sir4A$ strains, targeting of Ty5 to the telomeres and $HM$ loci is only partially impaired in $sir2A$ strains and occurs at levels approximately 50% of wild type (52). Furthermore, the loss of telomeric silencing caused by overexpressing the targeting domain cannot be restored by overexpressing Sir2p. In light of recent findings that Sir2p is a histone deacetylase (20, 32, 46), this suggests that the Ty5 integration apparatus does not sense Sir2p-mediated acetylation patterns. The targeting defect in $sir2A$ strains is likely rather a secondary consequence of perturbations in silent chromatin. In contrast, the loss of telomeric silencing caused by overexpressing the TD can be restored by additional $SIR3$ expression. The Sir4p C-terminus interacts directly with Sir3p, and it may be that overexpression of the TD disrupts this interaction, which, in turn, is stabilized by additional Sir3p. In that regard, it has been observed that loss of telomeric silencing caused by overexpression of the Sir4p C-terminus can be complemented by overexpressing Sir3p (14). Finally, it is important to note that the combined expression of Sir3p and a wild type TD causes a growth defect, indicating that although the TD may not interact directly with Sir3p, Sir3p does modulate the biological activity of the TD.
The targeting domain and Sir4p

Prior to carrying out the two-hybrid assays, no evidence distinguished the roles of Sir3p and Sir4p in Ty5 target specificity. The only notable difference was that Ty5 cDNA recombination increased more than 10-fold in sir4Δ strains and was only marginally affected in sir3Δ strains (52). In strains with the sir4-42 allele, which expresses a C-terminally truncated form of Sir4p (aa 1-1237), Ty5 integrates preferentially into the rDNA (52). Here we demonstrate that the Ty5 targeting domain interacts with the Sir4p C-terminus (aa 950 – 1358), suggesting that the relevant region of interaction is located between amino acids 950 and 1237. The Sir4p C-terminus interacts with many proteins, including Sir2p and Sir3p (40), Rap1p (5, 42), Sif2p (11) and Dis1p (51). Two hybrid interactions may result if one or more of these proteins serve as a bridge between the targeting domain and Sir4p. However, we have not observed two-hybrid interactions between integrase and other components of silent chromatin (data not shown). Furthermore, the observed in vitro binding between Sir4p and the targeting domain argues that these molecules interact directly. Nonetheless, because we used fusion proteins purified from yeast cells for these experiments, we cannot exclude the possibility that other factors co-purified or modified the TD to yield productive interactions.

Concluding remarks

Previous studies demonstrated that DNA bound in silent chromatin is inaccessible to proteins like HO endonuclease, restriction enzymes or transcription factors (33). However, the ability of Ty5 to integrate into silent chromatin suggests that this DNA is accessible to Ty5 integrase. Our model that target specificity results from simply tethering the
preintegration complex to chromatin may therefore require additional refinements. For example, the integration complex may induce changes in silent chromatin during integration, and the role of the targeting domain in such processes is an important area of future research.

An increased understanding of targeting mechanisms may make it possible to manipulate retroelement target site choice. It may be possible to change the integration preference of retrotransposons by replacing targeting domains with peptide motifs that interact with specific chromosomal proteins. Such engineered retrotransposons may become useful tools for studying chromatin organization and may provide novel methods for genome manipulation. Retroviral vectors are widely used for DNA delivery in human gene therapy; however, uncontrolled, random integration into the host genome is one of their major drawbacks. It may now be possible to better control retroviral integration site choice to improve their efficacy as vectors for gene delivery.

Acknowledgements:

We thank Phil Irwin for his expert technical assistance. J. Rine and D. Gottschling graciously provided plasmids used in this study. This work was supported by a grant from the American Cancer Society (RPG9510106MBC) to D.F.V. This is Journal Paper No. J-19154 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, Project No. 3383, and supported by Hatch Act and State of Iowa funds.
References:


Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system. Mol Cell Biol **15:**3813-22.

Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. Nature **403:**795-800.


23. **James, P., J. Halladay, and E. A. Craig.** 1996. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. Genetics **144:**1425-36.

Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. Science **266:**2002-6.


48. **Wilhelm, M., M. Boutabout, and F. X. Wilhelm.** 2000. Expression of an active form of recombinant Ty1 reverse transcriptase in *Escherichia coli*: a fusion protein containing the C-terminal region of the Ty1 integrase linked to the reverse transcriptase-RNase H domain exhibits polymerase and RNase H activities. Biochem J **32**:337-42.


Figure Legends:

**Figure 1.** The Ty5 targeting domain is located at the C-terminus of integrase. (A) Ty5 is 5375 bp in length. It expresses a full-length protein of 182 kD, which is processed by protease (PR) into Gag, integrase (IN) and reverse transcriptase (RT) (21). The cleavage sites, based on the mobility of mature proteins by SDS-PAGE, are shown by dashed lines. The black bar marks the position of TD. The *BspEI* and *PflMI* sites define the region of integrase used in the mutagenesis experiment. pIP19, pWW32 and pWW59 carry Ty5 elements that were modified by an RGS-H₆ tag. The tag replaced TD (pWW59) or was inserted either into the middle of IN (pWW32) or at the end of RT (pIP19) (21). (B) The modified Ty5 elements were expressed in yeast, and an anti-RGS-H₆ antibody was used to identify IN or RT on immunoblots. The partially processed and mature protein species are indicated. (C) The Ty5 integrase fragments used throughout this paper are shown. INC and inc are 258 aa long (the small letters indicate the version with the S1094L mutation). TD and td represent the wild type or mutant targeting domain plus three flanking amino acids from Ty5. (D) Western blots demonstrating that the wild type and mutant GBD fusion proteins are expressed at comparable levels in the test strains (YSB2 and UCC3505). GBD-INC and GBD-TD have molecular weights of approximately 47kD and 19kD, respectively.

**Figure 2.** The Ty5 targeting domain nucleates silent chromatin. (A) A cartoon depicting the tethered silencing assay. Yeast strains were used with deletions in two of the protein binding sites in *HMR*-E (A, E or B) and three copies of UAS₉ – binding sites for Gal4p (9). In the example depicted, the E and B binding sites are deleted (e, b), resulting in derepression of
transcription at HMR. Expression of a fusion protein between GBD and the Ty5 targeting domain (GBD-TD) is tested for its ability to recruit components of silent chromatin and restore silencing. Silencing is measured by expression of the adjacent TRP1 marker gene. (B) Silencing was established when GBD-INc and GBD-TD were tethered to the weakened HMR locus by the triple UASq. Serial, 10-fold dilutions of cells were plated onto control (SC-Ura) or test (SC-Ura-Trp) media to measure silencing of the TRP1 reporter gene at HMR. (C) A point mutation that abolishes targeting fails to restore silencing for both fusion proteins (GBD-inc and GBD-td) in both test strains.

**Figure 3.** Silencing conferred by the targeting domain is Sir-dependent. The assay system is described in the legend to Figure 2. In strains with deletions of SIR2, SIR3 or SIR4, GBD-TD fails to establish silencing at a weakened HMR locus. Silencing, however, does not require the SIR1 gene. The Ade⁺ phenotype of the sir2Δ, sir3Δ and sir4Δ strains confers their dark color; the sir1Δ stain is Ade⁺.

**Figure 4.** Overexpression of the Ty5 targeting domain disrupts telomeric silencing and loss of silencing is complemented by overexpression of Sir3p. Two reporter genes URA3 and ADE2 are located at telomeres VIIL and VR, respectively (15). URA3 expression was measured by growth of the yeast cells on selective media. ADE2 expression is indicated by colony color; when ADE2 is repressed, the colonies are red/pink. (A) Overexpression of GBD-TD disrupts telomeric silencing (i.e. restores URA3 expression) in contrast to its mutant form (GBD-td) and GBD alone. (B) Overexpression of Sir2p and Sir3p strengthen silencing, and Sir4p breaks silencing, as previously reported (22, 38). This is demonstrated in the GBD
control strains and GBD-td strains. When the targeting domain and SIR genes are overexpressed in the same strains, Sir3p, but not Sir2p or Sir4p, restores silencing.

**Figure 5.** The C-terminus of the Ty5 integrase interacts with the C-terminus of Sir4p. (A) Yeast two-hybrid assays reveal an interaction between GAD-INC and LexA-SIR4C. Liquid cultures expressing the various LexA and GAD proteins were serially diluted 10-fold and spotted onto plates. A positive two-hybrid interaction is measured by transcriptional activation of the HIS3 reporter, which allows for growth on selective media (SC-Trp-Leu-His with 5 mM 3-amino-1,2,4-triazole, 3AT). Activation of the HIS3 marker requires both SIR4C and the wild type targeting domain. (B) SIR4C interacts with GBD-TD in vitro. SIR4C was expressed and labeled with $^{35}$S-methionine by coupled transcription/translation. GBD-TD and GBD-td were expressed in yeast and immunoaffinity-purified with anti-GBD agarose beads. The top lanes indicate the amount of labeled SIR4C bound by the various GBD proteins. The bottom lanes are from an immunoblot performed with anti-GBD antibodies, and they indicate the levels of GBD proteins in the extract used for immunoaffinity purification.
Table 1. PCR mutagenesis defines the Ty5 targeting domain

<table>
<thead>
<tr>
<th>Element</th>
<th>Transposition efficiency ((10^{-5})^a)</th>
<th>Transposition fold decrease</th>
<th>Percent targeted transposition(^b)</th>
<th>Targeting fold decrease</th>
<th>No. of base changes(^c)</th>
<th>Amino acid sequence near S1094(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>13.3 ± 2.8</td>
<td>1.0</td>
<td>7.9%</td>
<td>1.0</td>
<td>0</td>
<td>LDSSPP</td>
</tr>
<tr>
<td>Mutants identified in the screen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ut3</td>
<td>3.43 ± 0.92</td>
<td>3.9</td>
<td>1.3%</td>
<td>6.1</td>
<td>4</td>
<td>SDSSPP</td>
</tr>
<tr>
<td>ut29</td>
<td>0.97 ± 0.11</td>
<td>13.7</td>
<td>1.4%</td>
<td>5.6</td>
<td>5</td>
<td>VDSSPP</td>
</tr>
<tr>
<td>ut5</td>
<td>1.93 ± 0.42</td>
<td>6.9</td>
<td>0.4%</td>
<td>19.8</td>
<td>5</td>
<td>LDL(^L)SSPP</td>
</tr>
<tr>
<td>ut35</td>
<td>0.52 ± 0.08</td>
<td>25.6</td>
<td>0%</td>
<td>NA</td>
<td>7</td>
<td>LD(^P)SPP</td>
</tr>
<tr>
<td>ut41</td>
<td>1.49 ± 0.27</td>
<td>8.9</td>
<td>0.2%</td>
<td>39.5</td>
<td>2</td>
<td>LD(^P)SPP</td>
</tr>
<tr>
<td>ut38</td>
<td>1.14 ± 0.22</td>
<td>11.7</td>
<td>1.1%</td>
<td>7.2</td>
<td>2</td>
<td>LDS(^L)PP</td>
</tr>
<tr>
<td>ut46</td>
<td>2.46 ± 0.68</td>
<td>5.4</td>
<td>0.5%</td>
<td>15.8</td>
<td>8</td>
<td>LD(^P)PP</td>
</tr>
<tr>
<td>ut15</td>
<td>0.44 ± 0.19</td>
<td>30.2</td>
<td>0%</td>
<td>NA</td>
<td>4</td>
<td>LDSS(^L)P</td>
</tr>
<tr>
<td>ut23</td>
<td>1.42 ± 0.46</td>
<td>9.4</td>
<td>2.5%</td>
<td>3.2</td>
<td>5</td>
<td>LDSS(^L)P</td>
</tr>
<tr>
<td>ut31</td>
<td>1.55 ± 0.02</td>
<td>8.6</td>
<td>1.8%</td>
<td>4.4</td>
<td>6</td>
<td>LDSSPQ</td>
</tr>
<tr>
<td>ut33</td>
<td>1.40 ± 0.06</td>
<td>9.5</td>
<td>2.1%</td>
<td>3.8</td>
<td>4</td>
<td>LDSSPQ</td>
</tr>
<tr>
<td>Mutants with single amino acid substitutions in the vicinity of S1094</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rut3</td>
<td>2.40 ± 0.10</td>
<td>5.5</td>
<td>0.7%</td>
<td>11.3</td>
<td>1</td>
<td>SDSSPP</td>
</tr>
<tr>
<td>pWW39</td>
<td>ND</td>
<td>2.1</td>
<td>2.1%</td>
<td>3.8</td>
<td>1</td>
<td>VDSSPP</td>
</tr>
<tr>
<td>rut41</td>
<td>2.12 ± 0.16</td>
<td>6.3</td>
<td>0.7%</td>
<td>11.3</td>
<td>1</td>
<td>LD(^P)SPP</td>
</tr>
<tr>
<td>rut38</td>
<td>2.25 ± 0.66</td>
<td>5.9</td>
<td>1.1%</td>
<td>7.2</td>
<td>1</td>
<td>LDS(^L)PP</td>
</tr>
<tr>
<td>rut46</td>
<td>2.11 ± 0.98</td>
<td>6.3</td>
<td>0.6%</td>
<td>13.2</td>
<td>1</td>
<td>LD(^P)PPP</td>
</tr>
<tr>
<td>rut15</td>
<td>2.64 ± 0.23</td>
<td>5.0</td>
<td>3.0%</td>
<td>2.6</td>
<td>1</td>
<td>LDSS(^L)</td>
</tr>
<tr>
<td>rut31</td>
<td>3.09 ± 0.89</td>
<td>4.3</td>
<td>2.0%</td>
<td>4.0</td>
<td>1</td>
<td>LDSSPQ</td>
</tr>
</tbody>
</table>
Data compiled from three independent experiments.

Number of transposition events to the target plasmid divided by total number of transposition events. For mutants indicated by 0%, transposition was too low to meaningfully determine the percent targeted transposition.

Base changes were identified by DNA sequencing of the BspEI - PflMI fragment from each of the mutants.

S1094 is in bold; missense mutations are underlined.
Table 2. Site-directed mutagenesis indicates that only four of the six amino acids in the Ty5 targeting domain are required for integration specificity

<table>
<thead>
<tr>
<th>Element</th>
<th>Percent targeted transposition</th>
<th>Amino acid sequence near the targeting domain</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type</td>
<td>7.9%</td>
<td>SPPSLDSSPPNTS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pWW40</td>
<td>7.0%</td>
<td>SPPSLÅSSPPNTS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pWW41</td>
<td>5.9%</td>
<td>SPPSLDSSAPNTS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pXW198</td>
<td>8.6%</td>
<td>AAPPALDSSPPNTS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pXW199</td>
<td>8.0%</td>
<td>SAPSLDSSPPNTS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pXW201</td>
<td>8.2%</td>
<td>SAASLDSSPPNTS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pXW200</td>
<td>8.6%</td>
<td>SPASLDSSPPNTS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pXW202</td>
<td>7.9%</td>
<td>SPPSLDSSPPNAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pWW42</td>
<td>7.8%</td>
<td>SPPSLDSSPPATS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aData compiled from three independent experiments.

bTargeting domain is in bold; missense mutations are underlined.
Figure 1. Xie et al.
A component of silent chromatin

<table>
<thead>
<tr>
<th>A</th>
<th>e</th>
<th>b</th>
<th>3 X UASg</th>
<th>TRP1</th>
</tr>
</thead>
</table>

B

<table>
<thead>
<tr>
<th>Construct</th>
<th>Strain</th>
<th>SC-Ura</th>
<th>SC-Ura-Trp</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBD-INC</td>
<td>aeB no UASg</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>aeB::3xUASg</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aeb::3xUASg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBD-TD</td>
<td>aeB::3xUASg</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aeb::3xUASg</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C

<table>
<thead>
<tr>
<th>Strain</th>
<th>Construct</th>
<th>SC-Ura</th>
<th>SC-Ura-Trp</th>
</tr>
</thead>
<tbody>
<tr>
<td>aeB::3xUASg</td>
<td>GBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GBD-INC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GBD-inc</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GBD-TD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GBD-td</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. Xie et al.
<table>
<thead>
<tr>
<th>Strain</th>
<th>Construct</th>
<th>SC-His-Ura</th>
<th>SC-His-Ura-Trp</th>
</tr>
</thead>
<tbody>
<tr>
<td>aeB::3xUASg, sir2::URA3</td>
<td>GBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GBD-TD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GBD-td</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aeB::3xUASg, sir3::URA3</td>
<td>GBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GBD-TD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GBD-td</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aeB::3xUASg, sir4::URA3</td>
<td>GBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GBD-TD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GBD-td</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aeB::3xUASg, sir1::URA3</td>
<td>GBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GBD-TD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GBD-td</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Xie et al.
Figure 4. Xie et al.
Figure 5. Xie et al.
CHAPTER III. KINASES REGULATE TARGETING OF THE YEAST RETROTRANSPONON TY5 TO HETEROCHROMATIN

A manuscript to be submitted to *Genes and Development*

**Weiwu Xie, Junbiao Dai and Daniel F. Voytas**

Department of Zoology & Genetics, Iowa State University, Ames, IA 50011

The yeast retrotransposon Ty5 preferentially inserts its cDNA into heterochromatin found at the telomeres and silent mating (*HM*) loci. We previously demonstrated that targeting is due to a large extent to a protein-protein interaction between a short domain (the targeting domain, TD) of Ty5 integrase and the silencing regulator Sir4p. The TD is only six amino acids in length (LDSSPP), and when it is expressed and immuno-purified from yeast cells, TD binds *in vitro* to the C-terminus of Sir4p (Sir4C). The TD, however, cannot bind to Sir4C when it is expressed and purified from *E. coli*, suggesting that it is post-translationally modified and that this modification is essential for binding. The modification is likely phosphorylation, because treatment of immuno-purified TD with Lambda protein phosphatase abrogates Sir4C binding. To further test if the two serines are phosphorylated, they were mutated individually and in combination to threonine or glutamate. TD can establish silencing when tethered to a crippled *HM* locus, and the mutant TDs established silencing to different degrees. We also introduced the single and double glutamate
substitutions into Ty5 and checked the target specificity of these mutants. Mutation of the second serine did not affect targeting, whereas mutation in the first serine hindered targeting moderately. When the two mutations were combined, targeting was still partially maintained, but was decreased relative to the single mutations. This data implies that the serines in the targeting domain are phosphorylated and that phosphorylation can be partially mimicked by a negatively charged amino acid. A yeast deletion collection was screened to identify the kinases that act on the targeting domain. A few candidate kinases were identified in two rounds of screening using a plasmid-based targeting assay. Dunlp, the DNA damage checkpoint kinase, affects Ty5 targeting and TD's ability to bind Sir4C. We suggest that Dunlp is the kinase that indirectly or redundantly modifies TD.

[Key Words: Retrotransposon; target specificity; heterochromatin; phosphorylation; kinases]

Introduction

A retrovirus encodes reverse transcriptase (RT) and integrase (IN) for its propagation. RT copies the retroviral RNA genome into cDNA. IN inserts the cDNA into the host genome. After integration, host resources are exploited to synthesize new virus particles until the host cell is ultimately depleted and killed. Retrotransposons are closely related to retroviruses. They also use RT and IN to multiply. However, a retrotransposon has to comprise with its
host for one reason: retrotransposons do not encode envelope genes as do retroviruses, so they can not escape the host cell. Retrotransposons can only survive if the host survives.

*Saccharomyces cerevisiae* accommodates five families of retrotransposons, designated Ty1-Ty5. More than 80% of Ty1-Ty4 elements are found inserted in the upstream regions of genes transcribed by RNA polymerase III (Pol III), such as tRNA genes (Kim et al. 1998). Eight of the native Ty5 insertions in *S. cerevisiae* and 95% of newly integrated Ty5 elements are associated with silent chromatin (also called heterochromatin) (Zou et al. 1996). Silent chromatin includes the telomeres and *HM* loci. Regions surrounding Pol III-transcribed genes and silent chromatin are gene poor, and the observed target specificity suggests that yeast retrotransposons evolved a mechanism to avoid killing their host by avoiding disruption of genes.

It has been shown that the Ty1 and Ty3 preintegration complexes (PIC) require Pol III transcription factors to target cDNA insertion (Kirchner et al. 1995; Devine and Boeke 1996). Similarly, Ty5 requires a functional domain of heterochromatin to target effectively to silent regions (Zou and Voytas 1997). Heterochromatin is initiated by cis-acting sequences, the silencers, which bind the proteins Rap1p, Abf1p, and the origin recognition complex (ORC). The silent information regulators (Sirs) are then bound and extend along the chromatin by binding to the deacetylated histone H3 and H4 N-terminal tails (reviewed by Lustig 1998). Silencing extends until a silencing boundary is reached (Bi and Broach 2001). Sir2p has lately
been shown to be a histone deacetylase (Imai et al. 2000; Landry et al. 2000; Smith et al. 2000), and Sir2p, Sir3p, and Sir4p are essential structural components of heterochromatin. Ty5 expresses a special domain in the end of the integrase, which is called the targeting domain (TD). TD interacts with one of the silent chromatin components, Sir4p, to guide its integration (Xie et al. 2001).

In this paper, we show that the targeting of Ty5 is efficiently regulated by the host. Our results indicate that in order to be functional, TD needs to be phosphorylated by host kinases. Thus, the integration pattern of Ty5 could readily be changed by altering the kinases' activity. This regulation may benefit both the element and its host. In case of a dramatic change in the environment, the kinase activity may be inhibited, and Ty5 insertions will be randomized. The consequence of random insertion will be to accelerate evolution of yeast, making it better suited to the environmental change.

**Results**

*In vitro binding of TD to Sir4p is blocked by protein phosphatase.*

We have previously characterized the Ty5-encoded targeting domain (TD) and shown that it interacts with the Sir4p C-terminus and guides the PIC to heterochromatin resulting in Ty5’s insertion preference. TD is only six amino acids long (LDSSPP), and the four bolded
residues are required for function, including the two serines (Xie et al. 2001). We have shown that when nine amino acids encompassing TD are expressed as a GBD fusion, they bind in vitro to the Sir4p C-terminus (Sir4C) (Figure 1 central panel). A mutant version of TD (td) with a leucine substituted for the first serine fails to bind Sir4C. In these experiments, the fusion protein was expressed and immuno-purified (using a GBD antibody) from yeast cells. However, when TD was fused to GST, then expressed and purified from the E. coli cells, it could not bind to Sir4C (Figure 1 left panel). We could not exclude the possibility that the upstream peptide (i.e. GST) blocked TD's function, although this was not the case when TD was fused to either GAD or lexA (Xie et al. 2001). The results suggest a difference between the TD from yeast cells and from E. coli cells. We hypothesize that the TD in yeast cells has been modified, perhaps through phosphorylation of the serines. We carefully examined our Western blots of the GBD-TD. Doublets of the bands were found always associated with the wild-type TD rather than the mutant TD or with GBD by itself (data not shown). This supports the modification hypothesis. To show TD is phosphorylated, the immuno-purified TD fusion protein from yeast was further treated with the λ protein phosphatase before mixing with Sir4C. The binding decreased to the background level (Figure 1 right panel).

*Amino acid substitutions suggest serine phosphorylation of TD.*

We have shown that a change in the four essential amino acids of TD, even a moderate change from leucine (1092) to valine, causes serious difference in Ty5 targeting (about a 4-fold
decrease based on our targeting assay) (Xie et al. 2001). Even so, we reasoned that the serines in TD might be able to be substituted by threonine (T) without a functional change in TD, if indeed TD is phosphorylated in vivo by certain serine/threonine kinases. Furthermore, the phosphorylation may be mimicked when the serines are substituted by a negatively charged amino acid, for example, by glutamate (E).

The interaction of TD and Sir4C has been monitored in vivo by several assays: TD establishes silencing when tethered to a crippled HMR locus, TD overexpression breaks telomeric silencing, and TD and Sir4C interact in two-hybrid assays (Xie et al. 2001). We first tested amino acid substitutions in TD using the tethered silencing assay. In this assay, two protein binding sequences in the E silencer were deleted; thus the adjacent marker gene is not silenced (Chien et al. 1993). When a UAS₆ sequence is inserted at the deletion site, GBD can bind to the UAS₆ and be recruited to the region. When GBD-TD fusion proteins are bound to UAS₆, they recruit Sir4p, and silencing of the downstream marker is reestablished (Xie et al. 2001). We substituted each serine of TD in the GBD fusions by either threonine or glutamic acid. For controls, we also replaced the serines with cysteines (C), which mimic the structure of serines but cannot be phosphorylated, or with glutamines (Q), which are similar to glutamate but without the negative charges. Whereas the wild-type TD established silencing 5-10 times more effectively than the control (GBD only) and the original TD mutant (S1094L), the T and E substitutions enhanced silencing 2-10 fold (Figure 2). The E substitutions of the second serine (1095) worked as well as the wild type. However the C
and Q substitutions did not affect silencing by our assay (Figure 3). These results strongly suggest that TD is modified by a kinase(s) and that the modification is responsible for its biological activity. The data also indicate that both of the serines could be phosphorylated, although we can not conclude that both phosphorylations are necessary for TD function. The addition of negative charges to TD by phosphorylation helped us explain an early observation that GBD-TD could automatically activate a downstream marker gene located outside of HMR (Gai and Voytas, unpublished data). It has been reported that a highly negatively charged domain could be used as a transcriptional activator (Ma and Ptashne 1987).

We were curious about the two conserved amino acids of TD other than the serines. They may be conserved for two possible reasons (or both): they may be needed for the TD and Sir4p interaction; they may be conserved for kinase recognition. If the second reason is true, after the phosphorylation has been mimicked by an E substitution, the amino acid sequence of the other conserved residues should not matter any more. In the tethered silencing assay, we tested the S1095E mutation in combination with previously identified mutations at L1092, S1094 and P1097. No mutants were able to establish silencing (data not shown). This suggests that L and P are needed for the TD/Sir4p interaction, or they are needed for a kinase to recognize and phosphorylate the other serine (S1094).

In addition to the tethered silencing assay, we tested the new TD mutants for their ability to
break telomeric silencing. These mutants were different from the wild type and they could not break silencing like the wild type TD (data not shown). Our lab has shown that loss of telomeric silencing is due to the degradation of Sir4p that results from the interaction between TD and Sir4p. The TD/Sir4C interaction is necessary but not sufficient for Sir4p turnover (Fuerst and Voytas, unpublished). Intact GBD function (binding to its DNA sequence) is also needed for the effect. Furthermore, the C-terminus of Ty5 integrase that includes TD and which has a stronger effect on tethered silencing, also fails to break telomeric silencing (Xie and Voytas, unpublished). Therefore, the telomeric silencing test is possibly not simply a measure of the TD/Sir4p interaction.

We also used the two-hybrid assay to measure the ability of the T and E substituted TDs to interact with Sir4C. These mutants showed the weakened interaction with Sir4C compared to the wild type TD, however the interaction is stronger compared to the original mutant (S1094L) and the negative control (data not shown).

We identified the original TD mutants using our plasmid-based targeting assay (Gai and Voytas 1998). In this assay, a plasmid containing *HMR* is used to attract Ty5 insertions. When Ty5 cDNA transposes, it could insert either into the genome or into the plasmid. On suitable growth media, a genomic insertion would give rise to a red or red sectored colony, and a plasmid insertion would give a white colony. The percentage of white colonies is a quantitative measure of Ty5 targeting. The wildtype Ty5 has approximately of 8-10% white
coli
cies, and the TD mutants have 0.5 to 3% white colonies (Xie et al. 2001).

To examine the effect of the E mutations on Ty5 targeting, three Ty5 mutants – two with a single E substitution and one with the double E substitution – were constructed and tested in the targeting assay. The results are listed in Table 1. Consistent with the tethered silencing assay, the S1095E mutation has no effect on Ty5 targeting, and the S1094E mutation showed a moderate decrease (from ~8% to ~5%). The double mutation has more severe effect on targeting (~3%). This data could be explained in either of two ways: It is possible that double phosphorylation of the serines is not good for Ty5 targeting; alternatively, the E mutation may not exactly mimic phosphorylation. In the latter situation, when one serine is substituted with E, the other one still undergoes modification. We favor the second explanation since the double E mutant retains some degree of targeting (i.e. it is better than any originally characterized targeting mutants (Xie et al. 2001).

*Kinases modify the targeting domain.*

To identify the kinases that modify Ty5 TD, a yeast deletion collection was screened (Research Genetics Inc.). There are 135 possible kinases in the yeast genome (Hunter et al. 2000), of which 109 kinase knockout strains are available. The knockout strains are in a BY4742 background ($MAT\_his3\_1$ $leu2\_0$ $lys2\_0$ $ura3\_0$). We decided to use the targeting assay for the screen, so we had to reconstruct the knockout strains to make them

suitable for the assay. Strain YDV578 was derived from YPH499 (MATα ura3-52 lys2-801 ade2-101 trp1-Δ63 his3-Δ200 leu2-Δ1) by knocking out RAD52 with a LEU2 marker. We mated this strain with the kinase knockout strains and sporulated the diploids. Offspring with the genotype Kan' Leu+ trp1 ade2 (ura3 his3) were selected. Their mating types were determined by mating with test strains; this also ensured that the strains were haploid. We found that five kinase knockout strains, 121E10 (Aste20), 125F7 (Afus3), 131B9 (Apkh2), 139A8 (Aste7) and 149G7 (Astell), were unmatable. We also did not get diploids from strain 143E10 (Actkl), due to its slow-growth and cold-sensitive phenotype. There were two strains, 123G5 (Apho85) and 147G3 (Apsl), that could not be sporulated, and four strains, 133H6 (Abckl), 143C1 (Adhf2), 171C10 (Ayps15) and 171D4 (Abud32), for which haploid offspring could not be identified. Hence, 97 new strains were constructed. Then the target plasmid (pXW72 with the HMR locus and marker genes TRP1 and ADE2, Gai & Voytas, unpublished) and the Ty5 plasmid (pNK254 with a URA3 marker (Zou et al. 1996)) were transformed to the strains. After induction of Ty5 expression, the transposition events were selected on His− media, and the percentage of white colonies was determined.

To make a control strain, wildtype BY4742 was mated to YDV578 and reconstructed as described above. The targeting assay was performed with both mating type a and _ control strains, which gave the expected 7-8% white colonies (Table 2). We found that there were 28 constructed kinase knockout (CKK) strains with significant changes in the assay, that is, less than 5% white colonies. Multiple (at least two, but mostly three) transformants were tested
for each of these strains. In most cases, the results were consistent among the parallel tests. However, there were a few strains with variable results, which may be due to some colonies losing the target plasmid more slowly, and hence being counted as white colonies. Another possibility is that regulation by the kinase is leaky. Among the 28 CKK strains, there were three strains, 101E6 (Δkin3), 103A11 (Δyml059c), and 117B8 (Δhsl1), with percentages of white colonies less than 2%, and three other strains, 114F11 (Δyck1), 129D8 (Δprkl), and 145B9 (Δfab1), with percentages between 2% and 3% (Table 2). These numbers were close to the number of white colonies observed for the elements with TD mutations (less than 3%).

It is unlikely that all the kinases identified are involved in modifying the Ty5 TD. We therefore further characterized the six mutants with the extreme phenotype by testing in vitro binding between TD and Sir4C. The experiments were performed as described above, except that GBD-TD was expressed and purified from the different kinase knockout strains (BY4742 background). We expected that the GBD-TD and Sir4C binding would be significantly decreased, according to the targeting assay data, but the results showed no obvious differences. We also tested targeting with the Ty5 S1095E mutant. The rationale for this experiment was that the mutation should bypass the requirement for phosphorylation, and so targeting of this element should not be affected in the kinase mutants. However, this was not the case (Table 2). Both sets of data, therefore, suggested that these kinase mutations affect the targeting assay rather than Ty5 targeting itself. This conclusion was confirmed for strain 101E6 (Δkin3). We sequenced eight Ty5 insertions in the Δkin3 strain to
find out where they inserted in the genome. They were all targeted: five of them were located in telomeric regions and three in the HM loci (data not shown).

To find out the kinases involved in Ty5 targeting, a second screen was performed with the 28 CKK strains. As described above, we inferred that the E mutations should restore the targeting phenotype if a kinase mutant is really involved in the phosphorylation of the corresponding serine. Unfortunately, the S1094E mutation is not quite useful for such experiments, since it targets at 5%, which is too close to most of the kinase knockout phenotypes (3-5%). Only the other E substitution (S1095E) is suitable for the screen; it targets at 8%. Eight CKK strains, 110G12 (ΔkinI), 115F2 (Δsat4), 117F6 (Δtpk3), 119D2 (Δmkkl), 124A9 (Δirs1), 136F12 (Δprr2), 137D4 (Δchk1) and 138F2 (Δdun1), are restored by the Ty5 mutation with percentages of white colonies averaging 6-12% (Table 2). Two strains 124A9 (Δirs1) and 138F2 (Δdun1) had consistent results in at least three individual tests, but the others had variations with percentages less than 7% or greater than 10%. All of these kinases are possible candidates for being directly or indirectly involved in the S1095 modification.

Confirming the kinase mutants

As mentioned above, the interaction between TD and Sir4p was observed in three in vivo experiments: tethered silencing at HMR, loss of telomeric silencing and two-hybrid assays.
We tested eight kinase mutants for how they affected each of these in vivo assays. To knockout the eight kinases in the appropriate test strains, primers matching the flanking regions of the kinases were designed to amplify the fragments of the original Kan^r knockout cassettes. The PCR products were then transformed into the strains and the Kan^r marker was used to select the knockouts, which were then confirmed by another round of PCR.

One group of the kinases appears to be regulators of silencing (Figure 4A). Surprisingly, we observed a derepression in telomeric silencing in three strains, Δtpk3, Δmkk1 and Δprr2. However, the results were independent of the targeting domain. Further work showed that Mkk1p strengthens silencing at telomeres, mating loci and the rDNA, and the others work only on telomeres (Dai and Voytas, unpublished). The Ty5 targeting alternation in these mutant strains possibly relates to the weakening of silencing. Nevertheless, these kinases may be involved in TD modification, because their mutations can be restored by the E substitution, and they have a slightly changed interaction in two-hybrid assays (Figure 4B).

The second group of kinases includes Irs1p and Kin1p. Loss of these proteins increased the tethered silencing and two-hybrid interactions (Figure 4B, C). This suggests that they play a role in TD modification. However, there is not a direct explanation to connect these results with the targeting assay results.

The third group of kinase knockouts had the results expected for a reduction in TD
phosphorylation. Dunlp, the DNA damage checkpoint kinase (Hunter et al. 2000), is required for TD to break telomeric silencing and to interact with Sir4C in the two-hybrid assays (Figure 4A, B). However, it has no obvious effect on tethered silencing (Figure 4C). This conflicting result may be explained by the fact that the tethered silencing is less sensitive or possibly Dunlp confers leaky regulation. To confirm the role of Dunlp in breaking telomeric silencing, an episomal copy of the gene was introduced to the knockout strain, and the phenotype was successfully restored (Data not shown). Another checkpoint kinase Chk1p (Liu et al. 2000) and the kinase Sat4p have similar but less observable effects in the two-hybrid assays (Figure 4B).

Discussion

The targeting domain is phosphorylated in vivo.

Ty5 encodes a short domain that guides integration site choice. When separated from integrase, this targeting domain (TD) interacts with Sir4p in vivo and in vitro. Several lines of evidence suggest that TD is phosphorylated and that phosphorylation is required for interaction with Sir4C in vivo and in vitro. First, the in vitro interaction between TD and Sir4C requires that TD is expressed in yeast cells. Second, TD binding to Sir4C is disrupted by treatment with a protein phosphatase. Finally, TD retains partial function when critical serines are replaced with threonines or glutamates. The Sir4p domain that interacts with TD
has also been characterized in our lab, and is comprised of two leucines, two tryptophans and one arginine residue (Fuerst, Zhu and Voytas, unpublished). Since positive residues often interact with negative residues, this Sir4p domain may directly interact with the phosphorylated form of TD.

Because two serines are involved, there are four possible modified forms of TD: forms with none, both or either one of the serines modified. Although our results suggest phosphorylation is critical for in vitro binding, we do not know which serine is modified. The substitution mutants suggest that either serine could be phosphorylated. We favor the hypothesis that both serines are modified for the following reasons: A) the targeting assay results showed that the two serine mutants have equal effects (Xie et al. 2001). This indicates that they play equivalent roles. B) the double E substitution mutant of Ty5 was still partially targeted (3% in the targeting assay), and this was significantly different from the serine mutants (0.5-1%) (Xie et al. 2001). If only one serine is phosphorylated, we expected the double mutant to target as efficiently as one of the serine mutants.

There are four kinds of experiments to demonstrate the interaction between TD and Sir4p. In the tethered silencing and loss of silencing assays, the binding partner of TD is the full-length Sir4p; in the other experiments, it is the Sir4p C-terminus (950-1358). Loss of silencing is not directly due to the TD/Sir4p interaction. Rather, the decrease in silencing is caused by the degradation of Sir4p, which depends on 1) the presence and accumulation of GBD-TD, 2)
the DNA binding ability of GBD, and 3) Ris1p and other Sir proteins (Fuerst and Voytas, submitted). The differences in the assays should be taken into account when considering inconsistencies in the behavior of the substitution mutants among different experiments. The threonine and glutamate substitutions did not disrupt telomeric silencing, although they tethered silencing to a level comparable to the wild type TD. This suggests that the phosphorylated serines are essential for Sir4p turnover. In tethered silencing and Ty5 targeting, the S1095E mutant functioned better than the S1094E mutant, and did as well as the wild type. However, the S1095E mutation weakened the ability to bind Sir4C in two-hybrid and in \textit{in vitro} binding assays (Xie and Voytas, data not shown). This may be due to the different versions of Sir4p involved in the interaction (full-length Sir4p vs. Sir4C). It may be because the serine phosphorylation is different from the glutamate so that the wildtype TD interacts better with both forms of Sir4p. Alternatively, we assume that different forms of modified TD coexist and interact with Sir4p. Since the different length of Sir4p may have a slightly varied structure at the binding site, the assumption may raise a possibility of a dynamic interaction. That is, Sir4p may bind to a form of TD first, then TD switches to another modified form, which consequently changes the binding to Sir4p and forces Sir4p to shift its structure. The structural change may then trigger the turnover of Sir4p. This hypothesis is attractive, since it connects the observation of TD-induced Sir4p degradation with an explanation of the accessibility Ty5 needed for integration to silent chromatin. Unfortunately, we do not have any evidence for this model yet.
Kinases regulate Ty5 targeting.

We applied our targeting assay to identify the kinases that directly or indirectly modify TD. The assay quantitatively measures the proportion of Ty5 integration events that occur into an episomal HMR locus. We expected that kinase mutations that affect silencing would cause false results. However, there is no report of kinases that affect silencing, except for mutations in MEK1, which is an essential gene and is not included in the collection. Surprisingly, 28 kinases out of the 97 tested (~ one third) significantly altered the results of the targeting assay. We went through a second screen, taking advantage of the S1095E mutant, which has the phosphorylation mimicked by a negatively charged amino acid, and which has wild type target specificity. Thus, the S1095E mutation should suppress the mutant phenotype of kinases that modify TD. Eight candidate kinases were found, among which three regulate silencing, and the others have varied effects on the TD/Sir4p in vivo interaction. Among the latter class, Dun1p is the best candidate for the kinase that modifies S1095; dun1 mutations significantly decrease the ability of GBD-TD to break silencing, and they abrogate the TD/Sir4C two hybrid interaction. However, we believe that Dun1p has a redundant activity and/or indirectly modifies TD. This conclusion is based on the observation that dun1 mutants only reduce Ty5 targeting moderately (to ~5%) and in the dun1 knock out strain, TD is still able to establish silencing when tethered to a crippled HMR locus. Further experiments are needed to determine whether and how the candidate kinases modify TD.
Irs1p is another strong candidate kinase based on the targeting assay data. Interestingly, irs1 mutations increased rather than decrease the interaction between TD and Sir4p, as shown by the tethered silencing and two-hybrid assays. To explain this, we consider the hypothesis of dynamic binding mentioned above. We think that the lack of a specific kinase may limit the TD modification to one form, which in this case, favors its function in tethered silencing and two-hybrid assays, but prevents its transition to a second form involved in targeting.

Many kinases are employed to respond to stimulation from the internal and external environment. Three of the eight candidate kinases are involved in stress response -- sat4 strains are sensitive to 1M NaCl (Mulet et al. 1999), Irs1p is involved in the protein kinase C pathway (Miyahara et al. 1998), and mkk1 strains are sensitive to glycerol and nitrogen starvation (Irie et al. 1993). DUN1 and CHK1 are checkpoint genes, and Tpk3p (Garrett and Broach 1989) is a subunit of the cAMP-dependent protein kinase complex. These latter genes are crucial to respond to intracellular signals. The functions of these kinases imply that Ty5 integration is regulated by the environment. It is possible that in certain circumstances, Ty5 residing in silent chromatin will be expressed by derepression, as suggested by the mkk1, tpk2 and prr2 mutant, then the progeny elements will integrate randomly. This will result in the "explosion" (sudden and notable multiplication) of the element and cause dramatic change
to the host genome. Therefore, the evolution of the organism will be speeded up to fit a new environment.

Materials and Methods

Plasmids and Strains

See Table 3.

In vitro Binding and In vivo Assays

The assays to demonstrate the interaction between Ty TD and Sir4p are described in our previous paper (Xie et al. 2001). To treat the immuno-precipitated GBD fusion proteins with λPPase, the final wash was performed with 1 X λPPase buffer instead of 1 X PBS. The beads were incubated at 30°C with 0.5μl (200 units) enzyme for 25 minutes before mixing with the labeled Sir4C.

Targeting Assay

The targeting assay was described previously (Gai and Voytas 1998). Here we used the target plasmid pXW72 with a TRP1 marker (Gai, unpublished) rather than pXW78 with a
Constructing the Kinase Knockout Strains

To construct YDV578, the RAD52 knockout plasmid pSM20 was digested by BamHI and directly transformed to strain YPH499. The transformants were selected on SC-L plates and confirmed by PCR.

YDV578 and the kinase knockout strains (in the BY4742 background) were grown on YPD plates overnight. Both types of strains were mated on YPD plates at 30°C for 4-6 hrs. Diploids were selected on SC-Leu-Trp plates and grown in liquid selective media overnight (to O.D.-600 less than 3). 1 ml of the cells was collected and washed once, then transferred to 1 ml of sporulation medium. The cultures were incubated at 30°C for 3 overnights with vigorous shaking. The cells were diluted 5 times, and 1 ml was transferred to new tubes. The cells were spun down and washed once. 100 μl of 0.5mg/ml zymolyase (in 1M sorbitol or water) was added to each tube followed by 30°C incubation overnight. The cells were collected, washed twice, and suspended in 1 ml of water. The cells were then diluted 100 times and plated (10 μl- 100 μl) to SC-Leu+G418+5FAA medium (Toyn et al. 2000).

After 4-5 days of growth (first 3 days at 30°C, then at RT), a few colonies turned red. Three to five (or more) colonies were picked for each strain and made into small patches or struck
out on SC-Leu+G418 plates. The plates were replicated to SC-Trp-Leu (+G418) to make sure they were Trp-. The colonies were also mated to the mating type test strains (YDV515 and YDV516) to determine the mating types and confirm that they were haploids.

References


Devine, S.E. and J.D. Boeke. 1996. Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. *Genes Dev* **10**: 620-33.


**Table 1.** Targeting assay results of Ty5 mutants.

<table>
<thead>
<tr>
<th>Ty5 mutants</th>
<th>Percentage in targeting assay</th>
<th>Sequence of TD</th>
</tr>
</thead>
<tbody>
<tr>
<td>pNK254(WT)</td>
<td>7.8%</td>
<td>LDSSPP</td>
</tr>
<tr>
<td>pWW79</td>
<td>7.7%</td>
<td>LDSEPP</td>
</tr>
<tr>
<td>rut38*</td>
<td>1.1%</td>
<td>LDSLPP</td>
</tr>
<tr>
<td>rut46*</td>
<td>0.6%</td>
<td>LDSPPP</td>
</tr>
<tr>
<td>pWW102</td>
<td>5.0%</td>
<td>LDESPP</td>
</tr>
<tr>
<td>pXW137</td>
<td>0.5%</td>
<td>LDLSPP</td>
</tr>
<tr>
<td>rut41*</td>
<td>0.7%</td>
<td>LDPSPPP</td>
</tr>
<tr>
<td>pWW103</td>
<td>3.7%</td>
<td>LDEEPP</td>
</tr>
</tbody>
</table>

* Data from our previous paper (Xie et al., 2000)
Table 2. Selected targeting assay results in the strains of kinase knockout.

<table>
<thead>
<tr>
<th>Strains</th>
<th>/pNK254 with wildtype Ty5*</th>
<th>/pWW79 with S1095E Ty5*</th>
</tr>
</thead>
<tbody>
<tr>
<td>4742α#</td>
<td>71/980 57/720 46/588</td>
<td>8/90 6/94</td>
</tr>
<tr>
<td>4742a#</td>
<td>31/421 43/417 19/304</td>
<td>18/255 24/317 17/193</td>
</tr>
<tr>
<td>101E6</td>
<td>1/208 2/403 0.5%</td>
<td>3/343 0/145 2/248</td>
</tr>
<tr>
<td>103A11</td>
<td>3/252 1/295 3/364 0.8%</td>
<td>4/288 2/223 1.2%</td>
</tr>
<tr>
<td>114F11</td>
<td>20/777 17/662 34/1244 2.6%</td>
<td>13/476</td>
</tr>
<tr>
<td>117B8</td>
<td>3/154 1/116 2/80 2/141 1.6%</td>
<td>3/60 1/81 1/82 2.2%</td>
</tr>
<tr>
<td>129D8</td>
<td>6/233 2/92 3/115 2.5%</td>
<td>3/99 2/53 1/49 3.0%</td>
</tr>
<tr>
<td>145B9</td>
<td>33/1174 16/692 26/906 2.7%</td>
<td>11/361 8/218 4/191 3.0%</td>
</tr>
<tr>
<td>110G12</td>
<td>13/339 13/399 9/360 3.2%</td>
<td>3/151 12/168 18/244 8/137 6/115 5.8%</td>
</tr>
<tr>
<td>115F2</td>
<td>9/262 7/232 7/245 3.1%</td>
<td>5/115 10/120 6/70 35/479 6/61 7.3%</td>
</tr>
<tr>
<td>117F6</td>
<td>38/1059 12/564 26/686 3.3%</td>
<td>18/289 25/365 13/303 5.9%</td>
</tr>
<tr>
<td>119D2</td>
<td>33/685 30/793 10/341 4.0%</td>
<td>25/314 28/458 19/353 12/150 4/196 6.1%</td>
</tr>
<tr>
<td>124A9</td>
<td>25/924 44/1254 31/731 3.4%</td>
<td>100/1311 49/684 42/417 7.9%</td>
</tr>
<tr>
<td>136F12</td>
<td>30/772 26/838 52/1091 4.0%</td>
<td>75/745 26/209 68/466 11.9%</td>
</tr>
<tr>
<td>137D4</td>
<td>37/899 50/1091 41/1081 4.2%</td>
<td>36/531 70/784 11/226 7.6%</td>
</tr>
<tr>
<td>138F2</td>
<td>38/782 44/911 56/1101 4.9%</td>
<td>17/233 13/154 32/349 8.4%</td>
</tr>
</tbody>
</table>

# These strains are reconstructed by crossing of BY4742 and YDV578. The mating types were marked by α and a.
* The numbers indicate "the number of white colonies/total colonies" from individual plates. The percentages were calculated by "the sum of white colonies" divided by "the sum of the totals".
<table>
<thead>
<tr>
<th>Plasmids</th>
<th>vector/insertion</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pWW54</td>
<td>pGEX-4T1/TD</td>
<td>Xie, et al. 2000</td>
</tr>
<tr>
<td>pWW55</td>
<td>pGEX-4T1/td(S1094L)</td>
<td>Xie, et al. 2000</td>
</tr>
<tr>
<td>pWW56</td>
<td>pCITE-2a(+)/Sir4C</td>
<td>Xie, et al. 2000</td>
</tr>
<tr>
<td>pWW73</td>
<td>pGBDU-C3/td(S1094T)</td>
<td>This study</td>
</tr>
<tr>
<td>pWW74</td>
<td>pGBDU-C3/td(S1095T)</td>
<td>This study</td>
</tr>
<tr>
<td>pWW75</td>
<td>pGBDU-C3/td(S1094E)</td>
<td>This study</td>
</tr>
<tr>
<td>pWW76</td>
<td>pGBDU-C3/td(S1095E)</td>
<td>This study</td>
</tr>
<tr>
<td>pWW79</td>
<td>pNK254(S1095E)</td>
<td>Zuo et al. 1997; This study</td>
</tr>
<tr>
<td>pWW82</td>
<td>pGBD-C3/td(S1094T)</td>
<td>This study</td>
</tr>
<tr>
<td>pWW83</td>
<td>pGBD-C3/td(S1095T)</td>
<td>This study</td>
</tr>
<tr>
<td>pWW84</td>
<td>pGBD-C3/td(S1094E)</td>
<td>This study</td>
</tr>
<tr>
<td>pWW85</td>
<td>pGBD-C3/td(S1095E)</td>
<td>This study</td>
</tr>
<tr>
<td>pWW86</td>
<td>pGBDU-C3/td(L1094V, S1095E)</td>
<td>This study</td>
</tr>
<tr>
<td>pWW87</td>
<td>pGBDU-C3/td(S1094L, S1095E)</td>
<td>This study</td>
</tr>
<tr>
<td>pWW88</td>
<td>pGBDU-C3/td(S1095E, P1097Q)</td>
<td>This study</td>
</tr>
<tr>
<td>pWW89</td>
<td>pGBDU-C3/td(S1094E, S1095E)</td>
<td>This study</td>
</tr>
<tr>
<td>pWW102</td>
<td>pNK254(S1094E)</td>
<td>Zuo et al. 1997; This study</td>
</tr>
<tr>
<td>pWW103</td>
<td>pNK254(S1094E, S1095E)</td>
<td>Zuo et al. 1997; This study</td>
</tr>
<tr>
<td>pJB82</td>
<td>pRS414/Dun1ORF+1kb flanking</td>
<td>This study</td>
</tr>
<tr>
<td>pSM20</td>
<td>pBR322/rad52::LEU2</td>
<td>L. Prakash and S. Prakash, University of Texas, Galveston)</td>
</tr>
</tbody>
</table>
## Table 3. (continued)

<table>
<thead>
<tr>
<th>Strains</th>
<th>genotype</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>W303-1B</td>
<td>MATα ade2-1 ura3-1 his3-11,15 leu2-3,112 trp1-1 can1-1</td>
<td>R. Rothstein</td>
</tr>
<tr>
<td>YSB2</td>
<td>W303-1B aeB::3xUASα hmr::TRP1 gal4::LEU1</td>
<td>Chien et al. 1993</td>
</tr>
<tr>
<td>UCC3505</td>
<td>MATα ade2-101 his3-Δ200 leu2-Δ1 lys2-801 trp1-Δ63 ura3-52 ppr1::HIS3 adh4::URA3-TEL-VIIl DIA5-1</td>
<td>Singer and Gottschling, 1994</td>
</tr>
<tr>
<td>BY4742</td>
<td>MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0</td>
<td>From Research Genetics</td>
</tr>
<tr>
<td>YNK313</td>
<td>YPH499 rad52::TRP1</td>
<td>Gai and Voytas, 1998</td>
</tr>
<tr>
<td>YPH499</td>
<td>MATα ura3-52 lys2-801 ade2-101 trp1-Δ63 his3-Δ200 leu2-Δ1</td>
<td>From Stratagene</td>
</tr>
<tr>
<td>YDV578</td>
<td>YPH499 rad52::LEU2</td>
<td>This study</td>
</tr>
<tr>
<td>YDV515</td>
<td>MATα thr4</td>
<td>Chapter 13 Current Protocols in Molecular Biology</td>
</tr>
<tr>
<td>YDV516</td>
<td>MATα thr4</td>
<td>Chapter 13 Current Protocols in Molecular Biology</td>
</tr>
<tr>
<td>L40</td>
<td>MATα his3-Δ200 trp1-901 leu2-3,112 ade2 lys2-801amlLYS2::(lexAop)α-HIS3 URA3::(lexAop)α-lacZ-GAL4 gal4 gal80</td>
<td>gift from Rolf Sternglanz</td>
</tr>
<tr>
<td>Oligos</td>
<td>sequence (5' to 3')</td>
<td>application</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>DVO690</td>
<td>AATTCTGGATTCATCGCCTCAAATACC</td>
<td>insert TD to pGBD vectors</td>
</tr>
<tr>
<td>TCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO691</td>
<td>GATCTGAGGTATTTGGAGGCGATGAATCC</td>
<td>insert TD to pGBD vectors</td>
</tr>
<tr>
<td>AAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO752</td>
<td>AATTCTTGATTiATCGCCTCAAATACCT</td>
<td>insert td (S1094L) to pGBD vectors</td>
</tr>
<tr>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO753</td>
<td>GATCTGAGGTATTTGGAGGCGATaAATCC</td>
<td>insert td (S1094L) to pGBD vectors</td>
</tr>
<tr>
<td>AAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO1443</td>
<td>AATTCTGGATaCATCGCCTCAAATACCT</td>
<td>insert td (S1094T) to pGBD vectors</td>
</tr>
<tr>
<td>TCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO1444</td>
<td>GATCTGAGGTATTTGGAGGCGATGtATCCA</td>
<td>insert td (S1094T) to pGBD vectors</td>
</tr>
<tr>
<td>AG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO1445</td>
<td>AATTCTGGATTCaCGCCTCAAATACCT</td>
<td>insert td (S1095T) to pGBD vectors</td>
</tr>
<tr>
<td>TCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO1446</td>
<td>GATCTGAGGTATTTGGAGGCGtTGATCCA</td>
<td>insert td (S1095T) to pGBD vectors</td>
</tr>
<tr>
<td>AG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO1447</td>
<td>AATTCTGGATGaATCGCCTCAAATACCT</td>
<td>insert td (S1094E) to pGBD vectors</td>
</tr>
<tr>
<td>TCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO1448</td>
<td>GATCTGAGGTATTTGGAGGCGATcATCCA</td>
<td>insert td (S1094E) to pGBD vectors</td>
</tr>
<tr>
<td>AG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO1449</td>
<td>AATTCTGGATTCagaGCCTCAAATACCT</td>
<td>insert td (S1095E) to pGBD vectors</td>
</tr>
<tr>
<td>TCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO1450</td>
<td>GATCTGAGGTATTTGGAGGCitTGAATCCA</td>
<td>insert td (S1095E) to pGBD vectors</td>
</tr>
<tr>
<td>AG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO1514</td>
<td>AATTGgTGGATTCagaGCCTCAAATACCT</td>
<td>insert td (L1092V, S1095E) to pGBD vectors</td>
</tr>
<tr>
<td>TCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO1515</td>
<td>GATCTGAGGTATTTGGAGGCitTGAATCCA</td>
<td>insert td (L1092V, S1095E) to pGBD vectors</td>
</tr>
<tr>
<td>dG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO1516</td>
<td>AATTCTGGATTlAgaGCCTCAAATACCT</td>
<td>insert td (S1094L, S1095E) to pGBD vectors</td>
</tr>
<tr>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO1517</td>
<td>GATCTGAGGTATTTGGAGGCitTaAATCCA</td>
<td>insert td (S1094L, S1095E) to pGBD vectors</td>
</tr>
<tr>
<td>AG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO1518</td>
<td>AATTCTGGATTCagaGCCTCaAAATACCT</td>
<td>insert td (S1095E, P1097Q) to pGBD vectors</td>
</tr>
<tr>
<td>TCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO1519</td>
<td>GATCTGAGGTATTTGAGGCitTGAATCCA</td>
<td>insert td (S1095E, P1097Q) to pGBD vectors</td>
</tr>
<tr>
<td>AG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVO1520</td>
<td>AATTCTGGATGaAgaGCCTCAAATACCT</td>
<td>insert td (S1094E, S1095E) to pGBD vectors</td>
</tr>
<tr>
<td>TCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>oligo</td>
<td>sequence (5' to 3')</td>
<td>application</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>DVO1521</td>
<td>GATCTGAGGTATTTTGAGGCCATcTtcATCCAAG</td>
<td>insert td (S1094E, S1095E) to pGBD vectors</td>
</tr>
<tr>
<td>DVO1487</td>
<td>CATCGTTTGAATTCAgaGCTTCCAAATACC</td>
<td>mutate Ty5 S1095E</td>
</tr>
<tr>
<td>DVO1488</td>
<td>GGTATTTGGAGGCCatGTATACCAACGATG</td>
<td>mutate Ty5 S1095E</td>
</tr>
<tr>
<td>DVO2109</td>
<td>CATCGTTGAGTgaATCGCTCTCCAAATACC</td>
<td>mutate Ty5 S1094E</td>
</tr>
<tr>
<td>DVO2110</td>
<td>ATTTTGAGGCCATtcATCCAAACGATGGAGG</td>
<td>mutate Ty5 S1094E</td>
</tr>
<tr>
<td>DVO2111</td>
<td>TCGTTGAGTgaAgaGCTTCCAAATACCTCA</td>
<td>mutate Ty5 S1094E, S1095E</td>
</tr>
<tr>
<td>DVO2112</td>
<td>ATTTTGAGGCCATtcATCCAAACGATGGAGG</td>
<td>mutate Ty5 S1094E, S1095E</td>
</tr>
<tr>
<td>DVO392</td>
<td>ATCTCCAAGAGAGTTGGG</td>
<td>confirm rad52::LEU2 knockout</td>
</tr>
<tr>
<td>DVO393</td>
<td>GCCCTAGCAGGTGCGGCC</td>
<td>confirm rad52::LEU2 knockout</td>
</tr>
<tr>
<td>DVO2454</td>
<td>GATAAGAGCCTGCTGAAATACC</td>
<td>Knockout dun1/YDL101C::Kan</td>
</tr>
<tr>
<td>DVO2455</td>
<td>CAATGCAGAGTGTGGCCGGTG</td>
<td>knockout dun1/YDL101C::Kan</td>
</tr>
<tr>
<td>DVO2456</td>
<td>GTCCTGAGAACATCAGAGCC</td>
<td>confirm dun1/YDL101C::Kan</td>
</tr>
<tr>
<td>DVO2521</td>
<td>gggaattcGAAGCCCTGCTGAAATACC</td>
<td>clone DUN1 to pJB82</td>
</tr>
<tr>
<td>DVO2522</td>
<td>gggaatcttagagcaagataatt</td>
<td>clone DUN1 to pJB82</td>
</tr>
<tr>
<td>DVO2457</td>
<td>TTTGCCGTACCTGCTGTCC</td>
<td>knockout isr1/YPR106W::Kan</td>
</tr>
<tr>
<td>DVO2458</td>
<td>AATCCTCAATGGGTTGGCAAC</td>
<td>knockout isr1/YPR106W::Kan</td>
</tr>
<tr>
<td>DVO2459</td>
<td>ATCGAGCTCTCCATGATGC</td>
<td>confirm isr1/YPR106W::Kan</td>
</tr>
<tr>
<td>DVO2460</td>
<td>CTCGGTGCTGGCGGATTG</td>
<td>knockout prr2/YDL214C::Kan</td>
</tr>
<tr>
<td>DVO2461</td>
<td>GCCAGCGCGTCAGGCGAG</td>
<td>knockout prr2/YDL214C::Kan</td>
</tr>
<tr>
<td>DVO2462</td>
<td>ATTTGAATTGAGAACAGCCAAG</td>
<td>confirm prr2/YDL214C::Kan</td>
</tr>
<tr>
<td>DVO2463</td>
<td>GCCAGCGCGTCAGGCGAG</td>
<td>knockout chk1/YBR274W::Kan</td>
</tr>
<tr>
<td>DVO2464</td>
<td>GGTTCCACACGGGCGGCGG</td>
<td>knockout chk1/YBR274W::Kan</td>
</tr>
<tr>
<td>DVO2465</td>
<td>ACCAAATACTATGCTCTGCTAAAG</td>
<td>confirm chk1/YBR274W::Kan</td>
</tr>
<tr>
<td>DVO2466</td>
<td>ATACCCTACCTGCTGGGC</td>
<td>knockout mkk1/YOR231W::Kan</td>
</tr>
<tr>
<td>DVO2467</td>
<td>CTCTAGGGGGTTATACCTGG</td>
<td>knockout mkk1/YOR231W::Kan</td>
</tr>
<tr>
<td>DVO2468</td>
<td>GTGTAAGAGACCTGGCGG</td>
<td>confirm mkk1/YOR231W::Kan</td>
</tr>
<tr>
<td>DVO2469</td>
<td>AGCATTTCTCTGCTGTTG</td>
<td>knockout sat4/YCR008W::Kan</td>
</tr>
<tr>
<td>DVO2470</td>
<td>CGATAATAACGGCGGCGCATG</td>
<td>knockout sat4/YCR008W::Kan</td>
</tr>
<tr>
<td>DVO2471</td>
<td>TAAACGCGCAGCTCCTGCGG</td>
<td>confirm sat4/YCR008W::Kan</td>
</tr>
<tr>
<td>DVO2472</td>
<td>TCCTGTGACTGCTAGG</td>
<td>knockout tpk3/YKL166C::Kan</td>
</tr>
<tr>
<td>DVO2473</td>
<td>TATCTGTTTTTCCCTGTTGG</td>
<td>knockout tpk3/YKL166C::Kan</td>
</tr>
<tr>
<td>DVO2474</td>
<td>CCTCATTAGATGCTTCCCG</td>
<td>confirm tpk3/YKL166C::Kan</td>
</tr>
<tr>
<td>DVO2475</td>
<td>AACAAGAATATGGCTGGG</td>
<td>knockout kin1/YDR122W::Kan</td>
</tr>
<tr>
<td>DVO2476</td>
<td>TGCTCTGTTAAACGGGCTGATG</td>
<td>knockout kin1/YDR122W::Kan</td>
</tr>
<tr>
<td>DVO2477</td>
<td>CCAACGCGCGCGCCGCAA</td>
<td>confirm kin1/YDR122W::Kan</td>
</tr>
<tr>
<td>DVO2010</td>
<td>CTGCAGCGAGAGACGTAAT</td>
<td>kanB for confirming all the KanMX4 knockouts</td>
</tr>
</tbody>
</table>
Figure 1. In vitro binding of TD and Sir4C suggest that TD is phosphorylated. Sir4C is expressed and $^{35}$S labeled by a transcription/translation kit. TD and its mutant (td) were either fused to GST, expressed and purified from E.coli cells (left 3 lanes), or fused to GBD, expressed and immuno-purified from yeast cells (middle and right). TD lost its binding to Sir4C when it was expressed in E. Coli (the 2nd lane) or after βPPase treatment (last lane).
Figure 2. The TD mutants with threonine or glutamate substitutions at either of the serines could be functional in reestablishing silencing at the crippled HMR locus in the strain YSB2. The cell cultures were grown to saturation and diluted to O. D. 0.5. Then, 5-fold series of dilution was made and 10 µl of each dilution was spotted onto the control (SC-Ura) and test (SC-Ura-Trp) plates.
Figure 3. The TD mutants with cysteine (C) or glutamine (Q) substitutions at either of the serines could not be functional in reestablishing silencing at the crippled HMR locus in the strain YSB2. The cell cultures were grown to saturation and diluted to O. D. 1. Then, 10-fold series of dilution was made and 10μl of each dilution was spotted onto the control (SC-Ura) and test (SC-Ura-Trp) plates.
CHAPTER IV. CONCLUSIONS

At the end of 1997, before I decided to join the Voytas laboratory to study the retrotransposon Ty5, a functional Ty5 element was found in *Saccharomyces paradoxus*, a relative of *S. cerevisiae*. The target specificity of this Ty5 element was well characterized; over 90% of native and *de novo* Ty5 insertions were found within silent chromatin. The integrity of silent chromatin was also demonstrated to be necessary for Ty5 targeting. When the transcriptional silencer at the *HMR* locus was impaired and hence silencing was defective, the frequency of Ty5 insertion to the non-silenced *HMR* locus was decreased. Mutations in genes encoding essential components of silent chromatin, such as *SIR2, SIR3,* and *SIR4*, caused global changes in Ty5 integration site choice. These data suggested a simple working model for the mechanism of targeting: the pre-integration complex (PIC) of Ty5 interacts with some feature of silent chromatin to guide integration.

At the time I joined the lab, a plasmid-based targeting assay was developed and used to screen mutagenized Ty5 elements to identify element-encoded targeting determinants. A missense mutation near the IN and RT border was found to cause a targeting defect. I continued this work and started by asking a simple question: could I found more of this kind of mutation? After more than five years of investigation, I describe in this thesis the information I gathered. In this section I summarize my findings, discuss related observations, and, more importantly, describe the questions that still puzzle me.
The Ty5 targeting domain is defined

The answer to my original question was yes: I could find more targeting mutants. Eleven additional point mutations were identified that dramatically changed Ty5 target specificity. All the mutations, including the original, occur at one of four amino acid residues that together span on a six amino acid stretch (LDSSPP, the four critical residues are underlined). We called this motif the targeting domain (TD). The other amino acid residues around and within the domain were also mutated, and it was found that they do not contribute to target specificity.

The targeting mutants decreased transposition efficiency

The mutants we found in our screen not only changed targeting, but they also reduced transposition efficiency. The more severe the targeting defect, the lower the transposition efficiency (see the chart below). There are two possibilities to explain this phenomenon. First, it is possible that integration outside of the preferred target sites is less efficient. In this case, TD might facilitate transposition. The second possibility is that the PICs of the mutant elements are still recruited to silent chromatin; however, they have less ability to be integrated. In the second scenario, it is assumed that silent chromatin provides a barrier for element insertion, as was observed for the inability of DNases and restriction endonucleases to digest DNA bound by silent chromatin. Therefore, in this second hypothesis, TD possibly has a role in opening heterochromatin rather than simply recruiting the PIC. This is
supported by the observation in the *in vitro* binding that the mutant IN C-terminus is able to bind Sir4 C-terminus, which suggests TD is redundant for recruiting PIC.

TD is encoded by IN not by RT

Ty5 expresses a polyprotein, and the boundaries that define the individual peptides released by protease cleavage (e.g. IN and RT) are not defined. We replaced TD with a 6XHis-tag, and then showed in Western blots that the tag was associated with IN and not with RT.
The C-terminus of IN is sensitive to mutation

After mutagenizing the C-terminus of integrase, we noticed that more than 60% of the elements lost or had lowered frequencies of transposition. One reason that could account for the result: stop codons or frameshift mutations were introduced. However, the methods we applied for PCR mutagenesis introduce mostly substitutions of A and T to G and C, thus they bias against the introduction of stop codons and frameshift mutations. On average, elements had 2-8 mutations. Therefore, it is unlikely that fatal mutations were the only reason for the observation that 60% of the elements were transposition defective. The result rather suggested that the C-terminus is involved in some important functions for the element other than targeting. One example, is nuclear localization and we have shown that the IN nuclear localization determinants reside within the C-terminus (Fuerst and Voytas unpublished). We also showed that a dozen of the mutants failed to recombine with homologous substrates in the genome, suggesting that cDNA was not made. This implies that the C-terminus is involved in functions other than integration, and support for this comes from work with other Ty elements.

TD nucleates silencing

HMR silencing is greatly weakened when two of the three protein binding sequences in the HMR-E silencer are deleted. Transcriptional silencing can be restored by tethering Sir1p to the defective HMR locus. Tethering TD to the locus also restores silencing, and this depends on Sir2p, Sir3p and Sir4p, but not Sir1p. Obviously, the role of Sir1p in silencing can be
replaced by TD. Sir1p is proposed to recruit Sir4p to establish silencing. TD was later shown to bind to Sir4p.

**Overexpression of GBD-TD fusion disrupts silencing**

Telomeric silencing is decreased when a 2μ plasmid expressing a GBD-TD fusion is present in the cells. This effect is due to the GBD-TD caused degradation of Sir4p. Degradation depends on the ability of GBD to bind DNA and also requires Rislp, a DNA dependent ATPase (Fuerst and Voytas unpublished). Overexpression of Sir3p, but not Sir2p and Sir4p, can overcome the GBD-TD-induced transcriptional derepression. In addition, a growth defect is observed in the strain in which GBD-TD and Sir3p are overexpressed.

**TD interacts with Sir4p**

The interaction of TD with Sir4C was demonstrated *in vivo* by two-hybrid assays and *in vitro* by affinity binding experiments. In both cases, the nine amino acid peptide (TD with three flanking residues) was able to bind Sir4C. However, a longer version of the IN C-terminus enhanced the interaction with Sir4p in two hybrid assays. This IN C-terminal fragment could bind to Sir4C *in vitro*, even if TD was mutated or deleted (Xie and Voytas unpublished). This data suggests that another region of IN might also interact with Sir4C.
TD is phosphorylated

Several lines of evidence suggest that the serines in TD are phosphorylated and that this modification is essential for TD function. We observed doublets of GBD-TD in Western blots, whereas versions of GBD-TD with serine substitutions did not form doublets. In addition, GBD-TD expressed and purified from yeast cells (and not \textit{E. coli}) could bind to Sir4C \textit{in vitro}. More importantly, the binding was sensitive to the lambda protein phosphatase. When the serines were replaced by threonine or glutamate, TD function was partially maintained.

Kinases involved in TD modification

We genetically modified 97 kinase knockout strains so that they could be used in the targeting assay to screen for kinases that modify TD. Eight kinase candidates were found that affect Ty5 targeting. Among these, Dun1p is the most likely candidate, because in \textit{dunl} strains, GBD-TD cannot break telomeric silencing or interact with Sir4C. However, Dun1p may have a redundant activity or indirectly modify TD, because GBD-TD can still establish silencing at a crippled \textit{HMR} locus in a \textit{dunl} background. It is therefore likely that other kinases are involved in modifying TD.

Some kinases affect the targeting assay but not Ty5 targeting
We inferred that a kinase that regulates silencing would affect our targeting assay. However only one essential kinase involved in silencing has been reported, and it was not included in our strain collection. We found three additional kinases that are important for silencing, but a total of 28 kinases significantly affected the results of the targeting assay. We further analyzed the effect of the kinase Kin3p. When \textit{KIN3} was knocked out, the resulting percentage in the assay was less than 1\%, similar to the Ty5 TD mutants. However, the distribution of Ty5 insertions in the genome was not affected. The reason for how this and the other kinases affect the targeting assay remains a mystery.

**The Ty5 targeting model is updated**

Ty5 integration preference is executed by integrase. Integrase has a special targeting domain (TD), which binds to Sir4p, a component of silent chromatin. Targeting is regulated by the host cell, because TD is not functional until it is modified by a yeast kinase.

The updated model described above looks simple enough. However, a lot more details need to be filled in. How does TD interact with Sir4p? Does a barrier exist for the element to gain access to DNA in silent chromatin? If so, how does the element break in? How is TD phosphorylated? Is one or both serines modified? What kinases do the job? How do they regulate Ty5 targeting, and how are the kinases regulated? These questions and others are waiting to be answered.