THE FUNCTION OF THE GAG PROTEIN DURING THE PROCESS OF NORMAL AND

RESTRICTED TY1 RETROTRANSPOSITION IN *SACCHAROMYCES*

by

JESSICA MITCHELL TUCKER

(Under the Direction of David J. Garfinkel)

ABSTRACT

The long terminal repeat (LTR) retrotransposon Ty1 of the budding yeast *Saccharomyces* is a mobile genetic element whose life cycle resembles retroviral replication. Ty1 elements contain *GAG* and *POL* genes, which are translated to produce Gag, the structural protein of virus-like particles, integrase, reverse transcriptase, and protease. This work focuses on the molecular interactions between host and retrotransposon gene products during the Ty1 life cycle in order to increase our understanding of Ty1 genomic RNA dynamics prior to VLP assembly and restriction of retrotransposon amplification in the host genome. Ty1 genome-length sense RNA serves as the template for reverse transcription and protein production in the cytoplasm. In the absence of Gag, Ty1 genomic RNA is not efficiently exported from the nucleus and is less stable in the cytoplasm, where it is degraded in the processing body. Ectopic

Gag expression restores nuclear-cytoplasmic trafficking and stability of Ty1 genomic RNA, as well as virus-like particle assembly foci called retrosomes. Interestingly, the formation of retrosomes and virus-like particles are also disrupted in cells with elevated genomic copy numbers of Ty1 via an intrinsic restriction mechanism referred to as copy number control. High Ty1 copy number results in a dramatic decrease in Ty1 mobility and is mediated by a newly discovered 22 kilodalton protein (p22), which is translated from a subgenomic Ty1 sense RNA and shares sequence with the C-terminal half of the Gag protein. p22 has a *trans*-dominant negative effect on Ty1 retrotransposition and is both necessary and sufficient for copy number control (CNC). p22 inhibits Ty1 by interaction with Gag and its expression leads to disrupted retrosomes and lower yields of properly assembled and maturation-competent virus-like particles. Mutations in Gag cause resistance to CNC via the exclusion of p22 from assembled virus-like particles, indicating that the presence of p22 in virus-like particles is inhibitory to Ty1 replication. By studying the role of Ty1 Gag in the retrotransposon life cycle, the goal of this work is to inform future studies regarding retroelement replication and restriction in a variety of eukaryotes.

INDEX WORDS: retroelement, retrotransposon, retrovirus, budding yeast, Ty1, LTR, restriction factor, virus-like particle

THE FUNCTION OF THE GAG PROTEIN DURING THE PROCESS OF NORMAL AND

RESTRICTED TY1 RETROTRANSPOSITION IN *SACCHAROMYCES*

by

JESSICA MITCHELL TUCKER

B.S., The University of Alabama, 2010

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2016

THE FUNCTION OF THE GAG PROTEIN DURING THE PROCESS OF NORMAL AND

RESTRICTED TY1 RETROTRANSPOSITION IN *SACCHAROMYCES*

by

JESSICA MITCHELL TUCKER

Major Professor: David J. Garfinkel

Committee: Stephen Hajduk
Micheal McEachern
Heather Steet

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2016

DEDICATION

This dissertation is dedicated to the loving memory of Laurie Susan Kosloff Mitchell, Jeffrey Scott Mitchell, and Virginia Tucker Mitchell.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Eukaryotic genomes vary wildly in size and complexity, yet they all have one thing in common: they are subject to the escapades of molecular parasites called mobile DNA elements. Since the discovery of mobile DNAs, or transposons, by Barbara McClintock in the 1940s (1), we now know how transposons move around in their host genomes and occupy the majority of the genomic space in some organisms. While transposon insertions can be harmful to the stability of the host genome, more recent work has revealed potential benefits to the host for maintaining transposons. The budding yeast, *Saccharomyces cerevisiae*, is an informative model for the study of mobile genetic elements because it has guarded its genome against rampant colonization by transposons. The transposon content in the reference *S. cerevisiae* genome is only 3%, even though this yeast lacks several intrinsic mechanisms and pathways used by higher eukaryotes to inhibit mobile DNAs. In addition, the transposons found in budding yeast are limited to one type: long terminal repeat, or LTR retrotransposons. LTR retrotransposons replicate in a similar manner to retroviruses, like human immunodeficiency virus (HIV), which dramatically impacts human health. This introduction serves to illuminate the importance and biology of LTR retrotransposons, highlighting the *Saccharomyces* Ty1 retrotransposon, which is the most studied

retrotransposon of its class. The impact of Ty1 research due to its similarities with other retrotransposons and retroviruses across Eukarya is emphasized.

Retroelements in eukaryotic genomes

Retroelements, also known as Class I transposons (2), are defined as mobile genetic elements that replicate through an RNA intermediate and include both retrotransposons and retroviruses. In order to copy and insert into host genomes, retroelements encode a reverse transcriptase that converts the RNA intermediate into a DNA copy that can then be integrated or inserted into host DNA. In general, retroviruses encode additional proteins compared to retrotransposons that allow them to be infectious, or move from one host cell to another. In contrast, retrotransposons replicate intracellularly and are not infectious. Retrotransposons are amplified throughout host genomes via a "copy-and-paste" mechanism, potentially threatening the host cell due to amplifying copy number and potential for mutagenesis. In addition to retroelements, eukaryotic genomes also contain DNA transposons (Class II transposons (2)), which are transposons that mobilize through DNA intermediates utilizing a "cut-and-paste" mechanism for replication. Although DNA transposons have been highly successful at populating both prokaryotic and eukaryotic genomes (reviewed in (3)), DNA transposons are tangential to the work discussed here and will not be described further. Compared to DNA transposons, retrotransposons and retroviruses are exceedingly abundant and thus seem to have been especially successful at replicating in eukaryotic cells (4, 5). This has been attributed to the unique posttranslational modification of eukaryotic RNA, including a 5' cap and 3' poly(A) tail, which provides characteristics such as stability and protection from exonucleases in the context of the long RNA

intermediates produced by retroelements (6). In addition, eukaryotic genomes are larger than prokaryotic genomes and may be more tolerant of new insertions, allowing amplifying retroelements to occupy this niche (7). In turn, large genome sizes are associated with expanding populations of transposons (7-9). Prevalence in eukaryotic genomes is common, but the extent and the proportion of active retroelements vary. An extreme example is maize (*Zea mays*), which contains over 400 families of retrotransposons occupying at least 75% of its genome (10). In contrast, retrotransposons comprise a mere 3% of the *S. cerevisiae* genome with only five distinct retrotransposon families, two of which are completely inactive (11, 12). Similar to other mammals and plants, at least 40% of the human genome is comprised of retroelements (13). The majority of retrotransposon sequences harbored in eukaryotic genomes are not replication-competent due to mutations that occur during replication or accumulate over evolutionary time.

Non-LTR retrotransposons in eukaryotic genomes

Retroelements can be classified into two groups based on the presence or absence of LTR sequences flanking the element (3, 14) (Table 1.1). Non-LTR retrotransposons are a diverse group and include arthropod R1/R2 elements, *Drosophila* I elements, long interspersed nuclear elements (LINEs), and short interspersed nuclear elements (SINEs) (15, 16). Autonomous non-LTR retrotransposons, such as LINEs, encode the protein machinery necessary for replication, while non-autonomous retrotransposons including SINEs rely on gene products from autonomous elements for replication. The human LINE-1 element will be described here as an example of an autonomous non-

LTR retrotransposon, but there are major differences among non-LTR elements that will not be discussed (16).

Human LINE-1: A representative non-LTR retrotransposon

The replication of an active, full length human LINE-1 (L1) element is shown in Figure 1.1. Expression of L1 by RNA polymerase II using an internal promoter (17) results in a 6 kilobase (kb) mRNA with a 5' and 3' untranslated region (UTR) and two non-overlapping open reading frames, called ORF1 and ORF2. ORF1 binds RNA and contains nucleic acid chaperone activity (NAC) (18-21), while ORF2 contains endonuclease and reverse transcriptase activities (22-25). Both ORF1 and ORF2 are required for L1 retrotransposition (24, 26). Interestingly, antisense promoter activity has been identified in certain L1 elements (27), and recently, an associated protein product from antisense RNA transcripts called ORF0 was identified (28). This protein is thought to be specific to primates, localizes to the nucleus and enhances the level of L1 retrotransposition.

Following transcription and nuclear export of L1 mRNA, translation and the formation of a ribonucleoprotein (RNP) occurs in the cytoplasm. Translation of ORF1 is cap-dependent and relies on high efficiency scanning through a 5' UTR of 900 nucleotides (nt) (29). ORF0 mRNAs are also capped and may initiate translation at the first methionine (28). Translation of ORF2 is more complicated and not well understood. ORF2 is translated from the same RNA template as ORF1, yet does not rely on translation mechanisms used by other retroelements, such synthesizing a polyprotein or utilizing an internal ribosome entry site (IRES) (26, 29). Translation of ORF2 is independent of its initiating methionine and the sequence between the ORF1 and ORF2

4

(called interORF), but does require translation through an upstream ORF (30). This suggests that ribosome scanning and an unconventional termination/reinitiation mechanism is required for translation of ORF2. The levels of ORF2 are extremely low when compared to levels of ORF1 in cells. After translation of ORF1 and ORF2, ORF1 is thought to trimerize and bind L1 mRNA, along with ORF2, *in cis* (23, 31, 32). These RNPs colocalize with stress granules, which are cytoplasmic RNP aggregates that form under certain stress conditions and may contain non-translating RNAs (33).

Replication of L1 and other non-LTR retrotransposons likely occurs via target site primed reverse transcription (TPRT), an elegant mechanism for reverse transcription first characterized using the R1 and R2 elements of insects (34). Transport of the L1 RNP to the nucleus is not well understood, but there is evidence that active nuclear import occurs based on the fact that L1 can retrotranspose in non-dividing cells (35), and that nuclear import has been observed with other non-LTR retrotransposons (36). The endonuclease activity within ORF2 is responsible for creating a nick in target DNA, leaving a 3'-OH that serves as the primer for reverse transcription using the non-LTR RNA as a template. The target site cleaved by the endonuclease within ORF2 is typically T-rich (which speaks to L1's promiscuous integration specificity), which anneals with the poly(A) tail of non-LTR RNA to prime reverse transcription (24, 37). Abortive reverse transcription happens often, leading to the insertion of L1 with a severe 5' truncation (26). 5' truncated L1 elements are found throughout the human genome (38). Only 100 out of 500,000 L1 elements in the human genome may still be active, with the vast majority inactive due to this truncation event or various amounts of mutation accumulation (39-41).

<u>Eukaryotic LTR retrotransposons replicate in a manner similar to retroviruses</u>

The replication of LTR-containing retrotransposons is strikingly different than non-LTR retrotransposons (Table 1.1; Figure 1.2). In fact, LTR retrotransposon replication closely resembles that of retroviruses, making the study of LTR retrotransposons relevant to the study of infectious retroviruses such as HIV. The typical LTR retrotransposon contains two ORFs: *GAG* and *POL* (Table 1.1). The *GAG* gene encodes the structural protein (Gag) of the virus-like particle (VLP) and *POL* encodes various enzymes required for replication, including protease (PR), reverse transcriptase (RT) and integrase (IN) (42). LTR retrotransposons form a VLP using the structural Gag protein as a capsid, and two copies of the retrotransposon RNA in the form of an RNA dimer are packaged inside the VLP. Once the VLP assembles, the encoded protease cleaves the *POL* encoded polyprotein, releasing the individual LTR retrotransposon enzymes. The RNA dimer genome inside the VLP is the template used by reverse transcriptase, along with a tRNA primer, to create a double-stranded DNA copy of the element. This complementary DNA product (cDNA) is integrated into the host genome via the activity of the integrase enzyme. LTR-retroelement integration site specificity is determined by a tethering mechanism involving integrase and specific nuclear proteins involved in DNA transactions such as transcription, splicing or DNA replication. The nuclear proteins that interact with integrase differ among LTR retrotransposons. The focus of this dissertation, the Ty1 LTR retrotransposon of *S. cerevisiae*, is the most studied element of this class and more detail regarding replication of LTR retrotransposons can be found in subsequent sections. In addition to Ty elements in

yeast, other examples of LTR retrotransposons include the *copia* and *gypsy* elements of *Drosophila melanogaster* and *sireviruses* in plants (43), among others (42).

<u>Endogenous retroviruses in multicellular organisms can be classified as LTR retrotransposons</u>

An endogenous retrovirus is an integrated retrovirus present in the germline and is often classified as an LTR retrotransposon. Most endogenous retroviruses have exogenous retrovirus counterparts (44). The replication of a simple exogenous retrovirus is similar to that outlined above for LTR retrotransposons, with the primary difference being its infectious ability to exit and enter host cells (Figure 1.2). Exiting and entering host cells requires a gene called *ENV*, which encodes the envelope protein of the retrovirus. Once the immature virion (the retrovirus equivalent of the VLP; contains unprocessed Gag and Gag-Pol, the dimeric genomic RNA, and is coated with Env proteins) assembles either in the cytoplasm or at the plasma membrane, the virion buds from the plasma membrane and then undergoes maturation extracellularly. Maturation produces an inner capsid core in the virion, which consists of a cleaved capsid (CA) and nucleocapsid (NC) domains of the Gag protein surrounding the genomic RNA and mature Pol enzymes. The mature virions then bind to permissive cells through specific interactions between Env and host receptors/coreceptors and fuse with host membranes. The nucleocapsid core of the mature virion is released into the cell, where reverse transcription and integration occur. A newly integrated retrovirus in the genome is called a provirus. If an exogenous retrovirus is able to infect and establish a provirus in a germ line cell, then it becomes vertically transmissible and becomes an endogenous retrovirus (44). The endogenous retrovirus retains the ability to replicate

identically to its exogenous retrovirus counterpart; however vertical transmission of this sequence can lead to the accumulation of mutations that render the endogenous retrovirus inactive. In fact, it is thought that almost all of the ~450,000 endogenous retroviruses present in the human genome are no longer active due to mutation (41). An exception to this is the HERV-K (HML2) family of endogenous retroviruses. Recently, a comparison a 2,500 human genomes revealed several polymorphic HERV-K insertion sites (45), further supporting previous studies implicating that this family remains active in humans (46-48). In contrast, there are several examples of active endogenous retroviruses in mice (49, 50), including ERV-L, intracisternal A particles (IAP) and MusD elements. Most endogenous retroviruses have either lost or acquired mutations in the *ENV* gene, and thus are not infectious and function like LTR retrotransposons if active.

Consequences of having retrotransposons in eukaryotic genomes

Retrotransposon activity can threaten host cell survival due to *de novo* transposition events, DNA recombination between repetitive elements, or toxic effects of retrotransposon transcripts. Retrotransposition can have a variety of effects on a genome that is largely defined by the specific replication characteristics of the retrotransposon. For instance, integration site specificity of a particular retrotransposon can dictate the likelihood of insertional mutagenesis. LINE-1 elements integrate the genome at short and degenerate sequences and thus their promiscuity results in their high association with insertional mutagenesis, including the disruption of open reading frames and gene expression from a specific locus (51, 52). In fact, the insertion of LINE-1 elements causes a number of human genetic diseases, including hemophilia B and Duchenne muscular dystrophy among others (53). In contrast, yeast Ty1 elements

8

have a specific integration pattern that guides most integrations to benign regions, such as upstream of genes transcribed by RNA polymerase III (described below). However, even with the integration specificity of Ty1, a minority of transposition events can disrupt or alter gene expression or coding potential by insertional mutagenesis (54-56).

Aside from insertional mutagenesis, the presence of multiple retrotransposon copies in a genome can lead to chromosomal instability and promote chromosomal rearrangements (57-59). Extensive analysis with Ty1 has demonstrated this phenomenon. Homologous recombination between Ty1 elements or their directly repeated LTRs mediates several different chromosomal rearrangements including inversion, deletion, and translocations of genomic sequences (55, 60-63). A common LTR-LTR recombination event involves looping out of *GAG* and *POL* coding sequences, leaving behind a solo LTR. Solo LTRs from Ty1 are present in the reference strain of *S. cerevisiae* at around 300 copies, versus a total of 32 remaining full-length Ty1 elements (Table 1.2) (11, 12). Additionally, an artificially inserted Ty1 element on the left arm of chromosome V has been used to study Ty1-mediated chromosomal rearrangements (62). The most common type of rearrangement observed is the loss of the arm of chromosome V in combination with the presence of this arm fused to an alternate Ty1 genomic location (a non-reciprocal translocation), consistent with an event mediated by homologous recombination between Ty1 elements. In addition to retrotransposition and DNA recombination, there are also examples of the transcription products of retrotransposons playing a role in human disease. RNA transcripts produced from Alu elements, a class of nonautonomous non-LTR elements in the human genome, are unregulated and present in high levels in the retinal cells from patients with age-related

macular degeneration (64). Elevated Alu RNAs are cytotoxic and are thought to induce an inflammatory response and cell death.

Retrotransposon movement can also positively affect eukaryotic genomes by enriching the phenotypic variation of an organism or specific cell lineages. Retrotransposon insertions within a gene or its promoter can disrupt gene expression. For example, most of the light-skinned grape varieties are derived from dark-skinned varieties that have a *Gret1* element insertion upstream of a gene involved in red plant pigment, or anthocyanin, biosynthesis (65). The merle coat pattern in dogs is displayed in heterozygotes carrying an incompletely dominant allele of a mammalian pigmentation gene (*SILV*) disrupted by a retrotransposon (66). A dominant allele of the *agouti* gene in mice, caused by an insertion of an IAP endogenous retrovirus, results in a range of coat color that is typically more yellow and mottled when heterozygous (67). Interestingly, the range of phenotypes seen is due to epigenetic marks that are established at the *agouti* gene during development (68). IAP sequences are typically silenced by DNA methylation (discussed below) and the presence of this retrotransposon in the *agouti* gene can direct methylation to this site and alter gene expression. In humans and rats, L1 elements are transcribed and transpositionally active in neuronal precursor cells (69-71), resulting in somatic mosaicism in the brain. This is hypothesized to be a source of neuronal plasticity but could be a risk factor for certain mental disorders such as schizophrenia in humans (72).

The coevolution of eukaryotes and retroelements has resulted in an increasing number of exaptation or domestication events involving retroelement sequences. In other words, retroelement genes or regulatory elements have been co-opted for the use

or benefit by the eukaryotic host. For example, the evolution of pregnancy and placental development in mammals is laced with contributions from retrotransposons and endogenous retroviruses. A gene required for placental development, specifically for the fusion of trophoblast cells, called *syncytin* is derived from the *ENV* genes of endogenous retroviruses (73). Additionally, ancient retrotransposon sequences seem to function as regulatory elements for the expression of a certain subset of genes required for uterine gene expression during pregnancy (74). Expression and activation of human endogenous retroviruses are also important for development through the establishment of pluripotency (75). Another recent study revealed the enrichment of transcription factor binding to endogenous retrovirus sequences near interferon-stimulated genes (76). Thus, it seems likely that endogenous retroviruses have been co-opted by the host to serve as enhancers to boost the innate immune response.

Another important example of domestication of retrotransposon sequences relates to the maintenance of telomere length in *Drosophila*. Telomere length is maintained in most eukaryotes by a genome-encoded reverse transcriptase enzyme called telomerase. However, the telomerase gene and the standard telomeric repeats (pentameric in insects) are not present in *Drosophila melanogaster* (77, 78). Instead, *Drosophila* telomeres are comprised of repeating arrays of the non-LTR retrotransposons TAHRE, TART, and HeT-A (79). Retrotransposition of these elements maintains telomeric ends, as they preferentially integrate at the ends of chromosomes. Therefore, these telomeric retrotransposons have replaced the need for telomerase to replicate chromosome ends.

Domesticated retroviral sequences can protect the host cell from exogenous retroviral infection. Two well-studied examples involve restriction of murine leukemia virus (MLV) infection in mice and Jaagsiekte sheep retrovirus (JSRV) infection in sheep by domesticated retroelement genes (80-85). Both of these exogenous retroviruses are restricted in certain strains or cell types due to the production of proteins in the host cells called restriction factors. In mice, a gene called friend virus susceptibility gene-1 (Fv1) is responsible for protection against early stages of MLV infection. In sheep, a chromosomal gene locus called enJS56A1 is required to inhibit late stage replication of JSRV. Both of these genes are derived from endogenous retroviral *GAG* genes and interact with exogenous Gag proteins to restrict infection. The mechanism of restriction will be discussed further in the next section. In addition to *GAG* sequences, *ENV* sequences have also been demonstrated to exert a protective effect in cells. Domesticated Env proteins are thought to block the binding of exogenous virions to host cell receptors, thus function via receptor interference. One example is the feline Refrex-1 restriction factor, which restricts feline leukemia virus (86).

Genome defense mechanisms used by eukaryotes to control retroelements

Because of the potential harm caused by excessive replication of retroelements, eukaryotes have developed transcriptional and posttranscriptional mechanisms to minimize the successful replication of both retrotransposons and retroviruses. One highly utilized mechanism of control is methylation of transposon DNA. Eukaryotic DNA is methylated at cytosine bases (typically when adjacent by a guanine residue, called a CpG site) by DNA methyltransferase enzymes to form 5-methylcytosine, usually resulting in the cessation of gene expression from that locus. DNA methylation in

12

mammals plays a role in silencing L1 and endogenous retrovirus expression and is established in the early embryo (87, 88). In *Neurospora crassa*, almost all DNA methylation is found in repetitive sequences, including LTR retrotransposons (89). In addition to modification of DNA bases, the modification of histone proteins also impacts expression of retroelements. Although DNA methylation is not required for silencing of endogenous retroviruses in mouse embryonic stem cells (90-92), histone modifications such as the methylation of lysine 9 on histone H3 (H3K9) or histone deacetylation are responsible for the silencing of these elements. In *N. crassa* and *Arabidopsis thaliana,* H3K9 modification is required prior to DNA methylation (93, 94). Historically, the silencing of mammalian SINE elements was thought to be a result of high amounts of DNA methylation observed at these loci. However, recent studies suggest that the preceding histone modification could be more important than the DNA methylation for expression of these elements (95, 96). Similarly, histone modification is important for silencing of retrotransposons in *Drosophila*, whose genomes contain low levels of DNA methylation (97-99). In contrast, budding yeast lack detectable levels of methylated DNA resulting from the action of DNA methylases (99, 100). In diploid *S. cerevisiae*, Ty1 expression is limited by the binding of the *MAT*a1/α2 repressor complex to an internal sequence in Ty1 (101). However in haploid budding yeast*,* Ty1 mRNA transcripts constitute about 1% of the total RNA and 10% of polyadenylated RNA, suggesting that transcriptional regulation of Ty1 elements is not a mechanism for controlling retrotransposition in haploid cells (102, 103).

Another retrotransposon defense mechanism in eukaryotes is mediated by RNA interference (RNAi) that ultimately results in the cleavage of the retroelement RNA and

13

subsequent degradation. RNAi involves the cleavage of a dsRNA substrate in the cell into small interfering RNAs (siRNA; 20-25 nt in length) via a ribonuclease called Dicer (104). The siRNAs associate with a PAZ and PIWI-containing protein called Argonaute, which guides the RNA-induced silencing (RISC) complex to cleave complementary target RNA (105, 106). Studies in *Drosophila*, *Arabidopsis*, the fission yeast *Schizosaccharomyces pombe*, and *S. castellii* (a distant relative of *S. cerevisiae*) revealed some of the small RNAs present in cells map to retrotransposon sequences, suggestive of a role for RNAi in targeting retrotransposon RNAs (107-110). The source of the dsRNA substrate cleaved by Dicer during RNAi is thought to arise from either hairpin structures in retrotransposon RNAs, or annealing of sense and antisense retrotransposon RNA transcripts. There are no *dicer* or *argonaute*-like genes encoded by *S. cerevisiae* and additionally no small siRNA-like species observed in this organism, indicating that the RNAi pathway is not functional (110). When *S. castellii* dicer (*DCR1*) and argonaute (*AGO1*) genes are expressed in *S. cerevisiae*, a number of siRNAs correspond to Ty1 sequences and cause a dramatic decrease in the level of Ty1 RNA.

In eukaryotes with specialized germ cells, retrotransposon RNAs are targeted for degradation through a separate RNAi pathway mediated by piwi-associated RNAs, or piRNAs. This pathway appears to specifically protect the germline from harmful insertions of retroelements. piRNAs were first discovered in *D. melanogaster* and are encoded together in a genomic locus called a piRNA cluster, which is transcribed into one long piRNA precursor (111, 112). The sequences of the processed piRNAs (20-30 nt) correspond to transposable element and repeat sequences (107). After processing, the mature piRNAs are loaded onto a group of proteins within the Argonaute family

called PIWI proteins. This complex then can cleave target transposon RNA. Interestingly, the cleaved transposon RNA can be further processed and used to generate new secondary piRNAs from the piRNA precursor, in what is referred to as a "ping-pong" mechanism (112, 113). In addition to this post-transcriptional gene silencing, piRNAs have also been implicated in transcriptional control involving histone methylation (97, 98, 114). piRNA clusters and piRNA-mediated transposon silencing also occurs in mammalian germ cells (115). However, piRNAs and PIWI proteins are not found in the genome of *S. cerevisiae* or its close relatives such as *S. paradoxus* (110).

In addition to RNAi, host proteins referred to as restriction factors inhibit replication of retrotransposons and retroviruses. Restriction factors are commonly expressed as part of innate immune response and often can be stimulated by interferon during virus infection. One well studied example of a host restriction factor that can restrict a wide variety of retroelements is apolipoprotein B mRNA editing enzyme and catalytic polypeptide-like 3G (APOBEC3G). APOBEC3G is a cytidine deaminase and was originally discovered as the target of the HIV encoded Vif protein (116). To restrict retroelements, APOBEC3G is associated with the reverse transcription complex within the virion or VLP and deaminates deoxycytidine residues on single-stranded DNA, resulting in deleterious mutations during replication and a nonfunctional provirus (117). APOBEC3G can also interfere with the progression of reverse transcription or prevent integration of cDNA by mutation of the viral sequence in a way that makes it unsuitable for integration (118, 119). Vif, in turn, functions by increasing APOBEC3G ubiquitinylation and degradation by the proteasome (120-122). The APOBEC family has

been shown to inhibit endogenous retroviruses in a deaminase-dependent manner, as evidenced by characteristic hypermutation present at these loci (123). APOBEC3G restricts other viruses, including DNA viruses such as parvovirus and herpesvirus in both a deaminase-dependent (causing hyperediting of the viral genome) and -independent manner (perhaps stalling DNA replication of the viral genome) (124, 125). L1 retrotransposition is inhibited in the presence of APOBEC3A and this inhibition requires functional cytidine deaminase activity, but a global hypermutation phenotype of L1 elements is not observed, perhaps due to the fact that L1 cDNA is protected during reverse transcription by remaining bound to L1 mRNA (126). Although the *S. cerevisiae* genome does not encode the APOBEC family of proteins, human APOBEC3G associates with VLPs, hypermutates Ty1 cDNA and inhibits retrotransposition when ectopically expressed (127-130). Interestingly, there are many examples of interferon-induced restriction factors in higher eukaryotes, including Bst2/Tetherin, MOV10, ZAP, SAMHD1; however, none are present in budding yeast.

One group of restriction factors especially relevant to my work is the capsid (CA)-binding restriction factors, including Fv1, tripartite motif 5 alpha (TRIM5α), myxovirus resistance B or 2 (MxB/Mx2), and the enJS56A1 restriction factor (80, 82, 131-134). TRIM5α functions during the early stages of retroviral replication by binding to the incoming capsid core and preventing steps prior to reverse transcription. Fv1 and MxB/Mx2 also bind the capsid core, but inhibit a step post reverse transcription but prior to integration. The first retroviral restriction factor identified, Fv1, protects mice from murine leukemia virus (MLV) infection and displays specific patterns of MLV-tropism (135). Restriction can be overcome by increasing the infective dose or altering MLV CA

residue 110, implicating CA as the target of restriction (82). Subsequently, many other residue changes have been identified that impart sensitivity to Fv1 and some of these same residues also allow for restriction by TRIM5α, a CA-binding factor that restricts HIV-1 (83, 84, 136-138). Structural information combined with known residues mutated in sensitive viral CA proteins indicates that exposed surfaces on CA comprise the interaction domain for these CA-binding factors (137-140). MxB/Mx2 was previously characterized as a factor involved in resistance to influenza A in mice (141), but was recently demonstrated to restrict HIV-1 (132-134). MxB/Mx2 interacts directly with HIV-1 CA, but mutations in CA that restore viral replication do not disrupt binding between MxB/Mx2 and CA (142), indicating that other cellular factors may be involved. A late-acting restriction factor, enJS56A1, is an endogenous retroviral locus in sheep that protects against Jaagsiekte sheep retrovirus (JSRV) infection. enJS56A1 inhibits replication of the retrovirus post integration and translation through a mechanism referred to as Jaagsiekte late restriction (80). The enJS56A1 locus encodes a *trans*-dominant Gag protein which is coexpressed and coassembled with the exogenous JSRV Gag. This coassembly interrupts the productive trafficking of the viral Gag to the pericentriole and targets the complex for degradation, eliminating further infection. In Chapters 3 and 4, a novel Gag-binding restriction factor called p22 is described that inhibits Ty1 retrotransposition in *Saccharomyces*. p22 functions during Copy Number Control (CNC; described below) and its interaction with Ty1 Gag is compared and contrasted to known CA-binding restriction factors. Interestingly, p22 is derived from the Ty1 *GAG* gene, just as Fv1 and enJS56A1 are derived from endogenous retroviral *GAG* genes.

Ty elements in *Saccharomyces*

The budding yeast retrotransposon Ty1 (Transposon yeast 1 (143, 144)) and other yeast retrotransposons have been studied for over 30 years, greatly increasing our understanding of eukaryotic transposons and retroviral replication. In fact, the term "retrotransposon" was coined by a Ty1 biologist (Jef Boeke) upon the discovery that Ty1 replicated via an RNA intermediate like retroviruses (143, 145). The reference laboratory strain of *S. cerevisiae*, S288C, was the first eukaryotic genome to be sequenced and is one of the major strains used for study of Ty elements (146, 147). *S. cerevisiae* and close relatives harbor only LTR retrotransposons that comprise five distinct families called Ty1-Ty5 (144, 148). Ty1, Ty2, Ty4 and Ty5 elements belong to the Ty1/*copia* class of LTR retrotransposons and Ty3 belongs to the Ty3/*gypsy* class (149, 150), with the order of genes in *POL* being the defining feature separating these two classes. Ty1 and Ty2 elements are the most similar among the families, with a single nucleotide difference in the LTRs, but divergent *GAG* and *POL* sequences (11). Ty2 may have been introduced into a recent ancestor of *S. cerevisiae* by horizontal transfer from *S. mikatae* (12, 151). In S288C, full length copies of Ty1-Ty5 are present with varying copy numbers (Table 1.2) (11, 12, 152). Interestingly, the copy numbers of solo LTRs (a result of LTR-LTR recombination) sequences from these families are much more abundant (Table 2). Of the five families, Ty1 and Ty2 comprise three quarters of the retrotransposon insertions in S288C and are both active (11, 12). Ty3 also appears to be active (Table 1.2). Ty4 and Ty5 are no longer active; the full length copies of Ty4 and Ty5 present in the genome are nonfunctional (11, 12, 152). In sequenced natural isolates of *S. cerevisiae* and its closest relative *S. paradoxus*, Ty

abundance is lower when compared to S288C, perhaps implicating natural selection as a factor responsible for maintaining a lower copy number of retrotransposons (151, 153, 154).

Ty1 gene structure

Ty1 is the most active and abundant retrotransposon in the S288C genome (146, 147) and is the focus of my work. An active, full length Ty1 element is about 5.9 kb in length and consists of *GAG* and *POL* genes that are bracketed by LTRs (334 basepairs (bp)) (Figure 1.3). Similar to retroviruses, the LTRs are important for expression and replication and are comprised of three regions: U3 (unique to the 3' end of the mRNA), R (repeat), and U5 (unique to the 5' end of the mRNA). *GAG* and *POL* (or historically, TyA and TyB) overlap for a stretch of 38 bp. The synthesis of the Pol precursor protein, described below, requires a programmed +1 ribosomal frameshifting event during protein synthesis that relies on specific Ty1 mRNA sequences. *GAG* encodes the Gag protein, required for the formation of VLPs and nucleic acid chaperone (NAC) activity, while *POL* encodes three enzymes in the Ty1/*copia* class-specific order: protease (PR), integrase (IN), and reverse transcriptase (RT). The ends of the Ty1 element consist of a conserved inverted dinucleotide sequence (5′-TG…CA-3′), which is important to integration. The entire replication cycle occurs inside the cell, as Ty1 does not encode an *ENV* gene (Figure 1.4).

Ty1 transcription

Ty1 is transcribed by RNA polymerase II and the resulting transcripts are capped and polyadenylated. Transcription to produce retrotransposition-competent Ty1 transcripts initiates in the 5' LTR and terminates in the 3' LTR, producing a Ty1 mRNA

(5.6 kb) with R domain sequences at the 5' and 3' ends (Figure 1.3) (155-157). The 5' LTR contains upstream activating sequences and two TATA boxes upstream of the transcription start site for Ty1 that are necessary for optimal Ty1 transcription (158, 159). Because most retroviral promoters reside in the 5' LTR, the Ty1 LTR was initially hypothesized to serve as the promoter for Ty1. The 5' LTR does exhibit weak promoter activity (160), but is not sufficient for transcription of Ty1. The required promoter and transcriptional regulatory region extends from the 5' LTR into *GAG* sequence, covering the first kb of Ty1 sequence (159, 161). Replacing the 5' LTR sequence with exogenous DNA sequence still results in transcription of Ty1, implicating that the internal promoter sequence is highly important for normal transcription of Ty1 (161).

There are a number of transcription factor binding sites within the Ty1 promoter element (162). Only a subset of these transcription factors play a strong role in modulating Ty1 transcription under normal growth conditions in haploid cells. These include two transcription factors of the Kss1 mitogen-activated protein kinase (MAPK) cascade, Tec1 and Ste12, which activate Ty1 transcription and bind within *GAG* (163, 164). Other transcription factors regulate Ty1 transcription under specific growth conditions, such as amino acid starvation (Gcn4) or adenylic nucleotide stress (Tye7) (165, 166). In diploid cells, Ty1 expression decreases about 10-fold due to the binding of the a1/α2 repressor complex encoded by the *MAT* locus (101). In addition to transcription factors, members of the Spt/Ada/Gcn5 (SAGA) histone acetyltransferase complex are involved with transcriptional regulation of Ty1 (167). Spt3 is a SAGA component that has been studied extensively in regards to Ty1 transcription. In a *spt3Δ* mutant, full length Ty1 mRNA is greatly reduced; instead, a shorter Ty1 transcript

becomes more prominent (168-170). This shorter transcript initiates 800 bp downstream of the transcription start site for retrotransposition competent Ty1 mRNA (168). Spt3 physically interacts with the TATA-binding protein (TBP) and is thought to recruit TBP to the SAGA complex (171-173). The interaction of Spt3 with TBP has been shown to modulate transcription, and in this case Spt3 is required for enhancing Ty1 transcription from the 5' LTR. In a *spt3Δ* mutant, Ty1 transcription initiates predominantly at a downstream site within *GAG*, which is probably a site that is not as heavily dependent on Spt3 for initiation as the site within the LTR. Importantly, this internally initiated Ty1 mRNA is also present in wild type *Saccharomyces* (Chapter 3) and can have an inhibitory effect on Ty1 retrotransposition. The internal promoter of Ty1 is likely involved with the production of this internally initiated RNA species (called Ty1i RNA) and may also act bidirectionally, as several Ty1 antisense RNA transcriptions are initiated in this region as well. Although these antisense RNAs were initially thought to be inhibitory to Ty1 retrotransposition (see below), there is more convincing data implicating the Ty1i RNA species in Ty1 inhibition.

Ty1 RNA polyadenylation occurs at the end of the R domain of the 3' LTR via 3'- end forming signals that reside in the U3 and R domains (157, 174). In whole cell extracts, it appears that only 15% of Ty1 RNA is polyadenylated (175), but whether this is due to low efficiency of cotranscriptional polyadenylation or to posttranscriptional removal of the poly(A) tail is undetermined. Regardless, Ty1 mRNA is a highly abundant transcript in yeast cells (102, 103). The abundance of Ty1 mRNA in haploid yeast may be related to its stability in the cytoplasm via the protection of Gag-binding in cytoplasmic VLP assembly sites called retrosomes (discussed below). In Chapter 2, I present data

presented suggesting that Ty1 mRNA is noticeably less stable in the absence of Gag protein and is subject to degradation by cellular RNA degradation machinery.

<u>Ty1 mRNA export</u>

The exported Ty1 mRNA serves as a template for two important functions in the cytoplasm: translation and reverse transcription. Ty1 mRNA is exported from the nucleus in a Mex67p-dependent pathway (175), but details regarding its export and subsequent cytoplasmic localization are not well understood. Mex67 is an ortholog of the human TAP protein and a nuclear export factor responsible for the export of the majority of RNA transcripts in yeast (176). Mex67 is encoded by an essential gene, but experiments utilizing a temperature sensitive allele, *mex67-5*, reveal that a small proportion of Ty1 mRNA can be seen in the nucleus when cells are grown at the nonpermissive temperature of 30° C (175). It is not known whether there is an alternate export pathway for Ty1 mRNA or a spatiotemporal partitioning of Ty1 mRNA serving the functions of a translation template or the genomic RNA (gRNA) used for reverse transcription. In the case of some retroviruses, these pools are partially distinguished by splicing of the retroviral RNA. For example, HIV RNA is differentially spliced to form mRNA for translation of some viral proteins, while unspliced transcripts primarily serve as the genomic RNA (gRNA), although this unspliced transcript is required for production of Pol proteins as well (177). The Gag protein of Rous sarcoma virus (RSV) has been shown to traffic to the nucleus and bind unspliced RSV RNA (178-182). The Gag-RNA complexes are exported and this is thought to "mark" this pool of RSV RNA for packaging as the gRNA. Ty1 RNA is not spliced and thus functions as gRNA and mRNA simultaneously. I explore the role of Ty1 proteins, especially Ty1 Gag, in the

22

nuclear export and downstream function of Ty1 mRNA in Chapter 2. Separately, recent studies implicating a role for the signal recognition particle (SRP) and ER translocation of Ty1 Gag-p49 suggest that the pool of Ty1 RNA utilized for translation versus packaging might be the same (183).

<u>Translation of Ty1 proteins</u>

Ty1 mRNA is translated to form two products, a 49 kilodalton (kDA) Gag protein (Gag-p49) and a 199 kDA Gag-Pol polyprotein (Gag-Pol-p199) (Figure 1.3). The methionine (ATG) initiation codon is near the end of the 5' LTR (nt 294) and the stop codon (TGA; nt 1614) for Gag-p49 is directly after the overlap region between *GAG* and *POL* (Figure 1.3). The majority of the time, the ribosome translates Gag-p49. The synthesis of Gag-Pol-p199 requires a +1 ribosomal frameshifting site on Ty1 mRNA. Ribosomal frameshifting events are exceedingly common in retrotransposons and retroviruses and typically occur due to the presence of a pseudoknot or a heptanucleotide slippery sequence within the mRNA. Both of these structural features cause the ribosome to pause, either due to the physical barrier imposed by the pseudoknot structure or a codon downstream of the slippery sequence that is decoded by a rare tRNA. During the pause, the ribosome can slip forward or behind to alter the reading frame. Ty1 uses a heptameric frameshifting sequence (CUU-AGG-C; nt 1596-1602) responsible for putting *POL* in frame with *GAG* to form the Gag-Pol polyprotein (184, 185). The "hungry" AGG codon is decoded by tRNA^{Arg}(CCU), which is encoded by *HSX1* (184). Unlike other tRNA genes, *HSX1* is only present in one copy in the genome and is expressed at low levels. The ribosome pauses at this codon waiting on the rare tRNA^{Arg}(CCU) and as a result, the tRNA^{Leu} slips forward, as it is capable of decoding

both CCU and UUA. When *HSX1* is deleted, the frequency of Ty1 frameshifting increases dramatically, indicating that the ribosomal pause at the AGG codon is required for frameshifting (186, 187). In haploid cells, Ty1 frameshifting occurs 5% of the time, resulting in a ratio of Gag-p49/Gag-Pol-p199 that is 20:1. In addition to the production of Gag and Gag-Pol, I co-discovered a novel Ty1 protein product, called p22, which is described in Chapter 3. p22 is translated from Ty1i RNA from one of two in-frame start codons in *GAG* coding sequence. The expression of this protein inhibits the Ty1 retrotransposition.

Recent work indicates that ribosome-nascent chain complexes that are translating Ty1 mRNA are a target for binding by the signal recognition particle (SRP) (183). The SRP is a chaperone required for trafficking of specific ribosome-nascent chain complexes to the endoplasmic reticulum (ER). Studies examining affinity-purified SRP complexes found evidence of associated Ty1 mRNA (183, 188). Nascent peptides that are produced from the SRP-associated ribosomes are translocated to the ER. The ER trafficking of Ty1 Gag appears to increase protein stability, as the inhibition of ER trafficking results in extremely low levels of Gag (183). If Gag does in fact traffic through the ER, then there is a yet undiscovered pathway for retrograde transport of Gag from the ER to the cytoplasm, where it is hypothesized to bind to the Ty1 mRNA loaded onto the SRP/ribosome complex. Thus, in this scenario, the nucleation of translating Ty1 RNA and Gag on SRP/ribosome complexes initiates the assembly of retrosomes and eventually VLPs. The characterization of a retrograde pathway for Gag trafficking from the ER to the cytoplasm will be extremely enlightening, as no cytoplasmic yeast proteins have been shown to undergo this form of trafficking. Additionally, this ER translocation

mechanism explains a long-standing observation that Ty1 proteins function *in trans* to package Ty1 mRNA (187, 189), whereas L1 proteins show *cis* preferences with respect to their transcript (31, 32).

<u>Retrosomes: putative sites for VLP assembly</u>

After translation, Ty1 protein precursors and mRNA colocalize in cytoplasmic foci referred to as retrosomes (Figure 1.4). Retrosomes were initially discovered via temperature sensitive yeast carrying mutations in an essential gene encoding Rbp1, the largest subunit of RNA polymerase II (175). Large poly(A)$^+$ RNA foci that colocalized with Ty1 mRNA (visualized by fluorescence *in situ* hybridization (FISH)) and Gag (immunofluorescence (IF)) were observed in *rbp1* mutants and these foci were named T bodies. When wild type cells were examined, poly(A)$^+$ foci disappear, but the T bodies (retrosomes) remained, indicating that a fraction of Ty1 RNA is normally concentrated in a specific location in the cytoplasm. Retrosomes are most apparent in cells grown between 20-25 ˚C, which are optimal temperatures for Ty1 retrotransposition (190-193). Several results suggest a positive association between the presence of retrosomes and Ty1 retrotransposition. For example, retrosomes are not observed when cells are grown at suboptimal temperatures for retrotransposition (> 30 ˚C) (194). Evidence has also been presented suggesting that the retrosome is the site of VLP assembly. Retrosomes are observable by FISH/IF in the S288C background, but transmission electron microscopy (TEM) reveals that these foci do not contain assembled VLPs (194). If S288C is transformed with a multicopy (2μ) plasmid carrying a galactose-inducible Ty1 element (pGTy1), the retrosomes observed by FISH/IF now contain VLPs, indicating that retrosomes observed by FISH/IF can contain VLPs under conditions optimal for

assembly. In addition, VLPs observed in cells expressing pGTy1 are not 100% dispersed and are typically grouped in large clusters, indicating a general nucleation point in the cytoplasm (145, 194). I explore the determinants of retrosome formation in Chapter 2. Importantly, we found that Ty1 Gag is not only required for the formation of retrosomes but increases the stability of Ty1 mRNA.

Retrosome appearance and morphology is dependent on several cellular Ty1 cofactors. For example, processing (P) body proteins, including Xrn1, Lsm1 and Pat1, are required for normal formation of Ty1 retrosomes (194). Retrosomes are distinct from P bodies, although some P body proteins colocalize with Ty1 retrosomes (193, 194). Deletion of these host genes results in the localization of Ty1 mRNA and Gag proteins into cytoplasmic puncta, rather than large foci. P body mutants contain normal levels of Ty1 mRNA and Gag and produce VLPs, but the VLPs are more dispersed and are less likely to be found in clusters. VLP clusters may be important for efficient retrotransposition, as these mutants exhibit a 10-200-fold decrease in Ty1 retrotransposition. P body proteins may also play roles in Ty1 replication other than the assembly of VLP clusters. For instance, Xrn1 is a 5'-3' cytoplasmic exonuclease that degrades RNA in P bodies and has been shown to influence the levels of Ty1 mRNA and Ty1i RNA (refer to Chapter 3) (130, 195). P body mutants also produce VLPs that lack normal levels of mature Pol products, which is important for post-VLP assembly functions involving reverse transcription and integration of Ty1 (194). Separately, the assembly of retrosomes is enhanced in mutants that are defective for various aspects of RNA polymerase II function, such as *rbp1*, *srb2Δ* and *srb5Δ* (subunits of Mediator complex), or *spt21Δ* (regulates transcription by RNA polymerase II) (175). As described

above, Ty1 mRNA associated with a ribosome/SRP complex may nucleate retrosomes, as gene deletions that disrupt Gag association with the Ty1 mRNA/ribosome/SRP results in the absence of retrosomes and conditions that slow SRP-mediated ER translocation enhance the formation of retrosomes (183).

<u>VLP assembly</u>

Ty1 VLPs are icosahedral and exhibit a wide range of diameters from 30-80 nanometers (nm) (Figure 1.5) (196-198). The shell of the Ty1 VLP is comprised of Gag-p49 and Gag-Pol-p199 proteins, but VLPs can form from in cells that just express Gag. Producing Ty1 Gag in *E. coli*, which lacks many orthologous cytoplasmic proteins found in eukaryotes, also results in VLP formation, indicating that Ty1 Gag has certain self-assembly properties (196, 199). VLPs have a spiky shell, with each spike hypothesized to represent a trimeric association of Gag proteins based on visualization of purified Ty1 VLPs subjected to negative staining/TEM or low resolution (38 Å) cryo-EM (198, 200, 201). Immunological analysis of assembled VLPs reveals that the Gag protein is oriented such that the N-terminus of the protein faces the outside of the particle, while the C-terminus faces the inside of the VLP (202). This orientation allows for the Pol proteins and a region of Gag responsible for NAC activities (discussed below) to be within the VLP. VLPs apparently contain pores that are visible by cryo-EM and allow entry of RNase A (14.7 kDA), but not Benzonase (30 kDA) (197, 201). These results also reinforce earlier work suggesting that VLPs are permeable to nucleic acids and dNTPs without the detergent activation step required for enveloped retroviruses (145, 203, 204).

Certain cellular transcripts are specifically packaged within the Ty1 VLP. Two of these species are important for the replication of Ty1, the Ty1 gRNA and the primer for reverse transcription, tRNA[iMet] (205-207). Like retroviruses, Ty1 gRNA is present in the VLP as a noncovalent dimer and likely interacts with the NAC region of Ty1 Gag (207, 208). Dimerization of retroelement RNA is responsible for the copackaging of two retroelement RNAs within each virion or VLP. The packaging and dimerization of HIV gRNA has been studied extensively and requires both a packaging sequence (psi) and a dimerization sequence (DIS). The DIS is a 6 nt palindromic sequence found in the 5' UTR of HIV RNA (209-211). This palindromic sequence is found within an apical loop of a stem loop structure and can interact by standard base pair interactions with the same apical loop of a second transcript, called "kissing loop" interactions. Although there is an abundance of data supporting a role for the DIS in dimerization, it is not sufficient for dimerization *in vivo* and HIV infectivity is not completely abolished in its absence, which is likely a reflection of the extensive RNA-RNA interactions that occurs between retroviral RNAs (212-214). The dimerization site for Ty1 remains to be determined. However, recent structural analysis of Ty1 RNA within VLPs using selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) led to the identification of three possible palindromic sites that could be involved in intermolecular RNA interactions within the VLP (215, 216). There is some evidence suggesting that binding of tRNA[iMet] to Ty1 RNA initiates dimerization (208). In an *in vitro* dimerization assay involving a mini-Ty1 5' RNA and a Ty1 Gag-derived peptide called TyA1-D (amino acids 299-401), Ty1 RNA dimerization only occurs in the presence of tRNA[iMet] (208). This observation has yet to be confirmed *in vivo*. The sequences required for HIV RNA packaging (called

the ψ site) include the 5' LTR up to the start codon of *GAG* and thus encompasses the DIS site. It is generally thought that dimerization of HIV RNA occurs prior to binding of Gag to the ψ site for packaging, but a recent study has demonstrated that HIV RNA dimerization does not occur until HIV RNA/Gag complexes reach the plasma membrane (217). The exact sequences required for Ty1 gRNA packaging are not well understood. Deletion analysis of Ty1 RNA mobilized via Ty1 helper proteins expressed *in trans* revealed important *cis*-acting sequences in the 5' LTR and the first 100 bp of *GAG* that are important for Ty1 gRNA packaging (218). These sequences overlap a predicted pseudoknot structure found in Ty1 gRNA; mutations that disrupt this structure result in decreased retrotransposition, but do not block Ty1 gRNA packaging (216, 219). A distinct packaging signal within Ty1 gRNA needs to be rigorously defined. Other RNAs are packaged in Ty1 VLPs as well, such as *HIS3*, *TRP1*, and tRNA[Ser] (218, 220), but the presence of some of these cellular transcripts in VLPs can be attributed to their cellular abundance (221). This means that Ty1 could introduce pseudogenes into the yeast genome and in fact this has been demonstrated *in vivo* with the *HIS3* transcript (222).

RNA dimerization and packaging, along with certain steps required for reverse transcription (described below), are mediated by a NAC activity initially defined using a synthetic peptide (TyA1-D) containing residues 299-402 from the C-terminus of Ty1 Gag (208). This region contains several stretches of basic residues that could mediate interactions with nucleic acids, but does not contain a zinc finger motif common to NC proteins. Part of this region is contained within a pfam domain called Retrotran_gag_2 (Pfam14223; previously referred to as UBN2), which is described in Chapter 4. Both a purified C-terminal half of Gag (amino acids 173-401; called the CTR) and TyA1-D can

act as a nuclear chaperone *in vitro* by binding RNA and can promote the annealing of tRNA[iMet] to the Ty1 RNA primer binding site (PBS) and the dimerization of Ty1 RNA (208, 223). When the CTR is truncated to amino acids 173-355 (sCTR), the protein no longer exhibits NAC activity, suggesting that sequences between amino acids 355-401 are required for NAC functions (223).

VLP maturation via Ty1 protease

Once RNAs are packaged and the VLPs are assembled, maturation of Ty1 proteins occurs via the activity of Ty1 PR. Maturation is initiated by an autocleavage event within the Gag-Pol-p199 precursor by the internal PR. This PR-specific autocleavage, which releases Pol from Gag in the context of Gag-Pol-p199 and occurs after histidine 401 (cleavage site is RAH_{401}-NVS), is required for any further PR-mediated cleavage (224). While the mechanism of PR-specific autocleavage of Gag-Pol has not been studied with Ty1, the mechanism has been studied with HIV PR. The active site of HIV PR is formed from two PR molecules assembled as a dimer and involves two aspartic acid residues, one donated from each PR monomer. Similarly, the initial autocleavage event requires dimerization of two Gag-Pol polyproteins (225, 226). Because an alignment of Ty1 PR and HIV PR reveal that each PR contains a single aspartic acid in the active site domain, it can be assumed that Ty1 PR also functions as a dimer (43). After autocleavage, PR mediates the cleavage at three other sites in no particular order: the C-terminus of Gag-p49 (identical to Gag-PR site in Gag-Pol-p199), the PR-IN site, and the IN-RT site (224, 227, 228) (Figure 1.6). These cleavage events release a small 4kD peptide from the C-terminus of Gag with no known function (leaving mature Gag-p45 in the VLP shell) and each of the mature Ty1 enzymes (Gag-p45, PR-p20, IN-p71, RT-

p63) into the interior of the VLP (229). Based on analogy with retroviruses, maturation is hypothesized to occur only after complete VLP assembly. This is supported by the observation that the diameter and T number of wild type VLPs is more similar to the structure observed from Ty1 particles generated from expression of Gag amino acids 1-440 (the size prior to cleavage by PR), rather than the smaller structures observed from Ty1 particles made from Gag 1-408 (similar in size to Gag post cleavage by PR) (201). However, I present evidence that some level of cleavage of Gag-p49 to Gag-p45 may occur prior to VLP assembly, as discussed in Chapter 4.

Reverse transcription and nuclear import of the Ty1 pre-integration complex

After VLP maturation, reverse transcription is initiated using the tRNA^{iMet} primer bound to the Ty1 PBS. The PBS is a 10 nt stretch found in Ty1 gRNA that is complementary to the 3' acceptor stem of the tRNA^{iMet} (205). Besides the PBS, tRNA^{iMet} binds two short Ty1 gRNA sequences called Box 0 and Box 1 and this interaction is required for reverse transcription *in vivo* (230, 231). Gag NAC activity is responsible for annealing of tRNA^{iMet} to the 5' end of Ty1 gRNA (208). Like with many retroviruses, reverse transcription is carried out by a complex formed between mature RT-p63 and IN-p71 (232, 233). Specifically, the N-terminus of RT interacts with the highly acidic C-terminus of IN and this interaction is required for RT activity (234). The first product of reverse transcription called the minus strand strong stop DNA is created via reverse transcription from the PBS to the 5' end of the Ty1 gRNA (Figure 1.7) (235). Next, the RNase H activity associated with RT degrades Ty1 gRNA that is complexed with minus strand strong stop DNA (236). There is a strand transfer event that relocates the minus strand strong stop DNA to the 3' end of Ty1 gRNA, where it binds to the complementary

R sequence. The entire minus strand is then reverse transcribed, followed by additional RNase H digestion of most of the Ty1 gRNA template. Two polypurine tracts (PPT) are not degraded and one bound near the 5' end of the minus strand serves as a template for the production of positive strand strong stop DNA, which includes sequences from the PPT through the entire 3' LTR. A second strand transfer event occurs, and the positive strand strong stop DNA binds to the 3' of the minus strand via interactions with R and U5 and primes the synthesis of the complete positive strand. Finally, the positive strand is used as a template to complete the 5' LTR of the double stranded Ty1 cDNA. This mechanism ensures that the flanking LTRs are restored when the Ty1 gRNA is reverse transcribed.

The newly formed Ty1 cDNA remains associated with IN and perhaps other proteins to form the pre-integration complex (PIC). The PIC has to gain access to the host nucleus in order to integrate Ty1 sequences, and the yeast nuclear membrane remains intact during mitosis. The Ty1 VLP is not likely to enter the nucleus intact, as there has been no evidence via TEM that VLPs are found in the nucleus. In addition, the size of the VLP (average of 60 nm) is large, making it difficult to traverse the nuclear pore. HIV can infect non-dividing cells and uses active import for nuclear entry of the HIV PIC, and Ty1 may use a similar strategy (237, 238). A nuclear localization signal (NLS) has been identified near the C-terminus of Ty1 IN (239-241). The IN NLS is bipartite and consists of two basic stretches separated by a long linker and likely depends on the classical nuclear import pathway via the protein importin-α (242).

Ty1 is integrated upstream of RNA polymerase III-transcribed genes

Integration of Ty1 cDNA into the host genome is catalyzed by the IN-p71 and this reaction has been studied extensively *in vivo* (243, 244) and *in vitro* (228, 245, 246). Conserved inverted dinucleotides at the termini of Ty1 cDNA (5′-TG…CA-3′) are required *in vivo* for Ty1 integration and retrotransposition (247). Ty1 IN likely makes a 5 bp staggered cleavage in host DNA and joins the cDNA 3' OH ends to the exposed 5' phosphate of the host DNA via a transesterification reaction. Cellular enzymes or possibly Ty1 RT fill in the 5' bp single stranded gaps flanking the newly integrated Ty1 sequences. A target site duplication of 5 bp is the signature for IN-mediated Ty1 insertions. Newly integrated Ty1 elements are found upstream of tRNA genes, or more generally genes transcribed by RNA Polymerase III (11, 248-250). High throughput sequencing of novel Ty1 insertions indicate a periodicity to the insertion pattern of Ty1 elements. Additionally, insertion sites are predicted to have high nucleosomal occupancy, and the periodicity observed can be explained by a preferential integration of Ty1 cDNA at a specific nucleosomal surface (249). Because Ty1 elements are found upstream of genes transcribed by RNA Polymerase III, it is hypothesized that integration specificity must be determined by interaction of Ty1 IN with a subunit of the RNA Pol III complex. The identification of binding partners with Ty1 IN with a demonstrable role in integration specificity has historically been challenging, but recently a RNA Poll III subunit called AC40 was shown to interact with Ty1 IN and direct Ty1 integration into these tRNA hotspots (251). When AC40 is absent, Ty1 integration still occurs, but integration lacks specificity, with a slight preference to integration at chromosomal ends. Other subunits of RNA Pol III have been implicated in Ty1

integration (252), but there is likely an abundance of important host factors that interact with Ty1 IN waiting to be characterized.

Ty1 Copy Number Control

As mentioned above, retrotransposons are held to less than 3% occupancy in *Saccharomyces* genomes, and Ty1 copy number is generally less than the 32 copies present in the reference strain (151, 154). Early studies indicated that transposition was a very rare event although Ty1 RNA levels are abundant (253). This is not attributed to the accumulation of mutations rendering the chromosomal Ty1 elements non-functional, as most of the 32 Ty1 elements in S288C are likely functional copies (11, 189). Paradoxically, Ty1 transpositional dormancy is overcome by the expression of pGTy1, which increases retrotransposition up to a 100-fold higher than the increase in Ty1 RNA levels (254). This interesting observation led to the hypothesis that there was likely a factor present in yeast responsible for control of Ty1 retrotransposition at a posttranscriptional level and that this factor could be titrated by overexpression of Ty1 (189, 255, 256).

This natural inhibition of Ty1 retrotransposition was studied further utilizing a strain of *S. paradoxus* that is "Ty1-less" due to the loss of Ty1 elements by LTR-LTR recombination (257, 258). Ty1 retrotransposition frequency can be measured using retrotranscript indicator genes (RIGs) placed within the transcribed portion of Ty1, and a common RIG used for the study of Ty1 is called *his3-AI* (Figure 1.8; see figure legend for details). When one marked copy of a Ty1 element, Ty1*his3-AI* (253), is integrated into the genome of the Ty1-less strain or present on a low copy vector, transposition is dramatically higher when compared with the reference strain (257). In fact, when the

Ty1-less *S. paradoxus* strain is repopulated with various copy numbers of Ty1 in the genome, a copy number-dependent inhibition of transposition is evident, with increasing copy numbers corresponding to decreasing levels of transposition. This phenomenon is also observed when a multi-copy 2μ Ty1 vector is introduced into the Ty1-less strain; transposition of the marked genomic Ty1 is decreased about 100-fold in the presence of the estimated 10 - 50 copies of Ty1 present due to the copy number of the 2μ plasmid in yeast. Because transposition decreases as copy number increases, this phenomenon is referred to as Copy Number Control, or CNC (257).

Since additional copies of Ty1 lead to CNC, it was hypothesized that a factor produced by the Ty1 element itself was responsible for CNC. Deletion analysis of a Ty1 element revealed that a region including part of the 5' LTR and the *GAG* ORF is required for the CNC phenotype (+238 thru +1702) (257). Initially, evidence was presented suggesting that several antisense RNAs (Ty1AS RNAs) produced from this region were responsible for CNC (259). The data supporting a role for Ty1AS RNAs in CNC included the cofractionation of Ty1AS RNAs with Ty1 VLPs and evidence that a deletion in the Ty1 5' LTR resulted in both the elimination of Ty1AS RNAs and a corresponding rescue in retrotransposition. Although these studies suggested that these transcripts were required for CNC, Ty1AS RNAs are not sufficient for CNC. Work presented in Chapter 3 indicates that the CNC region in Ty1 mapped by deletion analysis instead produces a subgenomic Ty1 RNA that encodes a truncated version of Ty1 Gag called p22 that inhibits Ty1 in a *trans*-dominant manner and fulfills all criteria to be the elusive CNC factor.

The major biochemical phenotype associated with CNC is the production of non-functional Ty1 VLPs (259). Ty1 VLPs produced in the presence of high Ty1 copy number (CNC$^+$ VLPs) have reduced levels of mature RT-p63 and PR-p20, and very low levels of mature IN-p71. As IN-p71 is required for both reverse transcription and integration as described above, the lack of mature IN-p71 is responsible for the defect in Ty1 retrotransposition under CNC. cDNA is not detectable in CNC$^+$ VLPs, indicating that reverse transcription is blocked. In fact, the first DNA product of reverse transcription, the minus strand strong stop DNA, is not detected in CNC$^+$ VLPs (216). In Chapter 3, the discovery of the CNC factor p22 is discussed in relation to these known CNC phenotypes and further experiments reveal the extent of the insults imparted by p22 to Ty1 replication. In Chapter 4, I explore the hypothesis that p22 inhibits Ty1 primarily based on an interaction with Ty1 Gag.

References

1.   Ravindran S. Barbara McClintock and the discovery of jumping genes. Proc Natl Acad Sci U S A 2012;109(50):20198-20199.

2.   Finnegan DJ. Eukaryotic transposable elements and genome evolution. Trends Genet 1989;5(4):103-107.

3.   Craig NL, Craigie R, Gellert M, Lambowitz AM. Mobile DNA II; 2002.

4.   Goodier J, Kazazian H. Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell 2008;135(1):23-35.

5.   Kazazian H. Mobile elements: drivers of genome evolution. Science 2004;303(5664):1626-1632.

6.   Boeke JD. The unusual phylogenetic distribution of retrotransposons: a hypothesis. Genome Res 2003;13(9):1975-1983.

7.   Lynch M, Conery J. The origins of genome complexity. Science 2003;302:1401-1404.

8.   Chenais B, Caruso A, Hiard S, Casse N. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. Gene 2012;509(1):7-15.

9.   Sun C, Shepard DB, Chong RA, Lopez Arriaza J, Hall K, Castoe TA, Feschotte C, Pollock DD, Mueller RL. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. Genome Biol Evol 2012;4(2):168-183.

10.   Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, Deragon J-M, Westerman RP, Sanmiguel PJ, Bennetzen JL. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genet 2009;5(11):e1000732.

11.   Kim J, Vanguri S, Boeke J, Gabriel A, Voytas D. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. Genome Res 1998;8(5):464-478.

12.   Carr M, Bensasson D, Bergman CM. Evolutionary genomics of transposable elements in Saccharomyces cerevisiae. PloS one 2012;7(11):e50978.

13.   Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nat Rev Genet 2009;10(10):691-703.

14.   Capy P, Bazin C, Higuet D, Langin T. Special issue: Evolution and impact of transposable elements - Preface. Genetica 1997;100(1-3):1-2.

15.   Beauregard A, Curcio MJ, Belfort M. The take and give between retrotransposable elements and their hosts. Annu Rev Genet 2008;42:587-617.

16.   Han JS. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. Mob DNA 2010;1(1):15.

17.   Swergold GD. Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol Cell Biol 1990;10(12):6718-6729.

18.   Hohjoh H, Singer MF. Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. EMBO J 1997;16(19):6034-6043.

19.   Kolosha VO, Martin SL. High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). J Biol Chem 2003;278(10):8112-8117.

20.   Martin SL, Bushman FD. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. Mol Cell Biol 2001;21(2):467-475.

21.   Holmes SE, Singer MF, Swergold GD. Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. J Biol Chem 1992;267(28):19765-19768.

22.   Ergun S, Buschmann C, Heukeshoven J, Dammann K, Schnieders F, Lauke H, Chalajour F, Kilic N, Stratling WH, Schumann GG. Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. J Biol Chem 2004;279(26):27753-27763.

23.   Doucet AJ, Hulme AE, Sahinovic E, Kulpa DA, Moldovan JB, Kopera HC, Athanikar JN, Hasnaoui M, Bucheton A, Moran JV, Gilbert N. Characterization of LINE-1 ribonucleoprotein particles. PLoS Genet 2010;6(10).

24.   Feng Q, Moran JV, Kazazian HH, Jr., Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 1996;87(5):905-916.

25.   Mathias SL, Scott AF, Kazazian HH, Jr., Boeke JD, Gabriel A. Reverse transcriptase encoded by a human transposable element. Science 1991;254(5039):1808-1810.

26.   Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH, Jr. High frequency retrotransposition in cultured mammalian cells. Cell 1996;87(5):917-927.

27.   Matlik K, Redik K, Speek M. L1 antisense promoter drives tissue-specific transcription of human genes. J Biomed Biotechnol 2006;2006(1):71753.

28.   Denli AM, Narvaiza I, Kerman BE, Pena M, Benner C, Marchetto MC, Diedrich JK, Aslanian A, Ma J, Moresco JJ, Moore L, Hunter T, Saghatelian A, Gage FH. Primate-specific ORF0 contributes to retrotransposon-mediated diversity. Cell 2015;163(3):583-593.

29.   Dmitriev SE, Andreev DE, Terenin IM, Olovnikov IA, Prassolov VS, Merrick WC, Shatsky IN. Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5' untranslated region of the human retrotransposon LINE-1 mRNA is strictly cap dependent rather than internal ribosome entry site mediated. Mol Cell Biol 2007;27(13):4685-4697.

30. Alisch RS, Garcia-Perez JL, Muotri AR, Gage FH, Moran JV. Unconventional translation of mammalian LINE-1 retrotransposons. Genes Dev 2006;20(2):210-224.

31. Kulpa DA, Moran JV. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. Nat Struct Mol Biol 2006;13(7):655-660.

32. Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD, Moran JV. Human L1 retrotransposition: cis preference versus trans complementation. Mol Cell Biol 2001;21(4):1429-1439.

33. Goodier JL, Zhang L, Vetter MR, Kazazian HH, Jr. LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. Mol Cell Biol 2007;27(18):6469-6483.

34. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 1993;72(4):595-605.

35. Kubo S, Seleme MC, Soifer HS, Perez JL, Moran JV, Kazazian HH, Jr., Kasahara N. L1 retrotransposition in nondividing and primary human somatic cells. Proc Natl Acad Sci U S A 2006;103(21):8036-8041.

36. Kinsey JA. Tad, a LINE-like transposable element of Neurospora, can transpose between nuclei in heterokaryons. Genetics 1990;126(2):317-323.

37. Cost GJ, Boeke JD. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. Biochemistry 1998;37(51):18081-18093.

38.   Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J*, et al.* Initial sequencing and analysis of the human genome. Nature 2001;409(6822):860-921.

39.   Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH, Jr. Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci U S A 2003;100(9):5280-5285.

40.   Sassaman DM, Dombroski BA, Moran JV, Kimberland ML, Naas TP, DeBerardinis RJ, Gabriel A, Swergold GD, Kazazian HH, Jr. Many human L1 elements are capable of retrotransposition. Nat Genet 1997;16(1):37-43.

41.   Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J*, et al.* Initial sequencing and analysis of the human genome. Nature 2001;409(6822):860-921.

42.   Havecker ER, Gao X, Voytas DF. The diversity of LTR retrotransposons. Genome Biol 2004;5(6):225.

43.   Peterson-Burch BD, Voytas DF. Genes of the Pseudoviridae (Ty1/copia retrotransposons). Mol Biol Evol 2002;19(11):1832-1845.

44.   Boeke JD, Stoye JP. Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements. In: Coffin JM, Hughes SH, Varmus HE, editors. Retroviruses. Cold Spring Harbor (NY); 1997.

45.   Wildschutte JH, Williams ZH, Montesion M, Subramanian RP, Kidd JM, Coffin JM. Discovery of unfixed endogenous retrovirus insertions in diverse human populations. Proc Natl Acad Sci U S A 2016.

46.   Barbulescu M, Turner G, Seaman MI, Deinard AS, Kidd KK, Lenz J. Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans. Curr Biol 1999;9(16):861-868.

47.   Belshaw R, Dawson AL, Woolven-Allen J, Redding J, Burt A, Tristem M. Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity. J Virol 2005;79(19):12507-12514.

48.   Hughes JF, Coffin JM. Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution. Proc Natl Acad Sci U S A 2004;101(6):1668-1672.

49.   Stocking C, Kozak CA. Murine endogenous retroviruses. Cell Mol Life Sci 2008;65(21):3383-3398.

50.   Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL. Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS Genet 2006;2(1):e2.

51.   Szak ST, Pickeral OK, Makalowski W, Boguski MS, Landsman D, Boeke JD. Molecular archeology of L1 insertions in the human genome. Genome Biol 2002;3(10):research0052.

52.   Richardson SR, Doucet AJ, Kopera HC, Moldovan JB, Garcia-Perez JL, Moran JV. The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. Microbiol Spectr 2015;3(2):MDNA3-0061-2014.

53.   Hancks DC, Kazazian HH, Jr. Active human retrotransposons: variation and disease. Curr Opin Genet Dev 2012;22(3):191-203.

54.   Gaisne M, Becam AM, Verdiere J, Herbert CJ. A 'natural' mutation in Saccharomyces cerevisiae strains derived from S288c affects the complex regulatory gene HAP1 (CYP1). Curr Genet 1999;36(4):195-200.

55.   Chaleff DT, Fink GR. Genetic events associated with an insertion mutation in yeast. Cell 1980;21(1):227-237.

56.   Silverman SJ, Fink GR. Effects of Ty insertions on HIS4 transcription in Saccharomyces cerevisiae. Mol Cell Biol 1984;4(7):1246-1251.

57.   Prak ET, Kazazian HH, Jr. Mobile elements and the human genome. Nat Rev Genet 2000;1(2):134-144.

58.   Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, Rosenzweig F, Botstein D. Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 2002;99(25):16144-16149.

59.   Garfinkel D. Genome evolution mediated by Ty elements in Saccharomyces. Cytogenet Genome Res 2005;110(1-4):63-69.

60.   Roeder GS, Farabaugh PJ, Chaleff DT, Fink GR. The origins of gene instability in yeast. Science 1980;209(4463):1375-1380.

61.   Dunham M, Badrane H, Ferea T, Adams J, Brown P, Rosenzweig F, Botstein D. Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. PNAS 2002;99(25):16144-16149.

62.   Chan JE, Kolodner RD. A genetic and structural study of genome rearrangements mediated by high copy repeat Ty1 elements. PLoS Genet 2011;7(5):e1002089.

63.   Scheifele LZ, Cost GJ, Zupancic ML, Caputo EM, Boeke JD. Retrotransposon overdose and genome integrity. Proc Natl Acad Sci U S A 2009;106(33):13927-13932.

64.   Kaneko H, Dridi S, Tarallo V, Gelfand BD, Fowler BJ, Cho WG, Kleinman ME, Ponicsan SL, Hauswirth WW, Chiodo VA, Karikó K, Yoo JW, Lee D-K, Hadziahmetovic M, Song Y, *et al.* DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. Nature 2011.

65.   Kobayashi S, Goto-Yamamoto N, Hirochika H. Retrotransposon-induced mutations in grape skin color. Science 2004;304(5673):982.

66.   Clark LA, Wahl JM, Rees CA, Murphy KE. Retrotransposon insertion in SILV is responsible for merle patterning of the domestic dog. Proc Natl Acad Sci U S A 2006;103(5):1376-1381.

67.   Duhl DM, Vrieling H, Miller KA, Wolff GL, Barsh GS. Neomorphic agouti mutations in obese yellow mice. Nat Genet 1994;8(1):59-65.

68.   Morgan HD, Sutherland HG, Martin DI, Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. Nat Genet 1999;23(3):314-318.

69.   Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV, Gage FH. Somatic

mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature

2005;435(7044):903-910.

70.   Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, Morell M,

O'Shea KS, Moran JV, Gage FH. L1 retrotransposition in human neural progenitor cells.

Nature 2009;460(7259):1127-1131.

71.   Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F,

Brennan PM, Rizzu P, Smith S, Fell M, Talbot RT, Gustincich S, Freeman TC, Mattick

JS, Hume DA*, et al.* Somatic retrotransposition alters the genetic landscape of the

human brain. Nature 2011;479(7374):534-537.

72.   Bundo M, Toyoshima M, Okada Y, Akamatsu W, Ueda J, Nemoto-Miyauchi T,

Sunaga F, Toritsuka M, Ikawa D, Kakita A, Kato M, Kasai K, Kishimoto T, Nawa H,

Okano H*, et al.* Increased l1 retrotransposition in the neuronal genome in schizophrenia.

Neuron 2014;81(2):306-313.

73.   Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY,

Edouard P, Howes S, Keith JC, Jr., McCoy JM. Syncytin is a captive retroviral envelope

protein involved in human placental morphogenesis. Nature 2000;403(6771):785-789.

74.   Lynch VJ, Nnamani MC, Kapusta A, Brayer K, Plaza SL, Mazur EC, Emera D,

Sheikh SZ, Grutzner F, Bauersachs S, Graf A, Young SL, Lieb JD, DeMayo FJ,

Feschotte C*, et al.* Ancient transposable elements transformed the uterine regulatory

landscape and transcriptome during the evolution of mammalian pregnancy. Cell Rep

2015;10(4):551-561.

75.   Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, Martin L, Ware CB, Blish CA, Chang HY, Pera RA, Wysocka J. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature 2015;522(7555):221-225.

76.   Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science 2016;351(6277):1083-1087.

77.   Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, *et al.* The genome sequence of Drosophila melanogaster. Science 2000;287(5461):2185-2195.

78.   Sahara K, Marec F, Traut W. TTAGG telomeric repeats in chromosomes of some insects and other arthropods. Chromosome Res 1999;7(6):449-460.

79.   Levis RW, Ganesan R, Houtchens K, Tolar LA, Sheen FM. Transposons in place of telomeric repeats at a Drosophila telomere. Cell 1993;75(6):1083-1093.

80.   Mura M, Murcia P, Caporale M, Spencer TE, Nagashima K, Rein A, Palmarini M. Late viral interference induced by transdominant Gag of an endogenous retrovirus. Proc Natl Acad Sci U S A 2004;101(30):11117-11122.

81.   Benit L, De Parseval N, Casella JF, Callebaut I, Cordonnier A, Heidmann T. Cloning of a new murine endogenous retrovirus, MuERV-L, with strong similarity to the human HERV-L element and with a gag coding sequence closely related to the Fv1 restriction gene. J Virol 1997;71(7):5652-5657.

82.   Kozak CA, Chakraborti A. Single amino acid changes in the murine leukemia virus capsid protein gene define the target of Fv1 resistance. Virology 1996;225(2):300-305.

83.   Stevens A, Bock M, Ellis S, LeTissier P, Bishop KN, Yap MW, Taylor W, Stoye
JP. Retroviral capsid determinants of Fv1 NB and NR tropism. J Virol
2004;78(18):9592-9598.

84.   Jung YT, Kozak CA. A single amino acid change in the murine leukemia virus
capsid gene responsible for the Fv1(nr) phenotype. J Virol 2000;74(11):5385-5387.

85.   Armezzani A, Arnaud F, Caporale M, di Meo G, Iannuzzi L, Murgia C, Palmarini
M. The signal peptide of a recently integrated endogenous sheep betaretrovirus
envelope plays a major role in eluding gag-mediated late restriction. J Virol
2011;85(14):7118-7128.

86.   Ito J, Watanabe S, Hiratsuka T, Kuse K, Odahara Y, Ochi H, Kawamura M,
Nishigaki K. Refrex-1, a soluble restriction factor against feline endogenous and
exogenous retroviruses. J Virol 2013;87(22):12029-12040.

87.   Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of
intragenomic parasites. Trends Genet 1997;13(8):335-340.

88.   Rowe HM, Trono D. Dynamic control of endogenous retroviruses during
development. Virology 2011;411(2):273-287.

89.   Selker EU, Tountas NA, Cross SH, Margolin BS, Murphy JG, Bird AP, Freitag M.
The methylated component of the Neurospora crassa genome. Nature
2003;422(6934):893-897.

90.   Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana
M, Lorincz MC, Shinkai Y. Proviral silencing in embryonic stem cells requires the
histone methyltransferase ESET. Nature 2010;464(7290):927-931.

91.   Hutnick LK, Huang X, Loo TC, Ma Z, Fan G. Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism. J Biol Chem 2010;285(27):21082-21091.

92.   Dong KB, Maksakova IA, Mohn F, Leung D, Appanah R, Lee S, Yang HW, Lam LL, Mager DL, Schubeler D, Tachibana M, Shinkai Y, Lorincz MC. DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. EMBO J 2008;27(20):2691-2701.

93.   Tamaru H, Selker EU. A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. Nature 2001;414(6861):277-283.

94.   Jackson JP, Lindroth AM, Cao X, Jacobsen SE. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature 2002;416(6880):556-560.

95.   Varshney D, Vavrova-Anderson J, Oler AJ, Cowling VH, Cairns BR, White RJ. SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. Nat Commun 2015;6:6569.

96.   Varshney D, Vavrova-Anderson J, Oler AJ, Cairns BR, White RJ. Selective repression of SINE transcription by RNA polymerase III. Mob Genet Elements 2015;5(6):86-91.

97.   Sienski G, Donertas D, Brennecke J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell 2012;151(5):964-980.

98.   Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Toth KF. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev 2013;27(4):390-399.

99.   Capuano F, Mulleder M, Kok R, Blom HJ, Ralser M. Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. Anal Chem 2014;86(8):3697-3702.

100.   Proffitt JH, Davie JR, Swinton D, Hattman S. 5-Methylcytosine is not detectable in Saccharomyces cerevisiae DNA. Mol Cell Biol 1984;4(5):985-988.

101.   Errede B, Company M, Hutchison CA, 3rd. Ty1 sequence with enhancer and mating-type-dependent regulatory activities. Mol Cell Biol 1987;7(1):258-265.

102.   Curcio M, Hedge A, Boeke J, Garfinkel D. Ty RNA levels determine the spectrum of retrotransposition events that activate gene expression in Saccharomyces cerevisiae. Mol Gen Genet 1990;220(2):213-221.

103.   Elder R, St John T, Stinchcomb D, Scherer S, Davis R. Studies on the transposable element Ty1 of yeast. I. RNA homologous to Ty1. II. Recombination and expression of Ty1 and adjacent sequences. Cold Spring Harb Symp Quant Biol 1981;45 Pt 2:581-591.

104.   Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 2001;409(6818):363-366.

105.   Hammond SM, Bernstein E, Beach D, Hannon GJ. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature 2000;404(6775):293-296.

106.   Tuschl T, Zamore PD, Lehmann R, Bartel DP, Sharp PA. Targeted mRNA degradation by double-stranded RNA in vitro. Genes Dev 1999;13(24):3191-3197.

107.   Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B, Gaasterland T, Meyer J, Tuschl T. The small RNA profile during Drosophila melanogaster development. Dev Cell 2003;5(2):337-350.

108.   Llave C, Kasschau KD, Rector MA, Carrington JC. Endogenous and silencing-associated small RNAs in plants. Plant Cell 2002;14(7):1605-1619.

109.   Reinhart BJ, Bartel DP. Small RNAs correspond to centromere heterochromatic repeats. Science 2002;297(5588):1831.

110.   Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, Bartel DP. RNAi in budding yeast. Science 2009;326(5952):544-550.

111.   Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr Biol 2001;11(13):1017-1027.

112.   Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 2007;128(6):1089-1103.

113.   Aravin AA, Hannon GJ, Brennecke J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 2007;318(5851):761-764.

114.   Klenov MS, Lavrov SA, Korbut AP, Stolyarenko AD, Yakushev EY, Reuter M, Pillai RS, Gvozdev VA. Impact of nuclear Piwi elimination on chromatin state in Drosophila melanogaster ovaries. Nucleic Acids Res 2014;42(10):6208-6218.

115.   Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 2006;442(7099):199-202.

116.   Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 2002;418(6898):646-650.

117.   Cullen BR. Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors. J Virol 2006;80(3):1067-1076.

118.   Guo F, Cen S, Niu M, Saadatmand J, Kleiman L. Inhibition of tRNA(3)(Lys)-primed reverse transcription by human APOBEC3G during human immunodeficiency virus type 1 replication. J Virol 2006;80(23):11710-11722.

119.   Mbisa JL, Barr R, Thomas JA, Vandegraaff N, Dorweiler IJ, Svarovskaia ES, Brown WL, Mansky LM, Gorelick RJ, Harris RS, Engelman A, Pathak VK. Human immunodeficiency virus type 1 cDNAs produced in the presence of APOBEC3G exhibit defects in plus-strand DNA transfer and integration. J Virol 2007;81(13):7099-7110.

120.   Marin M, Rose KM, Kozak SL, Kabat D. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. Nat Med 2003;9(11):1398-1403.

121.   Yu X, Yu Y, Liu B, Luo K, Kong W, Mao P, Yu XF. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. Science 2003;302(5647):1056-1060.

122.   Mehle A, Strack B, Ancuta P, Zhang C, McPike M, Gabuzda D. Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. J Biol Chem 2004;279(9):7792-7798.

123.   Anwar F, Davenport MP, Ebrahimi D. Footprint of APOBEC3 on the genome of human retroelements. J Virol 2013;87(14):8195-8204.

124.   Narvaiza I, Linfesty DC, Greener BN, Hakata Y, Pintel DJ, Logue E, Landau NR, Weitzman MD. Deaminase-independent inhibition of parvoviruses by the APOBEC3A cytidine deaminase. PLoS Pathog 2009;5(5):e1000439.

125.   Suspene R, Aynaud MM, Koch S, Pasdeloup D, Labetoulle M, Gaertner B, Vartanian JP, Meyerhans A, Wain-Hobson S. Genetic editing of herpes simplex virus 1 and Epstein-Barr herpesvirus genomes by human APOBEC3 cytidine deaminases in culture and in vivo. J Virol 2011;85(15):7594-7602.

126.   Richardson SR, Narvaiza I, Planegger RA, Weitzman MD, Moran JV. APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. Elife 2014;3:e02008.

127.   Schumacher AJ, Hache G, Macduff DA, Brown WL, Harris RS. The DNA deaminase activity of human APOBEC3G is required for Ty1, MusD, and human immunodeficiency virus type 1 restriction. J Virol 2008;82(6):2652-2660.

128.   Schumacher AJ, Nissley DV, Harris RS. APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast. Proc Natl Acad Sci U S A 2005;102(28):9854-9859.

129.   Dutko JA, Schafer A, Kenny AE, Cullen BR, Curcio MJ. Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases. Curr Biol 2005;15(7):661-666.

130.   Dutko JA, Kenny AE, Gamache ER, Curcio MJ. 5' to 3' mRNA decay factors colocalize with Ty1 gag and human APOBEC3G and promote Ty1 retrotransposition. J Virol 2010;84(10):5052-5066.

131.   Owens CM, Yang PC, Gottlinger H, Sodroski J. Human and simian immunodeficiency virus capsid proteins are major viral determinants of early, postentry replication blocks in simian cells. J Virol 2003;77(1):726-731.

132.   Goujon C, Moncorge O, Bauby H, Doyle T, Ward CC, Schaller T, Hue S, Barclay WS, Schulz R, Malim MH. Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. Nature 2013;502(7472):559-562.

133.   Kane M, Yadav SS, Bitzegeio J, Kutluay SB, Zang T, Wilson SJ, Schoggins JW, Rice CM, Yamashita M, Hatziioannou T, Bieniasz PD. MX2 is an interferon-induced inhibitor of HIV-1 infection. Nature 2013;502(7472):563-566.

134.   Liu Z, Pan Q, Ding S, Qian J, Xu F, Zhou J, Cen S, Guo F, Liang C. The interferon-inducible MxB protein inhibits HIV-1 infection. Cell Host Microbe 2013;14(4):398-410.

135.   Pincus T, Hartley JW, Rowe WP. A major genetic locus affecting resistance to infection with murine leukemia viruses. I. Tissue culture studies of naturally occurring viruses. The Journal of experimental medicine 1971;133(6):1219-1233.

136.   Yap MW, Nisole S, Lynch C, Stoye JP. Trim5alpha protein restricts both HIV-1 and murine leukemia virus. Proc Natl Acad Sci U S A 2004;101(29):10786-10791.

137.   Ohkura S, Stoye JP. A comparison of murine leukemia viruses that escape from human and rhesus macaque TRIM5alphas. J Virol 2013;87(11):6455-6468.

138.   Ohkura S, Goldstone DC, Yap MW, Holden-Dye K, Taylor IA, Stoye JP. Novel escape mutants suggest an extensive TRIM5alpha binding site spanning the entire outer surface of the murine leukemia virus capsid protein. PLoS Pathog 2011;7(3):e1002011.

139.   Mortuza GB, Dodding MP, Goldstone DC, Haire LF, Stoye JP, Taylor IA. Structure of B-MLV capsid amino-terminal domain reveals key features of viral tropism, gag assembly and core formation. Journal of molecular biology 2008;376(5):1493-1508.

140.   Soll SJ, Wilson SJ, Kutluay SB, Hatziioannou T, Bieniasz PD. Assisted evolution enables HIV-1 to overcome a high TRIM5alpha-imposed genetic barrier to rhesus macaque tropism. PLoS Pathog 2013;9(9):e1003667.

141.   Lindenmann J. Resistance of mice to mouse-adapted influenza A virus. Virology 1962;16:203-204.

142.   Fricke T, White TE, Schulte B, de Souza Aranha Vieira DA, Dharan A, Campbell EM, Brandariz-Nunez A, Diaz-Griffero F. MxB binds to the HIV-1 core and prevents the uncoating process of HIV-1. Retrovirology 2014;11:68.

143.   Boeke J, Garfinkel D, Styles C, Fink G. Ty elements transpose through an RNA intermediate. Cell 1985;40(3):491-500.

144.   Voytas DF, Craig NL. Ty1 and Ty5 of Saccharomyces cerevisiae. Mobile DNA II; 2002. p. 614-630.

145.   Garfinkel DJ, Boeke JD, Fink GR. Ty element transposition: reverse transcriptase and virus-like particles. Cell 1985;42(2):507-517.

146.   Mortimer RK, Johnston JR. Genealogy of principal strains of the yeast genetic stock center. Genetics 1986;113(1):35-43.

147.   Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, *et al.* Life with 6000 genes. Science 1996;274(5287):546, 563-547.

148.   Lesage P, Todeschini A. Happy together: the life and times of Ty retrotransposons and their hosts. Cytogenet Genome Res 2005;110(1-4):70-90.

149.   Sandmeyer SB, Aye M, Menees T, Craig N. Ty3, a position specific, gypsy-like element in Saccharomyces cerevisiae. Mobile DNA II; 2002. p. 663-683.

150.   Voytas DFB, J.D. Ty1 and Ty5 of Saccharomyces cerevisiae. In: Craig NLC, R.; Gellert, M.; Lambowitz, A.M. , editor. Mobile DNA II. Washington DC: ASM Press; 1991. p. 631-662.

151.   Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, Tsai IJ, Bergman CM, Bensasson D, O'kelly MJT, Van Oudenaarden A, *et al.* Population genomics of domestic and wild yeasts. Nature 2009;458(7236):337-341.

152.   Hug AM, Feldmann H. Yeast retrotransposon Ty4: the majority of the rare transcripts lack a U3-R sequence. Nucleic Acids Res 1996;24(12):2338-2346.

153.   Wilke C, Maimer E, Adams J. The population biology and evolutionary significance of Ty elements in Saccharomyces cerevisiae. Genetica 1992;86(1-3):155-173.

154.   Bleykasten-Grosshans C, Friedrich A, Schacherer J. Genome-wide analysis of intraspecific transposon diversity in yeast. BMC Genomics 2013;14:399.

155.   Elder RT, Loh EY, Davis RW. RNA from the yeast transposable element Ty1 has both ends in the direct repeats, a structure similar to retrovirus RNA. Proc Natl Acad Sci U S A 1983;80(9):2432-2436.

156.   Mules EH, Uzun O, Gabriel A. In vivo Ty1 reverse transcription can generate replication intermediates with untidy ends. J Virol 1998;72(8):6490-6503.

157.   Hou W, Russnak R, Platt T. Poly(A) site selection in the yeast Ty retroelement requires an upstream region and sequence-specific titratable factor(s) in vitro. EMBO J 1994;13(2):446-452.

158.   Dudley AM, Gansheroff LJ, Winston F. Specific components of the SAGA complex are required for Gcn4- and Gcr1-mediated activation of the his4-912delta promoter in Saccharomyces cerevisiae. Genetics 1999;151(4):1365-1378.

159.   Fulton AM, Rathjen PD, Kingsman SM, Kingsman AJ. Upstream and downstream transcriptional control signals in the yeast retrotransposon, TY. Nucleic Acids Res 1988;16(12):5439-5458.

160.   Servant G, Pennetier C, Lesage P. Remodeling yeast gene transcription by activating the Ty1 long terminal repeat retrotransposon under severe adenine deficiency. Mol Cell Biol 2008;28(17):5543-5554.

161.   Yu K, Elder RT. A region internal to the coding sequences is essential for transcription of the yeast Ty-D15 element. Mol Cell Biol 1989;9(9):3667-3678.

162.   Curcio MJ, Lutz S, Lesage P. The Ty1 LTR-Retrotransposon of Budding Yeast, Saccharomyces cerevisiae. Microbiol Spectr 2015;3(2):MDNA3-0053-2014.

163.	Laloux I, Dubois E, Dewerchin M, Jacobs E. TEC1, a gene involved in the activation of Ty1 and Ty1-mediated gene expression in Saccharomyces cerevisiae: cloning and molecular analysis. Mol Cell Biol 1990;10(7):3541-3550.

164.	Morillon A, Springer M, Lesage P. Activation of the Kss1 invasive-filamentous growth pathway induces Ty1 transcription and retrotransposition in Saccharomyces cerevisiae. Mol Cell Biol 2000;20(15):5766-5776.

165.	Servant G, Pinson B, Tchalikian-Cosson A, Coulpier F, Lemoine S, Pennetier C, Bridier-Nahmias A, Todeschini AL, Fayol H, Daignan-Fornier B, Lesage P. Tye7 regulates yeast Ty1 retrotransposon sense and antisense transcription in response to adenylic nucleotides stress. Nucleic Acids Res 2012;40(12):5271-5282.

166.	Morillon A, Benard L, Springer M, Lesage P. Differential effects of chromatin and Gcn4 on the 50-fold range of expression among individual yeast Ty1 retrotransposons. Mol Cell Biol 2002;22(7):2078-2088.

167.	Pollard KJ, Peterson CL. Role for ADA/GCN5 products in antagonizing chromatin-mediated transcriptional repression. Mol Cell Biol 1997;17(11):6212-6222.

168.	Winston F, Durbin KJ, Fink GR. The SPT3 gene is required for normal transcription of Ty elements in S. cerevisiae. Cell 1984;39(3 Pt 2):675-682.

169.	Winston F, Dollard C, Malone EA, Clare J, Kapakos JG, Farabaugh P, Minehart PL. Three genes are required for trans-activation of Ty transcription in yeast. Genetics 1987;115(4):649-656.

170.	Boeke JD, Styles CA, Fink GR. Saccharomyces cerevisiae SPT3 gene is required for transposition and transpositional recombination of chromosomal Ty elements. Mol Cell Biol 1986;6(11):3575-3581.

171.   Belotserkovskaya R, Sterner DE, Deng M, Sayre MH, Lieberman PM, Berger SL. Inhibition of TATA-binding protein function by SAGA subunits Spt3 and Spt8 at Gcn4-activated promoters. Mol Cell Biol 2000;20(2):634-647.

172.   Eisenmann DM, Arndt KM, Ricupero SL, Rooney JW, Winston F. SPT3 interacts with TFIID to allow normal transcription in Saccharomyces cerevisiae. Genes Dev 1992;6(7):1319-1331.

173.   Mohibullah N, Hahn S. Site-specific cross-linking of TBP in vivo and in vitro reveals a direct functional interaction with the SAGA subunit Spt3. Genes Dev 2008;22(21):2994-3006.

174.   Yu K, Elder RT. Some of the signals for 3'-end formation in transcription of the Saccharomyces cerevisiae Ty-D15 element are immediately downstream of the initiation site. Mol Cell Biol 1989;9(6):2431-2444.

175.   Malagon F, Jensen T. The T-body: a new cytoplasmic RNA granule in Saccharomyces cerevisiae. Mol Cell Biol 2008.

176.   Segref A, Sharma K, Doye V, Hellwig A, Huber J, Luhrmann R, Hurt E. Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)+ RNA and nuclear pores. EMBO J 1997;16(11):3256-3271.

177.   Swanson CM, Malim MH. Retrovirus RNA Trafficking: From Chromatin to Invasive Genomes. Traffic 2006;7(11):1440-1450.

178.   Garbitt-Hirst R, Kenney S, Parent L. Genetic evidence for a connection between Rous sarcoma virus gag nuclear trafficking and genomic RNA packaging. J Virol 2009;83(13):6790-6797.

179.   Gudleski N, Flanagan JM, Ryan EP, Bewley MC, Parent LJ. Directionality of nucleocytoplasmic transport of the retroviral gag protein depends on sequential binding of karyopherins and viral RNA. Proc Natl Acad Sci USA 2010;107(20):9358-9363.

180.   Parent LJ. New insights into the nuclear localization of retroviral Gag proteins. Nucleus 2011;2(2):92-97.

181.   Scheifele LZ, Kenney SP, Cairns TM, Craven RC, Parent LJ. Overlapping roles of the Rous sarcoma virus Gag p10 domain in nuclear export and virion core morphology. J Virol 2007;81(19):10718-10728.

182.   Scheifele LZ, Garbitt RA, Rhoads JD, Parent LJ. Nuclear entry and CRM1-dependent nuclear export of the Rous sarcoma virus Gag polyprotein. Proc Natl Acad Sci U S A 2002;99(6):3944-3949.

183.   Doh JH, Lutz S, Curcio MJ. Co-translational localization of an LTR-retrotransposon RNA to the endoplasmic reticulum nucleates virus-like particle assembly sites. PLoS Genet 2014;10(3):e1004219.

184.   Belcourt M, Farabaugh P. Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. Cell 1990;62(2):339-352.

185.   Sundararajan A, Michaud WA, Qian Q, Stahl G, Farabaugh PJ. Near-cognate peptidyl-tRNAs promote +1 programmed translational frameshifting in yeast. Mol Cell 1999;4(6):1005-1015.

186.   Kawakami K, Pande S, Faiola B, Moore D, Boeke J, Farabaugh P, Strathern J, Nakamura Y, Garfinkel D. A rare tRNA-Arg(CCU) that regulates Ty1 element ribosomal frameshifting is essential for Ty1 retrotransposition in Saccharomyces cerevisiae. Genetics 1993;135(2):309-320.

187.   Xu H, Boeke J. Host genes that influence transposition in yeast: the abundance of a rare tRNA regulates Ty1 transposition frequency. Proc Natl Acad Sci U S A 1990;87(21):8360-8364.

188.   del Alamo M, Hogan DJ, Pechmann S, Albanese V, Brown PO, Frydman J. Defining the specificity of cotranslationally acting chaperones by systematic analysis of mRNAs associated with ribosome-nascent chain complexes. PLoS Biol 2011;9(7):e1001100.

189.   Curcio M, Garfinkel D. Heterogeneous functional Ty1 elements are abundant in the Saccharomyces cerevisiae genome. Genetics 1994;136(4):1245-1259.

190.   Boeke J, Styles C, Fink G. Saccharomyces cerevisiae SPT3 gene is required for transposition and transpositional recombination of chromosomal Ty elements. Mol Cell Biol 1986;6(11):3575-3581.

191.   Paquin CE, Williamson VM. Temperature effects on the rate of ty transposition. Science 1984;226(4670):53-55.

192.   Lawler J, Haeusser D, Dull A, Boeke J, Keeney J. Ty1 defect in proteolysis at high temperature. J Virol 2002;76(9):4233-4240.

193.   Dutko JA, Kenny AE, Gamache ER, Curcio MJ. 5' to 3' mRNA decay factors colocalize with Ty1 gag and human APOBEC3G and promote Ty1 retrotransposition. Journal of Virology 2010;84(10):5052-5066.

194.   Checkley MA, Nagashima K, Lockett SJ, Nyswaner KM, Garfinkel DJ. P-body components are required for Ty1 retrotransposition during assembly of retrotransposition-competent virus-like particles. Molecular and Cellular Biology 2010;30(2):382-398.

195.   Berretta J, Morillon A. Pervasive transcription constitutes a new level of eukaryotic genome regulation. EMBO Rep 2009;10(9):973-982.

196.   Luschnig C, Hess M, Pusch O, Brookman J, Bachmair A. The gag homologue of retrotransposon Ty1 assembles into spherical particles in Escherichia coli. Eur J Biochem 1995;228(3):739-744.

197.   Burns N, Saibil H, White N, Pardon J, Timmins P, Richardson S, Richards B, Adams S, Kingsman S, Kingsman A. Symmetry, flexibility and permeability in the structure of yeast retrotransposon virus-like particles. Embo J 1992;11(3):1155-1164.

198.   Palmer K, Tichelaar W, Myers N, Burns N, Butcher S, Kingsman A, Fuller S, Saibil H. Cryo-electron microscopy structure of yeast Ty retrotransposon virus- like particles. J Virol 1997;71(9):6863-6868.

199.   Luschnig C, Bachmair A. RNA packaging of yeast retrotransposon Ty1 in the heterologous host, Escherichia coli. Biol Chem 1997;378(1):39-46.

200.   Uhde-Holzem K, Fischer R, Commandeur U. Characterization and diagnostic potential of foreign epitope-presenting Ty1 virus-like particles expressed in Escherichia coli and Pichia pastoris. J Mol Microbiol Biotechnol 2010;18(1):52-62.

201.   Al-Khayat H, Bhella D, Kenney J, Roth J, Kingsman A, Martin-Rendon E, Saibil H. Yeast Ty retrotransposons assemble into virus-like particles whose T- numbers depend on the C-terminal length of the capsid protein. Journal of molecular biology 1999;292(1):65-73.

202.   Brookman J, Stott A, Cheeseman P, Burns N, Adams S, Kingsman A, Gull K. An immunological analysis of Ty1 virus-like particle structure. Virology 1995;207(1):59-67.

203.   Eichinger D, Boeke J. The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: cell-free Ty1 transposition. Cell 1988;54(7):955-966.

204.   Eichinger DJ, Boeke JD. A specific terminal structure is required for Ty1 transposition. Genes Dev 1990;4(3):324-330.

205.   Chapman K, Bystrom A, Boeke J. Initiator methionine tRNA is essential for Ty1 transposition in yeast. PNAS 1992;89(April):3236-3240.

206.   Wilhelm M, Wilhelm F, Keith G, Agoutin B, Heyman T. Yeast Ty1 retrotransposon: the minus-strand primer binding site and a cis-acting domain of the Ty1 RNA are both important for packaging of primer tRNA inside virus-like particles. Nucleic Acids Res 1994;22(22):4560-4565.

207.   Feng Y, Moore S, Garfinkel D, Rein A. The genomic RNA in Ty1 virus-like particles is dimeric. J Virol 2000;74(22):10819-10821.

208.   Cristofari G, Ficheux, D., Darlix, J.L. The Gag-like Protein of the Yeast Ty1 Retrotransposon Contains a Nucleic Acid Chaperone Domain Analogous to Retroviral Nucleocapsid Proteins. Journal of Biological Chemistry 2000;275:19210-19217.

209.   Moore MD, Fu W, Nikolaitchik O, Chen J, Ptak RG, Hu WS. Dimer initiation signal of human immunodeficiency virus type 1: its role in partner selection during RNA copackaging and its effects on recombination. J Virol 2007;81(8):4002-4011.

210.   Chin MP, Rhodes TD, Chen J, Fu W, Hu WS. Identification of a major restriction in HIV-1 intersubtype recombination. Proc Natl Acad Sci U S A 2005;102(25):9002-9007.

211.	Skripkin E, Paillart JC, Marquet R, Ehresmann B, Ehresmann C. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. Proc Natl Acad Sci U S A 1994;91(11):4945-4949.

212.	Hill MK, Shehu-Xhilaga M, Campbell SM, Poumbourios P, Crowe SM, Mak J. The dimer initiation sequence stem-loop of human immunodeficiency virus type 1 is dispensable for viral replication in peripheral blood mononuclear cells. J Virol 2003;77(15):8329-8335.

213.	McBride MS, Panganiban AT. Position dependence of functional hairpins important for human immunodeficiency virus type 1 RNA encapsidation in vivo. J Virol 1997;71(3):2050-2058.

214.	St Louis DC, Gotte D, Sanders-Buell E, Ritchey DW, Salminen MO, Carr JK, McCutchan FE. Infectious molecular clones with the nonhomologous dimer initiation sequences found in different subtypes of human immunodeficiency virus type 1 can recombine and initiate a spreading infection in vitro. J Virol 1998;72(5):3991-3998.

215.	Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). J Am Chem Soc 2005;127(12):4223-4231.

216.	Purzycka KJ, Legiewicz M, Matsuda E, Eizentstat LD, Lusvarghi S, Saha A, Le Grice SF, Garfinkel DJ. Exploring Ty1 retrotransposon RNA structure within virus-like particles. Nucleic Acids Res 2012.

217.	Chen J, Rahman SA, Nikolaitchik OA, Grunwald D, Sardo L, Burdick RC, Plisov S, Liang E, Tai S, Pathak VK, Hu WS. HIV-1 RNA genome dimerizes on the plasma

membrane in the presence of Gag protein. Proc Natl Acad Sci U S A 2016;113(2):E201-208.

218.   Xu H, Boeke J. Localization of sequences required in cis for yeast Ty1 element transposition near the long terminal repeats: analysis of mini-Ty1 elements. Mol Cell Biol 1990;10(6):2695-2702.

219.   Huang Q, Purzycka KJ, Lusvarghi S, Li D, Legrice SF, Boeke JD. Retrotransposon Ty1 RNA contains a 5'-terminal long-range pseudoknot required for efficient reverse transcription. RNA 2013;19(3):320-332.

220.   Pochart P, Agoutin B, Fix C, Keith G, Heyman T. A very poorly expressed tRNA(Ser) is highly concentrated together with replication primer initiator tRNA(Met) in the yeast Ty1 virus-like particles. Nucleic Acids Res 1993;21(7):1517-1521.

221.   Maxwell P, Coombes C, Kenny A, Lawler J, Boeke J, Curcio M. Ty1 mobilizes subtelomeric Y' elements in telomerase-negative Saccharomyces cerevisiae survivors. Mol Cell Biol 2004;24(22):9887-9898.

222.   Derr LK, Strathern JN, Garfinkel DJ. RNA-mediated recombination in S. cerevisiae. Cell 1991;67(2):355-364.

223.   Nishida Y, Pachulska-Wieczorek K, Blaszczyk L, Saha A, Gumna J, Garfinkel DJ, Purzycka KJ. Ty1 retrovirus-like element Gag contains overlapping restriction factor and nucleic acid chaperone functions. Nucleic Acids Res 2015;43(15):7414-7431.

224.   Merkulov GV, Lawler JF, Jr., Eby Y, Boeke JD. Ty1 proteolytic cleavage sites are required for transposition: all sites are not created equal. J Virol 2001;75(2):638-644.

225. Pettit SC, Everitt LE, Choudhury S, Dunn BM, Kaplan AH. Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism. J Virol 2004;78(16):8477-8485.

226. Weber IT. Comparison of the crystal structures and intersubunit interactions of human immunodeficiency and Rous sarcoma virus proteases. J Biol Chem 1990;265(18):10492-10496.

227. Merkulov GV, Swiderek KM, Brachmann CB, Boeke JD. A critical proteolytic cleavage site near the C terminus of the yeast retrotransposon Ty1 Gag protein. J Virol 1996;70(8):5548-5556.

228. Moore SP, Garfinkel DJ. Expression and partial purification of enzymatically active recombinant Ty1 integrase in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 1994;91(5):1843-1847.

229. Lawler JF, Jr., Merkulov GV, Boeke JD. A nucleocapsid functionality contained within the amino terminus of the Ty1 protease that is distinct and separable from proteolytic activity. J Virol 2002;76(1):346-354.

230. Friant S, Heyman T, Bystrom A, Wilhelm M, Wilhelm F. Interactions between Ty1 retrotransposon RNA and the T and D regions of the tRNA(iMet) primer are required for initiation of reverse transcription in vivo. Mol Cell Biol 1998;18(2):799-806.

231. Friant S, Heyman T, Wilhelm M, Wilhelm F. Extended interactions between the primer tRNAi(Met) and genomic RNA of the yeast Ty1 retrotransposon. Nucleic Acids Res 1996;24(3):441-449.

232. Wilhelm M, Boutabout M, Wilhelm F. 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 2000;348 Pt 2:337-342.

233.   Wilhelm M, Wilhelm F. Cooperation between reverse transcriptase and integrase during reverse transcription and formation of the preintegrative complex of Ty1. Eukaryot Cell 2006;5(10):1760-1769.

234.   Wilhelm M, Wilhelm FX. Role of integrase in reverse transcription of the Saccharomyces cerevisiae retrotransposon Ty1. Eukaryot Cell 2005;4(6):1057-1065.

235.   Muller F, Laufer W, Pott U, Ciriacy M. Characterization of products of TY1-mediated reverse transcription in Saccharomyces cerevisiae. Mol Gen Genet 1991;226(1-2):145-153.

236.   Lauermann V, Boeke JD. Plus-strand strong-stop DNA transfer in yeast Ty retrotransposons. EMBO J 1997;16(21):6603-6612.

237.   Bukrinsky MI, Sharova N, Dempsey MP, Stanwick TL, Bukrinskaya AG, Haggerty S, Stevenson M. Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. Proc Natl Acad Sci U S A 1992;89(14):6580-6584.

238.   Bukrinsky MI, Stanwick TL, Dempsey MP, Stevenson M. Quiescent T lymphocytes as an inducible virus reservoir in HIV-1 infection. Science 1991;254(5030):423-427.

239.   Kenna M, Brachmann C, Devine S, Boeke J. Invading the yeast nucleus: a nuclear localization signal at the C terminus of Ty1 integrase is required for transposition in vivo. Mol Cell Biol 1998;18(2):1115-1124.

240.   McLane L, Pulliam K, Devine S, Corbett A. The Ty1 integrase protein can exploit the classical nuclear protein import machinery for entry into the nucleus. Nucleic Acids Res 2008;36(13):4317-4326.

241.   Moore S, Rinckel L, Garfinkel D. A Ty1 integrase nuclear localization signal required for retrotransposition. Mol Cell Biol 1998;18(2):1105-1114.

242.   Lange A, McLane LM, Mills RE, Devine SE, Corbett AH. Expanding the definition of the classical bipartite nuclear localization signal. Traffic 2010;11(3):311-323.

243.   Friedl AA, Kiechle M, Maxeiner HG, Schiestl RH, Eckardt-Schupp F. Ty1 integrase overexpression leads to integration of non-Ty1 DNA fragments into the genome of Saccharomyces cerevisiae. Mol Genet Genomics 2010;284(4):231-242.

244.   Sharon G, Burkett TJ, Garfinkel DJ. Efficient homologous recombination of Ty1 element cDNA when integration is blocked. Mol Cell Biol 1994;14(10):6540-6551.

245.   Braiterman LT, Boeke JD. In vitro integration of retrotransposon Ty1: a direct physical assay. Mol Cell Biol 1994;14(9):5719-5730.

246.   Moore SP, Garfinkel DJ. Correct integration of model substrates by Ty1 integrase. J Virol 2000;74(24):11522-11530.

247.   Friedl AA, Kiechle M, Maxeiner HG, Schiestl RH, Eckardt-Schupp F. Ty1 integrase overexpression leads to integration of non-Ty1 DNA fragments into the genome of Saccharomyces cerevisiae. Molecular genetics and genomics : MGG 2010.

248.   Devine S, Boeke J. Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. Genes Dev 1996;10(5):620-633.

249. Baller JA, Gao J, Stamenova R, Curcio MJ, Voytas DF. A nucleosomal surface defines an integration hotspot for the Saccharomyces cerevisiae Ty1 retrotransposon. Genome Res 2012;22(4):704-713.

250. Eigel A, Feldmann H. Ty1 and delta elements occur adjacent to several tRNA genes in yeast. EMBO J 1982;1(10):1245-1250.

251. Bridier-Nahmias A, Tchalikian-Cosson A, Baller JA, Menouni R, Fayol H, Flores A, Saib A, Werner M, Voytas DF, Lesage P. Retrotransposons. An RNA polymerase III subunit determines sites of retrotransposon integration. Science 2015;348(6234):585-588.

252. Cheung S, Ma L, Chan PH, Hu HL, Mayor T, Chen HT, Measday V. Ty1 Integrase Interacts with RNA Polymerase III-specific Subcomplexes to Promote Insertion of Ty1 Elements Upstream of Polymerase (Pol) III-transcribed Genes. J Biol Chem 2016;291(12):6396-6411.

253. Curcio M, Garfinkel D. Single-step selection for Ty1 element retrotransposition. Proc Natl Acad Sci U S A 1991;88(3):936-940.

254. Curcio M, Garfinkel D. Posttranslational control of Ty1 retrotransposition occurs at the level of protein processing. Mol Cell Biol 1992;12(6):2813-2825.

255. Farabaugh P. Post-transcriptional regulation of transposition by Ty retrotransposons of Saccharomyces cerevisiae. J Biol Chem 1995;270:10361-10364.

256. Fink G, Boeke J, Garfinkel D. The mechanism and consequences of retrotransposition. Trends in Gentics 1986(May):118-123.

257. Garfinkel D, Nyswaner K, Wang J, Cho J. Post-transcriptional cosuppression of Ty1 retrotransposition. Genetics 2003;165(1):83-99.

258.	Wilke C, Adams J. Fitness effects of Ty transposition in Saccharomyces cerevisiae. Genetics 1992;131(1):31-42.

259.	Matsuda E, Garfinkel DJ. Posttranslational interference of Ty1 retrotransposition by antisense RNAs. Proc Natl Acad Sci USA 2009;106(37):15657-15662.

260.	Garfinkel DJ, Tucker JM, Saha A, Nishida Y, Pachulska-Wieczorek K, Blaszczyk L, Purzycka KJ. A self-encoded capsid derivative restricts Ty1 retrotransposition in Saccharomyces. Curr Genet 2015.
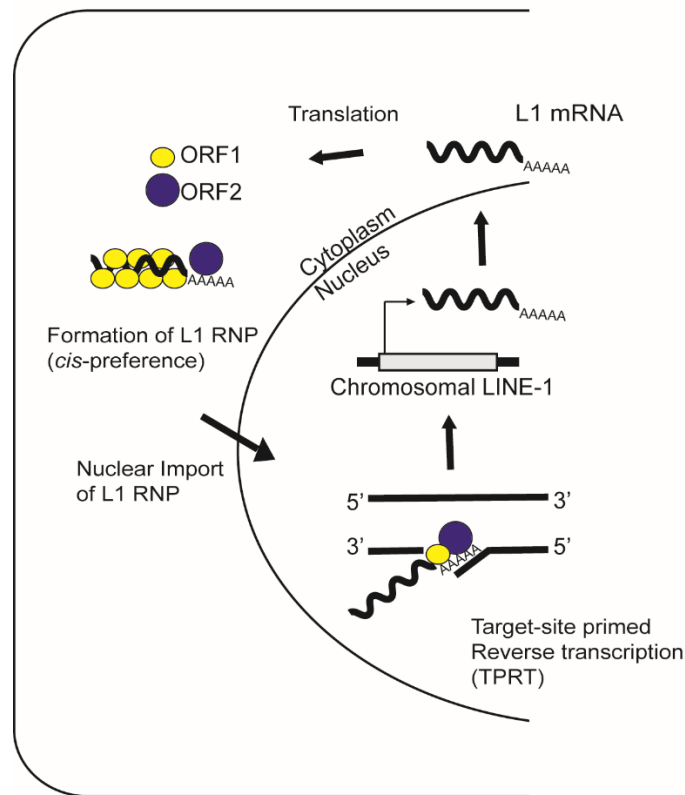
Figure 1.1 Replication of LINE-1 elements.

L1 is transcribed by RNA polymerase II and exported from the nucleus. L1 mRNA is translated to form ORF1 and ORF2 proteins, which bind to L1 mRNA based on *cis*-preference, i.e., L1 proteins bind the same mRNA used as a template for their translation. The L1 ribonucleoprotein complex (RNP) is imported into the nucleus. ORF2 catalyzes a single-strand nick in host DNA exposing a 3'-OH, which acts as a primer for reverse transcription of the L1 mRNA. This reaction is called target-site primed reverse transcription (TPRT).
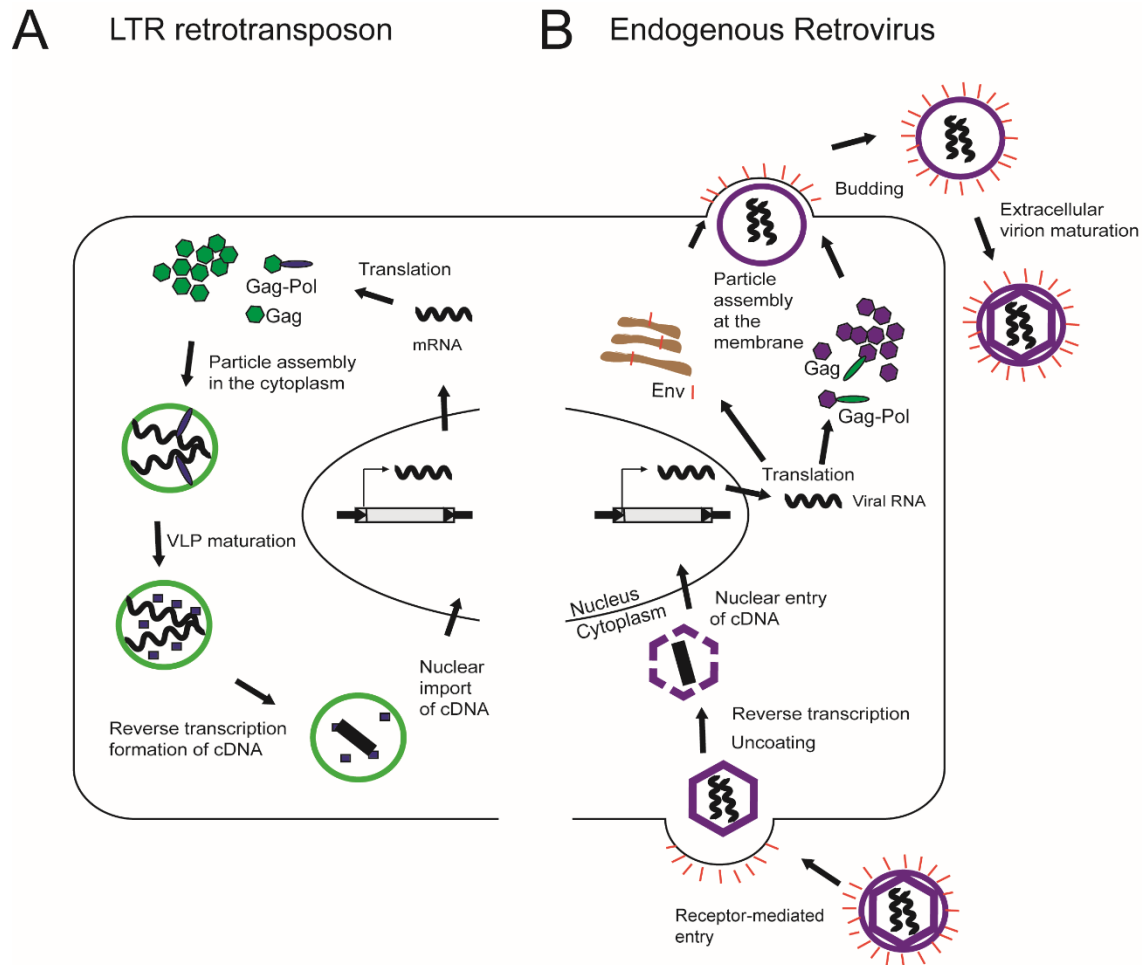
Figure 1.2 Replication of LTR retrotransposons. A general comparison of the life cycles of elements with LTRs. (A) The VLP (green circle) of the LTR retrotransposons is assembled and undergoes maturation (cleavage of Gag and Gag-Pol proteins) in the cytoplasm. The cDNA (black rectangle) formed from reverse transcription is imported into the nucleus and is integrated into the same genome from which it originated. (B). Endogenous retroviruses, which also contain LTRs, replicate similarly to their exogenous counterparts. Virions (purple circle) typically assemble at the plasma membrane, although there are some endogenous retroviruses that assemble in the cytoplasm. The Env protein traffics to the plasma membrane via the secretory pathway

71

and forms the outer membrane of the virion. Env is required for budding of the virion and maturation occurs outside of the producer cell. The virion enters a new cell based on interactions between Env and cell surface receptors, and the capsid core is released into the cytoplasm. The capsid core uncoats and reverse transcription of viral RNA occurs. The cDNA is either imported into the nucleus or gains access to host genome during disassembly of the nuclear membrane.
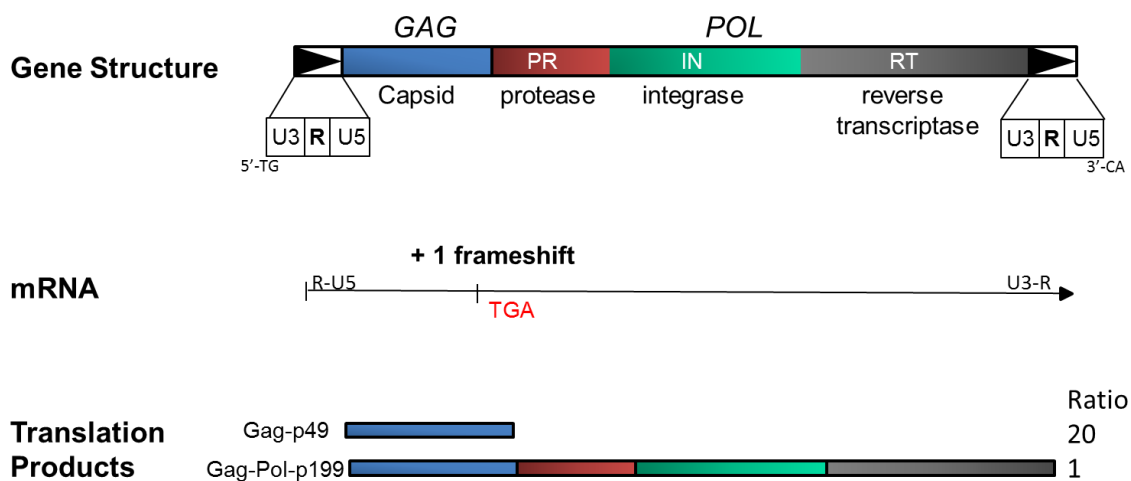
Figure 1.3   Gene structure, transcription, and translation of the Ty1 LTR retrotransposon. The LTRs of Ty1 (black arrows) are comprised of U3, R and U5 domains. Ty1 elements have an inverted dinucleotide sequence present at the extreme 5' and 3' ends. The LTRs flank two overlapping ORFs: *GAG* and *POL*. *GAG* encodes for the capsid protein called Gag. *POL* encodes for the enzymes protease (PR-p20), integrase (IN-p71), and reverse transcriptase (RT-p63). Ty1 is transcribed from the 5' R

domain to the 3' R domain, creating an mRNA with terminal R domains. Ty1 mRNA

contains a +1 ribosomal frameshifting site within the overlapping sequence of *GAG* and

POL and a stop codon directly thereafter. 95% of the time, Gag-p49 is produced from

Ty1 mRNA and the TGA stop codon is recognized. 5% of the time, a +1 ribosomal

frameshifting event occurs, resulting in the synthesis of Gag-Pol-p199. This results in a

ratio of Gag-p49 to Gag-Pol-p199 products of 20:1.
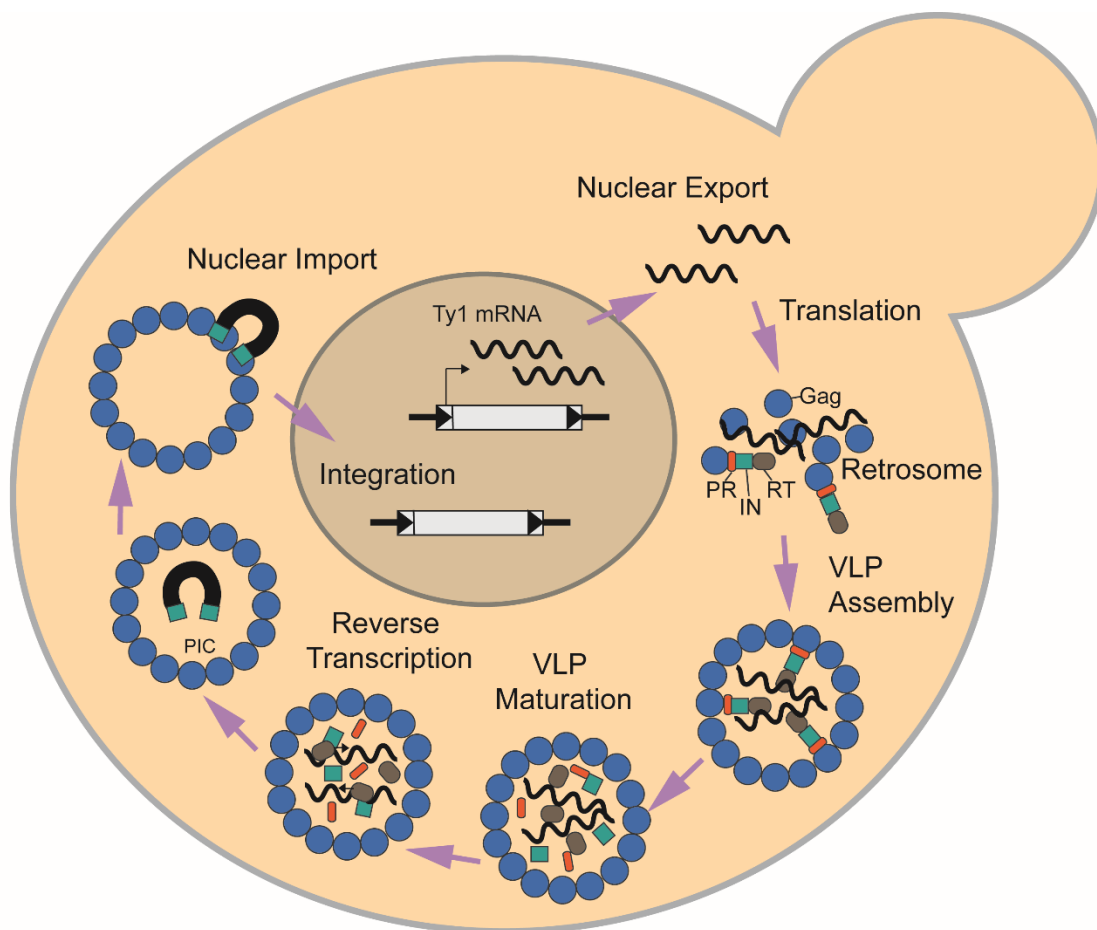
Figure 1.4 Ty1 retrotransposition. Ty1 is transcribed from by RNA polymerase II and

exported from the nucleus, where it is packaged into VLPs or translated to produce

Gag-p49 (blue circles) or Gag-Pol-p199, which requires a programmed +1 frameshift event near the end of *GAG*. Gag accumulates in the cytoplasm to become the capsid of Ty1 VLPs and also mediates nucleic acid chaperone functions during VLP assembly and reverse transcription. Prior to VLP formation, Gag and presumably Gag-Pol form mRNA/Gag foci termed retrosomes, which are sites where VLPs assemble. Ty1 mRNA is specifically packaged as a dimer into VLPs. Ty1 PR cleaves Gag-p49 and Gag-Pol-p199 precursors to form mature proteins, Gag-p45, PR-p23 (orange rectangle), IN-p71 (teal rectangle), and RT-p63 (gray oval). Like retroviruses, an IN/RT heteromer reverse transcribes Ty1 mRNA into a linear cDNA. A pre-integration complex (PIC) minimally containing Ty1 cDNA and IN is imported into the nucleus via a bipartite nuclear localization signal present in IN. Retrotransposition is completed by IN-mediated insertion into a new location in the host genome. Adapted from (260) with permission.
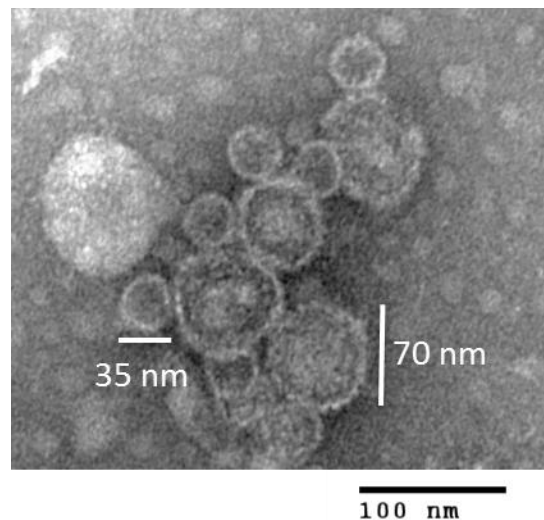
Figure 1.5 The variable diameter size of Ty1 VLPs. Ty1 VLPs were isolated and stained with 2% uranyl acetate and visualized by TEM. The size range in this prep was

from 35-70 nm in diameter, but VLPs have been reported to range from 30-80 nm in diameter.
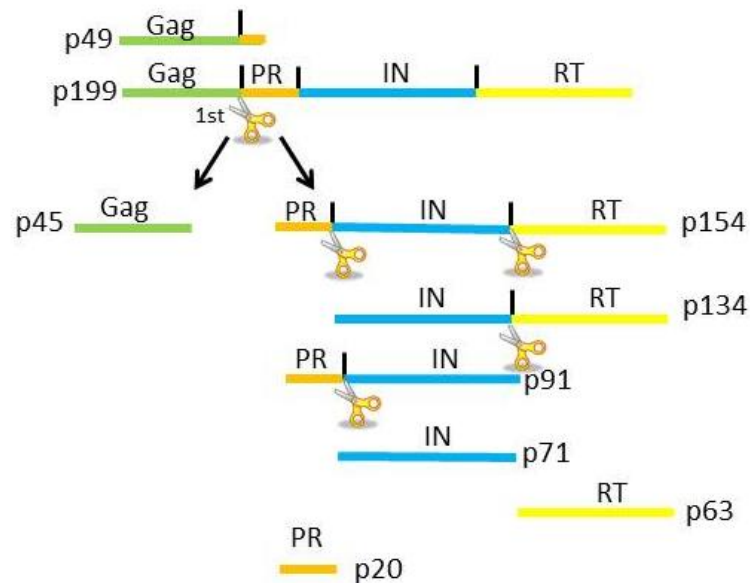
Figure 1.6 Proteolytic maturation pathway in Ty1. The two primary translation products of Ty1, Gag-p49 and Gag-Pol-p199, are initially cleaved at the Gag-PR cleavage site (1st). This produces mature Gag-p45 and precursor Pol-p154. The remaining two cleavage sites are cleaved in no particular order, resulting in the precursors (IN-RT-p134 and PR-IN-p91) and ultimately the mature Ty1 enzymes: IN-p71, RT-p63 and PR-p20.
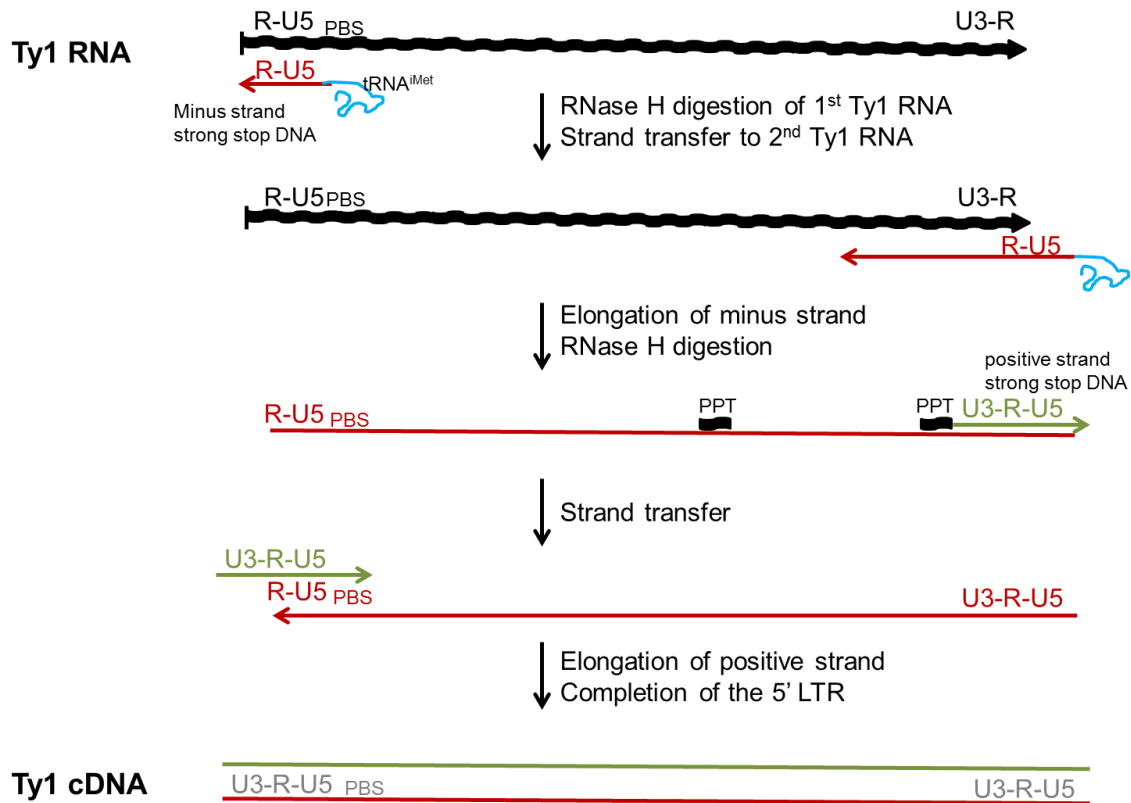
Figure 1.7 Ty1 reverse transcription. Ty1 RNA (black squiggle) serves as a template for the formation of Ty1 cDNA (red and green lines) and does not contain full length LTRs. Reverse transcription is primed by tRNA^{iMet} (blue) bound at the primer binding site (PBS) on Ty1 RNA. The first product of reverse transcription is the minus strand strong stop DNA, which is subsequently transferred to the 3' of the second Ty1 RNA strand packaged inside the VLP where it hybridizes via complementary R sequences. The minus strand is completed and Ty1 RNA:DNA complexes are digested via the RNase H activity of RT. A polypurine tract (PPT) that is not sensitive to RNase H digestion primes the synthesis of the positive strand strong stop DNA. The positive strand is completed after the transfer of the positive strand strong stop DNA to the 5' end of the minus

strand and hybridizes via complementary R-U5 sequences. The 5' LTR is restored using the positive strand as a template. The resulting cDNA (red and green lines) contains full length LTRs (U3-R-U5).
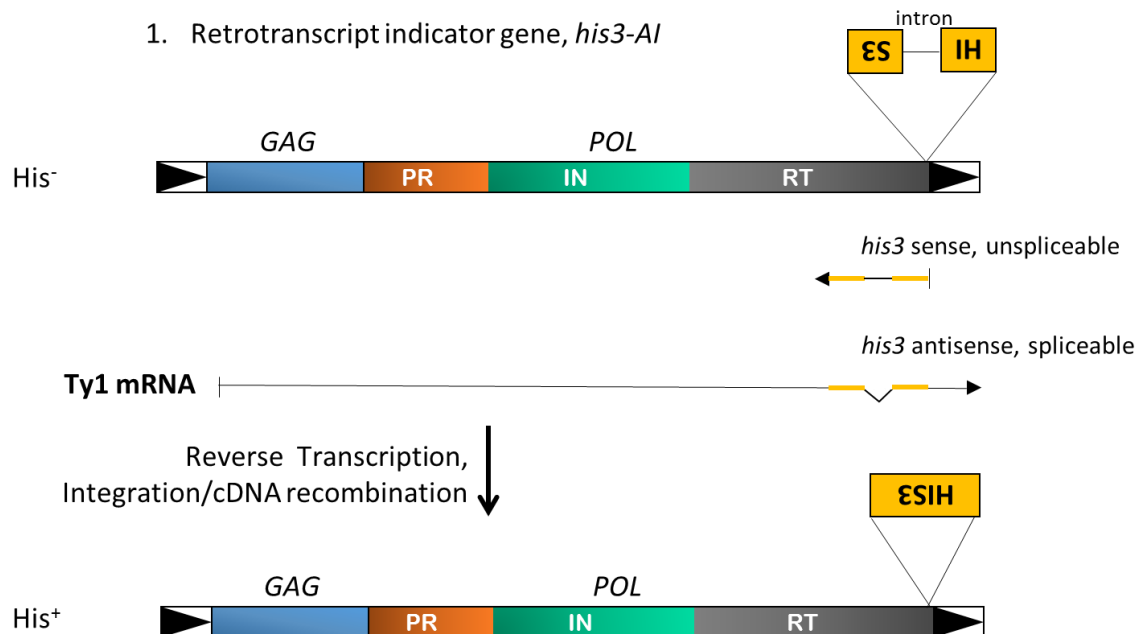
Figure 1.8 Ty1*his3-AI*, an element tagged with a retrotranscript indicator gene (RIG). RIGs consist of a selectable marker placed in an antisense orientation to the retroelement and an artificial intron which disrupts the ORF of the selectable marker. The *his3-AI gene* contains its own promoter, meaning that both Ty1 and *his3-AI* transcripts are produced. However, the artificial intron is only spliceable within the Ty1 transcript. If the Ty1 RNA undergoes splicing, packaging, reverse transcription and either integration (mediated by Ty1 IN-p71) or recombination (with pre-existing Ty1 elements) with the host genome, the cells are able to grow on media lacking histidine.

The number of His$^+$ cells divided by the total number of cells yields a number referred to

as Ty1 mobility and likely underestimates the actual Ty1 retrotransposition frequency.

<center>Tables</center>

Table 1.1 Non-LTR versus LTR retrotransposons
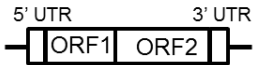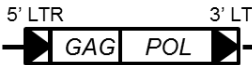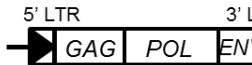
	LTR absent	**LTR present**	
	Non-LTR retrotransposons	LTR retrotransposons	Endogenous retroviruses
Examples	Arthropod R1/R2, *Drosophila* I element, LINEs, SINEs	Yeast Ty elements, *Drosophila* copia and gypsy	Human ERV-K, mouse IAP and MMTV
Gene Structure	5' UTR ORF1 ORF2 3' UTR	5' LTR GAG POL 3' LTR	5' LTR GAG POL ENV 3' LTR
Type of reverse transcription	Target-site site primed	Extrachromosomally primed; within a virus-like particle	Extrachromosomally primed; within a virion

Table 1.2 LTR retrotransposon content in S288C (data from (12))

	Solo LTR	Full length	Truncated	Total
Ty1	279	32	2	313
Ty2	31	13	2	46
Ty3 and Ty3p	58	2	0	60
Ty4	45	3	1	49
Ty5	14	1	0	15

CHAPTER 2

TY1 GAG ENHANCES THE STABILITY AND NUCLEAR EXPORT OF

TY1 MRNA[1]

Abstract

Retrotransposon and retroviral RNA delivery to particle assembly sites is essential for their replication. mRNA and Gag from the Ty1 retrotransposon colocalize in cytoplasmic foci, which are required for transposition and may be sites for virus-like particle (VLP) assembly. To determine which Ty1 components are required to form mRNA/Gag foci, localization studies were performed in a Ty1-less strain expressing galactose-inducible Ty1 plasmids (pGTy1) containing mutations in *GAG* or *POL*. Ty1 mRNA/Gag foci remained unaltered in mutants defective in Ty1 protease or deleted for *POL*. However, Ty1 mRNA containing a frameshift mutation (Ty1fs) that prevents the synthesis of all proteins accumulated in the nucleus. Ty1fs RNA showed a decrease in stability that was mediated by the cytoplasmic exosome, nonsense mediated decay, and the processing-body. Localization of Ty1fs RNA remained unchanged in an *nmd2Δ* mutant. When Gag and Ty1fs mRNA were expressed independently, Gag *provided in trans* increased Ty1fs RNA level and restored localization of Ty1fs RNA in cytoplasmic foci. Endogenously expressed Gag also localized to the nuclear periphery independent of RNA export. These results suggest that Gag is required for Ty1 mRNA stability, efficient nuclear export, and localization into cytoplasmic foci.

Introduction

The *Saccharomyces* Ty1 element belongs to a widely distributed group of transposable elements called long terminal repeat (LTR) retrotransposons or Pseudoviruses, which resemble retroviruses in genome organization and replication (1,

81

2) (see Figure 2.1A for a summary of Ty1 replication). Ty1 is transcribed by RNA polymerase II to yield an abundant 5.7 kb genomic RNA (gRNA) that can be translated or packaged into virus-like particles (VLPs). The translation products are Gag-p49 and Gag-Pol-p199, the latter being less abundant (~3%) because it is the product of a +1 ribosomal frameshift (3). The Gag-p49 and Gag-Pol-p199 precursor proteins form immature VLPs that undergo proteolytic processing by Ty1 protease (PR) to form mature Gag and Pol proteins and functional VLPs. Gag-p49 is cleaved by PR near its C-terminus to form Gag-p45, which is the major capsid protein. Gag-p45 contains a functional nucleic acid chaperone domain that is analogous to retroviral nucleocapsid proteins (4). The Gag-Pol-p199 precursor is also cleaved to form Gag-p45, in addition to the enzymes required for retrotransposition: PR, integrase (IN), and reverse transcriptase (RT). After processing of precursor proteins within VLPs, the encapsidated Ty1 gRNA is reverse transcribed into cDNA. To complete the replication cycle, Ty1 cDNA and IN are imported into the nucleus via a bipartite nuclear localization signal present on IN (5, 6), and the cDNA is integrated into the host genome.

Ty1 gRNA shares two key trafficking steps with its retroviral relatives (7). Both Ty1 and retroviral gRNAs are exported from the nucleus, translated, and packaged into VLPs or viral particles. Since retroviral gRNA can be spliced to form subgenomic mRNAs, a variety of mechanisms have evolved to allow nuclear export and cytoplasmic accumulation of unspliced gRNA, which is required for assembly of infectious particles. For example, HIV-1 encodes a master regulator called Rev, which inhibits splicing of mRNAs containing a Rev response element and promotes nuclear export of partially spliced forms of viral mRNAs and unspliced gRNA (8). HIV-1 Gag then binds

cytoplasmic gRNA, and this RNP complex is recruited to the plasma membrane for virus assembly, although it is not known if there is a specific cellular environment where Gag first encounters the gRNA (9–11). In contrast, *cis*-acting sequences flanking the *src* gene of Rous sarcoma virus (RSV) interact with host export proteins TAP and Dpb5, which modulate the accumulation of unspliced gRNA in the cytoplasm (12–14). Interestingly, RSV Gag enters the nucleus transiently, where it binds gRNA and promotes nuclear export (15–19). It has been suggested that this discrete population of RNA is selectively packaged. Unlike retroviruses, Ty1 mRNA is not spliced; therefore, it is also the gRNA (1). To date, it is not clear whether this pool of gRNA is differentiated into packaged RNA versus translated mRNA.

The trafficking and utilization of Ty1 mRNA during retrotransposition is unique in several respects. Given that Ty1 mRNA is a relatively stable transcript and accumulates to 5–10% of total mRNA in haploid cells, Ty1 proteins and VLPs are present in low amounts (20–25). Transposition events also occur at a low rate. However, overexpression of Ty1 from a *GAL1* promoter in a multicopy plasmid (pGTy1) results in visualization of VLPs by electron microscopy and more than a 10,000-fold increase in retrotransposition events. Interestingly, this occurs with only a 5–15 fold increase in Ty1 mRNA. Therefore, it can be inferred that the use of Ty1 gRNA as a template for translation and reverse transcription increases dramatically during galactose-induced expression of pGTy1.

Recent studies have begun to clarify this paradox. After export from the nucleus via the Mex67p pathway, Ty1 mRNA colocalizes with Gag in cytoplasmic granules, referred to as T-bodies, or mRNA/Gag foci, in about 30% of mitotic cells (26, 27). So far,

two general classes of genes that affect the formation or appearance of mRNA/Gag foci have been identified. The first class is involved in transcription by RNA polymerase II (pol II) (26). For example, mutations in the RNA pol II subunit Rpb1p, Mediator subunits Srb2p and Srb5p, or the Ty1 transcriptional regulator Spt21p do not markedly change Ty1 mRNA levels or transposition frequency yet enhance the formation of mRNA/Gag foci. The second class is genes required for RNA degradation or storage in processing (P) bodies (27, 28). Deletion of P-body components can decrease the number and alter the appearance of Ty1 mRNA/Gag foci, in addition to decreasing the levels of retrotransposition-competent VLPs, mature Ty1 proteins, and packaged gRNA. There is conflicting evidence that Ty1 mRNA/Gag foci occur in P-bodies (26–28). However, P-body components are important cofactors for Ty1 (27–29) and Ty3 retrotransposition (30) and Ty3 VLP assembly (30).

Here, we investigated the Ty1 components required to form mRNA/Gag foci. We provide evidence that Ty1 transcripts defective or deleted for PR, IN, and RT coding sequence still form foci with Ty1 Gag. These results show that either unprocessed (p49) or mature (p45) Gag is sufficient for forming mRNA/Gag foci. When Gag is absent, Ty1 RNA accumulates in the nucleus and the RNA level and stability decreases. When Gag is provided *in trans*, Ty1 RNA colocalizes with Gag in the cytoplasm and the RNA level returns to normal. Endogenously expressed Gag can also be detected in association with the nuclear envelope in wild type cells. Therefore, Gag enhances Ty1 mRNA stability, nuclear export and colocalization into cytoplasmic foci.

Materials and Methods

Yeast strains, plasmids, and growth conditions

Standard yeast genetic and microbiological procedures were used (64, 65). The Ty1-less *S. paradoxus* strains DG1768 (*MATα his3-Δ200hisG ura3 gal3* Spo⁻) (32), and DG2204 (*MATα his3-Δ200hisG ura3 trp1*) were used in this work. Strain DG1768 was transformed to uracil prototrophy with plasmids pGTy1, pGPOLΔ, pGPR⁻, and pGfs. DG2204 was transformed with both pBDG1130 (pGfs) and pMACB36 (pMPOLΔ). pGTy1 (pGTy1H3Cla) contains an active Ty1 element fused at its transcriptional initiation site to the *GAL1* promoter (66). pGPOLΔ(pBDG1130, pGTy1pol-Δ1702-5562) contains a deletion between BglII restriction enzyme (RE) sites, which are in PR and adjacent to the stop codon in *POL*, that removes most of *POL* (deletion of 1702-5562) (31). pGPR− (pJEF1266, pGTy1pr-1702SacI) contains an in-frame SacI linker (codons S-E-L-G) inserted into the PR domain, which renders PR enzymatically inactive (33, 34). pGfs (pBDG1112, pGTy1gag-ATGfs) had a single A nucleotide deleted adjacent to the start codon in GAG, which creates a -1 frameshift mutation and a BamHI RE site (31). With the exception of pMACB36, all plasmids are present on 2μ-based multicopy shuttle vectors carrying the *URA3* gene. The plasmid pMACB36 has the *MET25* promoter in place of *GAL1* and is present in a *TRP1*-based vector. To create pMACB36, the *GAL1* promoter was replaced by the *MET25* promoter using standard cloning procedures, and then the *URA3* marker was replaced by the *TRP1* marker by homologous recombination in yeast (67). Briefly, the *GAL1* promoter from pBDG1130 was removed by EcoRI and XhoI RE digest. The *MET25* promoter was amplified from

S. cerevisiae strain BY4742 using the following primers pMET25-F616-EcoRI and pMET25-R1000-XhoI (Table 2.2). This PCR fragment was then ligated into the cut BDG1130 vector to make plasmid pMACB33. To replace the *URA3* marker for the *TRP1* marker, pMACB33 was cut with StuI, which cuts only once in *URA3*. Homologous recombination between linearized pMACB33 and a PCR fragment containing the *TRP1* marker with *URA3* flanking sequences was carried out after transformation into strain DG1251 (*MATα his3-Δ200 ura3-167 trp1-hisG spt3-101*) and Trp⁺ Ura3⁻ recombinants were obtained. pMACB36 was recovered from a Trp⁺ Ura3⁻ recombinant by transformation of total yeast DNA into *E. coli* DH5α. *S. paradoxus* strain DG1768 deleted for *NMD2*, *SKI3*, *RRP6*, and *XRN1* were constructed as previously described (67, 68), using flanking or terminal gene specific primers conserved between *S. cerevisiae* and *S. paradoxus* isolates (69) (Table 2.2), and tools provided on the Welcome Trust Sanger Institute (http://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html) and *Saccharomyces* Genome databases (http://www.yeastgenome.org/). Correct deletants containing KanMX were verified by PCR and then transformed with pGTy1 or pGfs. DG2204 was transformed with pGfs and pRS424 (*TRP1*/2μ-vector) (70) or pMPOLΔ. Strains were grown in synthetic complete (SC) media lacking uracil (SC-Ura) containing 2% raffinose (pH 6.5) at 30°C until log phase, then placed in SC-Ura media containing 2% galactose and grown at 20°C for 14–16 hr for microscopic analyses, or 2–16 hr for Northern analyses. For shutting off *GAL1* expression, glucose (2% final concentration) was added to a 16 hr galactose culture, followed by incubation at 20°C for 1 hr. For Northern analysis of cells coexpressing pGfs and pMPOLΔ or pGfs and an empty

vector, cells were grown in SC-Ura plus 2% raffinose, diluted 25-fold into SC-Ura-Trp-Met plus 2% galactose, and grown for 48 hr at 20°C. For sequential expression of pGTy1fs and pMPOLΔ, cells were first grown in synthetic minimal media (SD) containing histidine (H), leucine (L), lysine (K), 2% raffinose, and 5 mM methionine at 20°C overnight. To first induce pGfs, cells were resuspended in SD+H+L+K containing 2% galactose and 5mM methionine and grown at 20°C for 22 hr. To express pMPOLΔ, cells were washed and resuspended in SD+H+K+L medium containing 2% galactose but lacking methionine, and then grown at 20°C for 2 hr. Isogenic *S. cerevisiae* strains containing the temperature-sensitive allele *mex67-5* (THJY2232; *MATa his3-Δ1 leu2-Δ0 trp1-ΔhisG ura3-Δ0 mex67-5*) or wild type *MEX67* (THJY2210) were kindly provided by F. Malagon and T. Jensen. Cells were grown in SC-Ura containing 2% glucose for 20 hr at 20°C, and shifted to 30°C for 1–2 hr. Cells were fixed at 30°C and analyzed by FISH-IF and confocal microscopy.

<u>Northern analysis</u>

Total RNA was isolated using the MasterPure yeast purification kit (Epicentre Biotechnologies (Madison, WI) (71). Northern blot hybridization and phosphorimage analysis were performed as previously described (31, 72). ^{32}P-labeled riboprobes were synthesized from Ty1 GAG and *ACT1* coding sequence using the MAXscript *in vitro* transcription kit (Applied Biosystems/Ambion, Woodland, TX) and [α-^{32}P]UTP (3000 Ci/mmol) (PerkinElmer, Waltham, MA).

<u>Oligonucleotide probes used for fluorescent in situ hybridization (FISH)</u>

Ty1 mRNA was detected with either a GAG or an RT oligonucleotide labeled at the 3′ end with digoxigenin (DIG)-dUTP using the DIG Oligonucleotide Tailing Kit, 2nd

Generation (Roche Applied Science, Indianapolis, IN). The GAG and RT oligonucleotides antisense TyA-322 or antisense RT-4219 (Table 2.2), respectively, were synthesized and HPLC-purified by Invitrogen (Carlsbad, CA).

Antibodies used for immunofluorescence microscopy (IF)

Gag was detected using a rabbit anti Ty1-VLP antibody (kindly provided by Alan Kingsman) at dilutions of 1:1000–2000 for IF. Either AF594 (for colabeling with Ty1 RNA) or AF488 (for colabeling with Nsp1) conjugated anti-rabbit (Invitrogen) was used at 1:100–200. DIG-labeled probes were detected using Fluorescein (FITC) conjugated sheep anti-DIG Fab fragment (Roche Applied Sciences) at 1 ng/μl. Nsp1p was detected using a mouse monoclonal antibody from Abcam (Cambridge, MA) at a dilution of 1:250–500. AF555 conjugated anti-mouse (Invitrogen) was used at 1:100–200.

FISH and IF microscopy

The FISH-IF procedure has been described previously (27). For most experiments, 80–200 cells per strain were scored to obtain quantitative data for localization of Ty1 RNA and Gag signal. For confocal analyses, the number of cells scored is mentioned in the text. Signals below background were not considered for quantification. Percentages of Ty1 mRNA and Gag foci were calculated from cells that were stained with Nsp1p antibody and DAPI. Percentage of Ty1 mRNA/Gag colocalization was calculated from cells containing mRNA/Gag overlap over the total number of cells containing Gag foci. Percentage of cytoplasmic localization of Ty1 mRNA was calculated from cells containing at least one visible foci of mRNA in the cytoplasm over the total number of cells containing Ty1 mRNA. Likewise, percentage of nuclear retention of Ty1 RNA was calculated from cells only containing Ty1

mRNA/Nsp1p/DAPI stain overlap over the total number of cells containing Ty1 mRNA. Fluorescence imaging was performed using a Nikon Eclipse E-1000 microscope equipped with a Nikon C-CU Universal condenser, an oil immersion Nikon Plan Fluor DIC 100x objective and a Hamamatsu Orca-ER c4742-95 cooled charge-coupled device (CCD) camera (Hamamatsu, Japan). The Openlab 5.0 software (Improvision, PerkinElmer, Waltham, MA) was used to acquire images and merge fluorescent images to give the composite images shown. For all experiments involving the Nsp1p antibody, image acquisition was carried out using a Zeiss Axio Observer microscope equipped with an AxioCam HSm Camera and analyzed with AxioVision v4.6 software (Carl Zeiss Microscopy, LLC, North America). Exposure times used to capture fluorescent and DAPI images were kept consistent throughout each experiment. Figures were constructed with Adobe Photoshop software (Adobe Systems, San Jose, CA).

Confocal images involving DAPI as a nuclear marker were performed using an Olympus FV1000 confocal microscope (Olympus America Inc., Center Valley, PA) using a 60X, 1.42 NA (numerical aperture), oil objective lens or a Zeiss LSM 510 confocal microscope (Carl Zeiss Microscopy, LLC, North America) using a 63X, 1.42 NA oil objective lens. The x, and y resolutions were 0.33, and z resolution was either 0.2 or 0.25 µm, respectively. The DAPI channel was acquired using a 405 nm laser and collecting emitted light from 420–480 nm. The green channel was acquired with a 488 nm laser and collecting light from 505–550 nm. The red channel was acquired with a 561 nm laser and collecting light above 570 nm. Confocal images involving Nsp1p as a nuclear marker were performed using a Zeiss LSM 510 confocal microscope (Carl Zeiss Microscopy, LLC, North America) using a 63X, 1.42 NA oil objective lens and a z

resolution of 0.25 µm. The green channel was acquired with a 488 nm laser and collecting light from 503–530 nm. The red channel was acquired with a 543 nm laser and collecting light above 560 nm. Images from both experiments were processed with MIPAV (Medical Image Processing, Analysis, and Visualization) software (http://mipav.cit.nih.gov). Noise was reduced in the green and red channels in the following way: each 2D slice image was filtered with a median filter using a 3×3 square kernel to reduce salt and pepper noise, each 2D slice image was filtered with a Gaussian filter using a standard deviation of 1 pixel to further reduce noise. Specific cells were cropped from the 3D color image in order to ease visual interpretation of these images. The contrast of the red, green, and blue channels was separately enhanced to clearly show Nsp1p, Ty1 Gag and RNA signals in final images. Intensity profiles along straight lines through Ty1 Gag and RNA signals were displayed in order to determine whether Gag and RNA signals actually colocalized with each other, and were not apparently overlapping because of blurring caused by the spatial resolution of the optical microscope. Gag foci were determined to be DAPI- or Nsp1-associated if Ty1 Gag signal was adjacent to, overlapping, or surrounded by the DAPI or Nsp1p signal.

Results and Discussion

POL products are not necessary for Ty1 mRNA/Gag localization into cytoplasmic foci

To determine the Ty1 components required to form Ty1 mRNA/Gag foci, we analyzed Ty1 mutants containing deletion or inactivation of *POL*-encoded products (Figure 2.1B). Galactose induction of pGTy1 mutants was performed in a *S. paradoxus* strain that lacks Ty1 elements to minimize any contribution from endogenous Ty1

expression (31, 32). Localization of Ty1 mRNA and Gag (p49 and p45) was analyzed by fluorescence in situ hybridization and immunofluorescence (FISH-IF) microscopy using a GAG-DIG probe (Figure 2.1B) and anti-VLP antiserum, respectively. As shown previously (27), multiple Ty1 mRNA and Gag foci appeared in about one third of cells overexpressing pGTy1 and colocalized in 96% of these cells (Figure 2.2A). Inactivating Ty1 PR (pGPR−) (33, 34) or deleting most of PR, and all of the IN and RT coding sequence (pGPOLΔ), but leaving the 3′ LTR intact did not alter the colocalization of mRNA/Gag foci. As expected, Ty1 mRNA and Gag signals in the Ty1-less control strain (DG1768) were indistinguishable from background levels (data not shown) (27). Since *POL* products are not required for localization of Ty1 mRNA into cytoplasmic foci, Gag-p49 may be the only Ty1 protein required. Our results also support and extend recent work suggesting that formation of mRNA/Gag foci precedes maturation of VLPs (35).

Ty1fs RNA remains in the nucleus in the absence of Gag

To examine if Gag is required for the localization of Ty1 mRNA into cytoplasmic foci, a single base was deleted adjacent to the *GAG* start codon in pGTy1 to create pGfs. This mutant produces full-length transcript (Ty1fs RNA) upon galactose-induction but does not synthesize detectable levels of Gag or Pol proteins (J. Wang and D.J. Garfinkel, unpublished results), and contains a premature termination codon (31). A Ty1-less strain containing pGfs was induced with galactose for 16 hr and subjected to FISH-IF analysis using a GAG-DIG probe, and antisera against VLPs and the nuclear pore component Nsp1p (36), which outlines the nuclear envelope (Figure 2.2B). Thirty-nine percent of the cells contained a Ty1fs gRNA focus, which localized within the boundaries of the Nsp1p stain and colocalized with the nuclear stain DAPI. The Ty1fs

RNA signal appeared fainter compared to Ty1 mRNA, which is consistent with previous work showing that the level of Ty1fs RNA decreases when expressed from its native promoter (31). As expected, there was only a background signal detected with the anti-VLP antiserum (data not shown). Confocal microscopy of these samples showed that Ty1fs mRNA localized within the nucleus (Figure 2.2C). Interestingly, FIV and HIV-1 gRNA also localize at the nuclear rim independent of Gag or a packaging signal (37).

Degradation and nuclear retention of Ty1fs RNA

To determine the level of Ty1fs RNA expressed from pGfs and identify the pathways used to degrade this abnormal transcript, total RNA from isogenic wild type and mutant strains defective in RNA degradation were subjected to Northern analysis (Figures 2.3-5, Table 2.1). The cells were induced for pGTy or pGfs expression for 2 and 16 hr to recapitulate the level of Ty1 gRNA observed in *S. cerevisiae* (21), and monitor changes in RNA level that may be dependent on the length of induction. Signals were normalized to *ACT1* (Actin) unspliced and spliced transcripts. As expected for a transcript containing a premature termination codon, the level of Ty1fs RNA decreased 5–10 fold in wild type cells induced for 2, 4, 8 or 16 hr (Figure 2.4). To quantify the difference in Ty1fs RNA stability, the level of the Ty1 mRNA and Ty1fs transcripts were determined after culturing cells in 2% glucose for 1 hr following induction for 16 hr (Figure 2.5), a treatment that quickly represses *GAL1* transcription The Ty1 and Tyfs transcripts decreased 1.25- and 12.5-fold, respectively, relative to the level of *ACT1* mRNA after glucose shut-off. Together, these results show that Ty1fs RNA is unstable when compared with the long-lived wild type transcript (25), and suggests that Gag may enhance the stability of Ty1 mRNA.

We examined mutants defective in nonsense-mediated decay (*NMD*; *nmd2Δ*), the cytoplasmic (*ski3Δ*) or nuclear exosome (*rrp6Δ*), and the P-body (*xrn1Δ*) for rescue of Ty1fs RNA when compared with Ty1 mRNA. Rescue was observed at 2 hr of galactose induction in the *nmd2Δ*, *xrn1Δ*, and *ski3Δ* mutants, but not in *rrp6Δ* (Figure 2.3). However, the relative amount of Ty1fs RNA decreased with the length of galactose induction in the *nmd2Δ*, *ski3Δ*, and *xrn1Δ* mutants. For example, Ty1fs RNA accumulation reached a maximum by 2 hr in the *nmd2Δ* mutant, and then decayed to the level present in the wild type *NMD2* strain (data not shown, Figure 2.5). The time-dependent difference in RNA decay suggests the possibility that each pathway degrades Ty1fs RNA at a different rate, a property that has been observed for other transcripts (38). In agreement with previous work (39, 40), exported Ty1fs RNA is recognized as an aberrant transcript containing a premature termination codon by NMD and then is targeted to the P-body for destruction. Also, Ty1fs RNA apparently is degraded by the cytoplasmic exosome (Figure 2.3). The lack of rescue observed in the *rrp6Δ* mutant suggests that Ty1fs RNA is not recognized as an aberrant transcript by the nuclear exosome. The level of wild type Ty1 mRNA remained about the same in all the mutants except the P-body exonuclease mutant *xrn1Δ* (Figure 2.3, Table 2.1). Surprisingly, the level of wild type Ty1 mRNA decreased 10-fold while Ty1fs RNA increased 3-fold after 2 hr of galactose-induction in the absence of *XRN1*, which resulted in an apparent overestimate of the rescue. One possibility is that the increased amount of a Ty1 or possibly a *GAL1* noncoding transcript observed in an *xrn1Δ* mutant represses *GAL1*-promoted Ty1 transcription via modifications in chromatin (41–43). However, this view fails to explain the increased amount of Ty1fs RNA in the *xrn1Δ*

mutant, which probably affects Ty1fs RNA stability more than its synthesis. Together, these results also raise the possibility that inefficient nuclear export of Ty1fs RNA occurs throughout galactose induction.

Since Ty1fs RNA is recognized by NMD, which occurs in the cytoplasm (40), we examined the localization of Ty1 RNA in *NMD2* wild type and *nmd2Δ* mutant strains after 16 hr of induction (Figure 2.6). In the *NMD2* strain, similar percentages of cells with cytoplasmic Ty1 mRNA (34%) or nuclear Ty1fs RNA (28%) foci were observed. Wild type Ty1 mRNA localized in foci in 39% of *nmd2Δ* mutant cells, and the foci were cytoplasmic. Ty1fs RNA foci were detected in 19% of *nmd2Δ* cells. However, these foci remained nuclear, since they colocalized with DAPI-stained material and were within the boundaries of the nuclear pore marker Nsp1p. In addition, comparable percentages of cells with nuclear Ty1fs RNA foci were observed in the presence (28%) or absence (19%) of NMD. Therefore, the failure to detect cytoplasmic Ty1fs RNA foci in an *nmd2Δ* mutant is not due to RNA turnover, but rather is associated with the absence of Gag.

<u>Gag restores Ty1fs RNA level and nuclear export</u>

We developed a system to distinguish Ty1fs RNA from mRNA encoding Gag-p49 (Figure 2.7). This system consists of expressing Ty1fs RNA from a *URA3*-based pGTy1 plasmid and Gag from the POLΔ mutant using the *MET25* promoter on a *TRP1*-based vector (pMPOLΔ). Thus, Ty1fs RNA expression is induced in the presence of galactose, whereas Gag mRNA expression from pMPOLΔ is repressed when methionine is present. These plasmids were introduced into a Ty1-less strain and expressed either concomitantly or sequentially. To determine if the expression of Gag *in trans* influences the level of full-length Ty1fs RNA (Figure 2.7A), Northern analysis was performed with

total RNA from cells expressing both pGfs and pMPOLΔ or pGfs and an empty vector for 48 hr, and a probe that detected both transcripts. The level of Ty1fs RNA increased 10-fold in the presence of Gag and was similar to the level of *GAG* transcript. Given that Ty1fs RNA is unstable (Figure 2.4), it is very likely that Gag restores Ty1fs RNA level by preventing its degradation, although early work suggested the possibility that Gag overexpression might stimulate endogenous Ty1 transcription (44).

Does coexpression of Gag *in trans* also restore nuclear export and formation of mRNA/Gag foci (Figure 2.7B)? We can distinguish which transcript is being expressed based on the probe used for FISH. The GAG-DIG probe detects both full-length Ty1fs and truncated POLΔ transcripts, whereas the RT-DIG probe only detects the Ty1fs mRNA (see Figure 2.1B). Cells were first grown in medium containing galactose and methionine for 22 hr to induce only the transcript from pGfs and then analyzed by FISH-IF. Ty1fs RNA was detected using an RT-DIG probe. Under these conditions, induction of pGfs led to 10% of cells containing Ty1fs RNA, out of which 67% contained RNA with nuclear retention. Only 2% of the cells contained Gag foci, which always colocalized with cytoplasmic RNA foci. These Gag foci probably result from leaky expression of the *MET25* promoter in pMPOLΔ even in the presence of methionine. This indicates that repression of pMPOLΔ by methionine was not complete but was sufficient to minimize nuclear export of Ty1fs RNA by Gag in most cells.

To allow expression of both plasmids, cells were washed and grown in medium containing galactose but lacking methionine for a period of 30 min to 2 hr (Figure 2.7B). At designated time points, cells were analyzed by FISH-IF using either an RT-DIG or GAG-DIG probe and anti-VLP antiserum. Between 30 min and 2 hr in the absence of

methionine, Gag was detected in 34–53% of the cells. FISH-IF analysis with the RT-DIG probe detected Ty1fs RNA in 31–40% of cells. In more than 88% of these cells Ty1fs RNA now localized into cytoplasmic foci. Colocalization of Ty1fs RNA with Gag was observed 25–30% of the time. Ty1fs RNA remained associated with the nucleus in only 7–12% of cells containing Ty1 RNA. These results show that Gag-p49 can act *in trans* to facilitate Ty1 RNA nuclear export. Failure to completely export Ty1fs RNA may be caused by lower expression from the *MET25* promoter relative to expression from *GAL1*, or possible defects in the export function of the Gag-p49 precursor relative to Gag-p45. When FISH-IF was performed using the GAG-DIG probe at 2 hr, Ty1 RNA from both pGfs and pMPOLΔ was present in foci in 22% of cells. Gag was detected in 40% of cells with 16% mRNA/Gag colocalization. At 2 hr, we observed 90% cytoplasmic localization of Ty1fs RNA when detected with the RT-DIG probe. However, when RNA was detected with the GAG-DIG probe, we observed only 54% cytoplasmic localization of Ty1 RNA in cells containing detectable transcripts. This suggests that Gag may prefer to export the full-length Ty1fs RNA rather than the truncated *POLΔ* transcript, which could be due to a sequence within *POL* or the RNA structure formed when *POL* is present. Taken together, these results suggest that Gag-p49 alone is necessary and sufficient to enhance nuclear export of Ty1 mRNA and form cytoplasmic mRNA/Gag foci.

Nuclear association of endogenous Gag in a conditional *mex67* mutant and wild type cells

Our results suggest that either Gag itself or some other Gag-dependent RNA binding protein may shuttle between the nucleus and the cytoplasm to promote Ty1

mRNA export, yet Gag is not present in the nucleus at steady state (26–28, 35). To address the possibility that the nuclear localization of Ty1 Gag is only transient, we analyzed endogenous Ty1 mRNA and Gag localization in *S. cerevisiae* cells containing a *mex67-5* temperature-sensitive allele, since *MEX67* is required for Ty1 mRNA nuclear export (26). Two experiments were performed in an attempt to trap Gag in the nucleus. Both experiments required a growth temperature that minimized RNA export while maintaining mRNA/Gag foci, since *mex67-5* cells become increasingly deficient in export at temperatures above 30°C, while mRNA/Gag foci begin to disperse into smaller puncta at 30°C (27). In the first experiment, cells were grown for 20 hr at 20°C, and then incubated for 1–2 hr at 30°C. Fixed cells were analyzed by FISH-IF and confocal microscopy (Figure 2.8). Fifty-eight cells were observed to have three different localizations for Gag: 1) cytoplasmic foci (Figure 2.8A, 19%: 11/58 cells), 2) scattered cytoplasmic puncta or no signal (data not shown) (65.5%: 38/58 cells), and 3) Close proximity to DAPI stained material (Figure 2.8B, 15.5%: 9/58 cells). Ty1 Gag, mRNA and DAPI stains were also visualized at multiple focal planes (0.2 or 0.25 µm intervals) in selected cells to generate a three-dimensional reconstruction. Analysis of individual planes of each Gag focus showed Gag associated with the periphery or possibly within the DAPI stained material (Figure 2.8B). Gag was usually observed associated with Ty1 mRNA regardless of its localization. However, a similar staining pattern was observed in an isogenic *MEX67* wild type strain (data not shown).

To confirm and extend the results obtained with DAPI staining to delineate the nucleus, cells were grown and analyzed under comparable conditions using Nsp1p and VLP antisera (Figure 2.9). To distinguish background signal from Gag foci, imaging was

performed alongside a Ty1-less strain (DG1768), which likely resulted in an underestimate of the number of cells containing Gag foci. Ty1 Gag and Nsp1p were visualized at multiple focal planes (0.25 μm intervals), to generate a three-dimensional reconstruction. Three Gag localization patterns were observed: 1) cytoplasmic foci, which were distinctly separate from the Nsp1p stain (Figure 2.9A, top panels, 3.4%: 22/655 cells), 2) scattered cytoplasmic puncta or no signal (data not shown, 89.6%: 587/655 cells), and 3) Nsp1p-associated foci, which were adjacent or overlapping with Nsp1p (Figure 2.9A, bottom panels, 7%: 46/655 cells). Although Ty1 Gag foci were present in close association with the nuclear envelope, we were unable to detect Gag within the nuclear envelope. When an isogenic *MEX67* wild type strain was analyzed under identical conditions, we observed a similar percentage of cells with cytoplasmic (4.4%: 16/363 cells) and Nsp1p-associated Gag foci (8.8%: 32/363 cells; Figure 2.9B), which reinforces the analysis described above using DAPI staining. In addition, wild type cells continuously grown at the permissive temperature displayed the same pattern as those shifted to 30°C (data not shown). Therefore, Ty1 Gag appears to collect near the nuclear envelope in 8% of mitotic cells, whether or not Ty1 mRNA export is inhibited. Although an association between Ty1 Gag and the nuclear envelope was not detected in previous studies (26–28), it is probably due to use of a DNA stain as a nuclear marker, rather than a marker for the nuclear envelope or nuclear pores. Similar localization has been reported for Ty3 Gag overexpressed from the *GAL10* promoter, but these were shown to consist of assembled VLPs at the nuclear pore (45). This is unlikely to be the case here, because VLPs are rarely observed by electron microscopy when chromosomal Ty1 elements are expressed from their endogenous promoters (24,

27). Our work supports the idea that Gag, prior to assembling into VLPs, can associate with the nuclear envelope and stabilize Ty1 mRNA as it exits the nucleus.

The nuclear export of specific unspliced mRNAs using a virally encoded nucleocytoplasmic shuttle protein has been described for several retroviruses (37, 46–50). With this mechanism, the shuttle protein contains a nuclear localization signal (NLS), a nuclear export signal (NES), an RNA-binding domain, and domains that bind karyopherins (e.g. importins and exportins), which are cellular factors involved in nucleocytoplasmic translocation. For example, RSV Gag possesses NLSs in its nucleocapsid and matrix (MA) domains, which can mediate nuclear entry in association with importin family members Imp-α and Imp-11 (16, 19, 49, 51). Once inside the nucleus, RSV Gag or MA binds unspliced gRNA (16, 52), which may promote Gag dimerization. The combination of RNA-binding and Gag-multimerization unmasks an NES present on the p10 domain of Gag (18, 53), resulting in export of mRNA/Gag complexes. In addition, binding of gRNA with Gag inhibits interactions with importins, and the viral RNA packaging site stimulates Gag-Crm1p interaction, which is necessary for nuclear export (16). Thus, the gRNA may regulate the ordered association of host import and export functions that mediate gRNA/Gag trafficking and viral assembly.

In budding yeast retrotransposons, mutations in Ty3 nucleocapsid (NC) that disrupt binding of Gag3 to Ty3 RNA lead to the accumulation of Gag3 in the nucleus (54). Furthermore, the *Schizosaccharomyces pombe* Tf1 Gag protein contains an NLS that is required for both transposition and an interaction with the nuclear pore protein Nup124p (55). Previous work indicates that Ty1 Gag is localized to the cytoplasm as mRNA/Gag foci or VLPs (26, 27, 35). At present, an NLS has not been identified in Ty1

Gag. Although Gag contains a putative NES (L-TF-L-YNT-F-Q-I between residues 217 and 226) that conforms to the Class 1D consensus for yeast (56), alanine scanning mutagenesis of this region compromised Ty1 transposition but failed to block nuclear export of Ty1 mRNA, Gag, or formation of mRNA/Gag foci (data not shown). Nuclear export of a model protein containing a NLS fused to GFP (56) also failed when the candidate Gag NES was added to the protein.

Since nuclear association of Ty1 Gag was not observed using the two-plasmid expression system, we hypothesized that Gag may have a transient nuclear localization. Our results suggest that Gag collects at the nuclear periphery, and enhances Ty1 mRNA export using host machinery. Given that Mex67p is required for Ty1 mRNA export (26), the Mex67p pathway may mediate export of mRNA/Gag complexes.

Although we could not show that Ty1 Gag enters the nucleus or contains a NES, other proteins involved in the nuclear export of RNA also exhibit a predominantly cytoplasmic localization. For example, Dbp5p, a DEAD-box ATPase in yeast localizes mainly to the cytoplasm and the cytoplasmic face of the nuclear pore (57), is involved in remodeling of RNPs prior to release from the nuclear pore and is required for normal export of poly(A) RNA (58–60). Although steady-state localization of Dpb5p is cytoplasmic, defects in the Ran or Crm1p pathways result in visual nuclear localization of Dbp5p (61). A human DEAD-box protein, DDX3, which is involved in Rev-mediated nuclear export of HIV-1 gRNA, localizes to the cytoplasm and the nuclear periphery, because it is constitutively exported by CRM-1 (62, 63). Hence, DDX3 is only detected in the nucleus when CRM-1 is inhibited by leptomycin B. Similar results are also

observed for RSV (19) and FIV Gag (50), with FIV Gag reported to localize to the nuclear periphery prior to the leptomycin B studies (37). Although FIV Gag nuclear shuttling is CRM1-sensitive, it seems to function independently of a classical NES (50). Thus, Ty1 Gag may also traffic into the nucleus without a classical NES, but we have yet to visualize nuclear-localized Gag. In addition, further examination by confocal microscopy of a possible overlap of Gag foci with the nucleus in *xrn1Δ* and *lsm1Δ* P-body mutants (27), reveals that Gag is present at the nuclear periphery (Checkley and Garfinkel, unpublished data).

What role does the nuclear interaction between Ty1 gRNA and Gag play during the process of retrotransposition? Like RSV and FIV gRNA, Ty1 gRNA is likely exported from the nucleus independently of Gag and undergoes multiple rounds of translation. As the level of Gag increases, which occurs during induction of pGTy1, it promotes export of Ty1 gRNA and may direct the transcript to sites of VLP assembly. In addition, Gag likely stabilizes Ty1 mRNA since Ty1fs RNA is unstable and coexpression of Gag restores Ty1fs RNA level. Nonetheless, early interactions between Ty1 gRNA and Gag immediately after nuclear export may facilitate the accumulation of cytoplasmic mRNA/Gag foci, which are sites where VLPs cluster and may be involved in VLP assembly (27). Although it remains to be determined where HIV-1 gRNA first interacts with Gag in the cytoplasm (17), an emerging theme from Ty1, RSV, foamy virus, and FIV interactions raise the possibility that HIV-1 Gag may interact with its gRNA soon after export.

In this study, we have shown that Ty1 Gag enhances Ty1 mRNA accumulation, nuclear export, and localization of this RNA into cytoplasmic foci. The development of a

two-plasmid expression system in the Ty1-less strain allowed us to separate the expression and detection of a full-length Ty1 mRNA from that of Gag in the same cell. This dual-plasmid system will be useful to help define the domains in Ty1 mRNA and Gag that interact and whether these sequences affect additional steps in the Ty1 replication cycle. This system also could be used to identify host cofactors involved in Gag-mediated Ty1 mRNA export and further our understanding of similar processes that occur with retroviral gRNAs.

References

1. Voytas DF, Boeke JD. Ty1 and Ty5 of Sacharomyces cerevisiae. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington, DC: ASM Press; 2002. pp. 614–630.

2. Peterson-Burch BD, Voytas DF. Genes of the Pseudoviridae (Ty1/copia retrotransposons) Mol Biol Evol. 2002;19(11):1832–1845

3. Kawakami K, Pande S, Faiola B, Moore DP, Boeke JD, Farabaugh PJ, Strathern JN, Nakamura Y, Garfinkel DJ. A rare tRNA-Arg(CCU) that regulates Ty1 element ribosomal frameshifting is essential for Ty1 retrotransposition in Saccharomyces cerevisiae. Genetics. 1993;135(2):309–320.

4. Cristofari G, Ficheux D, Darlix JL. The GAG-like protein of the yeast Ty1 retrotransposon contains a nucleic acid chaperone domain analogous to retroviral nucleocapsid proteins. J Biol Chem. 2000;275(25):19210–19217.

5. Kenna MA, Brachmann CB, Devine SE, Boeke JD. Invading the yeast nucleus: a nuclear localization signal at the C terminus of Ty1 integrase is required for transposition in vivo. Mol Cell Biol. 1998;18(2):1115–1124.

6. Moore SP, Rinckel LA, Garfinkel DJ. A Ty1 integrase nuclear localization signal required for retrotransposition. Mol Cell Biol. 1998;18(2):1105–1114.

7. Swanson C, Malim M. Retrovirus RNA trafficking: from chromatin to invasive genomes. Traffic. 2006;7(11):1440–1450.

8. Pollard VW, Malim MH. The HIV-1 Rev protein. Annu Rev Microbiol. 1998;52:491–532.

9. Jouvenet N, Simon SM, Bieniasz PD. Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. Proc Natl Acad Sci U S A. 2009;106(45):19114–19119.

10. Kutluay SB, Bieniasz PD. Analysis of the initiating events in HIV-1 particle assembly and genome packaging. PLoS Pathog. 2010;6(11):e1001200.

11. Moore MD, Nikolaitchik OA, Chen J, Hammarskjold ML, Rekosh D, Hu WS. Probing the HIV-1 genomic RNA trafficking pathway and dimerization by genetic recombination and single virion analyses. PLoS Pathog. 2009;5(10):e1000627.

12. LeBlanc JJ, Uddowla S, Abraham B, Clatterbuck S, Beemon KL. Tap and Dbp5, but not Gag, are involved in DR-mediated nuclear export of unspliced Rous sarcoma virus RNA. Virology. 2007;363(2):376–386.

13. Ogert RA, Lee LH, Beemon KL. Avian retroviral RNA element promotes unspliced RNA accumulation in the cytoplasm. J Virol. 1996;70(6):3834–3843.

14. Paca RE, Ogert RA, Hibbert CS, Izaurralde E, Beemon KL. Rous sarcoma virus DR posttranscriptional elements use a novel RNA export pathway. J Virol. 2000;74(20):9507–9514.

15. Garbitt-Hirst R, Kenney S, Parent L. Genetic evidence for a connection between Rous Sarcoma Virus Gag nuclear trafficking and genomic RNA packaging. Journal of Virology. 2009;83(13):6790–6797.

16. Gudleski N, Flanagan JM, Ryan EP, Bewley MC, Parent LJ. Directionality of nucleocytoplasmic transport of the retroviral gag protein depends on sequential binding of karyopherins and viral RNA. Proc Natl Acad Sci U S A. 2010;107(20):9358–9363.

17. Parent LJ. New insights into the nuclear localization of retroviral Gag proteins. Nucleus. 2011;2(2):92–97.

18. Scheifele L, Kenney S, Cairns T, Craven R, Parent L. Overlapping roles of the Rous Sarcoma Virus Gag p10 domain in nuclear export and virion core morphology. Journal of Virology. 2007;81(19):10718–10728.

19. Scheifele LZ, Garbitt RA, Rhoads JD, Parent LJ. Nuclear entry and CRM1-dependent nuclear export of the Rous Sarcoma Virus Gag polyprotein. Proc Natl Acad Sci U S A. 2002;99(6):3944–3949.

20. Boeke JD, Garfinkel DJ, Styles CA, Fink GR. Ty elements transpose through an RNA intermediate. Cell. 1985;40(3):491–500.

21. Curcio MJ, Garfinkel DJ. Posttranslational control of Ty1 retrotransposition occurs at the level of protein processing. Mol Cell Biol. 1992;12(6):2813–2825.

22. Curcio MJ, Hedge AM, Boeke JD, Garfinkel DJ. Ty RNA levels determine the spectrum of retrotransposition events that activate gene expression in Saccharomyces cerevisiae. Mol Gen Genet. 1990;220(2):213–221.

23. Elder RT, St John TP, Stinchcomb DT, Davis RW, Scherer S, Davis RW. Studies on the transposable element Ty1 of yeast. I. RNA homologous to Ty1. II. Recombination

and expression of Ty1 and adjacent sequences. Cold Spring Harb Symp Quant Biol. 1981;45(Pt 2):581–591.

24. Garfinkel DJ, Boeke JD, Fink GR. Ty element transposition: reverse transcriptase and virus-like particles. Cell. 1985;42(2):507–517.

25. Nonet M, Scafe C, Sexton J, Young R. Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. Mol Cell Biol. 1987;7(5):1602–1611.

26. Malagon F, Jensen TH. The T body, a new cytoplasmic RNA granule in Saccharomyces cerevisiae. Mol Cell Biol. 2008;28(19):6022–6032.

27. Checkley MA, Nagashima K, Lockett S, Nyswaner KM, Garfinkel DJ. P-bodies are required for Ty1 retrotransposition during assembly of retrotransposition-competent virus-like particles. Mol Cell Biol. 2010;30(2):382–398.

28. Dutko JA, Kenny AE, Gamache ER, Curcio MJ. 5′ to 3′ mRNA decay factors colocalize with Ty1 Gag and human APOBEC3G and promote Ty1 retrotransposition. J Virol. 2010;84(10):5052–5066.

29. Griffith JL, Coleman LE, Raymond AS, Goodson SG, Pittard WS, Tsui C, Devine SE. Functional genomics reveals relationships between the retrovirus-like Ty1 element and its host Saccharomyces cerevisiae. Genetics. 2003;164(3):867–879.

30. Beliakova-Bethell N, Beckham C, Giddings TH, Jr, Winey M, Parker R, Sandmeyer S. Virus-like particles of the Ty3 retrotransposon assemble in association with P-body components. RNA. 2006;12(1):94–101.

31. Garfinkel DJ, Nyswaner K, Wang J, Cho JY. Post-transcriptional cosuppression of Ty1 retrotransposition. Genetics. 2003;165(1):83–99.

32. Moore SP, Liti G, Stefanisko KM, Nyswaner KM, Chang C, Louis EJ, Garfinkel DJ. Analysis of a Ty1-less variant of Saccharomyces paradoxus: the gain and loss of Ty1 elements. Yeast. 2004;21(8):649–660.

33. Monokian GM, Braiterman LT, Boeke JD. In-frame linker insertion mutagenesis of yeast transposon Ty1: mutations, transposition and dominance. Gene. 1994;139(1):9–18.

34. Youngren SD, Boeke JD, Sanders NJ, Garfinkel DJ. Functional organization of the retrotransposon Ty from Saccharomyces cerevisiae: Ty protease is required for transposition. Mol Cell Biol. 1988;8(4):1421–1431.

35. Malagon F, Jensen TH. T-body formation precedes virus-like particle maturation in S. cerevisiae RNA biology. 2011;8(2):1–6.

36. Nehrbass U, Kern H, Mutvei A, Horstmann H, Marshallsay B, Hurt EC. NSP1: a yeast nuclear envelope protein localized at the nuclear pores exerts its essential function by its carboxy-terminal domain. Cell. 1990;61(6):979–989.

37. Kemler I, Meehan A, Poeschla EM. Live-cell coimaging of the genomic RNAs and Gag proteins of two lentiviruses. J Virol. 2010;84(13):6352–6366.

38. Parker R. RNA Degradation in Saccharomyces cerevisae. Genetics. 2012;191(3):671–702.

39. Sheth U, Parker R. Targeting of aberrant mRNAs to cytoplasmic processing bodies. Cell. 2006;125(6):1095–1109.

40. Kuperwasser N, Brogna S, Dower K, Rosbash M. Nonsense-mediated decay does not occur within the yeast nucleus. RNA. 2004;10(12):1907–1915.

41. Berretta J, Pinskaya M, Morillon A. A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in S. cerevisiae. Genes Dev. 2008;22(5):615–626.

42. Houseley J, Rubbi L, Grunstein M, Tollervey D, Vogelauer M. A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. Mol Cell. 2008;32(5):685–695.

43. van Dijk EL, Chen CL, d'Aubenton-Carafa Y, Gourvennec S, Kwapisz M, Roche V, Bertrand C, Silvain M, Legoix-Ne P, Loeillet S, Nicolas A, Thermes C, Morillon A. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature. 2011;475(7354):114–117.

44. Roth J, Kingsman S, Kingsman A, Martin-Rendon E. Possible regulatory function of the Saccharomyces cerevisiae Ty1 retrotransposon core protein. Yeast. 2000;16:921–932.

45. Beliakova-Bethell N, Terry LJ, Bilanchone V, DaSilva R, Nagashima K, Wente SR, Sandmeyer S. Ty3 nuclear entry is initiated by viruslike particle docking on GLFG nucleoporins. J Virol. 2009;83(22):11914–11925.

46. Renault N, Tobaly-Tapiero J, Paris J, Giron ML, Coiffic A, Roingeard P, Saib A. A nuclear export signal within the structural Gag protein is required for prototype foamy virus replication. Retrovirology. 2011;8(1):6.

47. Hope TJ. Viral RNA export. Chem Biol. 1997;4(5):335–344.

48. Cullen BR. Nuclear mRNA export: insights from virology. Trends Biochem Sci. 2003;28(8):419–424.

49. Parent LJ, Gudleski N. Beyond plasma membrane targeting: role of the MA domain of Gag in retroviral genome encapsidation. J Mol Biol. 2011;410(4):553–564.

50. Kemler I, Saenz D, Poeschla E. Feline Immunodeficiency Virus Gag is a nuclear shuttling protein. J Virol. 2012.

51. Butterfield-Gerson KL, Scheifele LZ, Ryan EP, Hopper AK, Parent LJ. Importin-beta family members mediate alpharetrovirus gag nuclear entry via interactions with matrix and nucleocapsid. J Virol. 2006;80(4):1798–1806.

52. Kenney S, Lochmann T, Schmid C, Parent L. Intermolecular interactions between retroviral Gag proteins in the nucleus. Journal of Virology. 2008;82(2):683–691.

53. Ako-Adjei D, Johnson MC, Vogt VM. The retroviral capsid domain dictates virion size, morphology, and coassembly of gag into virus-like particles. J Virol. 2005;79(21):13463–13472.

54. Larsen LS, Beliakova-Bethell N, Bilanchone V, Zhang M, Lamsa A, Dasilva R, Hatfield GW, Nagashima K, Sandmeyer S. Ty3 nucleocapsid controls localization of particle assembly. J Virol. 2008;82(5):2501–2514.

55. Dang VD, Levin HL. Nuclear import of the retrotransposon Tf1 is governed by a nuclear localization signal that possesses a unique requirement for the FXFG nuclear pore factor Nup124p. Mol Cell Biol. 2000;20(20):7798–7812.

56. Kosugi S, Hasebe M, Tomita M, Yanagawa H. Nuclear export signal consensus sequences defined using a localization-based yeast selection system. Traffic. 2008;9(12):2053–2062.

57. Schmitt C, von Kobbe C, Bachi A, Pante N, Rodrigues JP, Boscheron C, Rigaut G, Wilm M, Seraphin B, Carmo-Fonseca M, Izaurralde E. Dbp5, a DEAD-box protein

required for mRNA export, is recruited to the cytoplasmic fibrils of nuclear pore complex via a conserved interaction with CAN/Nup159p. EMBO J. 1999;18(15):4332–4347.

58. Snay-Hodge CA, Colot HV, Goldstein AL, Cole CN. Dbp5p/Rat8p is a yeast nuclear pore-associated DEAD-box protein essential for RNA export. EMBO J. 1998;17(9):2663–2676.

59. Tseng SS, Weaver PL, Liu Y, Hitomi M, Tartakoff AM, Chang TH. Dbp5p, a cytosolic RNA helicase, is required for poly(A)+ RNA export. EMBO J. 1998;17(9):2651–2662.

60. Tran EJ, Zhou Y, Corbett AH, Wente SR. The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA:protein remodeling events. Mol Cell. 2007;28(5):850–859.

61. Hodge CA, Colot HV, Stafford P, Cole CN. Rat8p/Dbp5p is a shuttling transport factor that interacts with Rat7p/Nup159p and Gle1p and suppresses the mRNA export defect of xpo1-1 cells. EMBO J. 1999;18(20):5778–5788.

62. Fornerod M, Ohno M, Yoshida M, Mattaj IW. CRM1 is an export receptor for leucine-rich nuclear export signals. Cell. 1997;90(6):1051–1060.

63. Yedavalli VS, Neuveut C, Chi YH, Kleiman L, Jeang KT. Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. Cell. 2004;119(3):381–392.

64. Guthrie C, Fink G. Guide to Yeast Genetics and Molecular Biology. Methods in Enzymology. 1991:194.

65. Sherman F, Fink GR, Hicks JB. Methods in Yeast Genetics. Cold Spring Harbor, NY: Cold Spring Harbor Press; 1986.

66. Garfinkel DJ, Mastrangelo MF, Sanders NJ, Shafer BK, Strathern JN. Transposon tagging using Ty elements in yeast. Genetics. 1988;120(2851484):95–108.

67. Manivasakam P, Weber SC, McElver J, Schiestl RH. Microhomology mediated PCR targeting in Saccharomyces cerevisiae. Nucleic Acids Research. 1995;23:2799–2800.

68. Wach A, Brachat A, Pohlmann R, Philippsen P. New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. Yeast. 1994;10(13):1793–1808.

69. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, Tsai IJ, Bergman CM, Bensasson D, O'Kelly MJ, van Oudenaarden A, et al. Population genomics of domestic and wild yeasts. Nature. 2009;458(7236):337–341.

70. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics. 1989;122(1):19–27.

71. Matsuda E, Garfinkel DJ. Posttranslational interference of Ty1 retrotransposition by antisense RNAs. Proc Natl Acad Sci U S A. 2009;106(37):15657–15662.

72. Nyswaner KM, Checkley MA, Yi M, Stephens RM, Garfinkel DJ. Chromatin-associated genes protect the yeast genome from Ty1 insertional mutagenesis. Genetics. 2008;178(1):197–214.
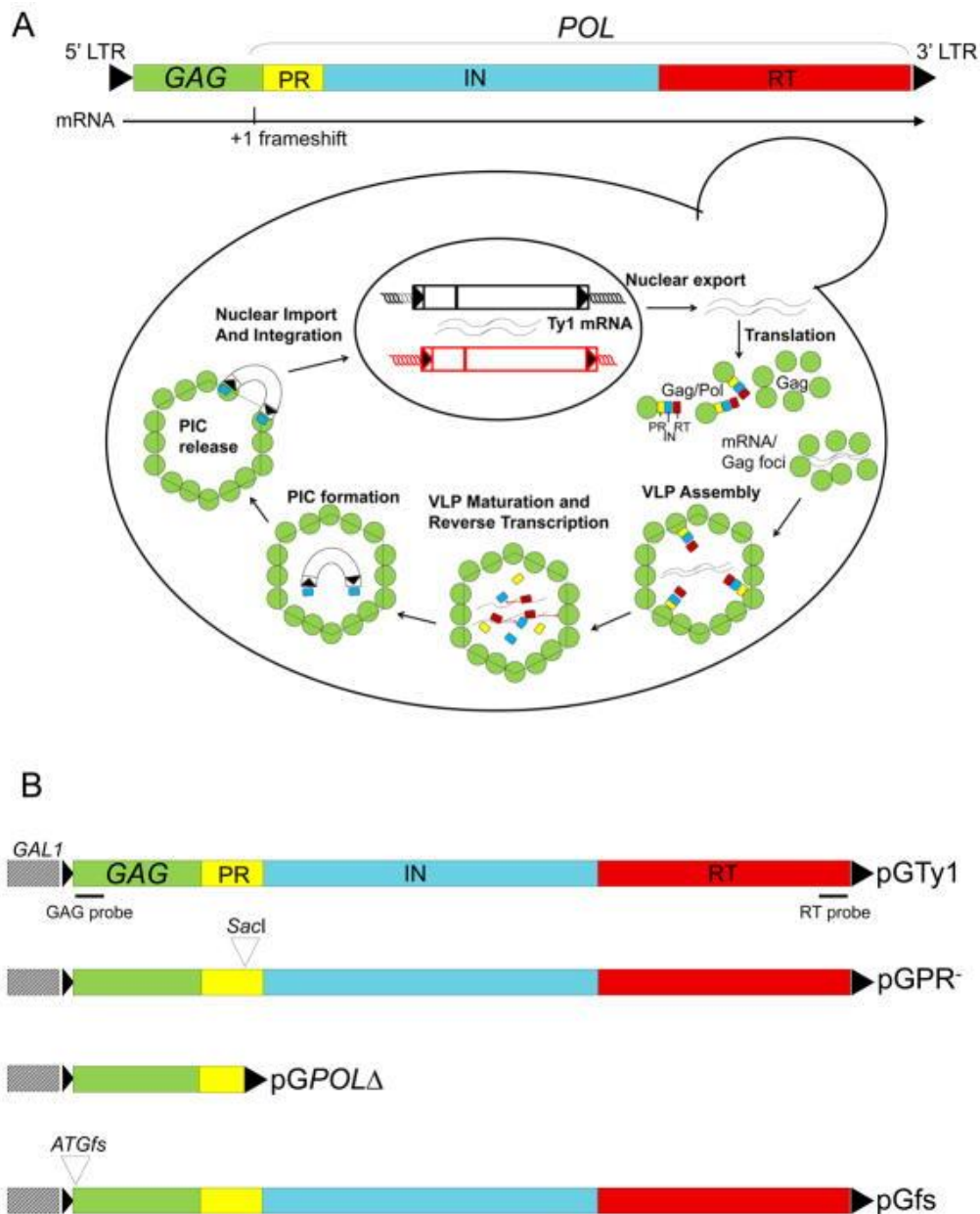
Figure 2.1. Ty1 retrotransposition. (A) The functional organization of Ty1 showing the Ty1 mRNA and the site of the +1 ribosomal frameshift is presented at the top. The Ty1 life cycle is depicted below. Ty1 is transcribed by RNA pol II to form a 5.7 kb transcript, which functions as either gRNA for packaging into VLPs or mRNA for translation into Gag-p49 or Gag-Pol-p199. Gag accumulates in the cytoplasm to become the capsid

protein of Ty1 VLPs. POL encodes the enzymes required for Ty1 transposition: PR, IN, and RT. Translation of POL requires a +1 ribosomal frameshift at the 3′ end of GAG yielding Gag-Pol-p199. Prior to VLP formation, Gag and presumably Gag-Pol form mRNA/Gag foci, which are sites where VLPs cluster and may be required for assembly. Ty1 mRNA is specifically packaged as a dimer into VLPs. Ty1 PR cleaves Gag-p49 and Gag-Pol-p199 precursors to form mature proteins within VLPs, such as Gag-p45 and IN-p71. An IN/RT heteromer reverse transcribes Ty1 mRNA into cDNA. A pre-integration complex (PIC) containing Ty1 cDNA and IN is imported into the nucleus via a nuclear localization signal present on IN. Retrotransposition is completed by IN-mediate insertion into a new location in the host genome. (B) pGTy1 plasmids. Overexpression plasmids contain Ty1 fused to the *GAL1* promoter in a 2µ-based multicopy shuttle plasmid. The position of the hybridization probes for FISH-IF is shown. The following pGTy1 mutants were used in this study: pGPR−, which renders PR enzymatically inactive, pGPOLΔ, which encodes an inactive PR and lacks RT and IN, and pGfs, which does not yield Ty1 proteins.
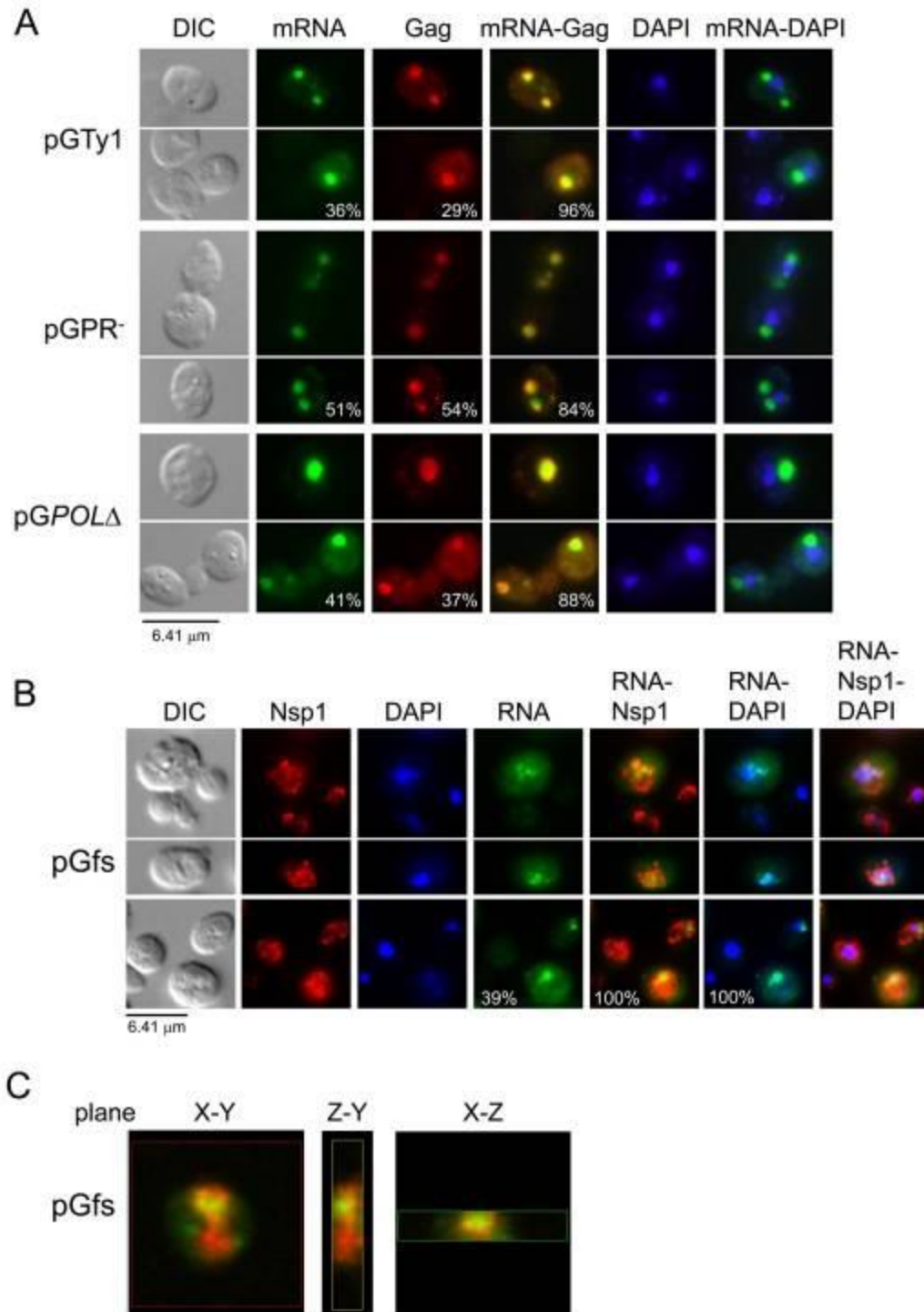
Figure 2.2 Localization of Ty1 mRNA into discrete cytoplasmic foci requires Gag. Cells lacking chromosomal Ty1 elements but containing wild type or mutant pGTy1 plasmids were induced with galactose for 16 hr at 20°C and visualized by differential interference

microscopy (DIC). Ty1 mRNA and Gag were detected by FISH-IF, and DNA was stained with DAPI. Ty1 mRNA was detected using a GAG-DIG probe and Gag was detected with anti-Ty1 VLP antiserum. **(A)** Ty1 mRNA and Gag colocalized in discrete cytoplasmic foci in cells expressing pGTy1 (wild type), and mutants pGPR⁻ and pG*POL*Δ. The merge of Ty1 mRNA/Gag and Ty1 mRNA/DAPI stains is shown. Percentages of cells from the same population containing cytoplasmic Ty1 mRNA and Gag foci, and the propensity of these to colocalize are indicated. **(B)** Nuclear retention of Ty1fs RNA. Anti-Nsp1p antibody was used to delineate the nuclear envelope. The merges of Ty1 RNA/Nsp1p/DAPI are on the right. Percentages of cells containing Ty1 RNA foci, Ty1 RNA foci within the boundaries of Nsp1p staining, and the percentages of Ty1 RNA/DAPI overlap are shown. Only Nsp1p positive cells were counted. **(C)** Nuclear localization of Tyfs RNA observed by confocal microscopy. The X-Y, Y-Z, and X-Z planes for one focal point are shown. Ty1fs RNA is shown in green and the nuclear envelope is in red.
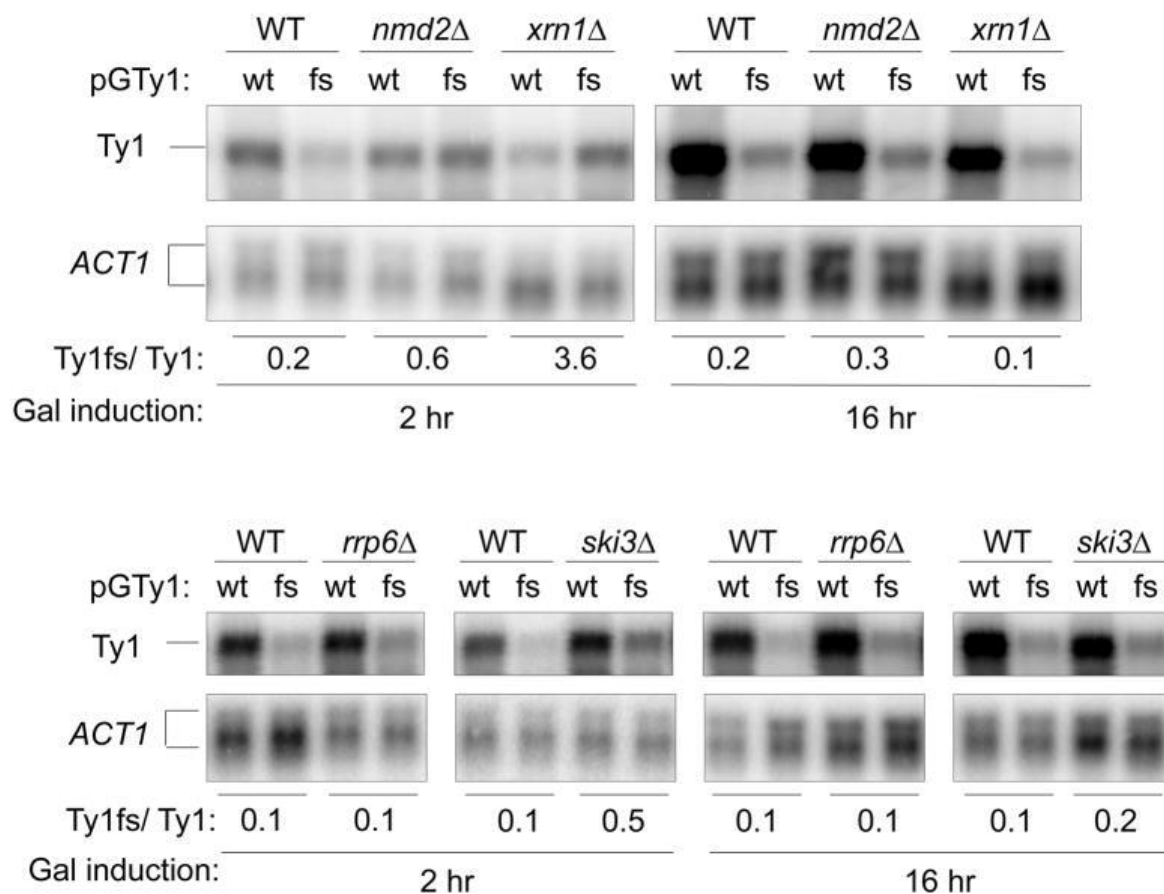
Figure 2.3. Level of wild type Ty1 (WT) and Ty1fs (fs) RNAs in *nmd2Δ, xrn1Δ, rrp6Δ,* and *ski3Δ* mutants. Total RNA was extracted from wild type (+) and mutant (Δ) strains after galactose induction for 2 and 16 hr. RNA was analyzed by Northern blot using ^{32}P-labeled Ty1 and *ACT1* riboprobes. Both the unspliced and spliced *ACT1* transcripts were quantified. The Ty1/*ACT1* ratios were normalized to the wild type control (Table 2.1), which is considered to be 1. The relative amount of Ty1fs RNA/Ty1 mRNA is shown at the bottom.
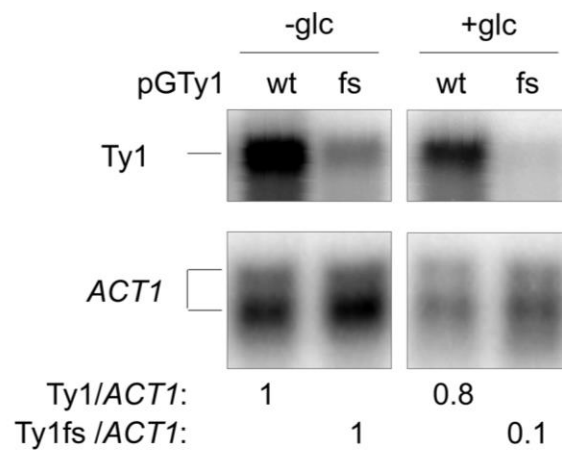
Figure 2.4. Stability of wild type Ty1 (+) and Ty1fs (fs) transcripts after glucose (Glc) shut off. Total RNA was extracted from wild type strains after galactose induction for 16 hr (lanes 1 and 2), followed by incubation for 1 hr in 2% glucose (lanes 3 and 4). RNA was subjected to Northern analysis as described in the Figure 2.3 legend.
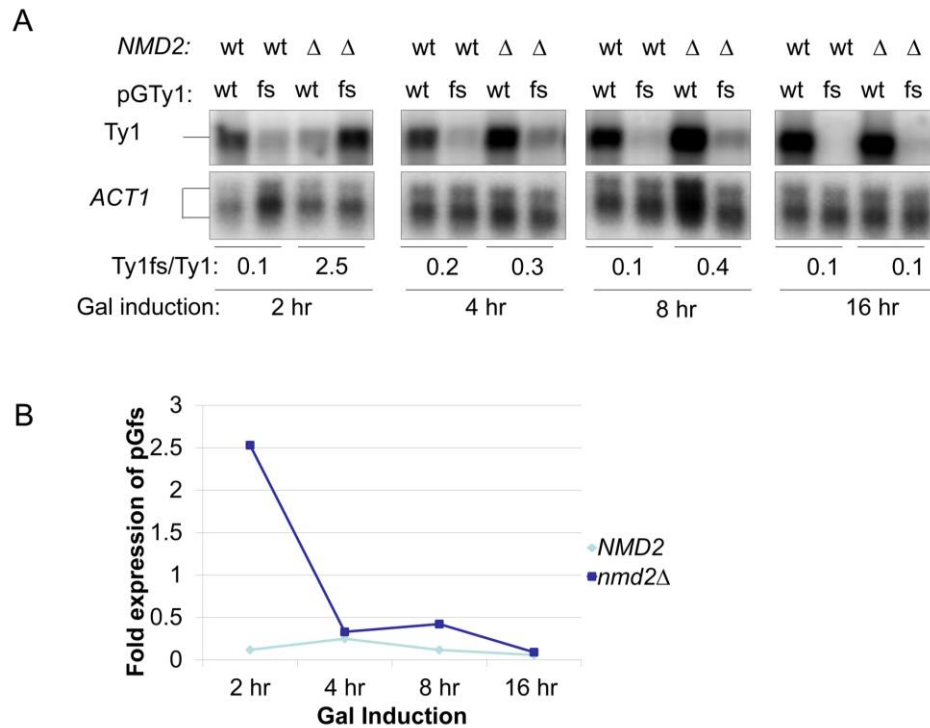
Figure 2.5. Induction of wild type Ty1 (+) and Ty1fs (fs) mRNA in an *nmd2Δ* mutant. (A) Total RNA was extracted from wild type (+) and *nmd2Δ* (Δ) strains after galactose induction for 2 (lanes 1-4), 4 (lanes 5-8), 8 (lanes 9-12) and 16 hr (lanes 13-16). RNA was analyzed as described in the Figure 2.3 legend. (B) The fold-expression of pGfs RNA normalized to wild type pGTy1 mRNA for *NMD2* (diamond) and *nmd2Δ* (square) strains is plotted against the length of galactose induction in hours.
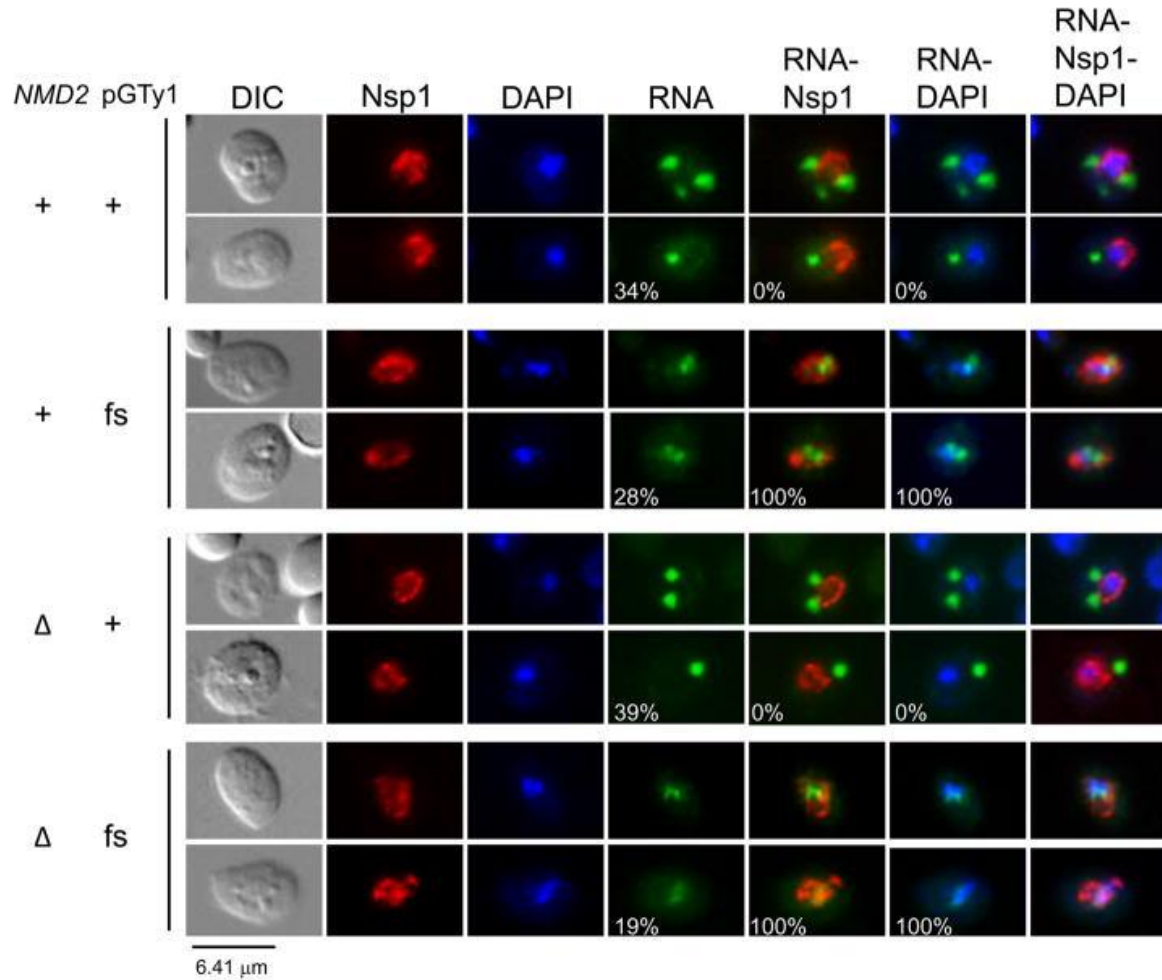
Figure 2.6. Nonsense Mediated Decay does not affect nuclear retention of Ty1fs RNA.
Ty1 RNA and Gag localization in an *nmd2Δ* mutant strain containing either wild type pGTy1 or pGfs were analyzed by FISH-IF. DNA was stained with DAPI, Ty1 RNA was detected using a GAG-DIG probe, and Nsp1p was detected with anti-Nsp1p antiserum. Representative images illustrating the percentages of cells containing Ty1 RNA foci (cytoplasmic or nuclear) and overlaps between Ty1 RNA-Nsp1p and Ty1 RNA-DAPI, are shown.
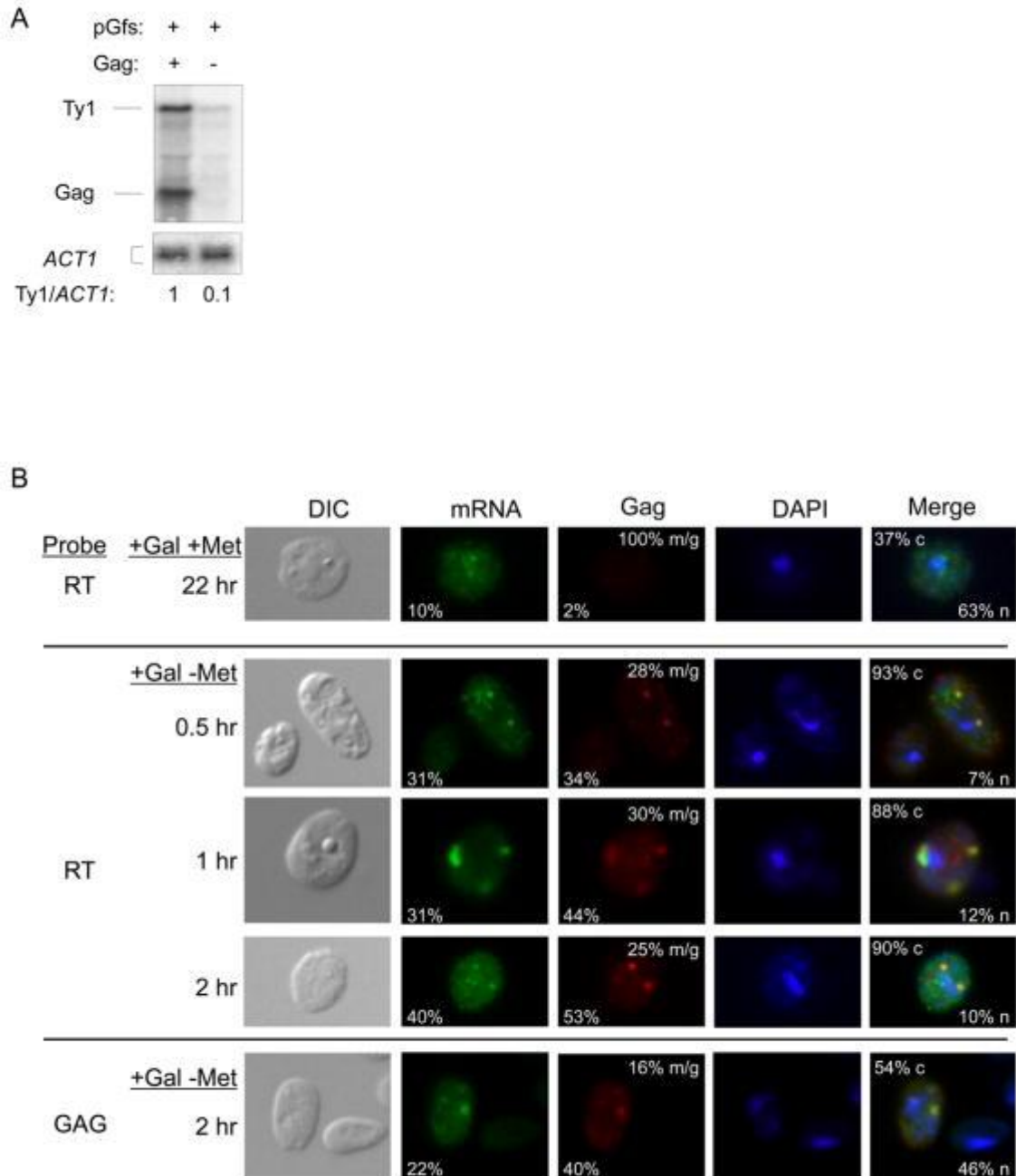
Figure 2.7. Gag provided *in trans* restores Ty1fs mRNA level, nuclear export, and formation of cytoplasmic mRNA/Gag foci. (A) Northern analysis of Ty1fs RNA in the presence or absence of Gag expressed from pMPOLΔ. Total RNA was extracted from strains expressing pGfs (expresses Ty1fs RNA) and Gag (pMPOLΔ: expresses a transcript encoding Gag) or pGfs and an empty vector (−) induced with galactose for 48

hr. RNA was hybridized with ^{32}P-labeled riboprobes corresponding to Ty1 GAG and *ACT1*. Note that the Ty1 GAG probe will detect both the Ty1fs RNA (5.7 kb) and GAG mRNA (2 kb) from pMPOLΔ. Ty1/*ACT1* ratios were estimated as described in Figure 2.3. (B) Ty1fs RNA dynamics when Gag is provided *in trans*. Cells containing pMPOLΔ and pGfs were grown in media containing galactose (+Gal) and methionine (+Met) to induce pGfs only, or lacking methionine (+Gal − Met) to induce both plasmids, for indicated time lengths prior to FISH-IF analysis. Ty1 mRNA was detected with either GAG-DIG or RT-DIG probes as indicated. Note that the GAG-DIG probe will detect both Ty1fs mRNA and truncated POLΔ mRNA, while the RT-DIG probe will only detect Ty1fs RNA (Figure 2.1B). Gag was detected with an anti-Ty1 VLP antiserum. Representative images shown illustrate the percentages of cells containing Ty1 mRNA or Gag foci, propensities for mRNA/Gag (m/g) overlap, and percentages of cytoplasmic (c) or nuclear (n) localization of Ty1 mRNA.
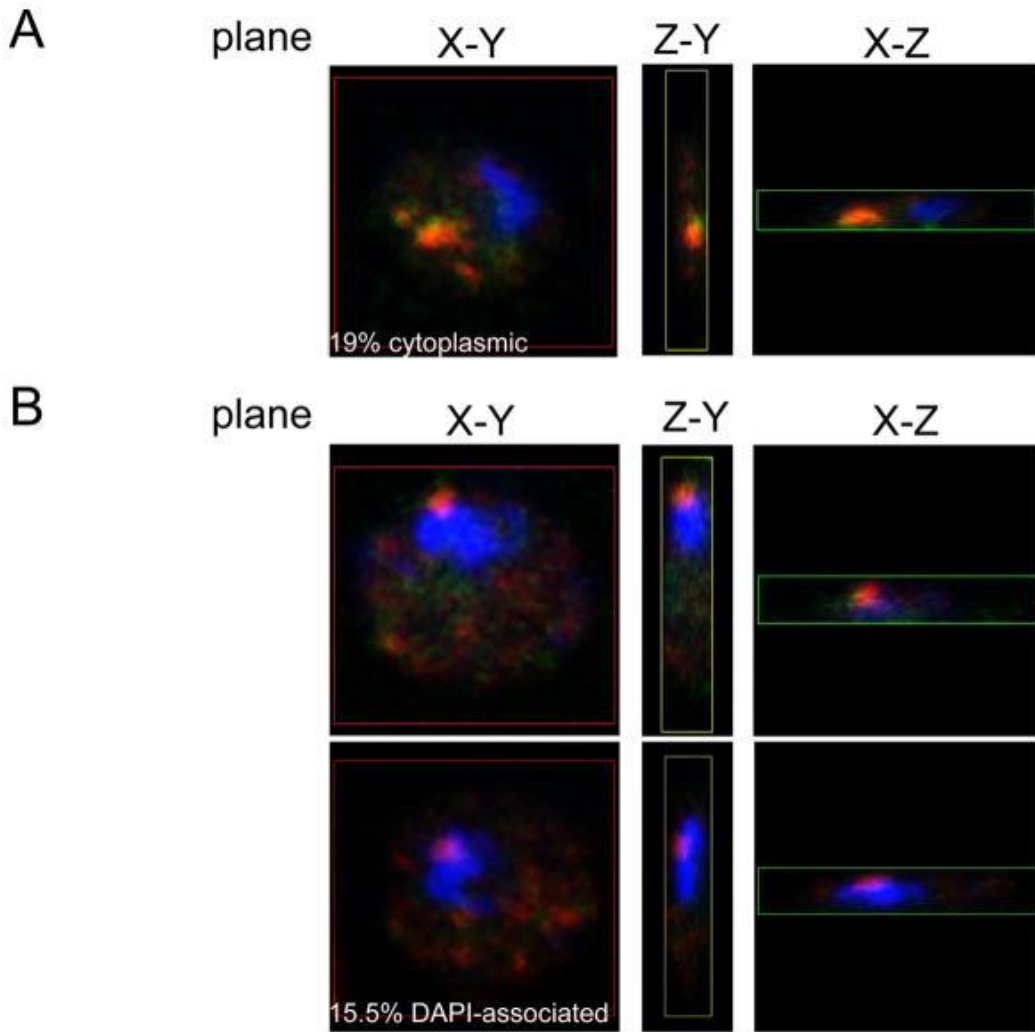
Figure 2.8. *Mex67-5* cells exhibit perinuclear mRNA-Gag foci. Temperature-sensitive *mex67-5* cells were grown for 20 hr at 20°C and then shifted to 30°C for 1–2 hr. Cells were processed for FISH-IF and images were taken with a confocal microscope. Ty1 mRNA was detected with an RT-DIG probe (shown in green) and Gag was detected with an anti-Ty1 VLP antiserum (shown in red). DNA was stained with DAPI (shown in blue). Representative images of the X-Y, Y-Z, and X-Z planes for one focal point of each cell with a type of Gag localization: (A) cytoplasmic, and (B) association with DAPI stained material.
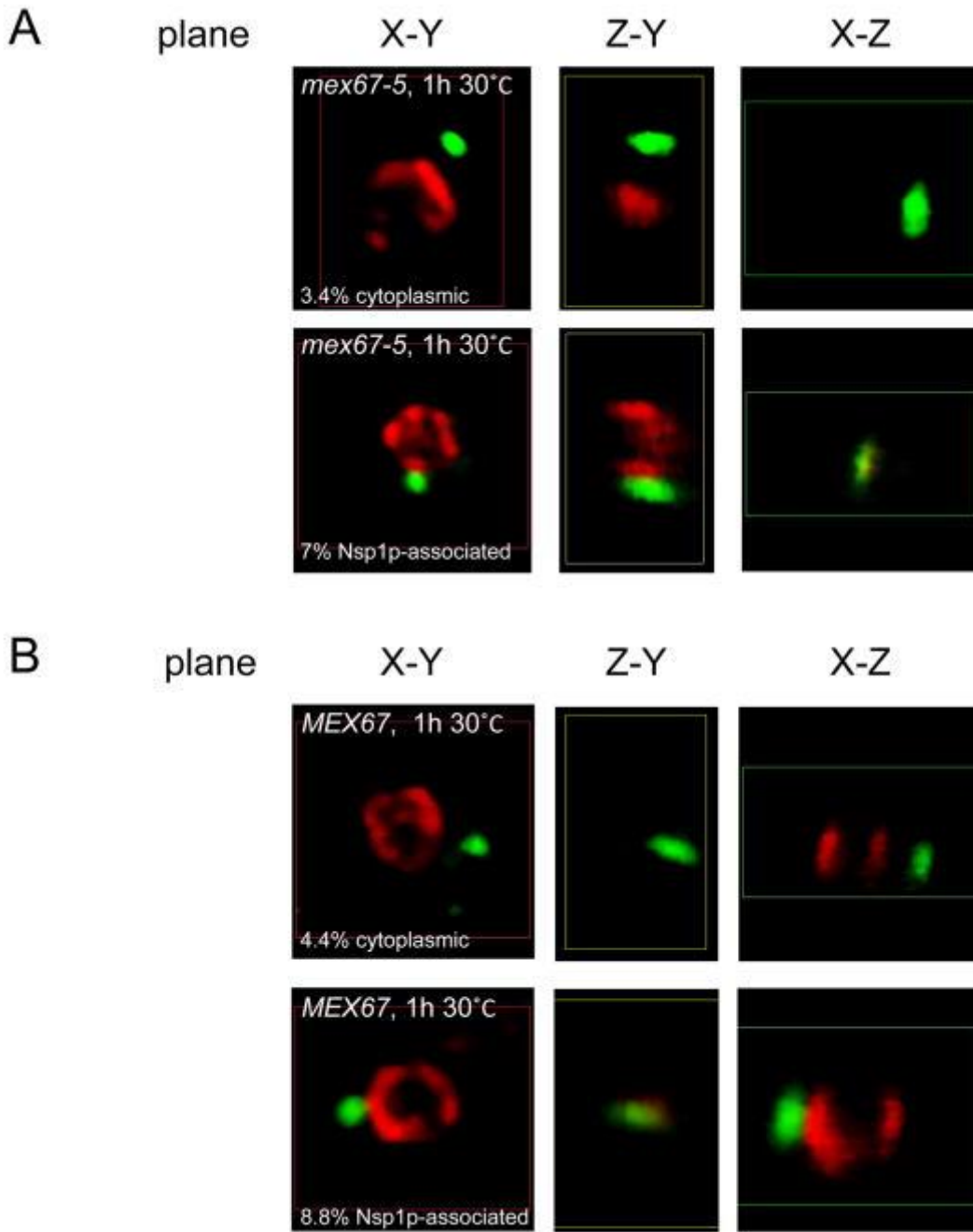
Figure 2.9. *Mex67-5* and wild type cells exhibit Gag foci in close association with the nuclear envelope. Temperature-sensitive *mex67-5* cells (A) and isogenic *MEX67* cells (B) were grown for 20 hr at 20°C and then shifted to 30°C for 1 hr. Cells were processed for IF and images were taken with a confocal microscope. Ty1 Gag was detected with

an anti-Ty1 VLP antiserum (shown in green) and Nsp1p was detected with anti-Nsp1p antiserum (shown in red). Representative images of the X-Y, Y-Z, and X-Z planes for one focal point of each cell were assembled in MIPAV. Examples of Gag localization in the cytoplasm or associated with Nsp1p are shown for each strain.

Tables

Table 2.1 Quantification of Northern hybridization signals from wild type, *nmd2Δ, xrn1Δ, rrp6Δ,* and *ski3Δ* strains. Relative level of wild type Ty1 and Ty1fs transcripts in *nmd2Δ, xrn1Δ, rrp6Δ, and ski3Δ* mutants. Northern hybridizations from wild type (WT) and mutant (Δ) strains (Figure 2.3) were subjected to phosphoimage analysis to estimate the level of each Ty1 transcript normalized to *ACT1*.

	Ty1/*ACT1*		Ty1fs/*ACT1*		Ty1fs/Ty1	
Gal induction (hr):	2	16	2	16	2	16
WT	1	1	1	1	0.2	0.2
nmd2Δ	0.8	0.7	3.1	1	0.6	0.3
xrn1Δ	0.1	0.6	3.1	0.4	3.6	0.1
WT	1	1	1	1	0.1	0.1
rrp6Δ	1.2	0.7	0.8	1.4	0.1	0.1
WT	1	1	1	1	0.1	0.1
ski3Δ	1.1	0.4	6.9	0.9	0.5	0.2

Table 2.2 Oligonucleotide sequences

Oligonucleotide   Sequence

pMET25-F616-EcoRI	CGGAATTCTTCGGATGCAAGGGTTCGAAT
pMET25-R1000-XhoI	CCGCTCGAGTGTATGGATGGGGGTAATAG
NMD2SparKanMXB	gcacgacttaaatacccgagcttggaatggcgaggaagtctttcccCGTACGCTGCAGGTCGACGGATC♦
NMD2SparKanMXT	ttagtcaaaagatcgttttaaaacaatctttttaatcttgtctcgATCGAT*GAATTC*GAGCTCGT♦
SKI3ScerA	TACTAATCAGACCATCCTCATCCTC
SKI3ScerD	GAACGACCTACTATTACCAAGCGAA
RRP6SparKanMXT	atggcttctaaaaatcaggatgtacttctatctagggtgaCGTACGCTGCAGGTCGACGGATC♦
RRP6SparKanMXB	aaatgacacattcttacctttggcggcaggtctcctctttttagctgcccATCGATGAATTCGAGCTCGT♦
Xrn1SparKanMX-W	ggaacacagattcctgagtttgataacttatacctggatatgaatCGTAGCTGCAGGTCGACGGATC♦
Xrn1SparKanMX-C	atgaattttttcaaatcttggctttccttttcattgacaacaataATCGATGAATTCGAGCYCGT♦
antisense TyA-322	GACTTCCTTAGAAGTAACCGAAGCACAGGCGCTACCATGAGAATTGGGTG
antisense RT-4219	TGATGTACGGTATTGGATTGCATGCCTGAGTCGTAAGTGTCAGGATGCTG

♦Open reading frame adjacent to initiation and stop codons: lower case, *KanMX*: upper case

CHAPTER 3

A *TRANS*-DOMINANT FORM OF GAG RESTRICTS TY1 RETROTRANSPOSITION

AND MEDIATES COPY NUMBER CONTROL[2]

Abstract

Saccharomyces cerevisiae and *S. paradoxus* lack the conserved RNA interference pathway and utilize a novel form of copy number control (CNC) to inhibit Ty1 retrotransposition. Although noncoding transcripts have been implicated in CNC, here we present evidence that a truncated form of the Gag capsid protein (p22) or its processed form (p18) is necessary and sufficient for CNC and likely encoded by Ty1 internal transcripts. Coexpression of p22/p18 and Ty1 decreases mobility more than 30,000-fold. p22/p18 cofractionate with Ty1 virus-like particles (VLPs) and affect VLP yield, protein composition and morphology. Although p22/p18 and Gag colocalize in the cytoplasm, p22/p18 disrupt sites used for VLP assembly. GST affinity pull-downs also suggest that p18 and Gag interact. Therefore, this intrinsic Gag-like restriction factor confers CNC by interfering with VLP assembly and function, and expands the strategies used to limit retroelement propagation.

Importance

Retrotransposons dominate the chromosomal landscape in many eukaryotes, can cause mutations by insertion or genome rearrangement and are evolutionarily related to retroviruses such as HIV. Thus, understanding factors that limit transposition and retroviral replication are fundamentally important. The present manuscript describes a retrotransposon-encoded restriction protein derived from the capsid gene of the yeast Ty1 element that disrupts virus-like particle assembly in a dose dependent manner. This form of copy number control acts as a molecular rheostat, allowing high levels of

retrotransposition when few Ty1 elements are present and inhibiting transposition as copy number increases. Thus, yeast and Ty1 have co-evolved a form of copy number control that is beneficial to both "host and parasite". To our knowledge, this is the first Gag-like retrotransposon restriction factor described in the literature and expands the ways restriction proteins modulate retroelement replication.

Introduction

Retrovirus-like retrotransposons and their long terminal repeat (LTR) derivatives inhabit the genomes of many organisms, including the budding yeast *Saccharomyces cerevisiae* and its closest relative *S. paradoxus*. The Ty1 family is active and related to the LTR retrotransposons Ty2-Ty5 in budding yeast (1). Variations in Ty1 copy number can be attributed to the relative rates of transposition, loss by LTR-LTR recombination or additional types of genome rearrangements, all of which can impact fitness (2-5). Ty1 resembles retroviruses in genome organization and replication (1). These elements consist of two overlapping open reading frames, *GAG* and *POL*, which are flanked by LTRs. Ty1 genomic RNA is translated or packaged as a dimer into virus-like particles (VLPs). The primary translation products are Gag (p49) and Gag-Pol (p199) precursors, the latter resulting from a +1 ribosomal frameshift during translation. Mature Gag (p45) is the major structural component of VLPs. *POL* encodes the enzymes required for proteolytic processing of Gag and Gag-Pol (protease; PR), cDNA integration (integrase; IN), and reverse transcription (reverse transcriptase; RT). Ty1 and Ty3 VLPs assemble within cytoplasmic foci, termed retrosomes or T-bodies, which contain Ty proteins and RNA (6-9). Once VLPs undergo maturation via the action of PR, Ty1 genomic RNA is

127

reverse transcribed to form a linear cDNA. A protein/DNA complex minimally containing Ty1 cDNA and IN are imported into the nucleus, where integration usually occurs near genes transcribed by RNA polymerase III.

S. cerevisiae and S. paradoxus laboratory strains and natural isolates contain fewer than 40 copies of Ty1 per haploid genome and several strains contain few if any elements (5, 10-14). Although budding yeast genomes characterized to date tend to have low Ty1 copy numbers, fertile S. cerevisiae strains containing more than 100 Ty1 insertions have been created artificially by numerous rounds of induction of a multi-copy plasmid containing an active Ty1 element (Ty1H3) fused to the GAL1 promoter (pGTy1) (15, 16). Host cofactor and restriction genes involved in modulating Ty1 retrotransposition are diverse and encompass different steps in the replication cycle, ranging from transcription to integration site preference (17-21). For example, SPT3 is required for transcription of full-length Ty1 mRNA (22) and encodes a component of the SAGA chromatin-remodeling complex (23), and XRN1 is an important Ty1 cofactor implicated in transcription (24, 25), assembly of functional VLPs (7, 8), and encodes a 5'- 3' exonuclease required for mRNA turnover (26).

Transposon-derived regulatory factors are critically important for keeping transposition at a low level. Forms of RNA interference affect the level or utilization of transposon mRNA, and the source of the interfering RNAs can be the transposons themselves (27). A unique form of copy number control (CNC) minimizes Ty1 transposition in S. cerevisiae and S. paradoxus (28) in the absence of dicer and argonaute genes that comprise a functional RNAi pathway in a distant species S. castellii (29, 30). Ty1 CNC is defined by a copy number dependent decrease in Ty1

128

retrotransposition and is especially robust in a "Ty1-less" strain of *S. paradoxus* (28) that may have lost Ty1 elements by LTR-LTR recombination (13). Ty1 CNC acts posttranslationally and *in trans*, can be overcome by pGTy1 expression, and is characterized by lower levels of mature IN and reverse transcripts (28, 31, 32). Reduced levels of endogenous Ty1 IN, PR, cDNA, and VLPs are also present in *S. cerevisiae* (33-35), which displays CNC (28). These results suggest that Ty1 produces a titrateable factor that inhibits transposition in a copy dependent manner.

Ty1 antisense (Ty1AS) RNAs have been implicated in silencing Ty1 expression by alterations in chromatin function (24) or when RNAi is reconstituted in *S. cerevisiae* (30). We reported evidence suggesting that Ty1AS RNAs interfere with Ty1 transposition posttranslationally (31). Inhibition occurs in a copy number dependent manner and the antisense transcripts map to a region within *GAG* that confers CNC (28). Deleting the common 3' end of the antisense transcripts abolishes CNC and decreases the level of Ty1AS RNAs. However, ectopic expression of individual antisense transcripts does not restore CNC, suggesting that either multiple antisense transcripts or additional factors are required (24, 31). Also, nuclease protection and structural probing analyses suggest that although Ty1AS RNAs specifically associate with VLPs from CNC$^+$ strains, these transcripts are not packaged into VLPs and do not interact with Ty1 mRNA (32).

Here, further characterization of the minimal Ty1 sequence that confers CNC has led to the discovery of p22, an N-terminal truncated form of Gag that is likely encoded by an internally initiated Ty1 mRNA. Importantly, p22 is both necessary and sufficient for CNC. Coexpression of p22 and Ty1 interferes with assembly of functional VLPs,

which is conceptually similar to the inhibition displayed by Gag-like restriction factors derived from endogenous retroviruses in mammals (36, 37).

Materials and Methods

Genetic techniques, media and strain construction

Strains are listed in Table 3.1. Strains repopulated with Ty1 elements were obtained following pGTy1 induction as described previously (28). Standard yeast genetic and microbiological procedures were used in this work (38).

Plasmids

All nucleotide information used here corresponds to Ty1H3 sequence (39) (Genbank M18706.1). pGPOLΔ derivatives of pGTy1 were generated by digestion with *Bgl*I and ligation. pBJM78, pBJM79 and pBDG1595 were constructed by overlap PCR using flanking primers (Ty335F, 5'-TGGTAGCGCCTGTGCTTCGGTTAC-3'; TyRP1, 5'-CATTGATAGTCAATAGCACTAGACC-3') and overlapping primers (DELC1071b, 5'-GGTATCAGATTCATTTTTTCAATACTTTTGGAAAGAATTTTC-3'; DELC1071c, 5'-GTATTGAAAAAATGAATCTGATACCCAAGAGGCAAACGAC-3'; ADDA1303b, 5'-GAACAGTTCATGCGACTGTCATATTTAGATGTCGATGACGTG-3'; ADDA1303c, 5'-CTAAATATGACAGTCGCATGAACTGTTCTTAGATATCCATGC-3; B-AUG1Ala-R, 5'-AAAGAATTTTCGCGATATCCGTATAATCAACG-3'; C-AUG1Ala-F, 5'-GGATATCGCGAAAATTCTTTCCAAAAGTATTG-3'; B-AUG2Ala-R, 5'-TATCAGATTGCGCTTTTTCAATACTTTTGG-3'; C-AUG2Ala-F, 5'-TGAAAAAGCGCAATCTGATACCCAAGAGGC-3') and Ty1H3 as template. Final PCR products were subcloned into pGPOLΔ using *Bst*XI and *Bgl*I restriction sites. Plasmid

pBDG1534 was generated from plasmid pBDG606 (pGTy1*his3-AI*/Cen-*URA3*) (18) by replacing the *URA3* marker for *TRP1*. Briefly, *TRP1* was amplified from BY4742 with primers containing flanking *URA3* sequence (20718uratrpfwd, 5'-ATGTCGAAAGCTACATATAAGGAACGTGCTGCTACTCATCAATTCGGTCGAAAAAA GAAA-3'; 20916uratrprev, 5'-AGCTTTTTCTTTCCAATTTTTTTTTTTTCGTCATTATAATATGCTTGCTTTTCAAAAGG C-3') and the PCR product was cotransformed into yeast with pBDG606 linearized within *URA3* with *Apa*I. Transformants were selected on SC–Trp and plasmids were verified phenotypically and by restriction mapping. Plasmid pBDG1565 was created by PCR-amplifying Ty1 *GAG* coding sequence (nt 1038-1613; EcoRIstartF, 5'-CATGTTTCGAATTCATGAAAATTCTTTCCAAAAGTATTG-3'; Xho1stopR, 5'-CATGTTTCCTCGAGTTAGTAAGTTTCTGGCCTAAGATGAAG-3) using Ty1H3 as a template and cloning into pYES2 (Life Technologies, Carlsbad, CA) using *Eco*RI and *Xho*I. Plasmid pBDG1568 was made in a similar manner as pBDG1565, except an initial PCR step was performed to insert V5 coding sequence (underlined) in-frame between Ty1 *GAG* nt 1442 and 1443 (V51442b, 5'-<u>CGTAGAATCGAGACCGAGGAGAGGGTTAGGGATAGGCTTACC</u>TATAACTTTGGGT TTGGT-3'; V51442c, 5'-<u>GGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACG</u>GCTCGGAATCCTCA AAAA-3'). For plasmid pBDG1571, *GAG* coding sequence cloned into pYES2 ended at nt 1496 (1496XhoI, 5'-CATGTTTCCTCGAGTTAGTGAGCCCTGGCTGTTTCG-3'). The GAG*PR mutation was created by mutating the Gag-PR cleavage site (RAHNVS) to AAGSAA (40) using overlapping primers (Gag*PRb, 5'-

AGCCGCTGCTGGATCCGCTGCTACATCTAATAACTCTCCCAGC-3'; Gag*PRc, 5'-GATGTAGCAGCGGATCCAGCAGCGGCTGTTTTCGATTTCGAAT-3'). To construct the *GAL1*-promoted GST-p18 protein fusion, the coding region for p18 (1038-1496) was amplified with *Xba*I and *Hind*III primer sets (1038XbaI, 5'-CTAGTCTAGACATGAAAATTCTTTCCAAAAGTATTG-3'; 1496XbaI, 5'-CCCAAGCTTTTAGTGAGCCCTGGCTGTTTTCG-3'). The PCR fragment was cloned into pEG(KT) (41) yielding pBDG1576. All plasmids generated by PCR cloning were verified by DNA sequencing. Phusion DNA polymerase, T4 DNA ligase, and restriction enzymes were obtained from New England BioLabs (Ipswich, MA).

Random mutagenesis and gap repair

The Ty1 CNC region was mutagenized by amplification with Taq DNA polymerase (ThermoFisher Scientific, Waltham, MA) using PCR forward primer

FP1 (5'-CTCCGTGCGTCCTCGTCTTCACC-3') and reverse primer RP1 (5'-CATTGATAGTCAATAGCACTAGACC-3'). Gel purified PCR product was cotransformed into DG2196 along with a multicopy pGTy1 plasmid gapped with *Xho*I and *Bst*EII. Gap-repaired transformants were selected on SC-Ura medium. Plasmids recovered from the CNC⁻ strains were introduced into DG2196 to verify loss of CNC and then subjected to DNA sequencing. Aligning mutant sequences with Ty1H3 using ClustalW2 identified point mutations.

Ribosome footprint profiling analysis of chromosomal Ty1 elements

Samples were prepared and ribosome footprint profiling (Ribo-seq) was performed as previously described (42). Briefly, *S. cerevisiae* strain Sigma 1278b (YWG025; MATa *ura3 leu2 trp1 his3*) was grown to OD600 ~1.0-1.1 at 30˚C in YPD media, spun down,

and resuspended in pre-warmed YPA (no glucose) media. After 3 hr in YPA media, cycloheximide was added to a final concentration of 0.1mg/ml and cells harvested by centrifugation. Cells were lysed in 1xPLB (20 mM HEPES-KOH, pH 7.4, 2 mM Mg(OAc)2, 100 mM KOAC, 1% Triton-X 100, 0.1 mg/mL cycloheximide, 3 mM DTT), and libraries prepared essentially as described (43). Reads were mapped to Ty1H3 using the STAR RNA-seq alignment (44), allowing for zero mismatches. The Ty1 reads represent a composite of all Ty1 elements in the genome, including partial elements such as solo LTRs. No attempts were made to sort multiple mapping reads. The abundance of 5'end reads are displayed over Ty1 using custom scripts available upon request. Libraries used for these analyses include NCBI GEO accession numbers SRX264202 and SRX366898 (Sigma Ribo-seq).

<u>Isolation of cDNA clones</u>

A *S. cerevisiae* cDNA expression library fused to the *GAL1* promoter on a centromere-based *URA3* vector (45) was introduced into DG2196. Approximately 5,000 primary transformants were replica plated to SC-Ura + 2% galactose and incubated for 3 days at 30°C. Colonies were then replica plated to SC-His-Ura and Ty1*HIS3* papillae were scored after incubation for 3 days at 30°C. Galactose induction was performed at a suboptimal temperature for transposition to sensitize the screen, since induction at 22°C resulted in too many Ty1*HIS3* papillae. Most transformants yielded about 5 Ty1*HIS3* mobility events/colony. Thirty-three transformants that had a lower level of Ty1 mobility were retested. Plasmids from 7 transformants were recovered in *E. coli* and sequenced from their 5' and 3' ends using *GAL1*- and vector-specific primers. *CCW12* (cell wall mannoprotein), *MSS4* (phosphatidylinositol-4-phosphate 5-kinase), *MRH1* (membrane

protein), *RGD1* (GTPase-activating protein), *TIR1* (cell wall mannoprotein), and *WHI5* (repressor of G1 transcription) were recovered as partial or complete cDNA clones, and were not studied further. One clone (pBDG1354) contained Ty1 sequences from nt 1042-5889 and conferred a strong *trans*-dominant negative inhibition of Ty1*his3-AI* mobility.

RNA isolation

Cultures were grown at $22^{o}C$ for 24 hr in SC or YEPD media. Total RNA was extracted using the MasterPure yeast RNA purification kit (Epicenter Biotechnologies, Madison, WI) following the manufacturer's protocol with minor modifications; 400 μl RNA extraction reagent and 200 μl of MPC protein reagent was used instead of 300 μl and 160 μl respectively. Poly(A)$^+$ RNA was isolated from $\sim$ 250 μg total RNA using the NucleoTrap mRNA purification kit (Clontech, Mountainview, CA).

Northern blotting

RNA was resolved on a 1.2% agarose-formaldehyde gel at 120 V for 2 hr and blotted onto Hybond-XL N (GE Healthcare, Little Chalfont, United Kingdom). Riboprobes were transcribed *in vitro* from Ty1 *GAG* and *ACT1* coding sequence using a MAXIscript kit (Life Technologies) and uniformly labeled with [α-^{32}P] UTP (3,000 Ci/mmol; Perkin Elmer, Waltham MA). Hybridization and phosphorimage analysis was carried out as previously described (19, 28).

5'RACE

200 ng of poly(A)$^+$ RNA was used for synthesis of cDNA library using the SMARTer PCR cDNA Amplification Kit (Clontech). This method is 5' cap independent and the

library included cDNA from all poly(A)$^+$ transcripts. Ty1 specific cDNA was amplified with the gene specific primer

GSP1_3389 (5'-GACATGGGAGCAAGTAAAGGAAC-3') and the universal primer mix from the supplier. RACE products were resolved on a 1% agarose gel. Gel purified DNA fragments were TA-cloned into pCR2.1-TOPO vector (Life Technologies). Plasmid DNA was subjected to DNA sequencing using Ty1 specific sequencing primer (Ty1new2rev; 5'GAGAATCATTCTTCTCATCACTCG-3').

qPCR

The number of Ty1A1123G transposition events in strain DG3798 was estimated by qPCR. Strains DG2196 (Ty1*his3-AI*), DG2512 (Ty1*his3-AI* + 9 additional Ty1 elements), and DG2511 (Ty1*his3-AI* + 12 additional Ty1 elements) were used as standards, based on results from Southern analysis (28) (Ahn and Garfinkel, unpublished results). Duplicate samples were subjected to qPCR using IQ SYBR Green Supermix (Bio-Rad Laboratories, Hercules CA) and two different primer pairs from Ty1 *POL* (4681F, 5'-GAAATTCAATATGACATACTTGGC-3'; + 4851R, 5'-GTTCATCCTGGTCTATATATAAAGA-3'; 3251F, 5'-GAGAAGTTGACCCCAACATATCTG-3'; + 3480R, 5'-TGTATGATTAGTCTCATTTTCAC-3').

Ty1*his3-AI* mobility

The frequency of Ty1*his3-AI* mobility was determined as described previously (28, 46) with minor modifications. For transposition assays involving strains containing pGPOLΔ, a single colony from a SC-Ura plate incubated at 30°C was resuspended in 1 ml of water and 5 µl of cells were added to quadruplicate 1 ml cultures of SC-Ura liquid

medium. The cultures were grown for 3 days at 22°C , washed, diluted and spread onto SC-Ura and SC-His-Ura plates to calculate Ty1 mobility. For mobility assays with strains repopulated with Ty1 elements, a single colony from a YEPD plate incubated at 30°C was diluted into 10 ml of water, and 1 μl of cell suspension was added to quadruplicate 1 ml YEPD cultures. The cultures were incubated for 2-3 days at 22°C, washed, diluted and then spread onto YEPD and SC-His plates. Plates were incubated for 4 days at 30°C. For mobility assays involving strains expressing pGTy1*his3-AI* and *GAL1*-p22 or related plasmids, a single colony was resuspended in 1 ml SC-Ura-Trp + 2% raffinose media and grown for 16 hr at 30°C, then diluted 25-fold into quadruplicate 1ml cultures of SC-Ura-Trp + 2% galactose. Cultures were grown at 22°C for 2 days, washed, diluted, and spread onto SC-Ura-Trp and SC-Ura-Trp-His plates. Qualitative Ty1*his3-AI* mobility assays were performed as described previously (28, 46). For qualitative mobility assays involving strains containing pGPOLΔ, single colonies patched onto SC-Ura plates were incubated at 22°C for 2 days. To detect Ty1*HIS3* mobility events, cells were replica-plated onto SC-Ura-His plates and incubated at 30°C for 3 days. For strains expressing pGTy1*his3-AI* and pGAL-p22 or related plasmids, single colonies patched onto SC-Ura-Trp plates were incubated for 2 days at 30°C. The resulting patches were replica-plated to SC-Ura-Trp + 2% galactose plates followed by incubation at 22°C for 2-4 days. To detect Ty1*HIS3* mobility events, galactose-induced cells were replica-plated to SC-Ura-Trp-His plates followed by incubation for 3 days at 30°C.

p18 antiserum

Ty1 (1068-1496) was amplified with primers (1068NdeI, 5'-CATGTTCCATATGCAATCTGATACCCAAGAGGCAA-3'; 1496XhoI, 5'-

CATGTTTCCTCGAGTTAGTGAGCCCTGGCTGTTTCG-3') and cloned into pET15bTEV vector (Novagen EMD, San Diego, CA). An 800 ml culture of *E. coli* BL21 (DE3) cells containing the expression plasmid in LB + 100 μg/ml ampicillin (Sigma-Aldrich, St. Louis, MO) were induced by 0.15 mM IPTG (Sigma-Aldrich) at 37°C. When cells reached an OD_{600} of 0.6-0.8, the temperature was reduced to 16°C, and incubated for an additional 24 hr. The cells were resuspended in 50 ml lysis buffer A (50 mM Phosphate Buffer pH 7.8, 1 M NaCl) and harvested by sonication. The His-tagged Ty1 product was purified with Talon affinity resin (Clontech) and eluted with 300 mM imidazole (Sigma-Aldrich). The elution product was dialyzed against storage buffer (10% glycerol, 1 M NaCl, and 25 mM Tris-HCl (pH 8.0)) overnight. A rabbit polyclonal antibody was raised against the truncated Ty1 Gag protein by Bio-synthesis Inc. (Lewisville, TX).

Protein isolation and immunoblotting

To detect protein expression from pGPOLΔ in the absence of galactose induction, 5 ml of SC-Ura medium was inoculated with a single colony and grown at 22°C for 24 hr. For coexpression of independent pGAL expression plasmids, 1 ml of SC-Ura-Trp + 2% raffinose was inoculated with a single colony and grown at overnight at 30°C. The overnight culture was diluted 25-fold into SC-Ura-Trp +2% galactose and grown for 2 days at 22°C. 5 ml of culture was processed by TCA extraction as described previously (47) except cells were broken by vortexing in the presence of glass beads, and 10 μl of the supernatant was separated by electrophoresis. For sucrose fractions, equal volumes of each fraction were analyzed. For P40 and VLP samples, 5 μg of P40 was used to detect p22/p18 and 10 μg of P40 was used for RT and IN. Samples were

separated on 10% (for RT and IN detection) or 15% (Gag p49/p45 and p22/p18 detection) SDS-PAGE gels. For optimal detection of p22/p18, proteins were transferred to PVDF membrane at 100 V for 90 min. The membranes were blocked in 5% milk/TBST (500 mM NaCl, 20 mM Tris-HCl, 0.1% Tween-20, pH 7.6) and then incubated with rabbit polyclonal antisera at the following dilutions: αp18; 1:5,000 in 2.5% milk/TBST, αRT/B8; 1:5,000 in TBST, and αIN/B2 1:2,500 in TBST (48). Immune complexes were detected with ECL reagent (GE Healthcare, Little Chalfont, United Kingdom).

VLP isolation

VLP purification from DG3739, DG3774 and DG3784 (Table 3.1) and reverse transcriptase assays were performed as described previously (49, 50) with the following modifications. Briefly, 40 ml SC-Ura-Trp + 2% raffinose cultures of strains used for VLP analysis were grown overnight at 30°C with shaking. Each culture was diluted 25-fold into 1 liter of SC-Ura-Trp + 3% galactose and grown at 21°C to OD_{600} 1-1.2. Cells were harvested by centrifuging at 6,000 rpm and homogenized with acid-washed glass beads in buffer B (15 mM KCl, 10 mM HEPES-KOH pH 7, 5 mM EDTA) containing protease inhibitor cocktail (0.125 mg/ml aprotinin, leupeptin, pepstatin A and 1.6 mg/ml PMSF). The crude lysate was centrifuged at 10,000 rpm and the supernatant loaded onto a step gradient of 20%, 30%, 45% and 75% sucrose in buffer B. The step gradient was centrifuged at 25,000 rpm in a SW28 rotor for 3 hr. Four ml of the gradient at the junction of the 30% and the 45% sucrose layers was withdrawn, diluted to 10% sucrose with buffer B and pelleted by centrifugation at 55,000 rpm in a Ti 70.1 rotor for 45 min. The resulting crude VLP pellet (P40) was suspended in buffer B and centrifuged

through a 20 - 60% continuous sucrose gradient in buffer B at 25,000 rpm in a SW41 rotor for 3 hr. The entire gradient was dripped into 19 equal fractions using an ISCO Foxy Jr. fraction collector (Lincoln, NE). All steps were carried out 4°C unless specified. Fractions were assayed for Ty1 reverse transcriptase activity as described previously (34, 49), except 10 μl samples were incubated with exogenous reverse transcriptase mix (50 mM Tris-HCl pH 8, 10 mM $MgCl_2$, 20 mM DTT, 15 μM dGTP, 10.7 μg poly(rC)/(dG)) and [α-^{32}P] dGTP (3,000 Ci/mmol; Perkin Elmer).

Electron microscopy

Three sucrose gradient fractions with the highest reverse transcriptase activity from DG3739 (fractions 5-7) and DG3774 (fractions 4-6) were pooled, diluted with buffer B and pelleted as described above. The sample was allowed to bind for 15 min to Formvar and carbon-coated 400-mesh copper grids. Grids were stained with 2% ammonium molybdate, pH 6.5 for 10 sec and visualized with a JEM-1210 Transmission Electron Microscope (JEOL USA Inc., Peabody, MA) equipped with an XR41C Bottom-Mount CCD Camera (Advanced Microscopy Techniques, Woburn, MA). Approximately 100 VLPs were analyzed to determine the percentage of closed versus open particles. VLP diameter was measured with closed VLPs only using ImageJ (51) and the two datasets were compared using an unpaired T test.

FISH/IF

Two ml SC-Ura-Trp + 3% raffinose cultures were inoculated with a single colony and grown 16 hr at 30°C. The overnight cultures were diluted 10-fold into SC-Ura-Trp + 3% galactose and grown at 22°C for 24-30 hr until an OD_{600} of 0.8-1.0 was reached. Formaldehyde was added directly to the culture at a final concentration of 4% and

allowed to fix for 1.75 hr. Processing of the cells for FISH/IF was performed as described previously (7). For Gag/p22-V5 colocalization experiments, primary antibodies were αVLP (rabbit polyclonal, 1:2,000, a kind gift from Alan and Susan Kingsman) and αV5 (Life Technologies, 1:4,000) and secondary antibodies used were α-rabbit-AF488 (Life Technologies, 1:200) and α-mouse-AF594 (Life Technologies, 1:400). Image acquisition was carried out using a Zeiss Axio Observer microscope equipped with an AxioCam HSm camera and analyzed with AxioVision v4.6 software (Carl Zeiss Microscopy, LLC, North America). Exposure times used to capture fluorescent and DAPI images were kept consistent throughout each experiment. Figures were constructed with Adobe Photoshop software (Adobe Systems, San Jose, CA).

GST-pulldown

One ml of SC-Ura + 2% raffinose was inoculated with a single colony at 30°C overnight and was then diluted 1:25 into 5 ml SC-Ura +2% galactose and grown for 2 days at 22°C. 2.5 ml of galactose-induced cells was suspended in 150 μl lysis buffer C (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, 0.1% Triton X-100, 1 mM PMSF, 1 μg/ml aprotinin, 0.5 μg/ml leupeptin and 1 μg/ml pepstatin A) and homogenized with the same volume of acid washed glass beads. The crude lysate was centrifuged at 10,000 rpm at 4°C for 10 min. and 500 μl supernatant containing 300 μg of protein was gently mixed with 20 μl glutathione-coated resin (GenScript, Piscataway, NJ) at 4°C for 2 hr. The resin was washed three times with 1 ml lysis buffer C and then suspended in 40 μl SDS loading buffer. After boiling for 10 min, 5-8 μl per lane were loaded onto a 12% SDS-polyacrylamide gel. Immunoblotting was performed as described above and membranes were incubated with mouse monoclonal antibodies

αGST/B-14 (Santa Cruz Biotech) at 1:1,000 or αTY tag (a kind gift from Stephen Hajduk) at 1:50,000 in TBST.

Results

An internal Ty1 sense-strand RNA is required for CNC

The CNC region of Ty1 spans the 5' UTR and all of *GAG*, and a multicopy pGTy1 expression plasmid confers CNC *in trans* even when *GAL1*-promoted transcription is repressed (28, 31). To identify sequences necessary for CNC (Figure 3.1A), a genetic screen for CNC⁻ mutations was performed in a Ty1-less *S. paradoxus* strain repopulated with a single chromosomal Ty1 insertion containing the selectable indicator gene *his3-AI* (46) (Table 3.1). Ty1*HIS3* insertions usually occur by retrotransposition following splicing of the artificial intron. Since Ty1*HIS3* cDNA can also undergo homologous recombination with genomic Ty1 elements or solo LTRs (52, 53), the term Ty1 "mobility" is used to describe both types of insertion. Ty1 mobility was followed using a qualitative papillation assay for His⁺ cells in a Ty1-less test strain containing a chromosomal Ty1*his3-AI* element and an empty vector, wild-type pGTy1 plasmid or randomly mutagenized pGTy1. Cells were grown under repressive conditions for *GAL1* expression. To generate point mutations, the CNC region was amplified using Taq DNA polymerase and PCR products were recombined into pGTy1 *in vivo* by gap repair. Approximately 3,500 pGTy1 recombinants were screened for loss of Ty1 CNC and recovered plasmids were reintroduced to confirm the CNC⁻ phenotype. Although pGTy1 plasmids with one to four base-changes in the CNC region were identified, only plasmids carrying single mutations (Figure 3.1A and Table 3.2A) were analyzed further.

To minimize the possibility that sequence changes outside of the gap-repaired region influence CNC and to facilitate molecular analyses, most of *POL* was deleted from the mutant pGTy1 plasmids to generate plasmid pGPOLΔ. Quantitative Ty1*his3-AI* mobility assays were performed with five mutants from the screen (Table 3.2A). Mutations T399C, ΔA1456 and A1296G conferred moderate decreases in CNC when compared with the CNC⁺ control, while T1108C and A1123G conferred low levels of CNC. Furthermore, the T1108C mutation affected CNC the most and was obtained from four independent isolates, suggesting T1108 is part of an important sequence motif involved in CNC. Since Ty1AS RNAs were reported to be necessary for CNC (31), Northern blotting was performed with total RNA from the five single mutants. All of the mutants except T399C contained a similar level of Ty1AS RNAs as that produced from a wild-type pGPOLΔ plasmid, when compared with the *ACT1* loading control (Figure 3.1B). Surprisingly, four of the five CNC⁻ mutations do not map in the Ty1AS RNA transcription units and instead are located in an adjacent segment of the CNC region (Figure 3.1A), and all change *GAG*'s coding potential (Table 3.2A).

A 5' truncated Ty1 sense RNA can be detected in wild type cells and is enriched in an *spt3* mutant (22, 51). A similar observation was reported for an *xrn1* mutant, where the RNA was termed Ty1SL (Ty1 short length RNA) (24). Therefore, the point mutations identified in the screen could map in a shorter Ty1 sense RNA that initiates in *GAG* independently of normal Ty1 transcription and this transcript could be involved in CNC. To determine if a shorter Ty1 sense RNA was produced from the pGPOLΔ plasmids, total RNA was subjected to Northern blotting using a strand-specific ³²P-labeled riboprobe from *GAG*. Cells containing pGPOLΔ and mutant derivatives were used in the

Northern blotting since deleting *POL* results in the synthesis of Ty1 transcripts that are clearly distinguishable from Ty1*his3-AI* RNA. All point mutants except T399C made a shorter sense-strand Ty1 RNA, termed Ty1 internal (Ty1i) RNA, whereas cells containing an empty vector control lacked this transcript (Figure 3.1C). Two additional mutants, Δ238-281 and Δ238-353 were derived in the pGPOLΔ context and included in this analysis. Originally described in Matsuda and Garfinkel (2009), pGTy1 plasmids with short deletions in the 5' LTR abolished CNC. The loss of CNC was attributed to a decrease in the level of the Ty1AS RNAs due to deletion of their 3' ends. However, the lack of detectable Ty1i RNA in the Δ238-281 and Δ238-353 mutants may now explain their CNC⁻ phenotype. These results also suggest that sequences near the 5' LTR, which contains the enhancer required for Ty1 transcription (5), may also be important for synthesizing Ty1i RNA.

Chromosomal Ty1A1123G insertions fail to confer CNC

The genetic screen identified several missense mutations in *GAG* that weakened CNC and were present on both Ty1 mRNA and the Ty1i transcript. To determine if CNC⁻ mutations impacted Gag function, full-length pGTy1 plasmids containing A1123G (Tyr277Cys) or A1296G (Thr335Ala) mutations were compared with wild-type pGTy1 for their ability to stimulate or trans-activate movement of a chromosomal Ty1*his3-AI* element (33). The A1296G mutation likely affects both CNC and transposition, since pGTy1A1296G expression did not stimulate Ty1 mobility. However, induction of pGTy1A1123G increased Ty1*his3-AI* mobility *in trans* to similar levels observed with wild-type pGTy1, suggesting that pGTy1A1123G encodes functional Gag yet is defective for CNC (data not shown). To determine if Ty1A1123G conferred CNC in a

natural chromosomal context (Figure 3.2), a strain with a single chromosomal Ty1*his3-AI* (Figure 3.2A) was repopulated with wild type (Figure 3.2B) or Ty1A1123G elements (Figure 3.2C). As expected, Ty1 mobility decreased 33-fold in a strain repopulated with 12 wild-type Ty1 elements when compared with the starting strain (Figure 3.2B, Table 3.2B). However, Ty1 mobility increased almost 5-fold in a strain containing 7 copies of Ty1A1123G, indicating that A1123G abolishes CNC without disrupting the function of Gag (Figure 3.2C, Table 3.2B). The separation of function phenotype displayed by Ty1A1123G raised the possibility that an altered form of Gag encoded by Ty1i RNA mediates CNC.

Expression of Ty1i RNA

Since multicopy Ty1 plasmids were used as the source of *trans*-acting factors required for CNC, it was important to determine if chromosomal elements also synthesized Ty1i RNA and truncated forms of Gag (Figure 3.3). To detect Ty1i RNA in repopulated *S. paradoxus* as well as *S. cerevisiae* strains, poly(A)$^+$ RNA was subjected to Northern blotting using a ^{32}P-labeled riboprobe from *GAG-POL* (nt 1266-1601) (Figure 3.3A). Three *S. paradoxus* strains were analyzed: the Ty1-less strain (lane C), a derivative repopulated with 38 Ty1 elements (lane 1) and an isogenic *spt3Δ* mutant (lane 2). Five *S. cerevisiae* strains were also analyzed: GRF167 (lane 3) and an isogenic *spt3Δ* mutant (lane 4), BY4742 (lane 5) and isogenic *spt3Δ* (lane 6) and *xrn1Δ* (lane 7) mutant derivatives. A discrete subgenomic Ty1 RNA of 4.9 kb was detected below the full-length transcript (5.7 kb) in all strains except repopulated *S. paradoxus* (Figure 3.3A, lane 1). The failure to detect a distinct transcript in this strain was unexpected, but may result from 5' heterogeneity of the 4.9 kb transcript. The 4.9 kb

Ty1 RNA comigrated with the truncated transcripts detected in *spt3Δ* and *xrn1Δ* mutants. To determine the 5' end of the 4.9 kb transcript in BY4742 and an isogenic *spt3Δ* mutant, poly(A)$^+$ RNA was subjected to cap-independent 5'-RACE (Figure 3.4). In both strains, the majority of the 5' ends from the 4.9 kb transcript mapped to nucleotide 1000 of Ty1H3 (Figure 3.4A). These results indicate that the 4.9 kb RNA observed in wild-type and an *spt3Δ* mutant share the same 5' ends. However, 5'-RACE analysis of the wild-type repopulated *S. paradoxus* strain showed heterogeneous amplification products (Figure 3.4B) rather than discrete bands, supporting the results from Northern blotting (Figure 3.3). Although our results suggest that the 4.9 kb Ty1 RNA contains the Ty1i transcript in *S. cerevisiae*, other truncated forms of Ty1 RNA may be present (54, 55).

Ty1i RNA encodes Gag proteins, p22 and p18

Two closely spaced AUG codons are present 38 (AUG1) and 68 (AUG2) nucleotides downstream of the transcription start site for Ty1i RNA and one or both may be utilized to initiate synthesis of a truncated form of Gag (Figure 3.1A). However, neither the predicted 22-kDa (p22) Gag-like protein nor its processed product (p18), if p22 is cleaved by Ty1 PR, have been reported to date, and a commonly used VLP antiserum (56) failed to detect p22/p18 reproducibly (data not shown). Therefore, we purified recombinant p18 and generated a new antiserum to determine if Ty1i RNA is translated to produce an N-terminal truncated form of Gag (Figure 3.3B). Whole cell extracts from the strains described above (Figure 3.3A) were immunoblotted with p18 antiserum to detect endogenous Gag and additional Gag-related proteins. As expected, normal levels of Ty1 Gag p49/p45 were detected in wild-type strains (Figure 3.3B, lanes

145

1, 3, and 5) while reduced levels were observed in *spt3Δ* (lanes 2, 4, and 6) mutants. Importantly, p22 was detected in the *spt3Δ* mutants (lanes 2, 4 and 6), whereas p18 was only detected in the wild-type strains (lanes 1, 3 and 5) and the *xrn1Δ* mutant (lane 7). The increase in p18 observed in the *xrn1Δ* mutant likely results from an increase in Ty1i RNA level, since Xrn1 is the major 5'-3' exonuclease involved in RNA decay. Taken together, our results not only suggest that Ty1i RNA encodes p22, but the striking relationship between expression of full-length Ty1 mRNA, and hence Ty1 PR, and detection of p22 versus p18 suggests that p22 is cleaved by Ty1 PR to form p18. Furthermore, processing of p22 to p18 raises the possibility that p22 associates with VLPs to gain access to PR. As expected, Gag proteins were not detected in the Ty1-less strain (lane C).

Ribosome footprint profiling reveals an internal AUG as a potential translation start for p22

To determine if the candidate AUG1 or AUG2 translation start sites on Ty1i RNA (Figure 3.1A) were present in genomic sequencing analyses, we turned to ribosome footprint profiling (Ribo-seq) (Figure 3.5). In Ribo-seq, ribosomes in the act of translating an mRNA are treated with RNase I, leaving a ~ 28 nt ribosome footprint, which is harvested for high throughput sequencing to provide a snapshot of the abundance and distribution of ribosomes on mRNAs (43). Yeast starved for glucose for 3 hr accumulate as much as 10% of Ribo-seq reads at the start codon of ORFs (42), providing a sensitive method for detecting initiation codons *in vivo*. We utilized a published dataset to analyze the Ribo-seq read distribution at the 5' end of Ty1i RNA (57). The most abundant read in this region corresponded to a ribosome footprint located on AUG1,

which is the first start codon downstream of the Ty1i transcription start site (Figure 3.4A). Also, the density of Ribo-seq reads increased downstream of AUG1, consistent with translation of the downstream ORF under glucose starvation. Additional mutational analysis of AUG1 and AUG2 will be required to verify the translation start of p22.

p22/p18 encoded by Ty1i RNA is necessary for CNC

To establish that p22/p18, rather than the Ty1i transcript itself, is responsible for CNC, we analyzed frameshift mutations in the pGPOLΔ construct (Figure 3.6A) for alterations in Ty1*his3-AI* mobility (Figure 3.6B, Table 3.2C-1), Ty1i RNA levels, (Figure 3.6C) and protein levels (Figure 3.6D). Cells containing pGPOLΔ decreased the mobility of a chromosomal Ty1*his3-AI* element up to 74-fold (Figure 3.6B, Table 3.2C-1) compared to an empty vector control, and produced Ty1i RNA (Figure 3.6C) and p22 (Figure 3.6D). Two frameshift mutations were placed downstream of AUG1 and AUG2 that introduce premature termination codons, ΔC1071 and +A1303. +A1303 was created to eliminate the possibility that downstream in-frame AUGs (AUG3 and AUG4, Figure 3.6A) could be utilized to produce a *trans*-dominant factor. Both frameshift mutations caused an increase in Ty1*his3-AI* mobility to almost the same level as that obtained in a strain lacking CNC (Figure 3.6B, Table 3.2C-1). Cells containing the mutant plasmids produced Ty1i RNA (Figure 3.6C), but not wild-type p22 (Figure 3.6D). The residual level of CNC conferred by the plasmids carrying the frameshift mutations may be caused by truncated protein synthesized prior to encountering the mutations; however, immunoblotting using the p18 antiserum did not detect these smaller proteins (Figure 3.6A and 3.6D). To fully eliminate protein production from AUG1 and AUG2, we replaced both initiation codons with the alanine codon GCG in pGPOLΔ. In cells

147

carrying pGPOLΔ-GCG1GCG2, transposition frequency was fully restored and about 2-fold higher than the ΔC1071 frameshift (Table 3.2C-2). These results show that AUG1 and/or AUG2 are necessary for CNC, and reinforces the observation that ΔC1071 confers a very low level of CNC. Taken together, our results identify p22 as a *trans*-dominant negative inhibitor of Ty1 retrotransposition and the intrinsic factor responsible for CNC.

<u>Ectopic expression of p22/p18 is sufficient to inhibit Ty1 movement</u>

To determine if p22/p18 reduces Ty1 transposition, a cDNA expression library (45) was screened for clones that inhibited Ty1*his3-AI* mobility, and p22/p18 was ectopically expressed from the *GAL1* promoter. One clone was obtained from the *GAL1*-driven cDNA library that contained Ty1 sequences 1042-5889 and inhibited chromosomal Ty1*his3-AI* mobility. The 5' end of the cDNA included AUG2 and 26 additional nucleotides upstream but did not contain AUG1. The 3' end terminated in the R region (3' LTR) of Ty1 RNA, which is the similar to the 3' ends mapped previously (15, 55). Therefore, an almost full length 4.9 kb Ty1i transcript from a chromosomal element was likely captured as this cDNA clone, and contains coding sequence for p22, as well as the *POL* coding sequence for PR, IN, and RT. When the cDNA clone and pGTy1*his3-AI* were coexpressed, Ty1 mobility decreased 570-fold when compared with a control strain expressing only pGTy1*his3-AI* (Table 3.2D). These results support the idea that a truncated Gag protein likely utilizing AUG2 inhibits Ty1 mobility, although initiation from AUG2 occurs less frequently than AUG1 based on ribosome profiling in the Sigma 1278b strain (Figure 3.5).

The following segments of *GAG* sequence starting with AUG1 were fused to the *GAL1* promoter on a multicopy expression plasmid and analyzed for *trans*-dominance (Table 3.2E) or protein expression (Figures 3.7 and data not shown) in the Ty1-less *S. paradoxus* strain: p22, p22 containing an internal V5 epitope, p18, and p22^{Gag*PR} containing a previously characterized mutation that disrupts Gag-PR cleavage by PR (40). Ty1*his3-AI* mobility decreased more than 32,000-fold in cells coexpressing pGTy1*his3-AI* and p22 or p22-V5 when compared with the control strain expressing pGTy1*his3-AI* (Table 3.2E). Both p22 and p18 are present, again suggesting that some p22 is incorporated into VLPs and cleaved by Ty1 PR (Figure 3.7). Coexpression of p18 and pGTy1*his3-AI* reduced Ty1 mobility to levels similar to those observed with p22 (Figure 3.7). *GAL1*-promoted expression of p22 or p18 also inhibited pGTy1*his3-AI* mobility in *S. cerevisiae* strains BY4742 and GRF167 (data not shown).

To determine if p22 alone inhibited Ty1 mobility, we coexpressed p22^{Gag*PR} and pGTy1*his3-AI* (Figure 3.7) in the Ty1-less *S. paradoxus* strain. Results from qualitative mobility assays indicated that p22^{Gag*PR} retained most if not all of its inhibitory function when compared with wild-type p22 or p18 and the empty vector control, even though processing of p22 to p18 was blocked (Figure 3.7). The level of p22^{Gag*PR} when expressed from the *GAL1* promoter was also comparable to that obtained with p22, p18, or full-length Gag. Together, these results show that p22 and p18 are potent *trans*-dominant inhibitors of Ty1 transposition.

p22/p18 cofractionates with VLPs and alters Ty1 proteins

One possibility to account for the dramatic decrease in Ty1 mobility is that p22 associates with assembling VLPs in the cell, leading to abnormal VLP function.

149

Therefore, crude VLP preparations from Ty1-less strains expressing pGTy1*his3-AI* alone (Figure 3.8A), p22 and pGTy1*his3-AI* (Figure 3.8B) or p22 alone (Figure 3.8C) were separated by centrifugation through 20-60% continuous sucrose gradients. Fractions were assayed for reverse transcriptase activity using an exogenous primer/template, and immunoblotted for Gag, IN, RT and p22/p18 (Figure 3.8A and 3.8B) or p22/p18 alone (Figure 3.8C). As expected, a peak of reverse transcriptase activity coincided with the highest concentrations of mature Gag, RT, and IN proteins in the strain expressing just pGTy1*his3-AI* (Figure 3.8A). When pGTy1*his3-AI* and p22 were co-expressed, Gag and p22/p18 displayed a similar fractionation pattern across the gradient (Figure 3.8B). p18 appeared to be the predominant form present in crude VLP preparations, which is likely due to processing by Ty1 PR in VLPs. To further investigate if the cofractionation of Gag and p22/p18 resulted from an association between VLPs and p22, rather than comigration of a protein complex containing p22 that had a similar density as VLPs, an identical fractionation was performed in a strain expressing only *GAL1*-promoted p22. When expressed alone, p22 was detected near the top of the gradient (Figure 3.8C), and therefore, had a different fractionation profile than that observed when pGTy1*his3-AI* and p22 were coexpressed (Figure 3.8B). Furthermore, we detected p18 in the CNC$^+$ VLPs (data not shown) used for structural probing of packaged Ty1 RNA (32). These results support an interaction between Ty1 VLPs and p22/p18.

When comparing the strain expressing pGTy1*his3-AI* (Figure 3.8A) with one expressing pGTy1*his3-AI* and p22 (Figure 3.8B), several differences in the fractionation patterns, protein composition and distribution, and reverse transcriptase activity were

evident. First, cells expressing only pGTy1*his3-AI* yielded a higher concentration of Gag, IN, RT and reverse transcriptase-catalyzed incorporation of [α^{32}P]-dGTP in the peak fractions. Second, cells coexpressing pGTy1*his3-AI* and p22 showed a broader distribution of Ty1 proteins and reverse transcriptase activity. Third, the VLPs formed in the presence of p22 had a lower level of [α^{32}P]-dGTP incorporation throughout the gradient. Fourth, Ty1 protein processing or stability was altered when pGTy1*his3-AI* and p22 were coexpressed. There was an accumulation of the PR-IN precursor (p91) and much less mature IN (p71), which is similar to results obtained previously (31). Ty1 RT (p63) now appeared as a doublet with an additional higher molecular weight protein that reacted with the RT antibody (Figure 3.8B, denoted by an asterisk). Fifth, Ty1 Gag appeared to undergo more proteolysis overall when p22 was present, as evidenced by multiple lower molecular weight Gag-related proteins, which cofractionated with full-length Gag. These unusual Ty1 proteins may result from aberrant processing by Ty1 PR, cleavage by a cellular protease, or from differences in posttranslational modification of Ty1 proteins brought about by a VLP-p22 interaction. Therefore, the mechanism of CNC involves differences in the physical and biochemical properties of VLPs assembled in the presence of p22.

p22/p18 changes VLP morphology

Since p22/p18 affected the fractionation of Ty1 VLPs and appearance of Ty1 proteins (Figure 3.8), we examined the size and morphology of VLPs assembled in the presence or absence of p22 by electron microscopy (Figure 3.9). Equivalent sucrose gradient fractions with the highest level of [α^{32}P]-dGTP incorporation (Figure 3.8) were

pooled, diluted, and concentrated by ultracentrifugation prior to staining with 2% ammonium molybdate. Ty1 VLPs formed in the absence of p22 (Figure 3.9A) were mostly intact with an average diameter of 37.4 ± 2.7 nm, and only 13% of wild-type VLPs appeared malformed. In contrast, almost half of Ty1 VLPs formed in the presence of p22 (Figure 3.9B) appeared open or incomplete, suggesting that these VLPs are either not formed properly or are less stable during sample preparation. The diameter of intact VLPs assembled in the presence of p22 was 39.2 ± 3.1 nm. Although the difference in diameters of the two batches of VLPs is statistically significant (P = 0.0005), further analyses will be required to determine if this difference is functionally relevant.

<u>p22-V5 disrupts pGTy1-induced retrosomes and colocalizes with Gag</u>

Since p22 altered the fractionation pattern and morphology of VLPs, and the processing or stability of Ty1 proteins, we examined whether p22 influenced the appearance of retrosomes, which are sites for VLP assembly. Ty1-less strains expressing p22-V5 and pGTy1*his3-AI* alone or together were subjected to indirect immunofluorescence (IF) and fluorescence *in situ* hybridization (FISH) to visualize retrosomes (Figure 3.10). VLP or V5 antibodies were used to detect Ty1 proteins and a *GAG*-DIG probe was used to detect full length Ty1 mRNA. The internal V5 tag in p22-V5 did not disrupt *trans*-dominance (Table 2E) and retrosome analysis of cells expressing untagged p22 was identical to that from strains expressing p22-V5. Three types of staining were observed: (1) large, distinct foci that co-stain for Ty1 mRNA and Gag were defined as retrosomes (R), (2) nondistinct, punctate staining for both Ty1 mRNA and Gag was termed "puncta" (P), and (3) lack of staining for Ty1 mRNA, Gag or

both was designated as "none". In cells containing puncta, colocalization between Ty1 mRNA and Gag could not be confidently determined in the majority of cells. In a control strain expressing pGTy1*his3-AI* alone, retrosomes were observed in 61% of cells, while only 7% of cells showed a punctate localization of Ty1 mRNA and Gag proteins (Figure 3.10A). When p22-V5 and pGTy1*his3-AI* were coexpressed, the percentage of cells containing normal retrosomes decreased to 18% while Ty1 puncta was observed in 31% of cells. Thus, p22-V5 disrupts Ty1 retrosomes in a large fraction of cells. In addition, cells were analyzed for Ty1 Gag and p22-V5 colocalization using VLP and V5 antibodies, respectively (Figure 3.10C and 3.10D). As expected, a similar percentage of cells exhibited retrosomes (61%) (Figure 3.10A) and Gag foci in the absence of p22-V5 (62%) (Figure 3.10C). In the presence of p22-V5, a comparable fraction of cells displaying Gag foci (28%) and puncta (42%) was observed (Figure 3.10D) when compared to the staining observed using FISH/IF analysis (Figure 3.10B). Interestingly, we detected colocalization of p22-V5 and Gag in almost 70% of Gag foci (Figure 3.10D inset). p22-V5 colocalized with endogenous retrosomes in *S. cerevisiae*, suggesting the possibility that p22 can associate with VLP pre-assembly intermediates (Mitchell and Garfinkel, unpublished results).

GST-pull downs support an interaction between Gag and p18

To provide additional evidence for an interaction between p22/p18 and Gag, a fusion protein consisting of p18 tagged at its N-terminus with Glutathione-S-transferase (41) was expressed from the *GAL1* promoter in BY4742 or a Ty1-less strain (Figure 3.11). Free glutathione-S-transferase was expressed alone as a negative control. Protein extracts were immunoblotted using antisera specific for GST, Gag p49/p45, or

Hts1 prior to mixing with the glutathione-coated resin (Input) or released from the GST complexes bound to resin after several washes with lysis buffer (Pull-down). Fusions between GST and full-length Gag were insoluble under a variety of conditions, and therefore, could not be analyzed further. The GST-p18 fusion protein was soluble under the conditions used for the pull-down; however, partial degradation of GST-p18 resulted in free GST protein. GST-p18 formed a complex containing Gag p45 and p49 encoded by the genomic Ty1 elements in BY4742, whereas GST expressed alone did not. Ty1 Gag-p18 complexes were also not detected in the Ty1-less strain. Hts1 was used to control for nonspecific trapping of cellular proteins in the Ty1 complexes, and as expected, was only detected in the input samples. Together, these results suggest that p18 and Gag interact.

Discussion

Here, we characterize a restriction factor derived from Ty1 *GAG* that confers CNC by perturbing VLP assembly and function. This unique form of transposon CNC (28) may have evolved after an ancestral *S. cerevisiae/paradoxus* lineage lost the evolutionarily conserved RNAi pathway used to silence Ty1 expression (29, 30). Noncoding antisense transcripts from Ty1 have been implicated in repressing transcription (24), RNAi in budding yeast (30), and CNC (31). The identification of mutations that abrogated both CNC and Ty1AS RNA expression implicated Ty1AS RNAs in CNC. Additionally, the association of Ty1AS RNAs with VLPs further supported models of AS RNA-based CNC. Here, we show additional mutations in the CNC region of Ty1 fail to confer CNC, yet do not perturb Ty1AS RNA expression. One *GAG*

154

mutation in particular abolished CNC but did not affect transposition or AS RNA production. The behavior of this separation of function mutation suggested that a Ty1 protein might contribute to CNC. Evaluation of these mutants, along with those previously reported helped reveal p22, a Gag-like restriction factor encoded by a 5' truncated sense RNA (Ty1i) that likely forms the basis of CNC. The role of Ty1AS RNAs in Ty1 CNC, if any, remains to be determined.

We detect differences in the transcripts encoding the p22 restriction factor and how these transcripts are utilized for protein synthesis. In *S. cerevisiae*, a 4.9 kb Ty1i RNA is detected in wild-type strains both in our work and in previous studies when poly(A)$^+$ RNA is subjected to Northern blotting (22, 54), but is rarely detected in numerous studies when total RNA is analyzed (7, 8, 18, 58-60). Perhaps the level of RNA degradation observed with the abundant 5.7 kb Ty1 genomic RNA obscures the 4.9 kb Ty1i transcript when total RNA is analyzed by Northern blotting, because we can detect a shorter Ty1i transcript produced from a pGPOLΔ plasmid with total RNA from *S. paradoxus*. Alternatively, it has been reported that only 15% of Ty1 mRNA transcripts are polyadenylated (9). Hence, it is possible that Ty1i RNA is readily detected by Northern blotting of poly(A$^+$) RNA because the majority of Ty1i transcripts are polyadenylated, whereas the majority of Ty1 mRNA is not. However, in an isogenic repopulated *S. paradoxus* strain, a discrete Ty1i transcript is not detected from chromosomal Ty1 elements even when poly(A)$^+$ RNA is analyzed by Northern blotting or cap-independent 5'-RACE. Ty1i RNA is present in both species when full-length Ty1 transcription is altered by deleting the Spt3 subunit of SAGA and related complexes (22, 23). Spt3 helps modulate the recruitment of the TATA-binding protein to the TATA box

155

of SAGA-dependent promoters (61-63), and therefore, can specify transcriptional initiation. However, the initiation site for Ty1i RNA within *GAG* predominates in an *spt3Δ* mutant, which is similar to the activation of cryptic intragenic promoters observed in a variety of chromatin and transcription-related mutants (64). Although our results are consistent with the idea that transcription of Ty1i RNA responds differently to the complexes containing Spt3, such as SAGA, in *S. cerevisiae* versus *S. paradoxus*, detailed functional comparisons between Spt3/SAGA from these species will be required to resolve this question.

Surprisingly, appreciable levels of p22/p18 are present in wild-type *S. paradoxus* repopulated with Ty1H3 in the absence of detectable 4.9 kb Ty1i RNA. This result raises the possibility that full-length Ty1 and Ty1i transcripts may utilize an internal ribosome entry site (65) upstream of AUG1 or AUG2 to drive synthesis of p22. Other mechanisms by which p22 could be translated from full-length Ty1 mRNA are leaky scanning, where scanning ribosomes sometimes initiate translation from an alternate AUG codon (66-69) or translation reinitiation in which translation starts at a downstream AUG after translation of an ORF situated upstream (70, 71). Although leaky scanning and translation reinitiation remain possible mechanisms, both require closely spaced AUGs. However, seven in-frame and seven out-of-frame AUGs are present in the 745 bases between the Gag initiation codon (nt 293) and p22 AUG1 (nt 1038), making leaky scanning or translation reinitiation unlikely. Alternatively, exceptional forms of translation initiation may not be required to synthesize p22 if heterogeneous Ty1i transcripts that contain AUG1 or AUG2 in the repopulated *S. paradoxus* strain remain translatable. In support of this view, we show that Ty1i RNA is a functional template for translation of

p22 in *S. paradoxus* and *S. cerevisiae spt3Δ* mutants in the absence of full length Ty1 mRNA. Although it is possible that there are two modes of p22 production in yeast (Ty1 mRNA and Ty1i RNA mediated), production of p22 from internal Ty1 RNA products alone is an attractive idea.

Once synthesized, p22 profoundly inhibits retrotransposition by altering VLP assembly and function. Earlier work as well as our mutational analysis of the CNC region demonstrates that Ty1 produces a *trans*-dominant inhibitor, now identified as p22, that decreases Ty1*his3-AI* mobility 20 to >340-fold depending on the relative expression of Ty1 and p22 (28, 31, 32, 72). However, when a cDNA derived from Ty1i RNA or p22 and Ty1*his3-AI* are coexpressed from the *GAL1* promoter in a Ty1-less strain, mobility decreases 570- and 32,000-fold, respectively, indicating that p22 is necessary and sufficient for inhibition. The extreme inhibitory effect and broad dynamic range raises the possibility that the process of retrotransposition is very sensitive to the level of p22, with increasingly severe defects appearing as the level of p22 increases. Conversely, the relative amount of Ty1 verses p22 expression can likely saturate the inhibitor, as is evident from previous studies utilizing *GAL1*-promoted Ty1 induction (15, 16, 33, 34). In fact, Ty1 "transpositional dormancy," which was described upon the discovery of Ty1 retrotransposition (15, 34) may result from an inhibitor that is saturated or overcome when Ty1 is induced via the *GAL1* promoter (73-75). The work presented here supports this hypothesis and identifies p22 as the intrinsic inhibitor at least partly responsible for Ty1 dormancy.

When crude VLPs from the Ty1-less strain expressing Ty1 and p22 are analyzed by sucrose gradient sedimentation, both p22 and its processed product p18

cofractionate with Ty1 VLPs. p22 does not exhibit the same fractionation pattern in the absence of pGTy1 expression. Furthermore, analysis of a p22^{Gag*PR} cleavage site mutant shows that p22 as well as p18 effectively inhibits transposition, and cleavage of p22 does not play a major role in CNC. The sucrose gradient fractions have also been assayed for reverse transcriptase activity and subjected to additional immunoblotting to detect Gag, IN, RT and p22/p18. Expression of p22 causes a moderate decrease in the level of reverse transcriptase activity when assayed using an exogenous primer/template, prevents the accumulation of mature IN, which reinforces previous work (31, 32), and broadens the peak containing VLP proteins. In addition, an overall degradation of Gag and the presence of aberrant RT proteins are indicative of proteolysis of the Gag-Pol precursor by Ty1 PR, increased proteolysis by cellular enzymes, or possible post-translational modifications. Furthermore, the excessive proteolysis of IN could explain the appearance of higher molecular weight, RT antibody-reactive proteins and the absence of mature IN. Our results suggest that p22 interacts with and inhibits VLP functionality during assembly or in association with fully formed VLPs and also is processed by Ty1 PR to form p18.

Since these results suggest that VLP structure may be altered by p22, peak sucrose gradient fractions have been concentrated and visualized by electron microscopy. Most of the VLPs (87%) isolated from the control strain lacking p22/p18 are completely spherical with similar curvatures, however, almost half (46%) of the VLPs formed in the presence of p22/p18 are aberrant and have an open or incomplete morphology. VLPs analyzed from CNC$^+$ cells containing much less p22/p18 do not appear malformed, but when extracts containing these VLPs are treated with the

endonuclease benzonase, less protection of packaged Ty1 mRNA is observed (32). Our results suggest that VLP integrity is compromised in the presence of higher levels of p22/p18, and that normal assembly of functional VLPs is inhibited by an interaction between Gag and p22.

To further investigate if Gag and p22 interact, cells expressing Ty1 and p22/p18 have been subjected to FISH/IF microscopy and GST-pull down analysis. The number of cells with aberrant retrosomes increases more than 3-fold when Ty1 and p22-V5 are coexpressed, and 70% of Gag foci also stain for p22-V5. In addition, GST pull down analysis suggests that endogenous Gag can interact with GST-p18. Although p22 engages Gag during active VLP assembly, p22 may also interact with Gag in endogenous retrosomes, which contain few if any VLPs (7). Ty1 *GAG* is necessary for retrosome formation (7, 76), and certain Ty3 *GAG* mutations alter retrosome appearance or location (77, 78). Interestingly, cellular mutations that alter retrograde movement of Gag from the endoplasmic reticulum (ER) destabilize Gag and abolish nucleation of retrosomes (79). Whether p22 enters the ER remains to be determined. Ty1 *GAG* mutations have been isolated that confer a *trans*-dominant negative phenotype (80-82) or affect VLP assembly (83), and some of these mutations map in p22. A synthetic peptide containing sequences within p22 also displays RNA chaperone activity (84), which is required for specific RNA transactions during the retroviral life-cycle such as virion assembly, RNA packaging, primer annealing, and reverse transcription (85). Thus, p22 may inhibit multiple functions carried out by Gag.

Certain retroelement genes have undergone purifying selection in mammals, suggesting that these elements have been domesticated or exapted by their host (86).

To date, domesticated *GAG* and *POL* genes have either evolved a new function used in normal cellular processes or have been incorporated into an innate defense pathway used to inhibit retroviral propagation. The prototypic Gag-like restriction factors Fv1 and enJS56A1 block replication of murine leukemia virus (MLV) and Jaagsiekte sheep retrovirus (JSRV), respectively, by interacting with viral proteins during infection (87-89), and share features in common with CNC of Ty1 by p22/p18. Fv1 is derived from the *GAG* gene of a member of the HERV-L family of human and murine endogenous retroviruses (87, 90, 91). Fv1 inhibits progression of the MLV life cycle following infection and reverse transcription, but prior to integration. Although the infecting viral Gag protein as well as Fv1 determines the level of restriction, an ordered assembly of Gag is required for efficient Fv1 binding (88, 92). Our results suggest that Ty1 Gag interacts with p22/p18; however, the polymerization state of Gag and p22 required for maximum restriction of retrotransposition remains an open question. In addition, p22/p18 affects VLP assembly and function, whereas Fv1 inhibits a different step in the replication cycle that occurs post-infection. Conceptually similar to MLV-Fv1 restriction, the sheep genome harbors about 20 copies of endogenous (en) JSRVs and these sequences are homologous with exogenous JSRV that can cause lung cancer. Certain endogenous copies have evolved a *trans*-dominant Gag protein enJS56A that like Ty1-p22 blocks replication at a step soon after protein synthesis. The JSRV-enJS56A interaction prevents Gag from entering into an endosome trafficking pathway, and results in aggregation and turnover by the proteasome (89, 93).

The MLV-Fv1 and JSRV-enJS56A restriction systems contain two components, raising the possibility of an arms race between the infecting retrovirus and the

domesticated chromosomal *GAG* gene (94). In contrast, the many retrotransposition-competent Ty1 elements inhabiting *Saccharomyces* genomes encode their own inhibitor, and therefore, must balance mutations altering p22 potency with those affecting *GAG* fitness. Since Ty1 *GAG* or p22 coding regions have not been detected as an exapted gene capable of inhibiting Ty1 movement, the graduated retrotransposition rate provided by CNC may benefit *Saccharomyces* and Ty1, as suggested by recent work relating increases in Ty1 copy number with longer chronological lifespan (95). The Ty1-p22 interaction appears to directly block assembly of functional VLPs in a dose dependent manner, and to our knowledge represents a novel and effective way to allow some but not rampant retroelement movement. Further understanding of the molecular events underlying Ty1 Gag-p22 interaction, including the characterization of CNC-resistant mutants and the role that cellular genes have in modulating p22 expression or function, should reveal additional similarities and differences between Ty1 and retroviral restriction factors.

References

1.   Voytas DF, Boeke JD. 2002. Ty1 and Ty5 of *Sacharomyces cerevisiae.*, p. 614-630. *In* Craig NL, Craigie R, Gellert M, Lambowitz AM (ed.), Mobile DNA II. ASM Press, Washington, DC.

2.   Dunham M, Badrane H, Ferea T, Adams J, Brown P, Rosenzweig F, Botstein D. 2002. Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae.* Proceedings of the National Academy of Sciences USA 99:16144-16149.

3.   Garfinkel DJ. 2005. Genome evolution mediated by Ty elements in *Saccharomyces.* Cytogenetic and Genome Research 110:63-69.

4.   Wilke C, Adams J. 1992. Fitness effects of Ty transposition in *Saccharomyces cerevisiae.* Genetics 131:31-42.

5.   Wilke CM, Maimer E, Adams J. 1992. The population biology and evolutionary significance of Ty elements in *Saccharomyces cerevisiae.* Genetica 86:155-173.

6.   Beliakova-Bethell N, Beckham C, Giddings TH, Winey M, Parker R, Sandmeyer S. 2006. Virus-like particles of the Ty3 retrotransposon assemble in association with P-body components. RNA 12:94-101.

7. Checkley MA, Nagashima K, Lockett SJ, Nyswaner KM, Garfinkel DJ. 2010. P-body components are required for Ty1 retrotransposition during assembly of retrotransposition-competent virus-like particles. Molecular and Cellular Biology 30:382-398.

8. Dutko JA, Kenny AE, Gamache ER, Curcio MJ. 2010. 5' and 3' mRNA decay factors colocalize with Ty1 Gag and human APOBEC3G and promote Ty1 retrotransposition. Journal of Virology 84:5052-5066.

9. Malagon F, Jensen TH. 2008. The T body, a new cytoplasmic RNA granule in *Saccharomyces cerevisiae.* Molecular and Cellular Biology 28:6022-6032.

10. Bleykasten-Grosshans C, Friedrich A, Schacherer J. 2013. Genome-wide analysis of intraspecific transposon diversity in yeast. BMC Genomics 14:399.

11. Carr M, Bensasson D, Bergman CM. 2012. Evolutionary genomics of transposable elements in *Saccharomyces cerevisiae.* PLoS ONE 7:e50978.

12. Liti G, Peruffo A, James SA, Roberts IN, Louis EJ. 2005. Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the *Saccharomyces sensu stricto* complex. Yeast 22:177-192.

13. Moore SP, Liti G, Stefanisko KM, Nyswaner KM, Chang C, Louis EJ, Garfinkel DJ. 2004. Analysis of a Ty1-less variant of *Saccharomyces paradoxus*: the gain and loss of Ty1 elements. Yeast 21:649-660.

14. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, Tsai IJ, Bergman CM, Bensasson D, O'kelly MJT, Van Oudenaarden A, Barton DBH, Bailes E, Nguyen AN, Jones M, Quail MA, Goodhead I, Sims S, Smith F, Blomberg A, Durbin R, Louis EJ. 2009. Population genomics of domestic and wild yeasts. Nature 458:337-341.

15. Boeke JD, Garfinkel DJ, Styles CA, Fink GR. 1985. Ty elements transpose through an RNA intermediate. Cell 40:491-500.

16. Scheifele LZ, Cost GJ, Zupancic ML, Caputo EM, Boeke JD. 2009. Retrotransposon overdose and genome integrity. Proceedings of the National Academy of Sciences USA 106:13927-13932.

17. Maxwell PH, Curcio MJ. 2007. Host factors that control long terminal repeat retrotransposons in *Saccharomyces cerevisiae*: implications for regulation of mammalian retroviruses. Eukaryotic Cell 6:1069-1080.

18. Dakshinamurthy A, Nyswaner KM, Farabaugh PJ, Garfinkel DJ. 2010. *BUD22* affects Ty1 retrotransposition and ribosome biogenesis in *Saccharomyces cerevisiae.* Genetics 185:1193-1105.

19. Nyswaner KM, Checkley MA, Yi M, Stephens RM, Garfinkel DJ. 2008. Chromatin-associated genes protect the yeast genome from Ty1 insertional mutagenesis. Genetics 178:197-214.

20. Risler JK, Kenny AE, Palumbo RJ, Gamache ER, Curcio MJ. 2012. Host co-factors of the retrovirus-like transposon Ty1. Mobile DNA 3:12.

21. Suzuki K, Morimoto M, Kondo C, Ohsumi Y. 2011. Selective autophagy regulates insertional mutagenesis by the Ty1 retrotransposon in *Saccharomyces cerevisiae*. Developmental Cell 21:358-365.

22. Winston F, Durbin KJ, Fink GR. 1984. The *SPT3* gene is required for normal transcription of Ty elements in *S. cerevisiae*. Cell 39:675-682.

23. Grant PA, Duggan L, Côté J, Roberts SM, Brownell JE, Candau R, Ohba R, Owen-Hughes T, Allis CD, Winston F, Berger SL, Workman JL. 1997. Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes & Development 11:1640-1650.

24. Berretta J, Pinskaya M, Morillon A. 2008. A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*. Genes & Development 22:615-626.

25. Haimovich G, Medina DA, Causse SZ, Garber M, Millán-Zambrano G, Barkai O, Chávez S, Pérez-Ortín JE, Darzacq X, Choder M. 2013. Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. Cell 153:1000-1011.

26. Parker R. 2012. RNA degradation in *Saccharomyces cerevisae.* Genetics 191:671-702.

27. Dumesic PA, Madhani HD. 2014. Recognizing the enemy within: licensing RNA-guided genome defense. Trends in Biochemical Sciences 39:25-34.

28. Garfinkel DJ, Nyswaner K, Wang J, Cho J-Y. 2003. Post-transcriptional cosuppression of Ty1 retrotransposition. Genetics 165:83-99.

29. Drinnenberg IA, Fink GR, Bartel DP. 2011. Compatibility with killer explains the rise of RNAi-deficient fungi. Science 333:1592.

30. Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, Bartel DP. 2009. RNAi in budding yeast. Science 326:544-550.

31. Matsuda E, Garfinkel DJ. 2009. Posttranslational interference of Ty1 retrotransposition by antisense RNAs. Proceedings of the National Academy of Sciences USA 106:15657-15662.

32. Purzycka KJ, Legiewicz M, Matsuda E, Eizentstat LD, Lusvarghi S, Saha A, Le Grice SFJ, Garfinkel DJ. 2013. Exploring Ty1 retrotransposon RNA structure within virus-like particles. Nucleic Acids Research 41:463-473.

33. Curcio MJ, Garfinkel DJ. 1992. Posttranslational control of Ty1 retrotransposition occurs at the level of protein processing. Molecular and Cellular Biology 12:2813-2825.

34. Garfinkel DJ, Boeke JD, Fink GR. 1985. Ty element transposition: Reverse transcriptase and virus-like particles. Cell 42:502-517.

35. Curcio MJ, Garfinkel DJ. 1999. New lines of host defense: inhibition of Ty1 retrotransposition by Fus3p and NER/TFIIH. Trends in Genetics 15:43-45.

36.   Sanz-Ramos M, Stoye JP. 2013. Capsid-binding retrovirus restriction factors: discovery, restriction specificity and implications for the development of novel therapeutics. Journal of General Virology 94:2587-2598.

37.   Spencer TE, Palmarini M. 2012. Endogenous retroviruses of sheep: a model system for understanding physiological adaptation to an evolving ruminant genome. Journal of Reproduction and Development 58:33-37.

38.   Guthrie C, Fink GR. 1991. Guide to Yeast Genetics and Molecular Biology, p. 933, Methods in Enzymology, vol. 194. Academic Press Inc., San Diego, CA.

39.   Boeke JD, Eichinger D, Castrillon D, Fink GR. 1988. The *Saccharomyces cerevisiae* genome contains functional and nonfunctional copies of transposon Ty1. Molecular and Cellular Biology 8:1432-1442.

40.   Merkulov GV, Lawler JF, Eby Y, Boeke JD. 2001. Ty1 proteolytic cleavage sites are required for transposition: all sites are not created equal. Journal of Virology 75:638-644.

41.   Mitchell DA, Marshall TK, Deschenes RJ. 1993. Vectors for the inducible overexpression of glutathione S-transferase fusion proteins in yeast. Yeast 9:715-722.

42.   Vaidyanathan PP, Zinshteyn B, Thompson MK, Gilbert WV. 2014. Protein kinase A regulates gene-specific translational adaptation in differentiating yeast. RNA 20:912-922.

43. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. 2009. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. Science 324:218-223.

44. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2012. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15-21.

45. Liu H, Krizek J, Bretscher A. 1992. Construction of a *GAL1*-regulated yeast cDNA expression library and its application to the identification of genes whose overexpression causes lethality in yeast. Genetics 132:665-673.

46. Curcio MJ, Garfinkel DJ. 1991. Single-step selection for Ty1 element retrotransposition. Proceedings of the National Academy of Sciences USA 88:936-940.

47. Lawler JF, Merkulov GV, Boeke JD. 2002. A nucleocapsid functionality contained within the amino terminus of the Ty1 protease that is distinct and separable from proteolytic activity. Journal of Virology 76:346-354.

48. Garfinkel DJ, Hedge AM, Youngren SD, Copeland TD. 1991. Proteolytic processing of Pol-TYB proteins from the yeast retrotransposon Ty1. Journal of Virology 65:4573-4581.

49. Eichinger DJ, Boeke JD. 1988. The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: cell-free Ty1 transposition. Cell 54:955-966.

50. Youngren SD, Boeke JD, Sanders NJ, Garfinkel DJ. 1988. Functional organization of the retrotransposon Ty from *Saccharomyces cerevisiae*: Ty protease is required for transposition. Molecular and Cellular Biology 8:1421-1431.

51. Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9:671-675.

52. Melamed C, Nevo Y, Kupiec M. 1992. Involvement of cDNA in homologous recombination between Ty elements in *Saccharomyces cerevisiae.* Molecular and Cellular Biology 12:1613-1620.

53. Sharon G, Burkett TJ, Garfinkel DJ. 1994. Efficient homologous recombination of Ty1 element cDNA when integration is blocked. Molecular and Cellular Biology 14:6540-6551.

54. Winston F, Dollard C, Malone EA, Clare J, Kapakos JG, Farabaugh P, Minehart PL. 1987. Three genes are required for trans-activation of Ty transcription in yeast. Genetics 115:649-656.

55. Elder RT, Loh EY, Davis RW. 1983. RNA from the yeast transposable element Ty1 has both ends in the direct repeats, a structure similar to retrovirus RNA. Proceedings of the National Academy of Sciences USA 80:2432-2436.

56. Adams SE, Mellor J, Gull K, Sim RB, Tuite MF, Kingsman SM, Kingsman AJ. 1987. The functions and relationships of Ty-VLP proteins in yeast reflect those of mammalian retroviral proteins. Cell 49:111-119.

57.   Arribere JA, Gilbert WV. 2013. Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. Genome Research 23:977-987.

58.   Curcio MJ, Hedge AM, Boeke JD, Garfinkel DJ. 1990. Ty RNA levels determine the spectrum of retrotransposition events that activate gene expression in *Saccharomyces cerevisiae*. Molecular and General Genetics 220:213-221.

59.   Servant G, Pennetier C, Lesage P. 2008. Remodeling yeast gene transcription by activating the Ty1 long terminal repeat retrotransposon under severe adenine deficiency. Molecular and Cellular Biology 28:5543-5554.

60.   Servant G, Pinson B, Tchalikian-Cosson A, Coulpier F, Lemoine S, Pennetier C, Bridier-Nahmias A, Todeschini A-L, Fayol H, Daignan-Fornier B, Lesage P. 2012. Tye7 regulates yeast Ty1 retrotransposon sense and antisense transcription in response to adenylic nucleotides stress. Nucleic Acids Research.

61.   Belotserkovskaya R, Sterner DE, Deng M, Sayre MH, Lieberman PM, Berger SL. 2000. Inhibition of TATA-binding protein function by SAGA subunits Spt3 and Spt8 at Gcn4-activated promoters. Molecular and Cellular Biology 20:634-647.

62.   Eisenmann DM, Arndt KM, Ricupero SL, Rooney JW, Winston F. 1992. Spt3 interacts with TFIID to allow normal transcription in *Saccharomyces cerevisiae*. Genes & Development 6:1319-1331.

63. Mohibullah N, Hahn S. 2008. Site-specific cross-linking of TBP *in vivo* and *in vitro* reveals a direct functional interaction with the SAGA subunit Spt3. Genes & Development 22:2994-3006.

64. Cheung V, Chua G, Batada NN, Landry CR, Michnick SW, Hughes TR, Winston F. 2008. Chromatin and transcription-related factors repress transcription from within coding regions throughout the *Saccharomyces cerevisiae* genome. PLoS Biology 6:e277.

65. Gilbert WV. 2010. Alternative ways to think about cellular internal ribosome entry. Journal of Biological Chemistry 285:29033-29038.

66. Harrison PM, Kumar A, Lang N, Snyder M, Gerstein M. 2002. A question of size: the eukaryotic proteome and the problems in defining it. Nucleic Acids Research 30:1083-1090.

67. Kochetov AV. 2008. Alternative translation start sites and hidden coding potential of eukaryotic mRNAs. BioEssays : news and reviews in molecular, cellular and developmental biology 30:683-691.

68. Kochetov AV, Sarai A, Rogozin IB, Shumny VK, Kolchanov NA. 2005. The role of alternative translation start sites in the generation of human protein diversity. Molecular genetics and genomics : MGG 273:491-496.

69. Wang XQ, Rothnagel JA. 2004. 5'-untranslated regions with multiple upstream AUG codons can support low-level translation via leaky scanning and reinitiation. Nucleic Acids Research 32:1382-1391.

70.   Kochetov AV. 2005. AUG codons at the beginning of protein coding sequences are frequent in eukaryotic mRNAs with a suboptimal start codon context. Bioinformatics 21:837-840.

71.   Porras P, Padilla CA, Krayl M, Voos W, Barcena JA. 2006. One single in-frame AUG codon is responsible for a diversity of subcellular localizations of glutaredoxin 2 in *Saccharomyces cerevisiae.* Journal of Biological Chemistry 281:16551-16562.

72.   Garfinkel DJ, Nyswaner KM, Stefanisko KM, Chang C, Moore SP. 2005. Ty1 copy number dynamics in *Saccharomyces*. Genetics 169:1845-1857.

73.   Curcio MJ, Garfinkel DJ. 1994. Heterogeneous functional Ty1 elements are abundant in the *Saccharomyces cerevisiae* genome. Genetics 136:1245-1259.

74.   Farabaugh P. 1995. Post-transcriptional regulation of transposition by Ty retrotransposons of *Saccharomyces cerevisiae*. The Journal of Biological Chemistry 270:10361-10364.

75.   Fink G, Boeke J, Garfinkel D. 1986. The mechanism and consequences of retrotransposition. Trends in Genetics 2:118-123.

76.   Malagon F, Jensen TH. 2011. T-body formation precedes virus-like particle maturation in *S. cerevisiae*. RNA Biology 8:184-189.

77.   Clemens K, Larsen L, Zhang M, Kuznetsov Y, Bilanchone V, Randall A, Harned A, DaSilva R, Nagashima K, McPherson A, Baldi P, Sandmeyer S. 2011. The

Ty3 Gag3 spacer controls intracellular condensation and uncoating. Journal of Virology 85:3055-3066.

78. Larsen LS, Beliakova-Bethell N, Bilanchone V, Zhang M, Lamsa A, Dasilva R, Hatfield GW, Nagashima K, Sandmeyer S. 2008. Ty3 nucleocapsid controls localization of particle assembly. Journal of Virology 82:2501-2514.

79. Doh JH, Lutz S, Curcio MJ. 2014. Co-translational localization of an LTR-retrotransposon RNA to the endoplasmic reticulum nucleates virus-like particle assembly sites. PLoS Genetics 10:e1004219.

80. Braiterman LT, Monokian GM, Eichinger DJ, Merbs SL, Gabriel A, Boeke JD. 1994. In-frame linker insertion mutagenesis of yeast transposon Ty1: phenotypic analysis. Gene 139:19-26.

81. Checkley MA, Mitchell JA, Eizenstat LD, Lockett SJ, Garfinkel DJ. 2013. Ty1 Gag enhances the stability and nuclear export of Ty1 mRNA. Traffic 14:57-69.

82. Monokian GM, Braiterman LT, Boeke JD. 1994. In-frame linker insertion mutagenesis of yeast transposon Ty1: mutations, transposition and dominance. Gene 139:9-18.

83. Martin-Rendon E, Marfany G, Wilson S, Ferguson DJ, Kingsman SM, Kingsman AJ. 1996. Structural determinants within the subunit protein of Ty1 virus-like particles. Molecular Microbiology 22:667-679.

84. Cristofari G, Ficheux D, Darlix J-L. 2000. The Gag-like protein of the yeast Ty1 retrotransposon contains a nucleic acid chaperone domain analogous to

retroviral nucleocapsid proteins. Journal of Biological Chemistry 275:19210-19217.

85.　Rein A, Datta SAK, Jones CP, Musier-Forsyth K. 2011. Diverse interactions of retroviral Gag proteins with RNAs. Trends in Biochemical Sciences 36:373-380.

86.　Kaneko-Ishino T, Ishino F. 2012. The role of genes domesticated from LTR retrotransposons and retroviruses in mammals. Frontiers in Microbiology 3:262.

87.　Best S, Le Tissier P, Towers G, Stoye JP. 1996. Positional cloning of the mouse retrovirus restriction gene Fv1. Nature 382:826-829.

88.　Hilditch L, Matadeen R, Goldstone DC, Rosenthal PB, Taylor IA, Stoye JP. 2011. Ordered assembly of murine leukemia virus capsid protein on lipid nanotubes directs specific binding by the restriction factor, Fv1. Proceedings of the National Academy of Sciences USA 108:5771-5776.

89.　Murcia PR, Arnaud F, Palmarini M. 2007. The transdominant endogenous retrovirus enJS56A1 associates with and blocks intracellular trafficking of Jaagsiekte sheep retrovirus Gag. Journal of Virology 81:1762-1772.

90.　Bénit L, De Parseval N, Casella JF, Callebaut I, Cordonnier A, Heidmann T. 1997. Cloning of a new murine endogenous retrovirus, MuERV-L, with strong similarity to the human HERV-L element and with a gag coding sequence closely related to the Fv1 restriction gene. Journal of Virology 71:5652-5657.

91. Qi CF, Bonhomme F, Buckler-White A, Buckler C, Orth A, Lander MR, Chattopadhyay SK, Morse HC. 1998. Molecular phylogeny of Fv1. Mammalian Genome 9:1049-1055.

92. Goldstone DC, Walker PA, Calder LJ, Coombs PJ, Kirkpatrick J, Ball NJ, Hilditch L, Yap MW, Rosenthal PB, Stoye JP, Taylor IA. 2014. Structural studies of postentry restriction factors reveal antiparallel dimers that enable avid binding to the HIV-1 capsid lattice. Proceedings of the National Academy of Sciences USA 111:9609-9614.

93. Arnaud F, Murcia PR, Palmarini M. 2007. Mechanisms of late restriction induced by an endogenous retrovirus. Journal of Virology 81:11441-11451.

94. Jern P, Coffin J. 2008. Effects of retroviruses on host genome function. Annual Review of Genetics 42:709-732.

95. VanHoute D, Maxwell PH. 2014. Extension of *Saccharomyces paradoxus* chronological lifespan by retrotransposons in certain media conditions is associated with changes in reactive oxygen species. Genetics 198:531-545.

96. Brachmann C, Davies A, Cost G, Caputo E, Li J, Hieter P, Boeke J. 1998. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast 14:115-132.

A

238

GAL1

AUG1 AUG2

Ty1

Ty1i

1000

T399C

T1108C
A1123G
A1296G

ΔA1456

POLΔ

Δ238-281
Δ238-353

Ty1AS RNAs

I
II
III

B

vector WT T399C T1108C A1123G A1296G ΔA1456 Δ238-281 Δ238-353

Ty1his3-AI -

Ty1i -

Figure 3.1 An internal Ty1i transcript is involved in CNC. (A) Functional organization of the Ty1 CNC region, which covers *GAG* and the beginning of *POL*. Location of the *GAL1* promoter (hatched rectangle), LTR (solid triangle), Ty1 transcripts (Figure 3.4), candidate initiation codons present on Ty1i RNA, and CNC[-] defective deletions and point mutations are noted. Ty1AS RNAs I, II and III are shown with dotted lines. Ty1AS RNAs share a common 3' end at nt 136 but have different 5'ends: nt 760 for II and 594 for III. The exact 5' end of Ty1AS RNA I has not been determined (31). (B) Total RNA from a Ty1-less strain with a single chromosomal Ty1*his3-AI* element containing empty vector, wild-type pGPOLΔ (DG2374), or mutant plasmids T399C (YAS73), T1108C (YAS69), A1123G (YAS71), A1296G (YAS72), ΔA1456 (YAS70) was analyzed by Northern blotting to detect Ty1AS RNAs. Cells were grown in glucose and Ty1 strand-

specific (nt 238-1702) and *ACT1* ^{32}P-labeled riboprobes were used. (C) Total RNA from the strains in (B), plus two additional strains containing mutant plasmids Δ238-281 (YAS74) and Δ238-353 (YAS75), was probed for Ty1i transcripts. Ty1*his3-AI* served as a loading control.
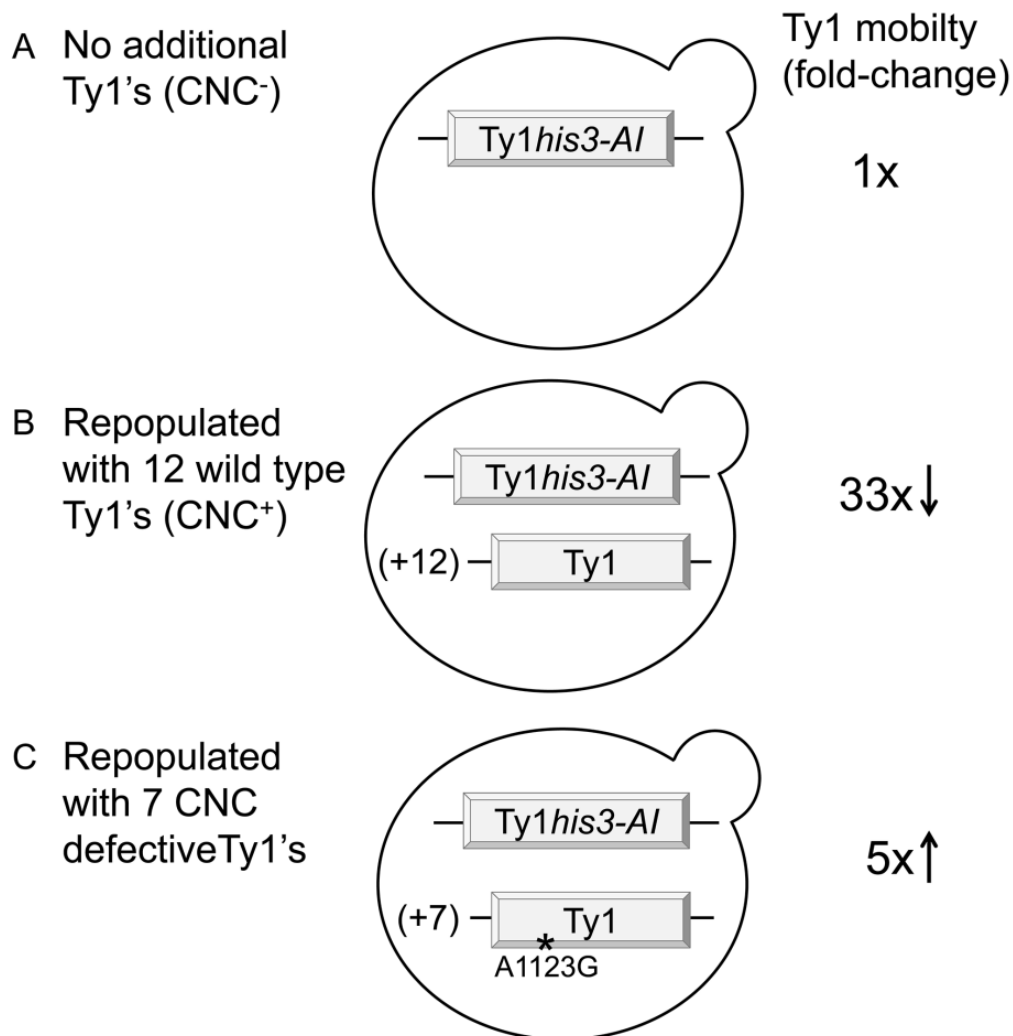
Figure 3.2   Chromosomal Ty1A1123G insertions do not confer CNC. Ty1-less *S. paradoxus* containing a single chromosomal Ty1*his3-AI* (A) was repopulated with unmarked, wild type (B) or A1123G (C) Ty1 elements. Genome repopulation with 12

wild type Ty1 elements resulted in an overall decrease in Ty1*his3-AI* mobility, while

repopulation with 7 CNC⁻ mutant Ty1A1123G elements resulted in an overall increase in

Ty1*his3-AI* mobility. Also refer to Table 2B.
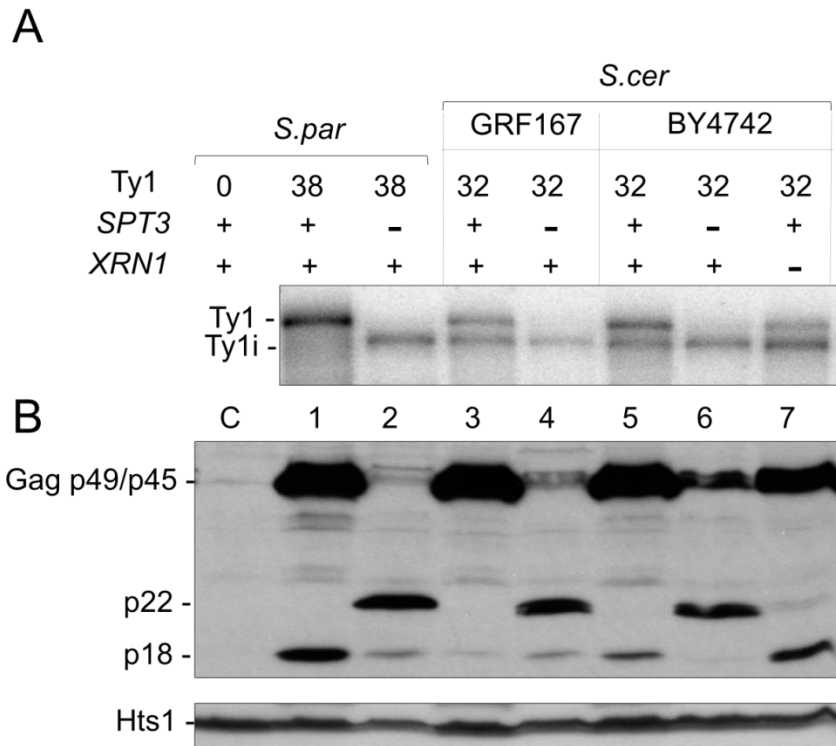
Figure 3.3 Detecting Ty1i RNA and p22/p18 from chromosomal Ty1 elements. (A)

Northern blotting of poly(A)⁺ RNA from *S. paradoxus* and *S. cerevisiae* (GRF167 and

BY4742) wild-type, *spt3Δ* (DG789 and DG2247) and *xrn1Δ* (MAC103) mutant strains. A

Ty1 ^{32}P-labeled riboprobe (nt 1266-1601) hybridized with full-length Ty1 and Ty1i

transcripts. (B) Total protein extracts were immunoblotted with the p18 antiserum to

detect full-length Gag p49/p45 and p22/18. A Ty1-less *S. paradoxus* strain (DG1768)

and cellular histidyl tRNA synthetase (Hts1) served as negative (lane C) and loading

controls, respectively.
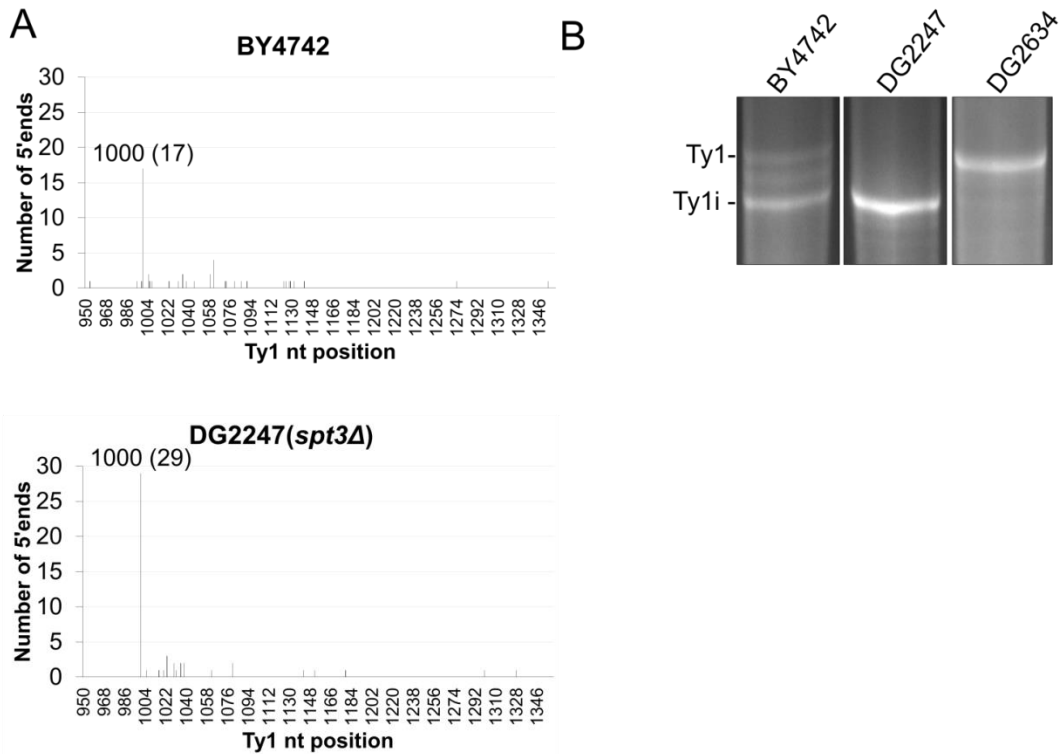
Figure 3.4 The major 5' end of the 4.9 kb Ty1i RNA maps to nt 1000. (A) Cap-independent 5'- RACE was performed with poly(A)$^+$ RNA from wild-type BY4742 and an isogenic *spt3Δ* mutant (DG2247). The number of 5' termini was plotted against the Ty1H3 sequence and the distribution of the termini are on the X and Y-axis respectively. The tallest peak represents the total number of 5' ends captured at nt 1000 and is shown in parentheses. (B) 5' RACE cDNA libraries from the wild type and *spt3Δ* strains mentioned above, and a repopulated *S. paradoxus* strain (DG2634) were amplified using a universal primer mix and a Ty1-specific primer GSP1_3389. The amplification reactions were separated by agarose gel electrophoresis to demonstrate the presence of cDNA products corresponding to the 5' ends of the full-length (5.7 kb) Ty1 and the truncated (4.9 kb) Ty1i RNAs.
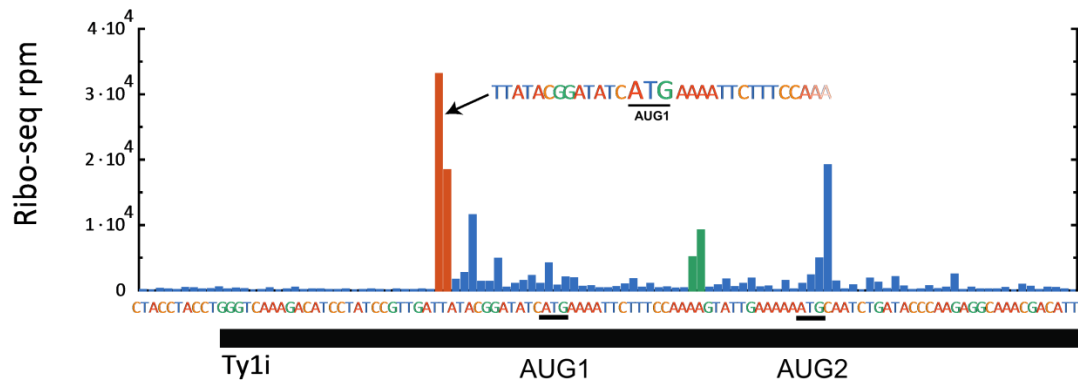
Figure 3.5   Whole genome analysis of internal translation initiation sites. Ribosome footprint profiling (Ribo-seq) was performed to detect translation initiation at internal AUG codons, two of which (AUG1 and AUG2, see Figure 3.1) are located immediately downstream of the Ty1i RNA transcription start site. Reads per million (rpm) Ty1-mapped reads were placed on the Ty1H3 sequence and the 5' end of ribosome footprints aligned downstream of the Ty1i transcription start are shown. Ribo-seq reads with 5' ends 12-13 nt upstream of AUG1 and AUG2 are highlighted in orange and green, respectively. ~12 nt downstream of the 5' end corresponds to the ribosomal P-site. Because these libraries were prepared with poly(A) tailing, the exact 3'-end of the footprint, and thus the footprint size at AUG1, is ambiguous but within the range of 26-30 nt, inclusive.

Figure 3.6 p22 is necessary for CNC. (A) Ty1 sequence present on pGPOLΔ illustrating the Ty1i RNA transcription start site (nt 1000), location of in-frame AUGs, and frameshift mutations (ΔC1071 and +A1303, black circles). Proteins encoded by wild-type (WT) or mutant plasmids are shown (wild-type sequence, solid; nonsense sequence, dashed) based on predicted usage of AUG1 by Ribo-seq (Figure 3.5). ΔC1071 and +A1303 are predicted to synthesize truncated p22 peptides of 11 and 89 residues, respectively, before encountering the frameshift mutation. (B) A *S. paradoxus* strain with a single chromosomal Ty1*his3-AI* carrying an empty vector (DG2411), pGPOLΔ (DG2374) or the mutant plasmids ΔC1071 (JM321) and +A1303 (JM320) were assessed for Ty1 mobility

using a qualitative assay. Cell patches grown on SC-Ura medium at 22°C were replica plated to SC-Ura-His medium to select for cells that contain at least one Ty1*HIS3* insertion. The number of His^{+} papillae that grew on SC-Ura-His medium is a read-out for Ty1 mobility. Also refer to Table 2C-1. (C) Total RNA from the strains described above was subjected to Northern blotting to detect Ty1*his3-AI* and Ty1i transcripts as described in Figure 3.1. The band labeled with an asterisk is a pervasive transcript approximately 4.5 kb in length, and contains both Ty1 and non-Ty1 sequences from the pGPOLΔ. The 'r' represents compression bands formed by two main species of ribosomal RNA in yeast, the 26S (3.8kb) and 18S (2kb) rRNAs. (D) Total cell extracts were analyzed for the presence of p22/p18 as described in Figure 3.3.
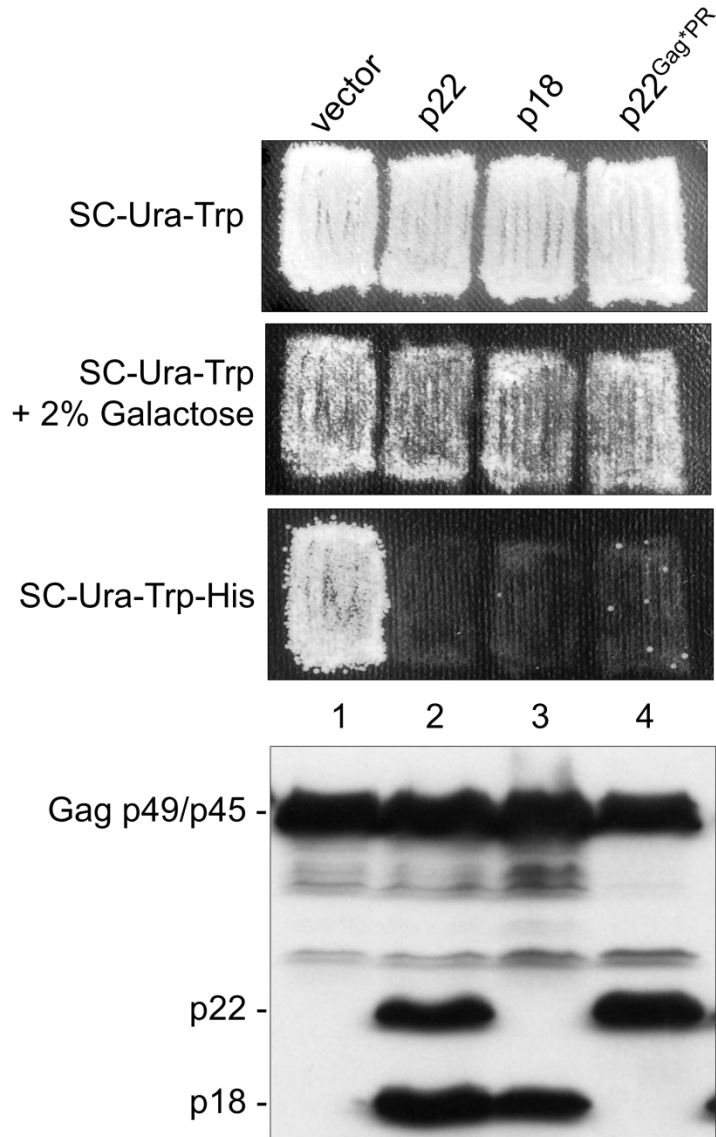
Figure 3.7 Cleavage of p22 to p18 does not disturb *trans*-dominant inhibition of Ty1 mobility. A mutant Gag-PR cleavage site, AAGSAA (Gag*PR) (40), was inserted into p22, replacing the normal Gag-PR cleavage site, RAHNVS. A Ty1-less strain containing pGTy1*his3-AI* and an empty vector (DG3739; lane 1), or *GAL1*-p22 (DG3774, lane 2), *GAL1*-p18 (DG3791, lane 3), and *GAL1*-p22^Gag*PR (JM399, lane 4) were analyzed for Ty1*his3-AI* mobility using a qualitative assay. Cell patches from a single colony were induced for pGTy1 expression by replica plating from SC-Ura-Trp to SC-Ura-Trp + 2%

183

galactose medium for 2 days at 22°C. To detect Ty1*his3-AI* mobility, galactose-induced

cells were replica plated to SC-Ura-His medium. Below is an immunoblot using total

cell extracts from the same strains and the p18 antiserum to detect Gag-p49/p45 and

p22/p18.
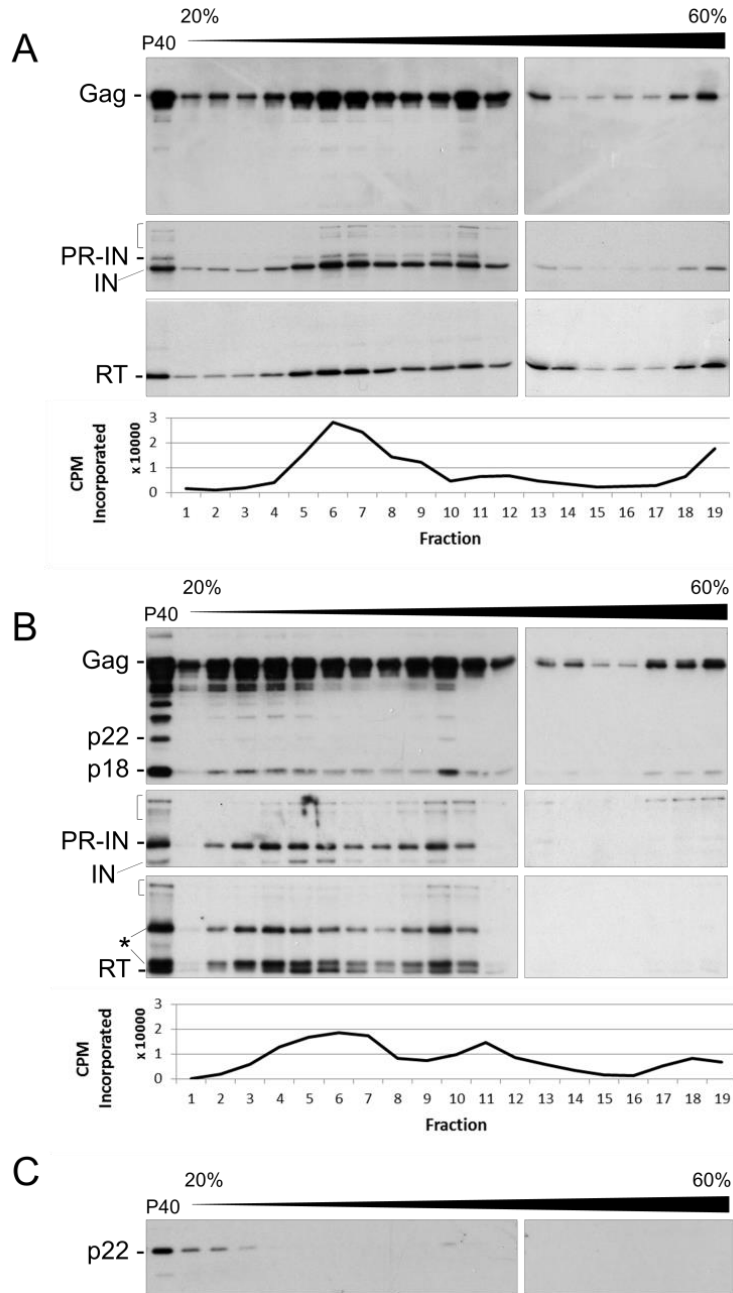
Figure 3.8   Cofractionation of p22/p18 with Ty1 VLPs.   Crude VLP pellets (P40) prepared from galactose-induced Ty1-less strains expressing pGTy1*his3-AI* alone (A; DG3739), pGTy1*his3-AI* and p22 (B; DG3774) or p22 alone (C; DG3784) were fractionated through a 20-60% continuous sucrose gradient. VLP pellets (P40) and equal volumes from collected fractions were analyzed by immunoblotting with p18

185

antiserum, IN and RT antisera. Ty1 proteins are labeled, brackets indicate known Ty1 processing intermediates, and asterisks indicate aberrant Ty1 proteins (estimated size: 65 and 90 kD). Reverse transcriptase activity was detected using an exogenous poly(rC)-oligo/(dG) template and [α^{32}P]-dGTP.

A   pGTy1*his3AI*, empty

Closed   Open
87%   13%

d= 37.4 ± 2.7 nm

B   pGTy1*his3AI*, p22

Closed   Open
55%   46%
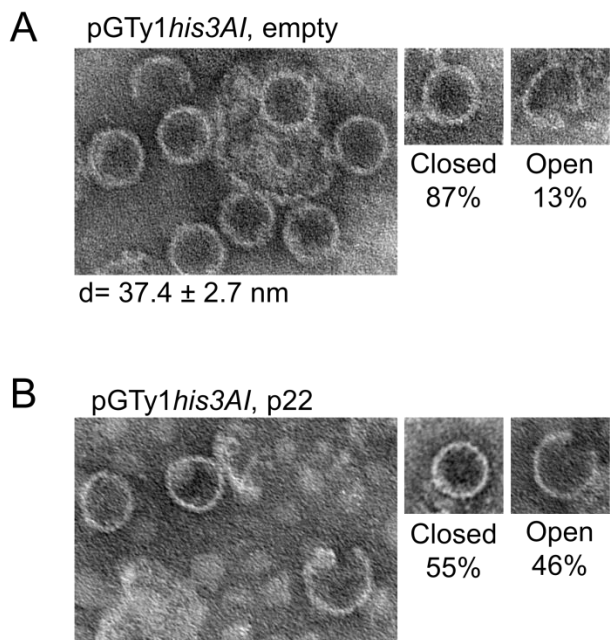
Figure 3.9 Electron microscopy of Ty1 VLPs assembled in the presence of p22/p18. VLP pellets were collected from sucrose gradient fractions with peak reverse transcriptase activity from experiments similar to those shown in Figure 3.8. VLPs from pGTy1*his3-A*I alone (A; DG3739) or pGTy1*his3-AI* and p22 (B; DG3774) were stained with 2% ammonium molybdate and examined by transmission electron microscopy. Approximately 100 VLPs were analyzed for closed versus open particles and representative images are shown. The diameter (d) was measured with closed VLPs only.

186
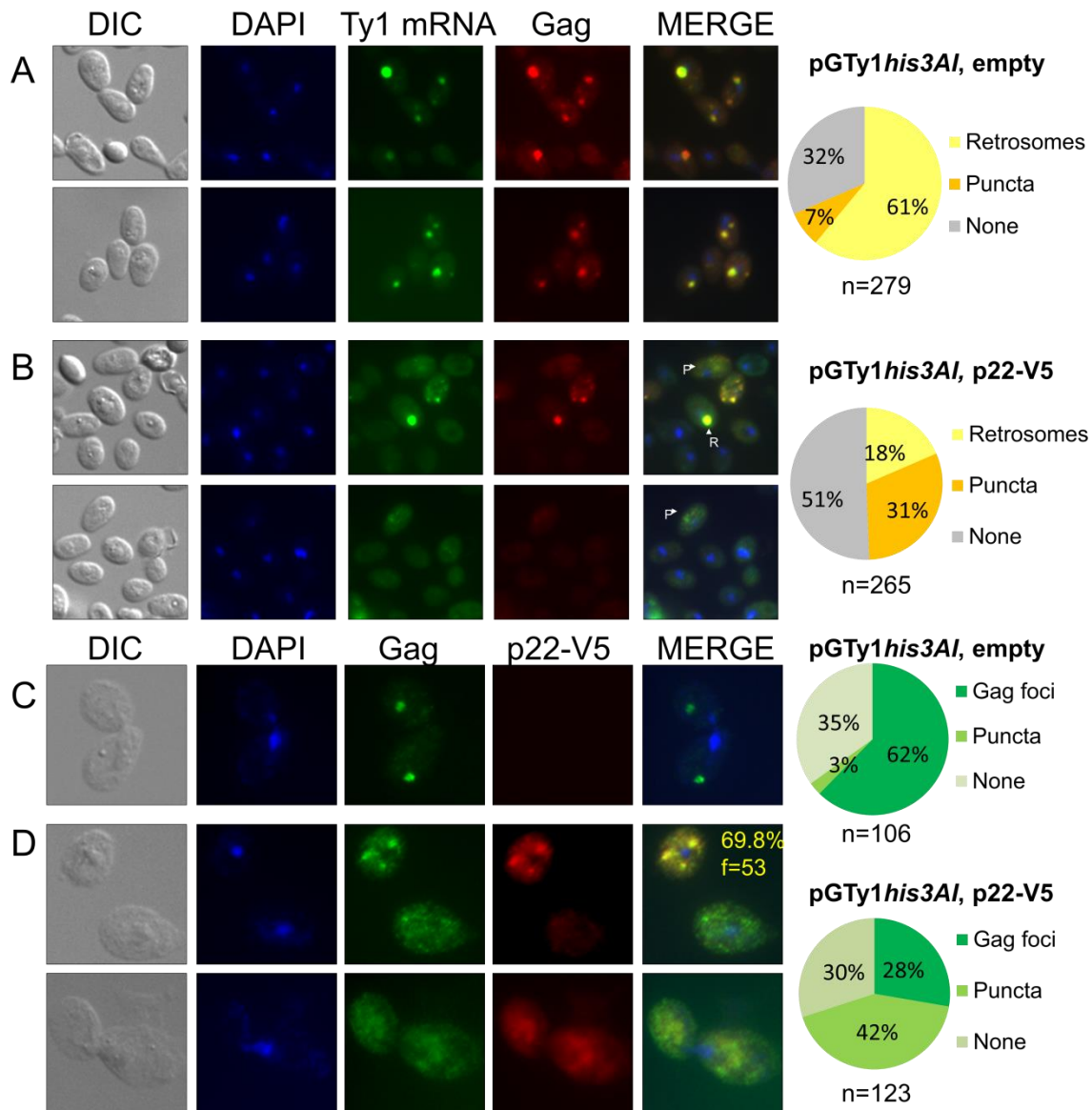
Figure 3.10   p22-V5 disrupts retrosomes and colocalizes with Gag. Ty1-less strains expressing pGTy1*his3-AI* alone (A, C; DG3739) or pGTy1*his3-AI* and p22-V5 together (B, D; JM367) were galactose-induced and analyzed for Ty1 mRNA and Gag colocalization via FISH/IF (panels A and B). Pie charts depict cells examined for the appearance of retrosomes (R), puncta (P), or no staining (None). Refer to the text for

additional details. (C and D) In a separate experiment, cells were analyzed for Ty1 Gag

and p22-V5 colocalization via IF using VLP and V5 antibodies, respectively. The

experiment in panel D was additionally analyzed for the percentage of Gag foci that

colocalize with p22-V5 (yellow, f = total Gag foci analyzed). For both experiments, DNA

was stained with DAPI and representative images are shown (n= number of cells

analyzed).
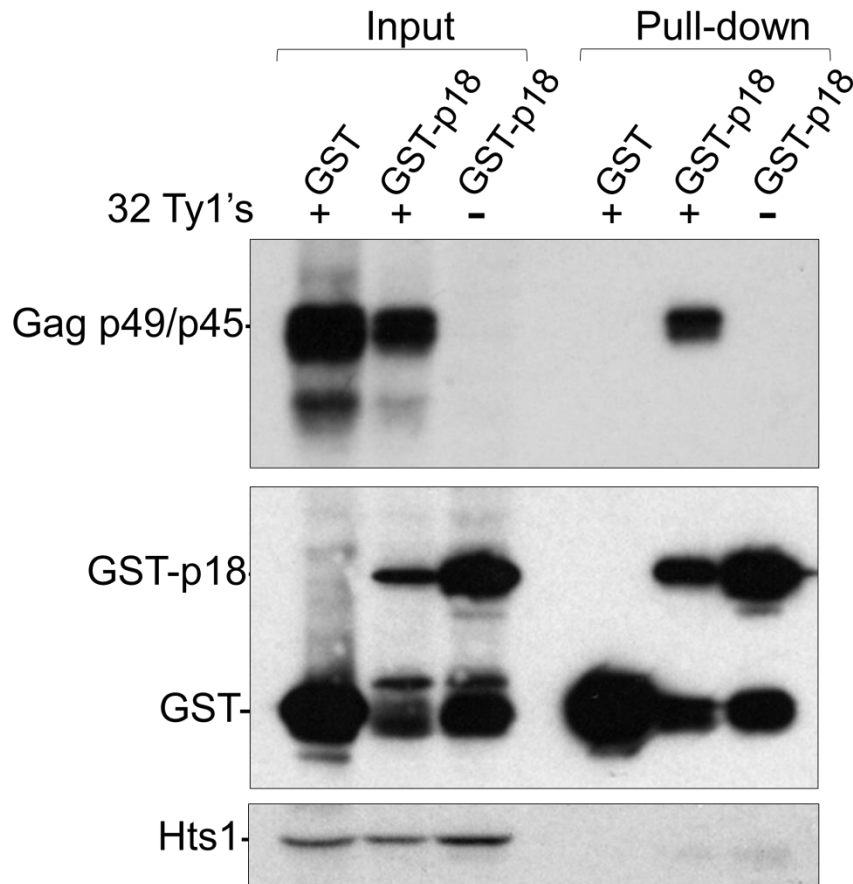
Figure 3.11 GST-p18 interacts with endogenous Ty1 Gag. Protein extracts (Input) from

BY4742 induced for expression of GST (DG3808) or GST-p18 (DG3809) were

incubated with glutathione-coated resin. Bound proteins were analyzed by

immunoblotting to detect Gag, GST-p18, and p18/Ty1 Gag complexes (Pull-down) after extensive washing with lysis buffer. A Ty1-less strain expressing GST-p18 (DG3810) and the presence of Hts1 served as negative controls. Gag was detected with TY tag monoclonal antibody, which recognizes p49/p45 but not p22/p18 due to the location of the epitope. GST proteins and Hts1 were detected with GST and Hts1 antibodies, respectively.

Tables

Table 3.1 Yeast strains

Strain	Genotype	Plasmid	Source
DG2196	*MATa*, *his3-Δ200hisG*, *ura3*, *trp1*, Ty1-less, Ty1*his3-AI(96)*	-	(28)
DG2254	DG2196	pGAL/2μ-*URA3*	(28)
DG2255	DG2196	pGTy1/2μ-*URA3*	(28)
DG2411	DG2196	empty/2μ-*URA3*	(28)
DG2374	DG2196	pBDG1130 (pGPOLΔ/2μ-*URA3*)	This study
YAS73	DG2196	pBAS39 (pGPOLΔ-T399C)	This study
YAS69	DG2196	pBAS38 (pGPOLΔ-T1108C)	This study
YAS71	DG2196	pBAS35 (pGPOLΔ-A1123G)	This study
YAS72	DG2196	pBAS36 (pGPOLΔ-A1296G)	This study

YAS70	DG2196	pBAS34 (pGPOLΔ-ΔA1456)	This study
YAS74	DG2196	pBAS43 (pGPOLΔ-Δ238-281)	This study
YAS75	DG2196	pBAS44 (pGPOLΔ-Δ238-353)	This study
JM321	DG2196	pBJM79 (pGPOLΔ-ΔC1071)	This study
JM320	DG2196	pBJM78 (pGPOLΔ- +A1303)	This study
DG2511	DG2196 +12 Ty1		This study
DG3856	DG2196	pBDG1595 (pGPOLΔ-GCG1GCG2)	This study
DG2512	DG2196 +9 Ty1		This study
DG3798	DG2196 +7 Ty1-A1123G		This study
DG1768	*MATα, his3-Δ200hisG, ura3,* Ty1-less		(28)
DG2533	DG1768, *Ty1–4253his3-AI*		(3)
DG2634	DG1768, *Ty1–4253his3-AI*, +37 Ty1		(31)
YEM515	DG2634, *spt3-ΔKanMX4*		E. Matsuda
GRF167	*MATα, ura3-167, his3-Δ200*		(15)

DG789	GRF167, *spt3-101*		(50)
BY4742	*MATα*, *his3-Δ1*, *leu2-Δ0*, *lys2-Δ0*, *ura3-Δ0*		(96)
DG2247	BY4742, *spt3-ΔKanMX4*		(31)
MAC103	BY4742, *xrn1/kem1-ΔKanMX4*		(7)
DG3582	DG1768, *trp1*		This study
DG3753	DG3582	p*GAL*/Cen-*URA3*, pBDG1534 (pGTy1*his3-AI*/Cen-*TRP1*)	This study
DG3751	DG3582	pBDG1354 (p*GAL*:1042-5889/Cen-*URA3*), pBDG1534	This study

Strain	Genotype	Plasmid	Source
DG3739	DG3582	pGAL-Yes2/2μ-*URA3*, pBDG1534	This study
DG3774	DG3582	pBDG1565 (p*GAL*-Yes2:1038-1616), pBDG1534	This study
DG3784	DG3582	pBDG1565, empty/Cen-*TRP1*	This study
JM367	DG3582	pBDG1568 (p*GAL*-Yes2:1038-V5-1616), pBDG1534	This study
DG3791	DG3582	pBDG1571 (p*GAL*-Yes2:1038-1496), pBDG1534	This study

JM399	DG3582	pBJM90 (p*GAL*-Yes2:1038-1616^{Gag*PR}), pBDG1534	This study
DG3808	BY4742	pEG(KT): p*GAL*:GST/2μ-*URA3*	This study
DG3809	BY4742	pBDG1576 (p*GAL*:GST-1038-1496)	This study
DG3810	DG3582	pBDG1576	This study

Table 3.2 Ty1*his3-AI* mobility

	Strain	Relevant genotype	Ty1*his3-AI* mobility x 10^{-6} (SD)	Fold decrease
A	DG2254	Ty1-less, Ty1*his3-AI*, p*GAL*/2μ	220 (69)	1
	DG2374	pGPOLΔ	7 (1.8)	31
	YAS73	pGPOLΔ-T399C (Gag: Ser36Pro)	24 (5.7)	9.2
	YAS69	pGPOLΔ-T1108C (Gag: Leu272Pro)	100 (27)	2.2
	YAS71	pGPOLΔ-A1123G (Gag: Tyr277Cys)	82 (11)	2.7
	YAS72	pGPOLΔ-A1296G (Gag: Thr335Ala)	42 (3.4)	5.2
	YAS70	pGPOLΔ-ΔA1456	31 (9.7)	7
B	DG2196	Ty1-less, Ty1*his3-AI*	120 (14)	1
	DG2511	+12 Ty1	3.6 (0.63)	33
	DG3798	+7 Ty1A1123G	570 (150)	4.8↑

C-1	DG2411	Ty1-less, Ty1*his3-AI*, empty/2µ	140 (30)	1
	DG2374	pGPOLΔ	2 (0.36)	74
	JM321	pGPOLΔ-ΔC1071	61 (39)	2.3
	JM320	pGPOLΔ- +A1303	58 (6.2)	2.4
C-2	JM321	pGPOLΔ-ΔC1071	65 (22)	2.1
	DG3856	pGPOLΔ-GCG1GCG2	137 (23)	1
D	DG3753	Ty1-less, pGTy1*his3-AI*/Cen, p*GAL*/Cen	21000 (2600)	1
	DG3751	pGAL:1042-5889	37 (8.8)	570
E	DG3739	Ty1-less, pGTy1*his3-AI*/Cen, pYES2/2µ	60000 (4700)	1
	DG3774	pYES2: 1038-1616	1.7 (0.57)	35,000
	JM367	pYES2: 1038-V5-1616	2 (0.65)	32,000

CHAPTER 4

THE TY1 RETROTRANSPOSON RESTRICTION FACTOR P22 TARGETS GAG[3]

Abstract

A novel form of copy number control (CNC) helps maintain a low number of Ty1 retrovirus-like transposons in the *Saccharomyces* genome. Ty1 produces an alternative transcript that encodes p22, a *trans*-dominant negative inhibitor of Ty1 retrotransposition whose sequence is identical to the C-terminal half of Gag. The level of p22 increases with copy number and inhibits normal Ty1 virus-like particle (VLP) assembly and maturation through interactions with full length Gag. A forward genetic screen for CNC-resistant (CNC[R]) mutations in Ty1 identified missense mutations in *GAG* that restore retrotransposition in the presence of p22. Some of these mutations map within a predicted UBN2 domain found throughout the Ty1/*copia* family of long terminal repeat retrotransposons, and others cluster within a central region of Gag that is referred to as the CNC[R] domain. We generated multiple alignments of yeast Ty1-like Gag proteins and found that some Gag proteins, including those of the related Ty2 elements, contain non-Ty1 residues at multiple CNC[R] sites. Interestingly, the Ty2-917 element is resistant to p22 and does not undergo a Ty1-like form of CNC. Substitutions conferring CNC[R] map within predicted helices in Ty1 Gag that overlap with conserved sequence in Ty1/*copia*, suggesting that p22 disturbs a central function of the capsid during VLP assembly. When hydrophobic residues within predicted helices in Gag are mutated, Gag level remains unaffected in most cases yet VLP assembly and maturation is abnormal. Gag CNC[R] mutations do not alter binding to p22 as determined by co-immunoprecipitation analyses, but instead, exclude p22 from Ty1 VLPs. These findings suggest that the CNC[R] alleles enhance retrotransposition in the presence of p22 by allowing productive Gag-Gag interactions during VLP assembly. Our work also expands

196

the strategies used by retroviruses for developing resistance to Gag-like restriction factors to now include retrotransposons.

Author Summary

The presence of transposable elements in the eukaryotic genome threatens genomic stability and normal gene function, thus various defense mechanisms exist to silence element expression and target integration to benign locations in the genome. Even though the budding yeast *Saccharomyces* lacks many of the defense systems present in other eukaryotes, including RNAi, DNA methylation, and APOBEC3 proteins, they maintain low numbers of mobile elements in their genome. In the case of the *Saccharomyces* retrotransposon Ty1, a system called copy number control (CNC) helps determine the number of elements in the genome. Recently, we demonstrated that the mechanism of CNC relies on a *trans*-acting protein inhibitor of Ty1 expressed from the element itself. This protein inhibitor, called p22, impacts the replication of Ty1 as its copy number increases. To identify a molecular target of p22, mutagenized Ty1 was subjected to a forward genetic screen for CNC-resistance. Mutations in specific domains of Gag, including the UBN2 Gag motif and a novel region we have named the CNC^R domain, confer CNC^R by preventing the incorporation of p22 into assembling virus-like particles (VLPs), which restores maturation and completion of the Ty1 life cycle. The mechanism of Ty1 inhibition by p22 is conceptually similar to Gag-like restriction factors in mammals since they inhibit normal particle function. In particular, resistance to p22 and the enJS56A1 restriction factor of sheep involves exclusion of the restriction factor during particle assembly, although Ty1 CNC^R achieves this in a way

that is distinct from the Jaagsiekte retrovirus escape mutants. Our work introduces an intriguing variation on resistance mechanisms to retroviral restriction factors.

Introduction

The Ty1 and Ty2 retrotransposons of *Saccharomyces* belong to the Ty1/*copia* group of long terminal repeat (LTR) retrotransposons which replicate in a manner analogous to retroviruses (1). Ty1 is the most abundant of five retrotransposon families (Ty1-Ty5) in the S288C reference genome of *Saccharomyces cerevisiae,* followed by the related Ty2 element (2, 3). Recently, Ty2 has been shown to outnumber Ty1 in some *Saccharomyces* genomes (2-4), but Ty1 remains the more widely studied retrotransposon (1). Ty1 contains two overlapping ORFs, *GAG* and *POL*, and many elements are transpositionally competent and transcriptionally active (5). An abundant full-length Ty1 mRNA is transcribed which serves as a template for translation and reverse transcription. Two translation products are produced: Gag (p49) and Gag-Pol (p199), of which the latter comprises only 5-10% of total translation products due to its production requiring a +1 ribosomal frameshifting event. Gag, Gag-Pol and Ty1 mRNA accumulate in the cytoplasm in distinct foci called retrosomes (6-9). Virus-like particles (VLPs) assemble from Gag and Gag-Pol proteins within retrosomes and encapsidate Ty1 mRNA, and tRNA^{iMet}, which is used to prime reverse transcription. VLP maturation occurs via the activity of the *POL*-encoded enzyme, protease (PR). Pol is cleaved from p199 via a PR-dependent autocatalytic event, followed by PR cleavage of Gag at its C-terminus (from p49 to p45) and Pol at two internal sites to form mature PR, integrase (IN), and reverse transcriptase (RT). Once maturation occurs, reverse transcription of

198

the packaged genomic Ty1 RNA forms a cDNA copy that is integrated into the host genome. Because Ty1 insertions can mutate cellular genes and mediate chromosome instability by homologous recombination with elements dispersed in the genome, it is beneficial to the host to control the process of retrotransposition (10-13).

Natural isolates of *S. cerevisiae* and its closest relative *S. paradoxus* maintain lower copy numbers of the Ty1 retrotransposon in their genomes compared to the reference laboratory strain S288C (2-4, 14), without the support of eukaryotic defense mechanisms such as RNAi or the presence of innate restriction factors like the APOBEC3 proteins (15-19). The maintenance of Ty1 copy number is due at least in part to a mechanism called copy number control (CNC), which was first observed in an isolate of *S. paradoxus* that lacks complete Ty1 elements but contains numerous solo-LTRs (20). The Ty1-less strain supports higher levels of Ty1 transposition compared to standard lab strains, as monitored using a Ty1 tagged with the *his3-AI* retrotransposition indicator gene (21). Additionally, Ty1*his3-AI* mobility decreases dramatically when the naive genome is repopulated with Ty1 elements (20). Introduction of a transcriptionally repressed Ty1 element on a multi-copy plasmid also inhibits Ty1*his3-AI* mobility. Based on these observations, CNC is conferred by a factor produced directly by the Ty1 element. The CNC phenotype, which includes decreased levels of transposition (20), the reduction of mature Ty1 RT and PR protein levels, and the absence of detectable mature IN (22), is dependent on the *GAG* open reading frame (20). Overexpression of Ty1 fused to a *GAL1* promoter on a multi-copy plasmid has been shown to override CNC, suggesting that CNC can be saturated (23-26).

Recently, we found that CNC functions through the protein product encoded by a subgenomic internally-initiated Ty1 sense transcript, called Ty1i (internal) RNA (26). Transcription of Ty1i RNA initiates within *GAG*, about 800 nucleotides downstream of the full-length, transposition-competent Ty1 mRNA. The first AUGs are in the same reading frame as Ty1 Gag, resulting in synthesis of a 22 kD protein (p22) that shares the same sequence as the C-terminal half of Gag (26, 27). This shared sequence includes the PR cleavage site, which is utilized within the inhibitory protein to form p18 (26). Ectopic co-expression of p22 or p18 with Ty1 dramatically inhibits Ty1 mobility. p22/p18 co-immunoprecipitates with Ty1 Gag and co-localizes with Ty1 Gag in the cytoplasm. Ectopic expression of p22/p18 disrupts normal retrosome formation and VLP assembly, followed by a block in maturation and reverse transcription within the VLPs that are able to form. In addition, p18 interferes with the nucleic acid chaperone (NAC) function of Gag, further disrupting Ty1 replication (27, 28). It is not clear which insult by p22/p18 is most destructive, but collectively these effects result in the strong inhibition of retrotransposition observed during CNC.

Retroelement restriction mechanisms have been aided by studying resistance mutations in retroviruses and/or sequence variation determining viral tropism. An example of particular relevance is the discovery that capsid (CA), a cleavage product of retroviral Gag, is the target of several restriction factors including Friend virus susceptibility factor-1 (Fv1), tripartite motif 5 alpha (TRIM5α), and myxovirus resistance protein 2 (Mx2), among others (29-33). Viruses that escape restriction by these factors typically carry mutations in the CA-encoding region of the *gag* gene (29, 34-38). In the case of Fv1 and TRIM5α, viral escape mutations disrupt binding between the restriction

factor and CA by altering CA surface residues (37-40), while Mx2 escape mutations in CA are not fully understood but likely alter the interactions between CA and host factors (32, 33, 41). While Fv1, TRIM5α, and Mx2 bind the incoming viral capsid during the early stages of retroviral infection, the sheep restriction factor enJS56A1 is known to interact with Jaagsiekte sheep retrovirus (JSRV) Gag at later stages when the integrated provirus is undergoing translation and particle assembly (42). Resistance to enJS56A1 is conferred by mutations in the signal peptide of the JSRV envelope glycoprotein, which is hypothesized to ultimately alter the ratio of JSRV to enJS56A1 Gag levels to favor JSRV particle production (43). Remarkably, Fv1 and enJS56A1 are both derived from endogenous retroelement *gag* genes (42, 44), similar to p22.

To further understand the mechanism of CNC, we carried out forward genetic screens for CNC-resistant (CNC[R]) Ty1 elements. Almost all of the CNC[R] elements contain missense mutations within *GAG* that map within predicted helices important for VLP assembly and maturation. Computational and functional analyses reveal three domains within the Ty1 Gag protein: TYA, CNC[R] and UBN2. All resistance mutations recovered map within the CNC[R] and UBN2 domains encoded by *GAG*. Importantly, several mutations are not present in p22 coding sequence, supporting the idea that p22 targets Gag to inhibit retrotransposition. Most CNC[R] mutations in *GAG* do not markedly alter Ty1 fitness or the interaction between Gag and p22, but prevent co-assembly of Gag and p22 into VLPs, which improves VLP maturation and progression through the retrotransposon life cycle.

Materials and Methods

Genetic techniques, media and strain construction

Strains are listed in S1 Table. Strains repopulated with Ty1 elements were obtained following pGTy1 induction as described previously (20). Ty1 insertions following repopulation experiments were estimated by Southern blotting as in (45). Standard yeast genetic and microbiological procedures were used in this work (46).

Plasmids

Refer to S3 Table for plasmid descriptions and sources. Directed mutagenesis was carried out by overlap PCR using the following primer sequences: W184Ab; 5'-ATGTTTTAACAGCATTTGGAAAGTCATTAGGTGAGGTTAAC; W184Ac, 5'-GACTTTCCAAATGCTGTTAAAACATACATCAAATTTTTAC; L252Rb, 5'-ATACTTTTGGATCTAATTTTCATGATATCCGTATAATCAACG; L252Rc, 5'-TCATGAAAATTAGATCCAAAAGTATTGAAAAAATGCAATCTG; IM248/9NRb, 5'-AAAGAATTTTCCTGTTATCCGTATAATCAACGGATAGGAT; IM248/9NRc, 5'-TATACGGATAACAGGAAAATTCTTTCCAAAAGTATTGAAA; LF339/40RDb, 5'-GGATATCTAAGTCCCGTTCAGCGACTGTCATATTTAGATG; LF339/40RDc, 5'-GTCGCTGAACGGGACTTAGATATCCATGCTATTTATGAAG; I343Kb, 5'-AAATAGCATGCTTATCTAAGAACAGTTCAGCGACTGTCAT; I343Kc, 5'-CTGTTCTTAGATAAGCATGCTATTTATGAAGAACAACAGG. For pBJM24, the plasmid markers were switched from *URA3* to *TRP1*, as described previously (26). Galactose-inducible centromere (*CEN*) vectors expressing p22-V5 were created by PCR amplification of Ty1-H3 p22 coding sequence 1038-1613 with the internal V5 tag and flanking *GAL1P* and *CYC1* TT sequences using pBDG1568 as a template (26) and

primers: cla1_galp, 5'-CATGTTTCATCGATACGGATTAGAAGCCGCCGAGC; cyc1ttrevSacII, 5'-CATGTTTCCCGCGGGAGTCAGTGAGCGAGGAAGC. The insert was cloned into an empty *URA3/CEN* vector (pRS416) using ClaI and SacII sites. The V5 tag is located between nucleotides 1442 and 1443(26). pTy2-917*his3-AI* (pBDG631) was constructed by digestion of pGTy917 with BglII and pBJC42 (*his3-AI*, pBDG619) with ClaI, fill-in synthesis of the linearized vector and *his3-AI* fragments using DNA polymerase I followed by blunt end-ligation. pGPOLΔ*d1* was constructed by BglII digestion and reclosure of pGTy1*his3-AId1* (kindly provided by Jef Boeke (47, 48) and Joan Curcio (49)), which deletes the majority of *POL*. pYES2-p45 was constructed by PCR using primers specific for the coding sequence of p45 and the amplification product was cloned into the multi-copy *GAL1*-promoted expression vector pYES2 (27). Recombinant plasmids were verified by restriction enzyme analysis or DNA sequencing.

CNC-resistance screen

Plasmid mutagenesis was performed by transforming 50 ng of pBDG606 (S3 Table) into XL-1 Red (Agilent Technologies) cells and sub-culturing transformants for 3-4 days at 37° C. Gap repair was performed with pBDG606 using mutagenized *GAG* template and AatII (upstream of *GAL1*P) and BstEII (within PR) sites, while *POL* mutagenesis was performed using BstEII and XbaI (within *his3-AI*) sites. Primers flanking these restriction sites (*GAG*: USAatII, 5'-ATAATACCGCGCCACATAGC ; RP1, 5'-CATTGATAGTCAATAGCACTAGACC; *POL*: USBsteIIf, 5'-GCACGACCTTCATCTTAGGC; 3pLTRrev, 5'-ATCAATCCTTGCGTTTCAGC) were used in a standard *Taq* (ThermoFisher Scientific, Waltham, MA) PCR reaction with Ty1-H3 as a template to mutagenize the area of interest at a low frequency. XL-1 Red

treated pBDG606 or DNA fragments for gap repair were transformed into YEM515 (see S1 Table) and plated onto SC-Ura. Transformants, were replica plated on SC-Ura + 2% galactose, incubated at 22° C for 1-2 days, and then replica plated on SC-Ura-His and incubated at 30° C for 2 days. Candidate plasmids were extracted, propagated in *E. coli*, transformed into YEM514 and YEM515 and retested for pTy1*his3-AI* mobility. Candidates with at least a 10-fold increase in retromobility in YEM515 were carried forward. After sequencing the CNC region of XL-1 Red treated plasmids, the *GAL1* and *GAG* segments were subcloned into wild type plasmid using AatII and Eco91I sites to eliminate other mutations present outside of the region of interest. In all cases, subcloned *GAG* mutations conferred a similar level of CNC^R as the primary isolates. For the gap repair screen, the entire region amplified by low fidelity PCR was sequenced.

Ty1*his3-AI* mobility

The mobility frequency of Ty1*his3-AI* was determined as described previously (21, 45) with the following modifications. For strains transformed with only pGTy1*his3-AI*, single colonies were grown at 30° C overnight in 1 ml of SC-Ura + 2% raffinose and then diluted 1:25 in quadruplicate 1 ml SC-Ura + 2% galactose cultures. Galactose cultures were grown at 22° C for 2 days, and cells were then washed, diluted and spread onto SC-Ura and SC-Ura-His plates. For strains transformed with both pGTy1*his3-AI* and p22-containing plasmids, similar methods were used for the assay except liquid and solid media also lacked tryptophan. For qualitative mobility assays with pGTy1*his3-AI*, cells were patched onto SC-Ura and grown at 30° C for 2 days. Cells were replica plated onto SC-Ura +2% galactose and incubated at 22° C for 2 days, followed by replica plating onto SC-Ura-His and incubation at 30° C until His$^+$ papillae

appeared. For transposition assays involving chromosomal Ty1*his3-AI*, a single colony was dissolved in 10 ml water. One microliter of diluted cells was added to quadruplicate 1 ml SC-Ura or YEPD cultures and grown 2-3 days until saturation. The cells were washed, diluted and spread onto SC-Ura or YEPD and SC-Ura-His or SC-His plates, and incubated at 30° C until colonies formed.

Protein isolation and immunoblotting

For strains carrying pGTy1*his3-AI*, 1 ml of SC-Ura + 2% raffinose was inoculated with a single colony and grown overnight at 30°C, then diluted 1:10 into SC-Ura + 2% galactose and grown at 22° C for 24 h. For growth in glucose, a dilution of 1:100 was used. To detect p22/p18, 5 ml of culture was processed by trichloroacetic acid (TCA) extraction as described previously (26). To detect all other Ty1 proteins, protein from 10 ml of culture was extracted as previously described (50) and 30 μg of protein was used for immunoblotting. Samples were separated on 10% (for RT and IN detection or to separate Gag-p49 and p45) or 15% (Gag-p49/p45 and p22/p18 detection) SDS-PAGE gels and immunoblotted as described previously (26). Antibody dilutions were as follows: anti-p18 1:5000 (26), anti-VLP 1:10,000, anti-RT 1:5,000, anti-IN 1:2500, anti-Hts1 1:40,000, anti-TY (BB2, UAB Epitope Recognition and Immunoreagent Core, Birmingham AL) 1:50,000, anti-V5 1:20,000 (Life Technologies, Carlsbad, CA).

Analysis of Ty1 Gag sequence

Ty1-H3 sequence (GenBank M18706.1) was submitted to the following online servers for secondary structure prediction: ITASSER (51-53) PredictProtein (54), Sable (55), PSIPRED (56), and SAMTO8(57). When comparing the secondary structure predictions, the results were consistent, with the same helices predicted by all five

servers. The boundaries of the helices varied slightly, but not by more than three residues. The I-TASSER results were chosen for display in Figure 4.3. Protein domains in Ty1 sequence were predicted using profile hidden Markov models (58) by scanning Ty1 Gag sequence against the Pfam database. Ty1 related sequences in UniProt were identified using HMMER (58) and aligned using CLUSTALW (59). Full alignment can be found in the supplemental data (Figure S4.1). Protein alignments were visualized using Jalview (http://www.jalview.org/) (60). ClustalX coloration was used with a conservation color increment of 35. The raw alignment file is provided as S1 File.

VLP isolation

VLPs were isolated as described previously (26), except the cells were induced in SC-Ura + 2% galactose for 24 h at 20˚ C. Two micrograms of final VLPs were immunoblotted to detect Gag, RT, and IN. A 1:2 dilution series was loaded to detect p18.

RNA isolation and northern blotting

Equivalent total cellular RNA and VLP RNA, as estimated by OD_{600} or total Gag protein respectively, was extracted using the MasterPure yeast RNA purification kit (Epicentre Biotechnologies, Madison, WI) and analyzed via Northern blotting as previously described (26).

V5 immunoprecipitation

Antibodies were crosslinked to resin using a Pierce Crosslink IP Kit (ThermoFisher Scientific) and following the manufacturer's instructions. For immunoprecipitations, a 25 ml culture was induced in SC-Ura-Trp + 2% galactose at 20˚ C for 24 h or until OD_{600} = 1.0. IP Lysis buffer was supplemented with 1 µg/ml aprotonin,

pepstatin and leupeptin and 1 mM PMSF. Cells were broken in IP Lysis buffer plus protease inhibitors by vortexing with glass beads. Equal amounts of protein were applied to Protein A/G agarose crosslinked with 2 µg of V5 Antibody (Life Technologies) and allowed to bind for 2 h at 4 °C. Beads were washed with IP Lysis buffer and eluted with 20 µl of elution buffer. 1/100 of the input and 1/2 of the pull-down material were loaded per lane. Beads not crosslinked to V5 antibody served as a negative control.

<u>Sucrose gradient sedimentation</u>

A 100 ml culture was induced in SC-Ura or SC-Ura-Trp + 2% galactose at 20° C for 24 h or until the culture reached an OD_{600} of 1. Cells were broken in 15 mM KCl, 10 mM HEPES-KOH, pH 7, 5 mM EDTA containing RNase inhibitor (100 U per ml), and protease inhibitors (16 µg/ml aprotinin, leupeptin, pepstatin A and 2 mM PMSF) in the presence of glass beads. Cell debris was removed by centrifuging the broken cells at 10,000 x g for 10 min at 4 °C. Five milligrams total protein in 300-500 µl of buffer was applied to a 7-47% continuous sucrose gradient and centrifuged using an SW41 rotor at 25,000 rpm (~100,000 x g) for 3 h at 4 °C. After centrifugation, 9 x 1.2 ml fractions were collected and 30 µg of the input and 15 µl of each fraction was immunoblotted to detect Ty1 proteins.

<div align="center">Results</div>

<u>Copy number control in an *S. paradoxus spt3Δ* mutant.</u>

We hypothesized that the generation of CNC[R] Ty1 mutants may identify a molecular target of p22. Since previous work implicated a physical interaction between Gag and p22 (26), isolating resistance mutations in *GAG* would suggest that this

interaction is important for CNC. To identify Ty1 mutants that are resistant to the effects of p22 and its processed form p18, we designed a system that allowed for simultaneous expression of wild type p22/p18 and a randomly mutagenized Ty1 *his3-AI* element fused to the regulated *GAL1* promoter carried on a low copy centromere-based (*CEN*) plasmid (pGTy1*his3-AI/CEN*). The purpose of using a low copy plasmid for pGTy1*his3-AI* expression was to minimize CNC saturation that occurs with overexpression of Ty1 on a multi-copy plasmid. In addition, the Ty1 copy number provided by a low copy centromere-based plasmid does not result in detectable CNC (22). Isogenic, repopulated Ty1-less *S. paradoxus* strains containing 1-38 copies of Ty1-H3 were analyzed, representing a wide range of Ty1 copy numbers naturally found in yeast (S1 Table) (2, 3, 14). All strains carried a deletion of *SPT3*, which encodes a transcription factor required for expression of full length Ty1 mRNA from nucleotide 238 (Ty1-H3, Genbank M18706.1) and the synthesis of Ty1 Gag and Gag-Pol (61). Ty1i RNA, which initiates internally at nucleotide ~1000, is still produced in *spt3Δ* mutants (26, 61, 62). pGTy1*his3-AI/CEN* provided Ty1 mRNA, Gag and Gag-Pol and the strains were analyzed for CNC (Figure 4.1). Ty1 mRNA produced from this plasmid contains the *his3-AI* indicator gene, allowing transposition levels to be analyzed by growth on media lacking histidine. As expected, increasing Ty1 copy number resulted in decreased Ty1 mobility, with the strongest decrease observed in the presence of 38 genomic copies of Ty1 (Figure 4.1A). These strains were immunoblotted for p22/p18 levels in the presence and absence of pGTy1*his3-AI* expression using p18 antiserum, which detects both Gag-p49/p45 and p22/p18 (26). Under both repressing (glucose) and inducing (galactose) growth conditions, p22 levels in cell extracts increased similarly with copy number

(Figure 4.1B). Because p22 was not detected in the lowest Ty1 copy strain (1 Ty1) containing pGTy1*his3-AI*, these results confirmed that pGTy1*his3-AI* does not produce detectable p22. It remains possible that increasing chromosomal copies of Ty1 stimulated p22 production from pGTy1*his3-AI*, but this seems unlikely considering that p22 levels do not increase in the 38 Ty1 copy strain containing pGTy1*his3-AI* compared to an empty vector in either growth condition (Figure 4.1B). Therefore, genomic Ty1-H3 elements, and not the pGTy1*his3-AI* mutant library, were the source of p22 in the screen. When pGTy1*his3-AI* expression was induced by galactose, Gag-p49/p45 was detected and the maturation of p22 to p18 was observed, supporting previous findings suggestive of p22 cleavage by Ty1 PR (26). In addition, mature RT (p63) and IN (p71) were present only in low copy number strains (Figure 4.1C), confirming another feature of the CNC phenotype (22). To further establish that cleavage of p22 was Ty1 PR-dependent, wild type or PR-defective pGTy1/*2μ* multi-copy plasmids were introduced into the 38 Ty1 copy strain (Figure 4.1D). As expected, neither Gag-p49 expressed from the PR⁻ Ty1 nor p22 expressed from genomic elements were cleaved to form mature products Gag-p45 and p18, respectively (Figure 4.1D).

CNC[R] mutations are in *GAG*.

To search for pGTy1*his3-AI* CNC[R] mutations, we utilized the 38 Ty1 copy strain described above, which produced the highest level of p22 of the strains tested (Figure 4.1B). pGTy1*his3-AI* was mutagenized by propagation in a mutator strain of *E. coli* (XL-1 Red, Agilent Technologies, Santa Clara, CA) and 20,000 transformants were screened for an increase in Ty1*HIS3* mobility following induction on medium containing galactose (see Materials and Methods*).* The CNC region (nucleotides 238-1702 (45))

was sequenced from pGTy1*his3-AI* plasmids that conferred an increase in Ty1 mobility when compared to wild type plasmid controls. Nine unique mutations were present in *GAG* (S2 Table; XL-1 Red). A restriction fragment encompassing the CNC region was subcloned from the mutant plasmids into a fresh pGTy1*his3-AI* vector to eliminate contribution of background mutations and no loss of CNC^R was observed. To avoid bias based on our mutagenesis method and to generate additional CNC^R mutations, *GAG* and *POL* were mutagenized separately by error-prone PCR, followed by gap-repair with pGTy1*his3-AI* in the 38 Ty1 copy strain. While *GAG* mutagenesis by PCR revealed 8 new CNC^R mutations from 500 colonies, *POL* mutagenesis revealed only 1 CNC^R candidate from 6,000 colonies (S2 Table). While most of the mutations (8 of 9) isolated via XL-1 Red mutagenesis were single base changes, 5 of 8 mutations recovered via PCR mutagenesis had more than one base change. Interestingly, all *GAG* mutations recovered with either method were missense, suggesting that they function at the level of the Gag protein. The only CNC^R pGTy1*his3-AI* plasmid isolated from *POL* mutagenesis contained two missense mutations within RT (D518G/V519A).

To quantify the level of resistance to p22/p18, the frequency of Ty1*his3-AI* mobility was determined for the mutants alongside wild type controls (Figure 4.2, Table S4.4). In the 38 Ty1 copy strain, most mutant plasmids produced mobility frequencies between 11- and 63- fold higher than wild type (Figure 4.2A). Four candidates from the CNC^R screen (Gag N183D, K186Q, I201T, and A273V) exhibited stronger resistance, ranging from 233- to 424-fold higher than wild type (Figure 4.2A). Because it was possible to obtain Ty1 mutations that globally increased transposition, rather than acting specifically in the presence of p22/p18, CNC^R mobility was also measured in the 1 Ty1

copy strain (Figure 4.2B). Importantly, all CNC^R mutants transposed at similar or decreased levels compared to wild type in the absence of p22/p18, indicating that we obtained CNC-dependent mutations. Three CNC^R mutants, Gag P173L, Gag K250E, and RT D518G/V519A, showed a decrease in Ty1*his3-AI* mobility in the absence of p22/p18, indicating that these mutations negatively impacted Ty1 fitness (Figure 4.2B). Percent recovery of Ty1*his3-AI* mobility with CNC^R plasmids was calculated by dividing Ty1*his3-AI* mobility in the presence or absence of p22 (Percent CNC Recovery, Figure 4.2C). As expected, the three CNC^R mutations with decreased fitness showed higher percent recovery, with K250E and RT D518G/V519A at >100%, due to the fact that these plasmids result in higher transposition frequencies in the presence of p22 than in its absence. Since overall Ty1*his3-AI* mobility was extremely low, further studies were not performed with these mutants. The remaining four mutations conferring >10% CNC^R include those resulting in Gag amino acid changes N183D, K186Q, I201T, and A273V. These elements exhibited 20-30% recovery, indicating that while the strongest CNC^R mutations dramatically increase transposition in the presence of p22 (Figure 4.2A), they are only partially resistant to the effects of p22. Note that both classes of recovery should be expected since the mutant screen balanced transposition fitness and CNC^R. Consequently, we focused on the N183D, K186Q, I201T, and A273V *GAG* mutations, since they conferred the strongest levels of resistance recovered with no effect on fitness of Ty1.

In an effort to increase CNC^R, cells containing double mutant pGTy1*his3-AI-K186Q/I201T* or pGTy1*his3-AI-A273V/I201T* were tested for Ty1*his3-AI* mobility. Gag K186Q/I201T was defective for transposition and was not studied further. Gag

A273V/I201T was able to transpose, but experienced decreased fitness. In the 38 Ty1 copy strain, the levels of Ty1*his3-AI* mobility with A273V/I201T was lower than either single mutant, but still 85-fold higher than wild type (Figure 4.2A). In the 1 Ty1 copy strain, A273V/I201T exhibited Ty1*his3-AI* mobility at 8% of wild type levels (Figure 4.2B). Interestingly, the double mutant did exhibit increased CNC recovery (45%), but at the expense of overall Ty1 mobility (Figure 4.2C). The loss of fitness in the double mutants reinforces the idea that mutations conferring a high level of CNC[R] may have been missed in the screen since they compromise Ty1 fitness. In addition, Ty1 Gag may be genetically fragile since it cannot tolerate wholesale alterations in sequence, a feature that is also observed with HIV-1 CA (63).

CNC[R] mutations N183D, K186Q, I201T, and A273V function in a genomic context.

To determine if the CNC[R] elements are resistant to p22/p18 in a genomic context, wild type or CNC[R] pGTy1*his3-AI*/*CEN* plasmids were expressed in Ty1-less *S. paradoxus*, and cells with 1-2 wild type or CNC[R] Ty1*his3-AI* genomic insertions were identified. Next, an empty vector or a p22-producing plasmid, pTy1-ATGfs (S3 Table), was introduced into these strains. In the absence of p22 (Table 1, empty vector), genomic CNC[R] Ty1 elements N183D, K186Q, and I201T transposed at similar levels to the respective wild type control (<2-fold change), confirming that they do not globally increase Ty1*his3-AI* mobility in a chromosomal context. In contrast, CNC[R] Ty1 A273V displayed a 4.3-fold increase in Ty1 mobility compared to wild type in the presence of empty vector, indicating that A273V may not be CNC-dependent in all contexts. This may be due to the fact that A273V is the only mutation tested that maps within *GAG* and p22 coding sequence, thus changes in both proteins could be affecting Ty1 mobility.

Dramatic differences in CNC were observed when the wild type and CNC^R Ty1*his3-AI* elements are challenged with p22 (Table 1, pTy1-ATGfs). While p22 expression decreases wild type Ty1*his3-AI* mobility 56- to 120- fold, CNC^R Ty1*his3-AI* mobility is partially resistant to the effects of p22, decreasing 2- to 13-fold.

A key feature of CNC is a decrease in Ty1 mobility of a single genomic Ty1*his3-AI* in the presence of elevated Ty1 copies (45). In contrast, additional chromosomal copies of CNC-defective Ty1 elements, which are elements that do not produce p22/p18 but retain the ability to undergo retrotransposition, increase the level of Ty1*his3-AI* mobility (26). To determine how chromosomal CNC^R elements influence Ty1*his3-AI* mobility (Table 2), *S. paradoxus* containing a wild type Ty1*his3-AI* genomic insertion was repopulated with unmarked CNC^R elements carrying the N183D, K186Q, I201T, and A273V mutations. It is important to note that Ty1*his3-AI* RNA is not preferentially packaged *in cis* (23) and can serve as the genomic RNA in mixed particles containing wild type and CNC^R Gag, with the latter likely being produced in excess due to increased genomic copy number. Compared with the starting strain, Ty1*his3-AI* mobility in strains repopulated with +14 and +21 wild type Ty1 elements decreased 31- and 620-fold, respectively. Repopulation with +14-20 CNC^R elements did not alter Ty1*his3-AI* mobility, supporting the idea that CNC^R mutations relieve the inhibitory effects of p22 produced by the additional chromosomal elements. However, the fact that additional CNC^R elements do not stimulate Ty1*his3-AI* mobility probably reflects the partial resistance phenotype imparted by the CNC^R mutations.

<u>CNC^R residues map within predicted helical domains of Gag.</u>

Since little is known about the structure of Ty1 or other LTR-retrotransposon Gag proteins, we submitted Ty1 Gag protein sequence for secondary structure prediction using I-TASSER (51-53), and several other structural prediction servers (see Materials and Methods). These analyses predicted that a central portion of Ty1 Gag contains nine helical regions (Figure 4.3A, gray boxes), which overlap previously identified conserved Gag domains A, B and C of Ty1/*copia* family of retrotransposons (64). Using profile-based methods, we identified two annotated protein families (Pfam) within Ty1 Gag called TYA (TYA transposon protein, PF01021) and UBN2 (gag-polypeptide of LTR copia-type, PF14223) (Figure 4.3A). The TYA domain is found strictly in yeast and corresponds to an unstructured region in the N-terminal half of Gag between residues 17-114. The UBN2 domain maps to the C-terminal half of Gag between residues 245-356, roughly overlapping Ty1/*copia* conserved Gag domains B and C (64), and is represented across multiple plant and fungal species. Of the 11 single *GAG* mutations that impart resistance to p22, 9 mapped within the helical domains, with 4 mapping within the UBN2 domain. Mutations outside of the UBN2 domain clustered between amino acids 170-220, which we refer to here as the CNC^R domain (Figure 4.3A). The CNC^R domain contains sequences belonging to Ty1/*copia* Gag conserved domain A (64), which is characterized by an invariant tryptophan residue corresponding to Ty1 Gag W184. Interestingly, CNC^R alleles lie in close proximity to the W184 codon.

Ty1-H3 Gag (Uniprot P08405) sequence was used in a profile hidden Markov model search to identify closely related Gag proteins and an alignment was generated to highlight variations in amino acid sequence in CNC^R (Figure 4.3B) and UBN2 (Figure

4.3C) domains (refer to Figure S4.1 for a full alignment). Redundant and partial Gag sequences were purged and the Gag sequence of the Ty2-917 element (GenBank KT203716), which was isolated as a spontaneous *HIS4* mutation (65), was added to the hits. In total, we generated a multiple alignment of 15 sequences representing Ty1 and Ty1-like Gag proteins from 9 different yeast strains in the *Saccharomycetaceae* family (66, 67). While substitutions of CNC^R residues do not naturally occur in known Ty1 Gag sequences from *S. cerevisiae*, substitutions in all but one of the CNC^R residues (E287) are found in the alignment of non-Ty1 Gag proteins, including those from Ty2 elements, the second most abundant Ty1/*copia* retrotransposon found in the *S. cerevisiae* S288C genome (2, 3), and Ty1-like elements present in *Saccharomyces kudriavzevii* and *Lachancea kluyveri* (Figures 4.3B-C). Most substitutions are different than those recovered in our CNC^R screen, with the exception of D180N and T218A (Figure 4.3B). Considering all 10 CNC^R residues altered in our screen, 6 of these are not conserved from Ty1 to Ty2.

Ty2 is not influenced by Ty1 or Ty1-like CNC

To determine if Ty2 undergoes CNC, we analyzed the retrotransposition-competent Ty2-917 element (68). Unlike pGTy1-H3, which confers CNC when *GAL1*-promoted Ty1 mRNA transcription is repressed, a multi-copy pGTy2-917 plasmid does not inhibit Ty1*his3-AI* mobility (45). This result suggests that a p22-like protein is either not produced by Ty2-917 or does not affect Ty1 movement. A transcriptionally silent pGTy2-917 also did not affect Ty2-917*his3-AI* mobility, demonstrating that Ty2-917 is not under a Ty1-like form of CNC (S5 Table). In fact, a transcriptionally active Ty2-917 carried on a multi-copy plasmid stimulated Ty2-917*his3-AI* mobility 1.5-fold. Whether all

Ty2 elements respond the same way as Ty2-917 will require further investigation. Considering the relationship between Ty1 CNC^R mutations and Ty2 residues in the CNC^R domain (Figure 4.3B), we introduced an empty vector or the p22 expression plasmid into a strain containing Ty2-917*his3-AI* to determine if Ty2 mobility was sensitive to inhibition by p22 (S5 Table) (26). A decrease in Ty2-917 mobility of less than 2-fold was observed in cells expressing Ty1-p22, supporting the idea that Ty2-917 is not sensitive to Ty1 CNC.

Substitutions at hydrophobic residues in Gag helices disturb Ty1 protein maturation.

Since CNC^R residues map to putative helical domains in Gag, a series of mutations were made in hydrophobic residues within several predicted helices (Figure 4.3A) and their impact on Ty1 transposition and protein levels was analyzed when the mutant elements were expressed from the *GAL1* promoter (Figure 4.4). We substituted the invariant tryptophan residue found in Ty1/copia Gag proteins to alanine (W184A, helix 1) and tested several published Gag mutations designed to interrupt hydrophobic faces of Gag helices (IM248/249NR, L252R, both in helix 4; LF339/340RD, I343K, both in helix 9) (69). All helix substitutions abolished Ty1*his3-AI* mobility when expressed in both 1 and 38 Ty1 copy strains (-p22 and +p22, respectively; Figure 4.4A). Mature RT was not detected in whole cell extracts from the 1 or 38 Ty1 copy strains expressing mutant Ty1, indicating that Gag-Pol maturation did not occur (Figure 4.4B). When helix 1 (W184A) or helix 4 (IM248/249NR and L252R) was perturbed, Gag was stable and present in both immature (p49) and mature forms (p45), while p22 maturation to p18 was similar or slightly decreased. In contrast, Gag-p49 and p22 from helix 9 substitutions (LF339/340RD and I343K) did not undergo maturation.

We analyzed Gag W184A (helix 1), L252R (helix 4) and I343 K (helix 9) for VLP assembly by sedimentation of total protein extracts through 7-47% sucrose gradients (Figure 4.4C). This analysis was performed in the 1 Ty1 copy strain to prevent further disturbance by p22 during VLP formation. Wild type Gag migrated primarily as larger complexes, which are probably comprised of assembled VLPs (Figure 4.4C, fractions 5-9). W184A and L252R VLP assembly was perturbed, as Gag was present in every fraction of the sucrose gradient. There was less cleavage of Gag-p49 to p45 compared to wild type and both forms were present in each fraction. More cleavage of Gag-p49 to p45 was visible in the higher percent sucrose fractions, suggestive of Ty1 PR activity in these fractions. Gag I343K, which did not exhibit Gag cleavage (Figure 4.4B), remained at the top of the sucrose gradient and did not form higher order structures (Figure 4.4C).

CNC[R] mutations alter Ty1 protein maturation.

CNC is associated with altered abundance and maturation of Ty1 proteins, including loss of mature RT and IN (20, 22). Disturbing Gag helices in which CNC[R] mutations mapped also affected Ty1 protein maturation (Figure 4.4). Therefore, Ty1 protein levels produced by CNC[R] pGTy1*his3-AI* elements were examined. Cell extracts were immunoblotted for Gag-p49/p45 and p22 using the 38 Ty1 copy strain (Figure 4.5A). When wild type pGTy1*his3-AI* was expressed, there was slightly more p18 than p22. Strikingly, expressing the four CNC[R] pGTy1*his3-AI* plasmids resulted in a lower level of p18, indicating less cleavage of p22 and/or decreased stability of p18. Mature RT levels were also recovered in the CNC[R] strains, suggesting that VLP maturation improved (Figure 4.5B).

Ty1 VLPs were isolated from the 38 Ty1 copy strain expressing wild type pGTy1*his3-AI* or pGTy1*his3-AI-I201T* to determine levels of Ty1 protein and RNA within assembled particles (Figure 4.6). Equal amounts of VLP preparations were immunoblotted for the detection of Gag, RT and IN. Gag protein levels were similar between wild type and I201T VLPs (Figure 4.6A). Since wild type VLPs represent different stages of maturation, the samples contained Ty1 precursors Gag-PR-IN-RT (Gag-Pol; p199), PR-IN-RT (p154), IN-RT (p134), and PR-IN (p91). In addition, RT antibodies reacted with two bands of unknown origin around 65 and 90 kD, which we reported previously (Figure 4.6B, asterisks) (26). As expected from previous work (22), mature IN (p71) was not detected from wild type VLPs isolated from the 38 Ty1 copy strain (Figure 4.6C). I201T VLP maturation occurred more efficiently, as indicated by increases in mature IN (Figure 4.6C) and a decrease in the unknown RT-reactive proteins in I201T VLPs (Figure 4.6B, asterisks). To determine if the increase in Ty1 mature protein products was due to less p22/p18 present in VLPs, dilutions of wild type and I201T VLPs were immunoblotted with p18 antiserum (Figure 4.6D). p18, rather than p22, was the main protein observed in wild type VLPs, likely due to cleavage by Ty1 PR within VLPs (26). The level of p18 within I201T VLPs was lower than that observed in wild type VLPs, raising the possibility that less p18 within assembled CNC[R] VLPs results in increased Ty1 protein maturation or stability.

Ty1 RNA packaging, as demonstrated by protection from digestion when whole cell extracts are treated with the nuclease benzonase (70), is markedly decreased in the presence of p22/p18 (27). To determine if a CNC[R] mutation functions by increasing the level of RNA packaged into VLPs, RNA extracted from purified WT and I201T VLPs was

subjected to Northern blotting (Figure 4.6E). Total cellular RNA was examined to control for Ty1 mRNA expression. Wild type and I201T RNA extracts from cells or purified VLPs contained similar levels of Ty1*his3-AI* transcript, suggesting that CNC[R] does not function through the enhancement of Ty1 RNA packaging, at least in the case of I201T.

Interestingly, p22/p18 shares sequence with two regions implicated in Ty1 nucleic acid transactions: the NAC region of Gag (amino acid residues 299-401) and the N-terminus of PR (known as p4 in Gag) that participates in reverse transcription (27, 28, 48). The first region was extensively examined using recombinant mature p18, which lacks p4 sequence (27). Recombinant p18 possesses NAC activity and can bind Ty1 RNA, but truncated versions that lack NAC activity still inhibit Ty1 retrotransposition, suggesting that NAC activity is dispensable for p22/p18 function (27). To test whether PR/p4 is implicated in CNC, we measured the mobility of a single genomic Ty1*his3-AI* element in presence of transcriptionally repressed wild type pGPOLΔ Ty1 plasmids (26), or derivatives carrying altered p4 regions. Wild type pGPOLΔ plasmids reduced Ty1*his3-AI* mobility by 150-fold compared to an empty *GAL1* plasmid (S6 Table). The pGPOLΔ*d1* plasmid, which carries a deletion in PR/p4 that blocks successful reverse transcription (48, 71), and pYES2-p45 lacking p4 (27) reduced Ty1*his3-AI* mobility by 140- and 160-fold, respectively. These results are supported by the observations that ectopic expression of mature p18 alone, which does not contain PR/p4 sequence, inhibits pGTy1*his3-AI* mobility, and expression of p22/p18 or p22 mutant for the PR cleavage site exhibit similar levels of inhibition (26, 27). Together, our results show that the PR/p4 region is not required for CNC.

To determine if the low level of p18 detected in I201T VLPs was related to altered binding of p22/p18 with wild type versus Gag I201T, pGp22-V5 and wild type pGTy1*his3-AI* or pGTy1*his3-AI-I201T* were co-expressed in a Ty1-less strain. Endogenous Gag produced by chromosomal Ty1 elements and GST-p18 have been shown to interact via a GST pull-down assay (26). Functional p22-V5, which carries an internal V5 tag that is present in both p22 and p18, was expressed from a low copy *CEN* plasmid to maximize CNC^R imparted by the Gag mutations. Quantitative mobility assays revealed that wild type Ty1*his3-AI* mobility decreased 783-fold in the presence of p22-V5, while Ty1*his3-AI-I201T* mobility only decreased 5-fold, confirming the CNC^R phenotype (Table 3). Utilizing the V5 tag on p22/p18, co-immunoprecipitations (co-IP) were performed from total cell extracts and analyzed for the level of Gag. We detected co-IP of p22-V5/p18-V5 with wild type Gag (Figure S4.2) which confirms previous pull-down analyses with p18 tagged with GST (26). p22-V5/p18-V5 co-immunoprecipitated I201T (Figure S4.2A), K186Q (Figure S4.2B) or wild type Gag with comparable efficiencies.

I201T VLP assembly excludes p22.

To track Gag and p22 and p18 independently during VLP assembly, the fractionation pattern of Gag was examined by sucrose gradient sedimentation as in Figure 4.4C using total protein extracts from cells expressing wild type pGTy1*his3-AI* or pGTy1*his3-AI-I201T* alone or co-expressed with pGp22-V5 (Figure 4.7). In the absence of p22-V5, wild type and I201T Gag-p49/p45 assembled into VLPs and migrated to fractions 6-9 (Figures 4.7A, B). In the presence of p22-V5, the fractionation pattern of wild type Gag was more dispersed, as reported previously (26) (Figure 4.7C). While

Gag was present throughout the gradient, it was found at the highest concentration in fractions 4-9. In contrast, p22-V5 when co-expressed with wild type pGTy1*his3*-AI collects as both p22-V5 and p18-V5 and was present in the highest concentration at the top of the gradient (Figure 4.7D). More p18-V5 co-sedimented with wild type Gag than p22-V5, but p22-V5 and p18-V5 were detected in all fractions. The co-sedimentation of wild type Gag and p18-V5 supports the idea that cleavage by Ty1 PR occurs in complexes migrating to the lower half of the gradient. Surprisingly, p18-V5 was also present at the top of the gradient, which contains most of the soluble proteins in the extract. Considering that the introduction of the internal V5 tag does not alter the requirement of Ty1 PR for cleavage (Figure S4.3) p22-V5 may be cleaved by Ty1 PR outside of fully assembled VLPs, perhaps in the Gag complexes present in retrosomes. Alternatively, p22-V5 may be cleaved within VLPs, but not all p22-V5/p18-V5 remains stably associated with the particles. Regardless, our results suggest that a fraction of p22-V5/p18-V5 co-assembles with wild type VLPs. Expression of pGTy1*his3-AI-I201T* and pGp22-V5 resulted in a Gag fractionation pattern similar to that observed in the absence of p22-V5 (Figure 4.7E). Interestingly, p22-V5/p18-V5 was detected at the top of the gradient but did not co-sediment with I201T VLPs (Figure 4.7F). These results suggest that the CNC^R conferred by I201T results from the exclusion of p22-V5/p18-V5 from VLPs, perhaps during a step in the assembly process. However, the fact some p18-V5 is produced in these cells suggests that the restriction factor does gain access to PR.

<u>Discussion</u>

To understand the mechanism of inhibition of Ty1 retrotransposition by the self-encoded restriction factor p22, we isolated and characterized Ty1 element CNC^R mutants. All but one of the recovered resistance mutations mapped within *GAG* and altered Gag protein sequence. More than half of the mutations mapped outside of p22 coding sequence, including the three strongest CNC^R mutations recovered (N183D, K186Q, and I201T). Importantly, most CNC^R mutations do not reflect simple gain-of-function since the mutations do not increase Ty1 mobility in the absence of p22 (Figure 4.2B). These results, along with the observations that the mutant centromere-based pGTy1*his3-AI* plasmids do not produce detectable p22 levels (Figure 4.1B) or confer CNC (22) due to their low copy number, demonstrate that Gag is the primary molecular target of p22. Although we focused on *GAG* mutations as they represent the vast majority of resistance mutations recovered, one CNC^R mutant contains two sequence changes (D518G/V519A) in the Ribonuclease H (RH) domain of RT (D518G/V519A) within *POL* that dramatically affected Ty1 fitness in the absence of p22 (0.4% recovery of wild type mobility). The RH domain of RT is responsible for degrading the RNA template during reverse transcription (reviewed in (72)), and the decrease in Ty1 mobility is probably due to the fact that D518 is a conserved residue predicted to be involved in metal chelation (73). Mutations resulting in decreased Ty1 RT activity, including active site mutations within the polymerization domain or host mutations that inhibit RT activity by altering cytoplasmic manganese levels can be suppressed by mutations in the RH domain (73, 74). This suppression has been attributed to allosteric communication between the RT polymerization domain and the RH domain (73). Full-

length cDNA is not detectable in cells undergoing CNC (22), because there is a low level of the initial reverse transcription product minus strand strong-stop DNA (75). The failure of RT is likely a downstream effect of the alteration in VLP maturation and the absence of mature IN (22, 26). Although the resistance mutation in RH may bypass the primary defect imposed by p22, it may promote a conformation of RT/RH that allows a low level of activity.

The data presented here greatly extends previous work suggesting that a Gag/p22 interaction is central to the mechanism of CNC (26, 27). Strains undergoing CNC experience a decrease in Ty1 retrotransposition as Ty1 copy number increases (20, 22, 26). A decrease in mobility of a genomic Ty1*his3-AI* element was not observed when additional genomic copies carry CNC[R] mutations, indicating that mutations in Gag, including those that map outside of p22 coding sequence, can relieve CNC in a genomic context. We have not identified resistance mutations with greater than 30% recovery of Ty1 mobility without additionally affecting the fitness of the element (Figure 4.2). Similarly, combining CNC[R] mutations resulted in a loss of Ty1 fitness, rather than a combinatorial increase in resistance. These results illustrate the delicate balance between resistance to p22 and overall fitness of Ty1. A similar tradeoff between resistance and fitness exists for mutations in HIV-1 CA that confer resistance to the restriction factor TRIM5α (40). The inability to achieve complete resistance to TRIM5α is attributed to the genetic fragility of HIV-1 CA, meaning that it is highly sensitive to mutation, and the fact that TRIM5α can bind multiple surfaces on the CA lattice (40, 63). Similarly, some CA mutations conferring resistance to the Mx2 restriction factor also have a negative impact on HIV infectivity (41). Ty1 Gag is a multifunctional protein but

223

unlike retroviral Gag is not cleaved into the functionally distinct proteins such as matrix, CA, and nucleocapsid (NC). Even so, Ty1 Gag is responsible for executing the same functions as retroviral CA and NC. Thus our inability to obtain fully resistant Ty1 elements strongly suggests some of the same rules concerning genetic fragility apply to Ty1 Gag, namely that its function is very sensitive to mutation. Secondly, p22 may bind multiple surfaces of the VLP during different stages of assembly, making it difficult to attain full resistance by mutating Gag at only one or two residues. A third consideration is that the surfaces or protein domains that interact with p22 may be the same or overlap with domains important for Gag function. In support of this idea, we recovered Gag mutations, P173L and K250E, that confer CNC^R, yet negatively impacted Ty1 fitness in the absence of p22 (Figure 4.2). K250 is located within predicted helix 4, an amphipathic helix important for VLP maturation (Figure 4.3A) and perturbation at this site may prevent p22-mediated effects, perhaps by altering VLP assembly and maturation dynamics. However, this alteration in VLP function was not efficient in the absence of p22. Like other infectious agents, the presence of Ty elements in *Saccharomyces* has resulted in positive selection for certain host genes, suggesting there is an ongoing "genetic conflict" or evolutionary arms race between Ty elements and their host (76). In our screen, we recovered mutations that mapped within Gag but not p22 coding sequence and found it difficult to recover mutations that fully restored Ty1 retrotransposition in the presence of p22. Gag and p22 share coding sequence in the natural setting, and this is likely to influence a Ty1-p22 arms race for the adaptation of Ty1 to inhibition via p22.

Our bioinformatic analysis of Ty1 Gag revealed 9 predicted helical stretches and two Pfam domains: TYA in the N-terminal half and UBN2 in the C-terminal half of the protein (Figure 4.3A). Four CNC^R mutations mapped within UBN2; although no CNC^R mutations were isolated in TYA. UBN2 is a Gag sequence motif that is found in Ty1/*copia* retrotransposons across plants and fungi. Because UBN2 can be recognized by profile-based methods across a wide range of organisms, this domain is likely involved in a conserved function of Gag. Because known mutations that affect VLP assembly fall within this domain (69), it is reasonable to hypothesize that the UBN2 domain is involved in VLP assembly. UBN2 also overlaps with, but does not fully encompass, the NAC region of Gag (Figure 4.3A) (28). Recent work demonstrates that p18 interferes with Gag NAC function (27). It would be interesting to investigate whether CNC^R mutations modulate NAC activity of Gag, although we showed that I201T VLPs do not exhibit enhanced levels of Ty1 RNA packaging (Figure 4.6E). Additionally, only V336I and Q350R S395L mapped within the NAC region (Table S4.2, Figure 4.3A), suggesting that CNC^R does not primarily alter Gag NAC functions. The remaining CNC^R mutations mapped to a central region of Gag, which we named the CNC^R domain. Predicted helix 1 within the CNC^R domain was frequently mutated in our screen and overlaps with conserved domain A present in all Ty1/*copia* elements that surrounds an invariant tryptophan (Gag W184 for Ty1) (64). Ty2 Gag differs from Ty1 Gag at many positions and some CNC^R mutations are present within Ty2 sequences, raising the possibility that Ty2 is naturally resistant to Ty1 CNC. Ty1 and Ty2 are closely related retrotransposons based on their near identical LTR sequences, with a single nucleotide deletion defining Ty2 LTR sequences (2, 77). Phylogenetic analyses suggest that *S.*

225

cerevisiae recently acquired Ty2 elements from *S. mikatae* by horizontal transfer (3, 78). We showed that Ty2-917 is neither under CNC by Ty2-917 nor inhibited by *GAL1*-promoted expression of p22 from Ty1. CNC[R] residues are also altered in a Tsk1 element from *L. kluyveri* (79) and a Ty1-like element present in S. *kudriavzevii* and hybrids thereof (3, 80). Further mutational analyses of specific CNC[R] residues within these elements will be required to address the role of naturally occurring CNC[R] residues in Ty1 and Ty2. In addition, the overlap between Gag and p22 coding sequence may create specificity even within Ty1 elements. It will be interesting to determine if Ty1 elements in natural *Saccharomyces* isolates confer CNC on Ty1-H3 or whether they have evolved specificity for elements in their native genomes.

Though the structural role of CNC[R] residues within Ty1 Gag's helical domains remains to be determined, the mutational analyses presented here define the importance of key hydrophobic residues within these predicted regions. Mutation of the conserved W184 residue within helix 1 to alanine resulted in complete loss of transposition and formation of mature Pol proteins, as well as abnormal VLP assembly (Figure 4.4). Other Gag helix mutations we tested were previously analyzed in the context of a truncated Gag protein containing amino acids 1-381, which is deleted for 21 residues at the C-terminus of p45 and lacks a complete NAC domain (28, 69). The truncated Gag is still able to form particles and thus was used to address the role of certain residues in particle assembly. In the context of 1-381 particles, L252R was reported to completely disrupt particle assembly, while LF339-340RD has no effect. IM248/249NR and I343K both alter the migration of 1-381 particles through a sucrose gradient and assemble into "giant" particles when visualized by negative staining and

TEM. These mutations had not been characterized in the context of a full length Ty1 element nor analyzed for effects on Ty1 mobility. Therefore, our work showed that altering these hydrophobic residues within helices severely hinders Ty1 mobility (Figure 4.4). Our results regarding particle assembly with these Gag substitutions differ from previous work, as L252R was capable of forming higher order complexes and I343K was not (Figure 4.4C). These conflicting results could be due to the differences in multimerization and VLP assembly using full length Gag compared to truncated 1-381 Gag used by others (69). Importantly, since most CNC^R mutations mapped to helices required for normal Ty1 transposition, protein processing and VLP assembly, our results indicate that inhibition by p22 disturbs a central function of Gag.

Mutations exhibiting the highest level of CNC^R (N183D, K186Q, I201T, and A273V) were associated with reduced levels of p18 (Figure 4.5), which is cleaved from p22 near its C-terminus by Ty1-PR, and there was less co-assembly between p22 and Gag I201T compared to wild type Gag (Figure 4.7). These results suggest that the reduction in p18 levels in the presence of the CNC^R VLPs may result from diminished access to PR. More p22/p18 associated with purified wild type VLPs when compared with I201T VLPs (Figure 4.6D). When total cellular protein from cells co-expressing p22-V5 and wild type or Ty1*his3-AI-I201T* was analyzed by sucrose gradient sedimentation (Figure 4.7), the majority of p22-V5 and p18-V5 remained in fractions containing soluble protein rather than in fractions containing VLPs or higher order assembly intermediates. Because Ty1 PR-mediated processing is thought to occur only within assembled VLPs and cleavage of p22 and p22-V5 was Ty1 PR-specific (Figures 4.1D and S4.3), p22/p18 may be capable of moving in and out of the VLP. Perhaps this occurs by diffusion of

p22/p18 through VLP pores, which are permeable to ribonuclease A (15.7 kD) but not to benzonase (30 kD) treatment *in vitro* (81-83). Consequently, p22/p18 may still be within the acceptable size limit to enter and exit VLP pores. Alternatively, once p22/p18 co-assembles with Ty1 proteins in VLPs and maturation is initiated, p22/p18-containing VLPs may be subject to dissociation and degradation. Although we did observe a modest shift in Ty1 Gag fractionation towards the top of the gradient in the presence of p22/p18, Gag was not concentrated in the first two fractions with p22/p18. Lastly, our results raise the possibility that Ty1 PR may function outside of stably assembled VLPs, perhaps in assembly intermediates present in retrosomes, which are cytoplasmic foci containing Ty1 mRNA and proteins (7, 9, 84). In support of this idea, few if any VLPs are detected in cells containing retrosomes resulting from endogenous Ty1 expression, VLP assembly increases dramatically when Ty1 is overexpressed from a strong promoter, and assembly occurs within retrosomes (7, 25, 85). Recent work also shows that steady state Gag expressed from endogenous Ty1 elements does not co-migrate with unprocessed Gag-p49 (84), suggesting that Gag cleavage can occur in the absence of detectable VLPs. Finally, several earlier studies demonstrate the presence of mature p45 resulting from endogenous Ty1 expression (8, 24, 86-88). Together, our results suggest that p22 cleavage may occur in the same spatiotemporal environment as pre-VLP Gag cleavage.

We observed varying degrees of p22 cleavage and/or p18 stability in the presence of altered Ty1 Gag proteins. While p22/p18 levels were comparable to WT in Gag L252R, p18 was not detected in Gag LF339-340RD and Gag I343K. CNC[R] mutations (N183D, K186Q, I201T, and A273V) and the helix-altering Gag W184A and

IM248-249NR were associated with decreased levels of p18, but we cannot distinguish if these changes represent a decrease in p22 cleavage or a reduction in p18 stability. It is interesting to consider that some loss-of-function changes in Gag (W184A and IM248-249NR) and the gain-of-function CNC^R mutations both result in decreased p18 levels. Perhaps p22 cleavage is a read-out for both Ty1 PR activity, which can be affected by several different situations, such as Gag:Gag-Pol ratio and particle assembly (89-92), or access of the p22 substrate to Ty1 PR. Whereas the helix mutations alter VLP assembly, the resistance mutations likely affect access to PR, since p22 is excluded from CNC^R VLPs. Thus, reduced p22 cleavage can occur in both loss-of-function and gain-of-function contexts.

Although the CNC^R mutations in Gag might affect Gag/p22 binding, co-IP experiments performed using standard washing conditions did not support a simple interaction between p22 and Gag involving CNC^R residues (Figure S4.2). In addition, sucrose gradient fractionation indicated that most p22-V5/p18-V5 was present in the fractions containing soluble proteins and did not co-sediment with VLPs. Perhaps p22 is capable of binding several forms of Gag, whether monomeric, small assembly intermediates, or intact VLPs, and perhaps these interactions inhibit VLP assembly or maturation with different potencies. If the crucial binding substrate of p22 is multimeric and represents a minority of Gag molecules present in the cell, co-IP analysis may not show differences in binding. Interestingly, retroviral CA-binding restriction factors TRIM5α and the Gag-derived Fv1 bind to their CA target after polymerization of the lattice (93, 94). We are considering that the interaction between p22 with

polymerized/assembled Gag alone may be the defining and initial insult to Ty1 replication.

Retroviral studies involving sensitivity and escape from host restriction factors show similarities to the Ty1-p22 system. Mx2 restriction of HIV-1 is thought to involve inhibition of viral uncoating and/or nuclear entry and requires Mx2-CA binding (41, 95). However, known Mx2 escape mutations in the CA gene do not significantly alter binding between Mx2 and CA (41), which demonstrates that viral escape mutations can promote replication in ways distinct from the disruption of restriction factor-target binding. In the case of the resistant provirus enJSRV26, increasing the levels of enJSRV26 Gag expression relative to the restriction factor enJS56A1 Gag protein is enough to allow JSRV replication in sheep (43). Increased expression of enJSRV-26 Gag is achieved by mutation of the signal peptide in the envelope glycoprotein, which modulates proviral gene expression. Similarly, increasing the level of Ty1 expression can overcome CNC (45). We favor the hypothesis that understanding how p22 is excluded from CNC^R VLPs is central to understanding CNC. Since the steady state level of Gag was unaffected in CNC^R mutants (Figure 4.5A), perhaps the ratio of Gag:p22 is specifically higher within retrosomes comprised of CNC^R Gag.

In summary, we have shown that mutations in Gag confer resistance to the p22 restriction factor produced by Ty1 during CNC. These mutations are beneficial only in the presence of p22 and do not globally increase Ty1 mobility. CNC^R mutations allow for VLP maturation, which may be the step in Ty1 replication most sensitive to CNC, by excluding p22 from assembling particles. Identification of the Gag multimerization states that bind p22 and host factors that modulate Gag assembly, in combination with

studies examining VLP assembly dynamics and structure, especially regarding the newly identified Gag domains, will deepen our understanding of retroelement control.

Acknowledgments

References

1. Curcio MJ, Lutz S, Lesage P. The Ty1 LTR-retrotransposon of budding yeast. Microbiology spectrum. 2015;3(2):1-35.

2. Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. Genome research. 1998;8(5):464-78.

3. Carr M, Bensasson D, Bergman CM. Evolutionary genomics of transposable elements in Saccharomyces cerevisiae. PloS one. 2012;7(11):e50978.

4. Bleykasten-Grosshans C, Friedrich A, Schacherer J. Genome-wide analysis of intraspecific transposon diversity in yeast. BMC genomics. 2013;14:399.

5.   Voytas DF, Craig NL. Ty1 and Ty5 of Saccharomyces cerevisiae. Mobile DNA II2002. p. 614-30.

6.   Beliakova-Bethell N, Beckham C, Giddings T, Winey M, Parker R, Sandmeyer S. Virus-like particles of the Ty3 retrotransposon assemble in association with P-body components. RNA. 2006;12(1):94-101.

7.   Checkley MA, Nagashima K, Lockett SJ, Nyswaner KM, Garfinkel DJ. P-body components are required for Ty1 retrotransposition during assembly of retrotransposition-competent virus-like particles. Molecular and cellular biology. 2010;30(2):382-98.

8.   Dutko JA, Kenny AE, Gamache ER, Curcio MJ. 5' to 3' mRNA decay factors colocalize with Ty1 gag and human APOBEC3G and promote Ty1 retrotransposition. Journal of Virology. 2010;84(10):5052-66.

9.   Malagon F, Jensen TH. The T body, a new cytoplasmic RNA granule in Saccharomyces cerevisiae. Molecular and cellular biology. 2008;28(19):6022-32.

10.   Dunham M, Badrane H, Ferea T, Adams J, Brown P, Rosenzweig F, et al. Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. PNAS. 2002;99(25):16144-9.

11.   Garfinkel D. Genome evolution mediated by Ty elements in Saccharomyces. Cytogenet Genome Res. 2005;110(1-4):63-9.

12.   Wilke C, Adams J. Fitness effects of Ty transposition in Saccharomyces cerevisiae. Genetics. 1992;131(1):31-42.

13.   Wilke C, Maimer E, Adams J. The population biology and evolutionary significance of Ty elements in Saccharomyces cerevisiae. Genetica. 1992;86(1-3):155-73.

14.   Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, et al. Population genomics of domestic and wild yeasts. Nature. 2009;458(7236):337-41.

15.   Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, et al. RNAi in budding yeast. Science. 2009;326(5952):544-50.

16.   Drinnenberg IA, Fink GR, Bartel DP. Compatibility with killer explains the rise of RNAi-deficient fungi. Science. 2011;333(6049):1592.

17.   Harris RS, Hultquist JF, Evans DT. The restriction factors of human immunodeficiency virus. The Journal of biological chemistry. 2012;287(49):40875-83.

18.   Malim MH, Bieniasz PD. HIV Restriction Factors and Mechanisms of Evasion. Cold Spring Harbor perspectives in medicine. 2012;2(5):a006940.

19.   Zheng YH, Jeang KT, Tokunaga K. Host restriction factors in retroviral infection: promises in virus-host interaction. Retrovirology. 2012;9:112.

20.   Garfinkel DJ, Nyswaner K, Wang J, Cho JY. Post-transcriptional cosuppression of Ty1 retrotransposition. Genetics. 2003;165(1):83-99.

21.   Curcio M, Garfinkel D. Single-step selection for Ty1 element retrotransposition. Proc Natl Acad Sci U S A. 1991;88(3):936-40.

22.   Matsuda E, Garfinkel DJ. Posttranslational interference of Ty1 retrotransposition by antisense RNAs. Proc Natl Acad Sci U S A. 2009;106(37):15657-62.

23.   Curcio MJ, Garfinkel DJ. Heterogeneous functional Ty1 elements are abundant in the Saccharomyces cerevisiae genome. Genetics. 1994;136(4):1245-59.

24.   Curcio MJ, Garfinkel DJ. Posttranslational control of Ty1 retrotransposition occurs at the level of protein processing. Molecular and cellular biology. 1992;12(6):2813-25.

25.   Garfinkel D, Boeke J, Fink G. Ty element transposition: Reverse transcriptase and virus-like particles. Cell. 1985;42:502-17.

26.   Saha A, Mitchell JA, Nishida Y, Hildreth JE, Ariberre JA, Gilbert WV, et al. A trans-Dominant Form of Gag Restricts Ty1 Retrotransposition and Mediates Copy Number Control. J Virol. 2015;89(7):3922-38.

27.   Nishida Y, Pachulska-Wieczorek, K., Błaszczyk, L., Saha, A., Gumna, J., Garfinkel, D.J., Purzycka, K.J. Ty1 retrovirus-like element Gag contains overlapping restriction factor and nucleic acid chaperone functions. Nucleic Acids Res. 2015:In press.

28.   Cristofari G, Ficheux D, Darlix JL. The GAG-like protein of the yeast Ty1 retrotransposon contains a nucleic acid chaperone domain analogous to retroviral nucleocapsid proteins. The Journal of biological chemistry. 2000;275(25):19210-7.

29.   Kozak CA, Chakraborti A. Single amino acid changes in the murine leukemia virus capsid protein gene define the target of Fv1 resistance. Virology. 1996;225(2):300-5.

30.   Owens CM, Yang PC, Gottlinger H, Sodroski J. Human and simian immunodeficiency virus capsid proteins are major viral determinants of early, postentry replication blocks in simian cells. J Virol. 2003;77(1):726-31.

31.   Liu Z, Pan Q, Ding S, Qian J, Xu F, Zhou J, et al. The interferon-inducible MxB protein inhibits HIV-1 infection. Cell host & microbe. 2013;14(4):398-410.

32.   Goujon C, Moncorge O, Bauby H, Doyle T, Ward CC, Schaller T, et al. Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. Nature. 2013;502(7472):559-62.

33.   Kane M, Yadav SS, Bitzegeio J, Kutluay SB, Zang T, Wilson SJ, et al. MX2 is an interferon-induced inhibitor of HIV-1 infection. Nature. 2013;502(7472):563-6.

34.   Stevens A, Bock M, Ellis S, LeTissier P, Bishop KN, Yap MW, et al. Retroviral capsid determinants of Fv1 NB and NR tropism. J Virol. 2004;78(18):9592-8.

35.   Yap MW, Nisole S, Lynch C, Stoye JP. Trim5alpha protein restricts both HIV-1 and murine leukemia virus. Proc Natl Acad Sci U S A. 2004;101(29):10786-91.

36.   Jung YT, Kozak CA. A single amino acid change in the murine leukemia virus capsid gene responsible for the Fv1(nr) phenotype. J Virol. 2000;74(11):5385-7.

37.   Ohkura S, Goldstone DC, Yap MW, Holden-Dye K, Taylor IA, Stoye JP. Novel escape mutants suggest an extensive TRIM5alpha binding site spanning the entire outer surface of the murine leukemia virus capsid protein. PLoS pathogens. 2011;7(3):e1002011.

38.   Ohkura S, Stoye JP. A comparison of murine leukemia viruses that escape from human and rhesus macaque TRIM5alphas. J Virol. 2013;87(11):6455-68.

39.   Mortuza GB, Dodding MP, Goldstone DC, Haire LF, Stoye JP, Taylor IA. Structure of B-MLV capsid amino-terminal domain reveals key features of viral tropism, gag assembly and core formation. Journal of molecular biology. 2008;376(5):1493-508.

40.   Soll SJ, Wilson SJ, Kutluay SB, Hatziioannou T, Bieniasz PD. Assisted evolution enables HIV-1 to overcome a high TRIM5alpha-imposed genetic barrier to rhesus macaque tropism. PLoS pathogens. 2013;9(9):e1003667.

41. Fricke T, White TE, Schulte B, de Souza Aranha Vieira DA, Dharan A, Campbell EM, et al. MxB binds to the HIV-1 core and prevents the uncoating process of HIV-1. Retrovirology. 2014;11:68.

42. Mura M, Murcia P, Caporale M, Spencer TE, Nagashima K, Rein A, et al. Late viral interference induced by transdominant Gag of an endogenous retrovirus. Proc Natl Acad Sci U S A. 2004;101(30):11117-22.

43. Armezzani A, Arnaud F, Caporale M, di Meo G, Iannuzzi L, Murgia C, et al. The signal peptide of a recently integrated endogenous sheep betaretrovirus envelope plays a major role in eluding gag-mediated late restriction. J Virol. 2011;85(14):7118-28.

44. Benit L, De Parseval N, Casella JF, Callebaut I, Cordonnier A, Heidmann T. Cloning of a new murine endogenous retrovirus, MuERV-L, with strong similarity to the human HERV-L element and with a gag coding sequence closely related to the Fv1 restriction gene. J Virol. 1997;71(7):5652-7.

45. Garfinkel DJ, Nyswaner K, Wang J, Cho J-Y. Post-transcriptional cosuppression of Ty1 retrotransposition. Genetics. 2003;165(1):83-99.

46. Guthrie C, Fink GR, editors. Guide to Yeast Genetics and Molecular Biology. San Diego, CA: Academic Press Inc.; 1991.

47. Merkulov GV, Swiderek KM, Brachmann CB, Boeke JD. A critical proteolytic cleavage site near the C terminus of the yeast retrotransposon Ty1 Gag protein. J Virol. 1996;70(8):5548-56.

48. Lawler JF, Jr., Merkulov GV, Boeke JD. A nucleocapsid functionality contained within the amino terminus of the Ty1 protease that is distinct and separable from proteolytic activity. J Virol. 2002;76(1):346-54.

49.   Curcio MJ, Kenny AE, Moore S, Garfinkel DJ, Weintraub M, Gamache ER, et al. S-phase checkpoint pathways stimulate the mobility of the retrovirus-like transposon Ty1. Molecular and cellular biology. 2007;27(24):8874-85.

50.   Braiterman LT, Monokian GM, Eichinger DJ, Merbs SL, Gabriel A, Boeke JD. In-frame linker insertion mutagenesis of yeast transposon Ty1: phenotypic analysis. Gene. 1994;139(1):19-26.

51.   Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nature methods. 2015;12(1):7-8.

52.   Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature protocols. 2010;5(4):725-38.

53.   Zhang Y. I-TASSER server for protein 3D structure prediction. BMC bioinformatics. 2008;9:40.

54.   Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, et al. PredictProtein--an open resource for online prediction of protein structural and functional features. Nucleic Acids Res. 2014;42(Web Server issue):W337-43.

55.   Adamczak R, Porollo A, Meller J. Combining prediction of secondary structure and solvent accessibility in proteins. Proteins. 2005;59(3):467-75.

56.   Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. Journal of molecular biology. 1999;292(2):195-202.

57.   Karplus K. SAM-T08, HMM-based protein structure prediction. Nucleic Acids Res. 2009;37(Web Server issue):W492-7.

58.   Eddy SR. Accelerated Profile HMM Searches. PLoS computational biology. 2011;7(10):e1002195.

59.   Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22(22):4673-80.

60.   Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189-91.

61.   Winston F, Durbin KJ, Fink GR. The SPT3 gene is required for normal transcription of Ty elements in S. cerevisiae. Cell. 1984;39(3 Pt 2):675-82.

62.   Winston F, Dollard C, Malone EA, Clare J, Kapakos JG, Farabaugh P, et al. Three genes are required for trans-activation of Ty transcription in yeast. Genetics. 1987;115(4):649-56.

63.   Rihn SJ, Wilson SJ, Loman NJ, Alim M, Bakker SE, Bhella D, et al. Extreme genetic fragility of the HIV-1 capsid. PLoS pathogens. 2013;9(6):e1003461.

64.   Peterson-Burch BD, Voytas DF. Genes of the Pseudoviridae (Ty1/copia retrotransposons). Molecular biology and evolution. 2002;19(11):1832-45.

65.   Roeder GS, Farabaugh PJ, Chaleff DT, Fink GR. The origins of gene instability in yeast. Science. 1980;209(4463):1375-80.

66.   Kurtzman CP. Phylogenetic circumscription of Saccharomyces, Kluyveromyces and other members of the Saccharomycetaceae, and the proposal of the new genera Lachancea, Nakaseomyces, Naumovia, Vanderwaltozyma and Zygotorulaspora. FEMS yeast research. 2003;4(3):233-45.

67.   Kurtzman CP, Robnett CJ. Phylogenetic relationships among yeasts of the 'Saccharomyces complex' determined from multigene sequence analyses. FEMS yeast research. 2003;3(4):417-32.

68.   Curcio MJ, Sanders NJ, Garfinkel DJ. Transpositional competence and transcription of endogenous Ty elements in Saccharomyces cerevisiae: implications for regulation of transposition. Molecular and cellular biology. 1988;8(9):3571-81.

69.   Martin-Rendon E, Marfany G, Wilson S, Ferguson DJ, Kingsman SM, Kingsman AJ. Structural determinants within the subunit protein of Ty1 virus-like particles. Molecular microbiology. 1996;22(4):667-79.

70.   Teysset L, Dang VD, Kim MK, Levin HL. A long terminal repeat-containing retrotransposon of Schizosaccharomyces pombe expresses a Gag-like protein that assembles into virus-like particles which mediate reverse transcription. J Virol. 2003;77(9):5451-63.

71.   Lawler JF, Jr., Merkulov GV, Boeke JD. Frameshift signal transplantation and the unambiguous analysis of mutations in the yeast retrotransposon Ty1 Gag-Pol overlap region. J Virol. 2001;75(15):6769-75.

72.   Wilhelm FX, Wilhelm M, Gabriel A. Reverse transcriptase and integrase of the Saccharomyces cerevisiae Ty1 element. Cytogenet Genome Res. 2005;110(1-4):269-87.

73.   Yarrington RM, Chen J, Bolton EC, Boeke JD. Mn2+ suppressor mutations and biochemical communication between Ty1 reverse transcriptase and RNase H domains. J Virol. 2007;81(17):9004-12.

74.   Uzun O, Gabriel A. A Ty1 reverse transcriptase active-site aspartate mutation blocks transposition but not polymerization. J Virol. 2001;75(14):6337-47.

75.   Purzycka KJ, Legiewicz M, Matsuda E, Eizentstat LD, Lusvarghi S, Saha A, et al. Exploring Ty1 retrotransposon RNA structure within virus-like particles. Nucleic Acids Res. 2012.

76.   Sawyer SL, Malik HS. Positive selection of yeast nonhomologous end-joining genes and a retrotransposon conflict hypothesis. Proc Natl Acad Sci U S A. 2006;103(47):17614-9.

77.   Jordan IK, McDonald JF. Evidence for the role of recombination in the regulatory evolution of Saccharomyces cerevisiae Ty elements. Journal of molecular evolution. 1998;47(1):14-20.

78.   Liti G, Peruffo A, James SA, Roberts IN, Louis EJ. Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the Saccharomyces sensu stricto complex. Yeast. 2005;22(3):177-92.

79.   Neuveglise C, Feldmann H, Bon E, Gaillardin C, Casaregola S. Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts. Genome research. 2002;12(6):930-43.

80.   Dunn B, Richter C, Kvitek DJ, Pugh T, Sherlock G. Analysis of the Saccharomyces cerevisiae pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. Genome research. 2012;22(5):908-24.

81.	Brookman JL, Stott AJ, Cheeseman PJ, Burns NR, Adams SE, Kingsman AJ, et al. An immunological analysis of Ty1 virus-like particle structure. Virology. 1995;207(1):59-67.

82.	Burns NR, Saibil HR, White NS, Pardon JF, Timmins PA, Richardson SM, et al. Symmetry, flexibility and permeability in the structure of yeast retrotransposon virus-like particles. The EMBO journal. 1992;11(3):1155-64.

83.	HA AL-K, Bhella D, Kenney JM, Roth JF, Kingsman AJ, Martin-Rendon E, et al. Yeast Ty retrotransposons assemble into virus-like particles whose T-numbers depend on the C-terminal length of the capsid protein. Journal of molecular biology. 1999;292(1):65-73.

84.	Doh JH, Lutz S, Curcio MJ. Co-translational localization of an LTR-retrotransposon RNA to the endoplasmic reticulum nucleates virus-like particle assembly sites. PLoS Genet. 2014;10(3):e1004219.

85.	Mellor J, Fulton S, Dobson M, Wilson W, Kingsman S, Kingsman A. A retrovirus-like strategy for expression of a fusion protein encoded by yeast transposon Ty1. Nature. 1985;313(5999):243-6.

86.	Dakshinamurthy A, Nyswaner KM, Farabaugh PJ, Garfinkel DJ. BUD22 affects Ty1 retrotransposition and ribosome biogenesis in Saccharomyces cerevisiae. Genetics. 2010;185(4):1193-205.

87.	Conte D, Jr., Barber E, Banerjee M, Garfinkel DJ, Curcio MJ. Posttranslational regulation of Ty1 retrotransposition by mitogen-activated protein kinase Fus3. Molecular and cellular biology. 1998;18(5):2502-13.

88.   Malagon F, Jensen TH. T-body formation precedes virus-like particle maturation in S. cerevisiae. RNA Biol. 2011;8(2):184-9.

89.   Xu H, Boeke JD. Host genes that influence transposition in yeast: the abundance of a rare tRNA regulates Ty1 transposition frequency. Proc Natl Acad Sci U S A. 1990;87(21):8360-4.

90.   Kawakami K, Pande S, Faiola B, Moore DP, Boeke JD, Farabaugh PJ, et al. A rare tRNA-Arg(CCU) that regulates Ty1 element ribosomal frameshifting is essential for Ty1 retrotransposition in Saccharomyces cerevisiae. Genetics. 1993;135(2):309-20.

91.   Youngren SD, Boeke JD, Sanders NJ, Garfinkel DJ. Functional organization of the retrotransposon Ty from Saccharomyces cerevisiae: Ty protease is required for transposition. Molecular and cellular biology. 1988;8(4):1421-31.

92.   Garfinkel DJ, Hedge AM, Youngren SD, Copeland TD. Proteolytic processing of pol-TYB proteins from the yeast retrotransposon Ty1. J Virol. 1991;65(9):4573-81.

93.   Hilditch L, Matadeen R, Goldstone DC, Rosenthal PB, Taylor IA, Stoye JP. Ordered assembly of murine leukemia virus capsid protein on lipid nanotubes directs specific binding by the restriction factor, Fv1. Proc Natl Acad Sci U S A. 2011;108(14):5771-6.

94.   Stremlau M, Perron M, Lee M, Li Y, Song B, Javanbakht H, et al. Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. Proc Natl Acad Sci U S A. 2006;103(14):5514-9.

95.   Schulte B, Buffone C, Opp S, Di Nunzio F, Augusto De Souza Aranha Vieira D, Brandariz-Nunez A, et al. Restriction of HIV-1 Requires the N-terminal Region of

MxB/Mx2 as a Capsid-Binding Motif but not as a Nuclear Localization Signal. J Virol. 2015.

96.   Boeke JD, Eichinger D, Castrillon D, Fink GR. The Saccharomyces cerevisiae genome contains functional and nonfunctional copies of transposon Ty1. Molecular and cellular biology. 1988;8(4):1432-42.

97.   Sikorski R, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics. 1989;122:19-27.

98.   Garfinkel D, Mastrangelo M, Sanders N, Shafer B, Strathern J. Transposon tagging using Ty elements in yeast. Genetics. 1988;120(1):95-108.

99.   Christianson TW, Sikorski RS, Dante M, Shero JH, Hieter P. Multifunctional yeast high-copy-number shuttle vectors. Gene. 1992;110(1):119-22.

100.   Garfinkel DJ, Mastrangelo MF, Sanders NJ, Shafer BK, Strathern JN. Transposon tagging using Ty elements in yeast. Genetics. 1988;120(1):95-108.
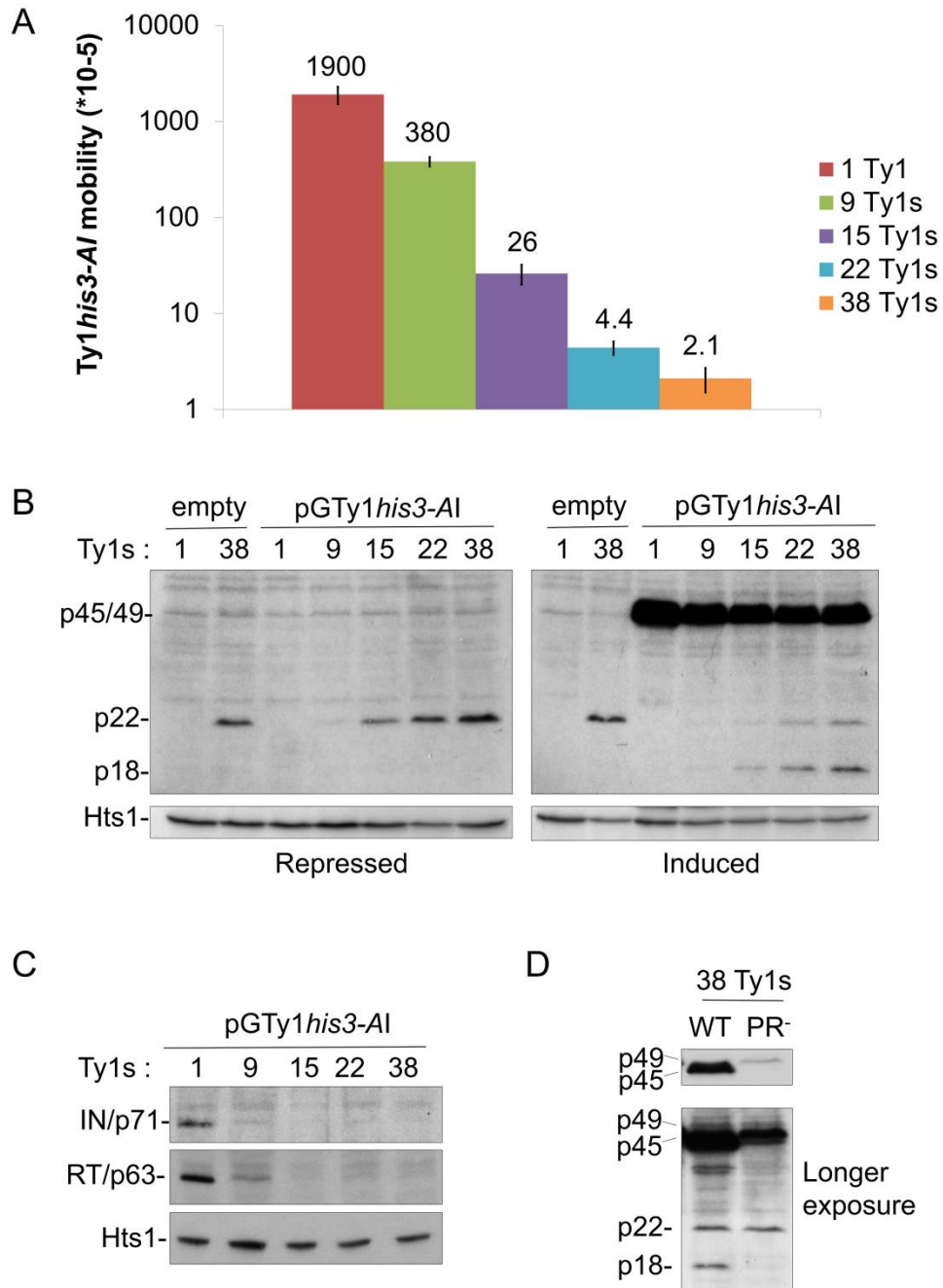
Figure 4.1   Ty1 mobility and protein levels in cells repopulated with Ty1 elements. A

Ty1-less *S. paradoxus* strain repopulated with increasing copies of genomic Ty1-H3

elements [YEM514 (1 Ty1), YEM568 (9 Ty1s), YEM570 (15 Ty1s), YEM572 (22 Ty1s),

and YEM515 (38 Ty1s)] and containing pGTy1*his3-AI*/*URA3*/*CEN* (pBDG606) were analyzed for Ty1*his3-AI* mobility and protein levels. All strains were deleted for *SPT3*, which blocks Ty1 mRNA, but not Ty1i RNA, expression from genomic loci. (A) A quantitative Ty1*his3-AI* mobility assay was performed with galactose-induced cells. Numbers represent Ty1*his3-AI* mobility events per cell and bars represent standard deviations. Mobility assays were repeated at least three times and representative results are shown. (B) Cells were grown for 24 h under repressing (glucose) or inducing (galactose) conditions for pGTy1*his3-AI* expression. 1 and 38 Ty1 copy strains containing an empty vector were used as controls. TCA-precipitated protein extracts were immunoblotted with p18 antibody, which recognizes full length immature and mature Gag (p49/p45) and p22/p18, and Hts1 antibody, which served as a loading control. Separation of Gag-p49 and p45 does not occur under these electrophoresis conditions (see Materials and Methods) and thus are bracketed and collectively labeled as Gag-p49/p45. (C) Cells were induced for pGTy1 expression for 24 h and protein extracts were immunoblotted with IN, RT, or Hts1 antibody. (D) The 38 Ty1 copy strain containing WT or PR-defective (SacI linker-1702 (91)) pGTy1/2µ plasmids were induced with galactose for 24 h. TCA-precipitated extracts were immunoblotted with p18 antibody.
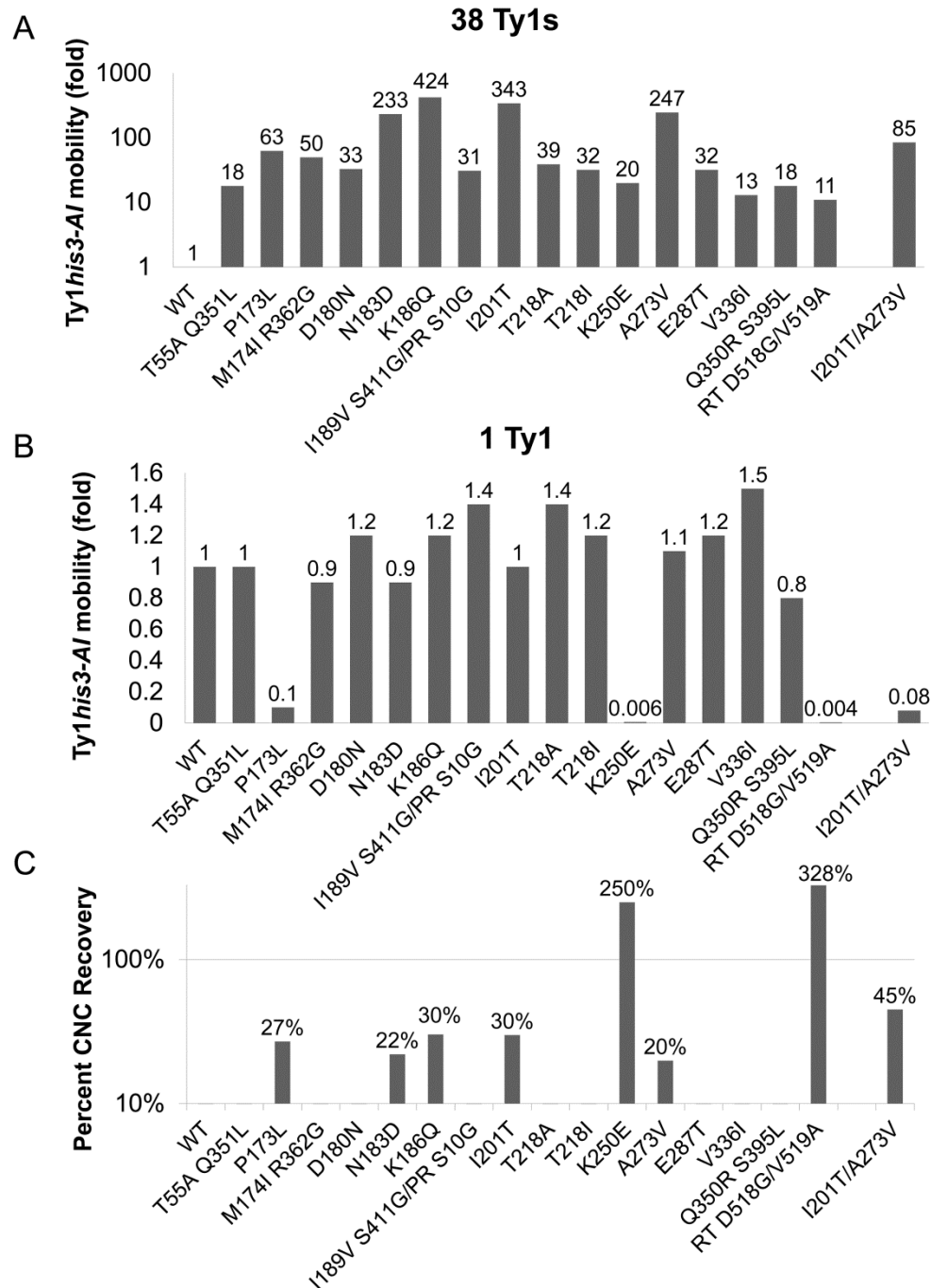
Figure 4.2 Ty1 *GAG* mutations are specific and partially resistant to CNC. *S. paradoxus* repopulated with 38 (YEM515; A) or 1 (YEM514; B) genomic copies of Ty1-H3 and containing wild type or CNC^R pGTy1*his3-AI* plasmids were analyzed for Ty1*his3-AI* mobility. Ty1*his3-AI* mobility for each mutant was normalized to wild type and fold

changes are represented. Raw data can be found in S4 Table. (C) Percent CNC recovery was calculated by dividing Ty1*his3-AI* mobility in the 38 Ty1 copy strain by mobility in the 1 copy strain. Recoveries above 10% are shown.

Figure 4.3 Predicted helical structure of Ty1 Gag and location of CNC^R substitutions. (A) Ty1 Gag sequence (UniProt P08405) was analyzed for secondary structure prediction and domain identification. Predicted helical regions are represented by gray boxes below and labeled 1-9. Two Pfam domains, TYA (PF01021, residues 17-114) and UBN2 (PF14223, residues 245-356), are present in Ty1 Gag. Several CNC^R residues clustered between residues 170-220, which we define as the CNC^R domain. Other features include the Ty1 p22 protein (mapped underneath Ty1 Gag), the PR cleavage sites in Ty1 Gag and p22 proteins (H402/N403 in Gag, scissors) (47), a region that exhibits nucleic acid chaperone (NAC) activity (black bar) (28), and important Gag amino acids (black circles): a highly conserved tryptophan residue in Ty1/copia (W184) (64), and hydrophobic residues previously shown to be important for assembly (IM248/249, L252, LF339/340, and I343) (69). (B, C) An alignment of yeast Ty1-like Gag sequences was generated with ClustalW and visualized with Jalview using the ClustalX color scheme. The CNC^R (B) and UBN2 (C) domains were chosen for display. Species, UniProt annotations, and element type are listed to the left and include Ty1 and Ty2 elements from *S. cerevisiae* (Scer), Ty1-like elements from *S. kudriavzevii* (Skud) (3, 80) and the Tsk1 element from *Lachancea kluyveri* (Lklu) (79). Ty1-H3 and Ty2-917 were included, however, these elements were isolated as spontaneous retrotransposition events and are unique from known genomic elements (65, 96). Ty1 Gag residue coordinates and CNC^R substitutions (above), predicted helical regions (gray boxes, below) and important residues (below) from (A) are labeled.
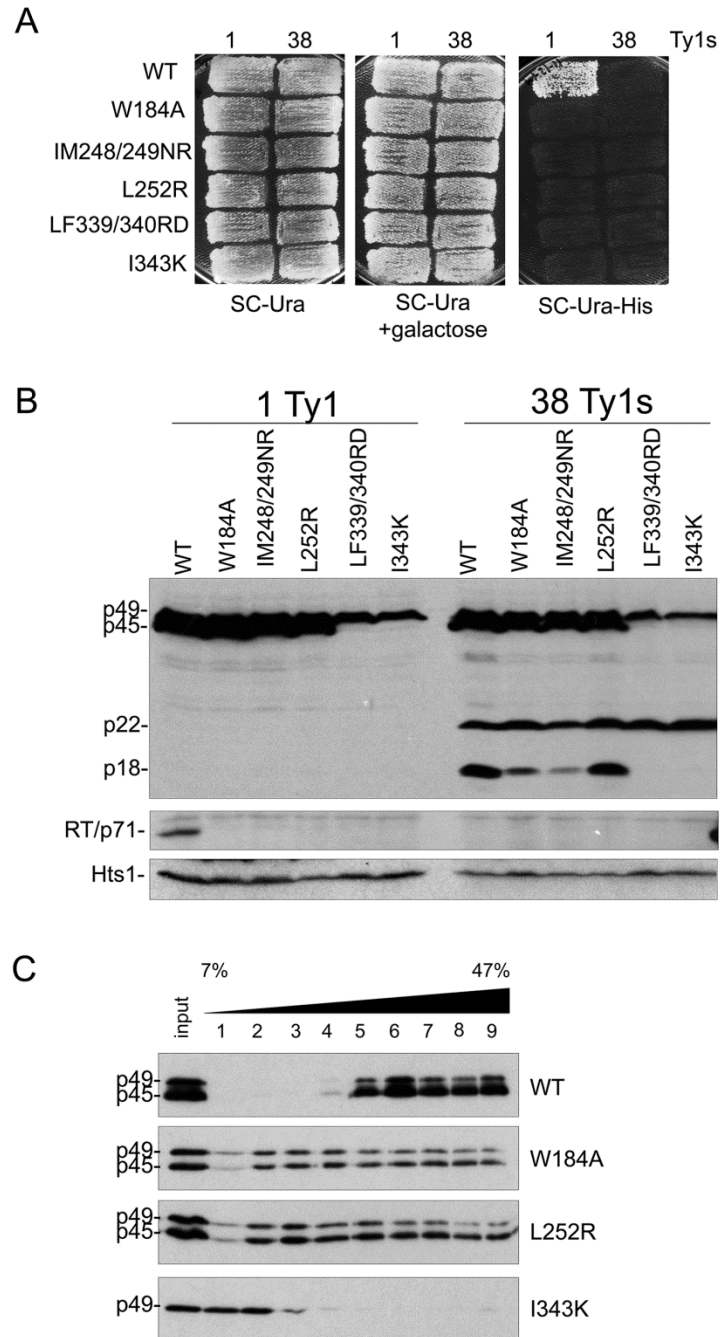
Figure 4.4   Alterations in Gag helical domains disrupt transposition and Ty1 protein cleavage. *S. paradoxus* repopulated with 1 (YEM514) or 38 (YEM515) genomic copies of Ty1-H3 and containing wild type or helix mutant pGTy1*his3-AI* plasmids were analyzed for Ty1*his3-AI* mobility using a qualitative plate assay (A) and for Ty1 protein

levels (B). (A) Cells grown on SC-Ura were induced for pGTy1*his3-AI* expression and retrotransposition by replica plating to SC-Ura + 2% galactose and incubating at 22° C for 2 days. Cell patches were replicated to SC-Ura-His and grown at 30 °C until His[+] papillae appeared. Each His[+] colony contains at least one Ty1*HIS3* insertion. (B) Cells were induced in liquid SC-Ura + 2% galactose medium for 24 h and TCA-extracted proteins were immunoblotted with p18, RT, and Hts1 antibodies. (C) Cell extracts from 1 Ty1 copy strain expressing helix mutant pGTy1*his3-AI* plasmids were centrifuged through a 7-47% continuous sucrose gradient. Equal volumes of each fraction were immunoblotted with TY antibody, which recognizes full length Gag (p49/p45).
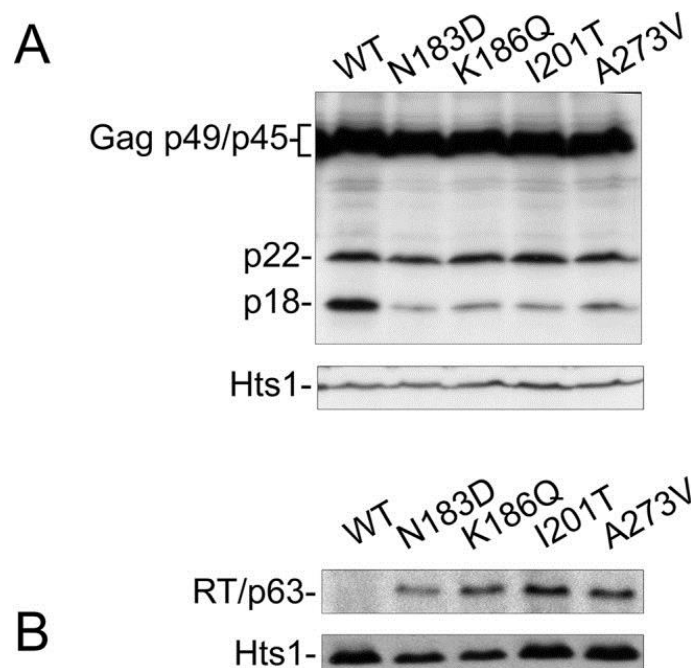
Figure 4.5 CNC[R] mutations N183D, K186Q, I201T, and A273V alter Ty1 protein levels. *S. paradoxus* repopulated with 38 genomic copies of Ty1-H3 (YEM515) containing wild type or CNC[R] pGTy1*his3-AI* plasmids were analyzed for Ty1 protein levels. Cells were induced in liquid SC-Ura + 2% galactose for 24 h. (A) TCA-extracted proteins were

immunoblotted with p18 and Hts1 antibodies. (B) Whole cell extracts were immunoblotted with RT and Hts1 antibodies.
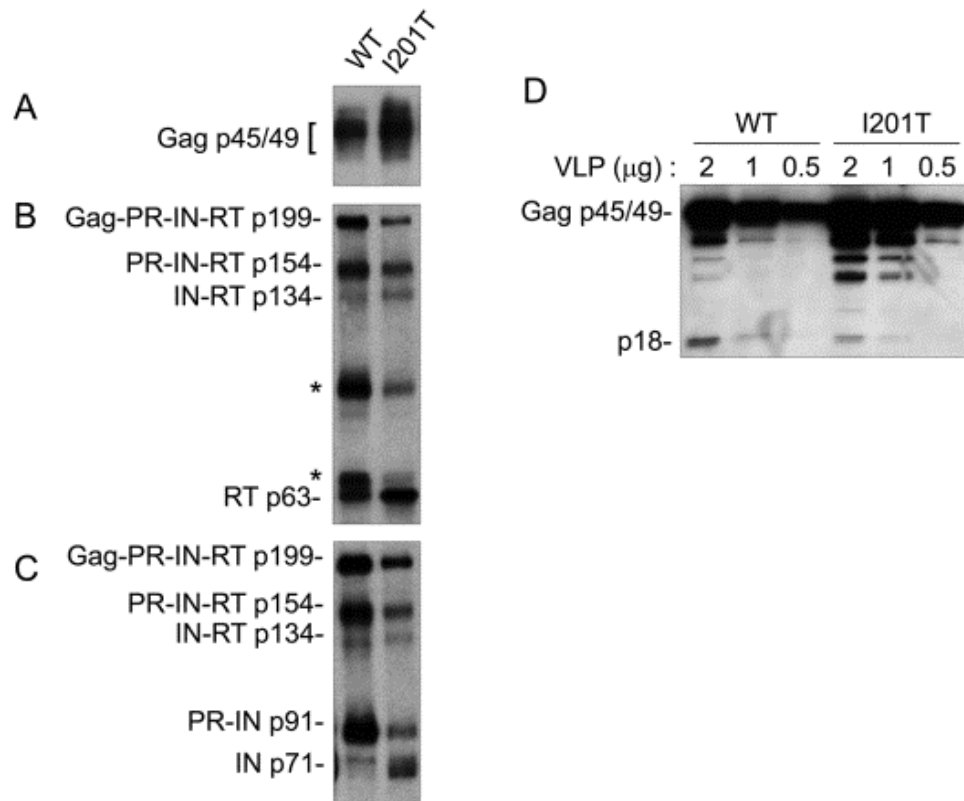
Figure 4.6 I201T CNC[R] VLPs have increased IN and RT protein levels and less p18. VLPs were isolated from *S. paradoxus* repopulated with 38 genomic copies of Ty1-H3 (YEM515) containing wild type or CNC[R] pGTy1*his3-AI-I201T* that were induced for expression. Equal amounts (2 μg) of VLPs were immunoblotted with (A) VLP, (B) RT, and (C) IN antibodies. (D) Dilutions of WT and I201T VLPs were analyzed with p18 antibody. (E) Northern blotting of total RNA and VLP RNA was performed using Ty1 (nucleotides 335-550) and *ACT1* ^{32}P-labeled riboprobes. Note that *ACT1* transcripts are not detected in purified VLPs.
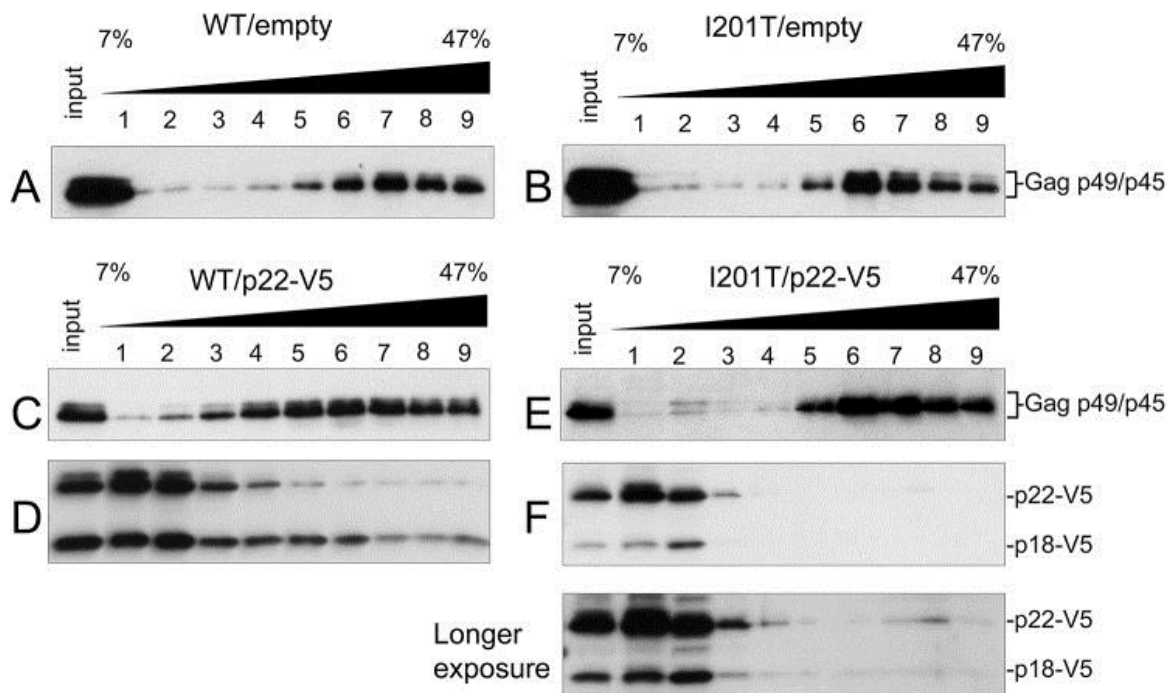
Figure 4.7 I201T CNC^R VLP assembly excludes p22-V5. Protein extracts (input) from *S. paradoxus* strains (DG3508) co-expressing WT (pBDG1534) or pGTy1*his3-AI-I201T* (pBJM24) and an empty vector (pRS416; A, B) or p22-V5 (pBJM93; C-F) were centrifuged through a 7-47% continuous sucrose gradient. Equal volumes of each fraction were immunoblotted with p18 (A-C, E) and V5 antibodies (D, F).

Table 1. Genomic CNC-resistant Ty1*his3-AI* mobility

Strain	# Ty1*his3-AI*	empty vector (-p22)	pTy1-ATGfs (+p22) X 10^{-6} (SD)	Fold decrease in Ty1*his3-AI* mobility
DG3710	2 WT	95 (12)	1.7 (0.3)	56
DG3716	2 N183D	160 (20)	37 (15)	4.3
DG3713	1 WT	110 (14)	1.5 (0.3)	73
DG3725	1 K186Q	83 (29)	22 (8.3)	4
DG3713	1 WT	38 (12)	0.32 (0.14)	120
DG3722	1 I201T	37 (7.0)	19 (4.0)	2
DG3713	1 WT	23 (6.0)	0.24 (0.04)	98
DG3719	1 A273V	100 (14)	7.6 (1.3)	13

Table 2. Ty1*his3-AI* mobility in CNC-resistant Ty1 repopulated strains

Strain	# Ty1	x 10^{-5} (SD)	Fold change in Ty1*his3-AI* mobility
DG2533	0	3.1 (0.9)	1
YEM13	+14 WT	0.1 (0.03)	↓31
YEM14	+21 WT	0.005 (0.002)	↓620
JM487	+18 K186Q	3.5 (0.7)	↑1.1
JM506	+20 K186Q	2.7 (0.4)	↓1.1
JM484	+14 I201T	2.8 (0.8)	↓1.1
JM503	+19 I201T	4.9 (1.8)	↑1.6

Table 3. CNC-resistant pGTy1*his3-AI* mobility when co-expressed with p22-V5

DG3508 (Ty1-less)	pGTy1*his3-AI* mobility x 10^{-4} (SD)	Fold decrease
+ pGTy1*his3-AI*, +empty	160 (19)	1
+pGTy1*his3-AI*, +pGp22-V5	0.2 (0.03)	783
+ pGTy1*his3-AI-I201T*, +empty	310 (34)	1
+ pGTy1*his3-AI-I201T*, +pGp22-V5	61 (19)	5

TYA

CNC^R

UBN2
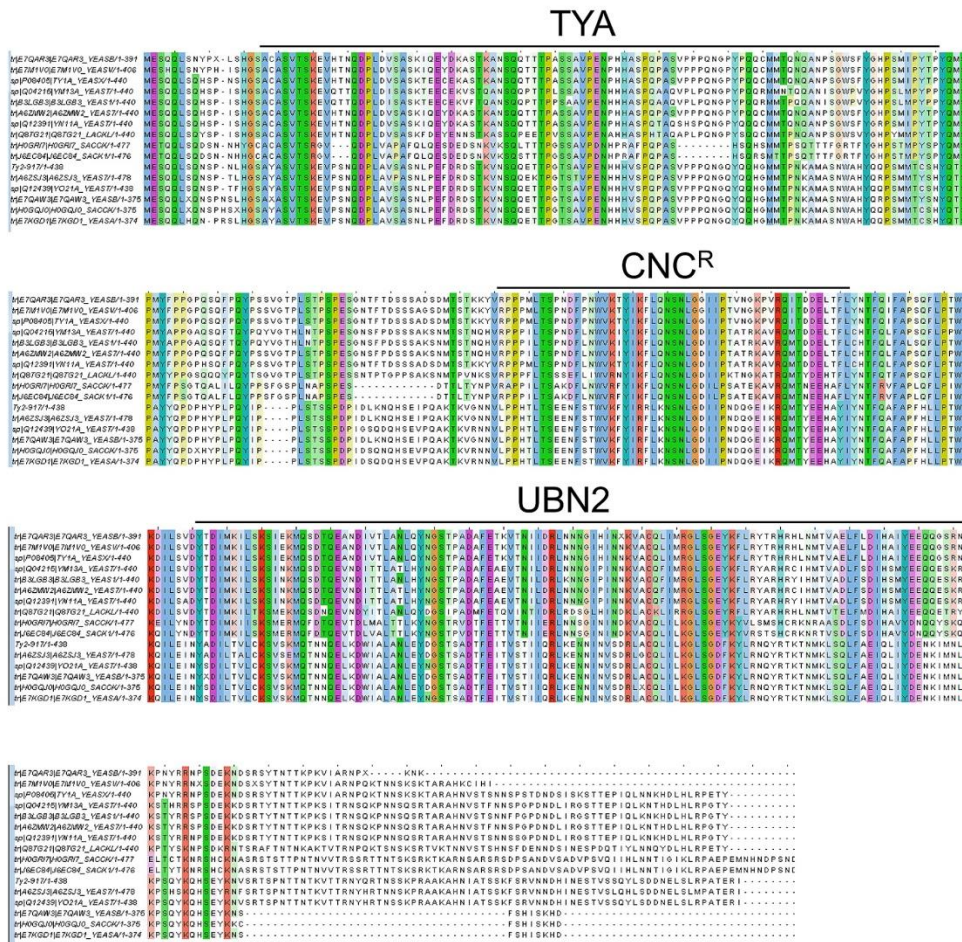
Figure S4.1 Full alignment of yeast Ty1-like Gag sequences. The alignment was generated with ClustalW and visualized with Jalview using the ClustalX color scheme. Uniprot annotations are included to the left and the following domains, which are described in the main text, are labeled: TYA (PF01021), CNC^R, and UBN2 (PF14223). Ty1-H3 (P08405) and Ty2-917 were included, however, these elements were isolated as spontaneous retrotransposition events and are unique from known genomic elements (65, 96).
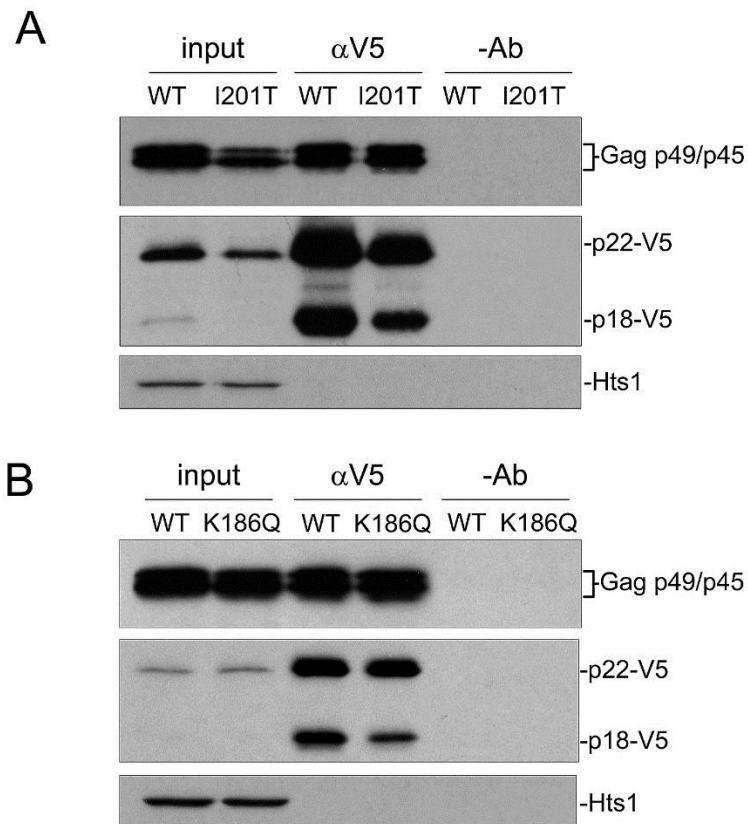
Figure S4.2 WT and I201T Gag co-immunoprecipitate with p22-V5. Protein extracts (input) from Ty1-less *S. paradoxus* strains (DG3508) co-expressing WT (pBDG1534) or pGTy1*his3-AI-I201T* (pBJM24) and p22-V5 (pBJM93) were incubated with Protein A/G Agarose beads crosslinked to V5 antibody for 2 hours at 4 ˚C. After washing, bound proteins were eluted and immunoblotted with p18 and V5 antibodies. Beads not crosslinked to V5 antibody and Hts1 were used as controls.(22)
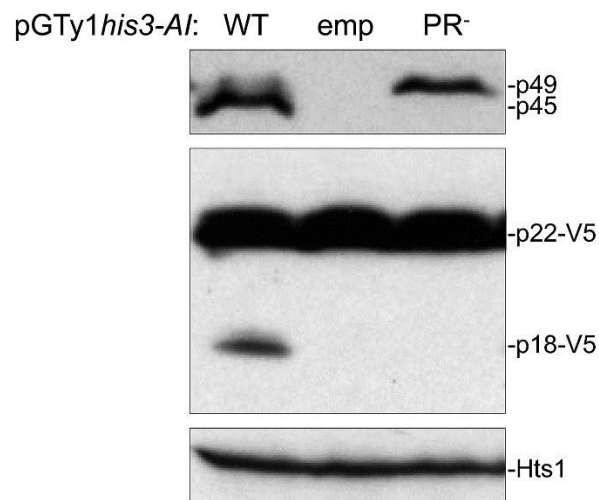
Figure S4.3 Cleavage of p22-V5 to p18-V5 is Ty1 PR-specific. *S. paradoxus* (DG3508) expressing p22-V5 (pBJM93) was transformed with empty vector (pRS414), wild type (pBDG1534), or PR-defective (PR⁻) pGTy1*his3-AI* (pBDG1606). TCA-precipitated extracts were immunoblotted with p18, V5, and Hts1 antibodies. PR was inactivated via a SacI linker insertion at the BglII site in Ty1-H3 (91).

Supplemental Tables

Table S4.1 Yeast Strains.

Strain	Relevant Genotype	Genomic Ty1s			Source
YEM514	*MATα*, *his3-Δ200hisG*, *ura3*, *spt3-ΔKanMX4*	Ty1-less, 4253*his3-AI*	+1	Ty1-	(22)
YEM570	YEM514	+7 Ty1			E. Matsuda

YEM568	YEM514	+14 Ty1	E. Matsuda
YEM572	YEM514	+21 Ty1	E. Matsuda
YEM515	YEM514	+37 Ty1	(22)
DG3582	*MATα, his3-Δ200hisG, ura3, trp1*	Ty1-less	(26)
DG3710	DG3582	+2 Ty1*his3-AI*	This study
DG3716	DG3582	+2 Ty1*his3-AI-N183D*	This study
DG3713	DG3582	+1 Ty1*his3-AI*	This study
DG3725	DG3582	+1 Ty1*his3-AI-K186Q*	This study
DG3719	DG3582	+1 Ty1*his3-AI-A273V*	This study
DG3722	DG3582	+1 Ty1*his3-AI-I201T*	This study
DG2533	*MATα, his3-Δ200hisG, ura3*	Ty1-less, +1 Ty1-4253*his3-AI*	(11)
YEM13	DG2533	+14 Ty1	E. Matsuda
YEM14	DG2533	+21 Ty1	E. Matsuda
JM487	DG2533	+18 *Ty1-K186Q*	This study
JM506	DG2533	+20 *Ty1-K186Q*	This study
JM484	DG2533	+14 *Ty1-I201T*	This study

JM503	DG2533	+19 *Ty1-I201T*	This study
DG3508	DG3582, *spt3-ΔKanMX4*	Ty1-less	This study
DG3664	DG3582	Ty2-917*his3-AI*	This study

Table S4.2. CNC-resistant mutations.

Amino acid change	Mutagenesis
Gag T55A Q351L	PCR-based (*GAG*)
Gag P173L	XL-1 Red
Gag M174I R362G	PCR-based (*GAG*)
Gag D180N	XL-1 Red
Gag N183D	XL-1 Red
Gag K186Q	XL-1 Red
Gag I189V S411G/ PR S10G	PCR-based (*GAG*)
Gag I201T	XL-1 Red
Gag T218A	XL-1 Red
Gag T218I	XL-1 Red
Gag K250E	PCR-based (*GAG*)
Gag A273V	XL-1 Red
Gag E287T	XL-1 Red and PCR-
Gag V336I	PCR-based (*GAG*)
Gag Q350R S395L	PCR-based (*GAG*)
RT D518G/V519A	PCR-based (*POL)*

Table S4.3 Plasmids.

Plasmid	Description	Markers	Source
pBDG606	pGTy1*his3-AI*	*URA3*/*CEN*	(86)
pRS416	Empty	*URA3*/*CEN*	(97)
pGTy1-H3Cla	pGTy1	*URA3*/*2μ*	(98)
pJEF1266	pGTy1pr-1702SacI	*URA3*/*2μ*	(91)
pRS426	Empty	*URA3*/*2μ*	(99)
pTy1-ATGfs	Ty1-ATGfs	*URA3*/*2μ*	(45)
pBJM42	pBDG606-*W814A*	*URA3*/*CEN*	This study
pBJM60	pBDG606-*IM248-249NR*	*URA3*/*CEN*	This study
pBJM43	pBDG606-*L252R*	*URA3*/*CEN*	This study
pBJM61	pBDG606-*LF339-340RD*	*URA3*/*CEN*	This study
pBJM62	pBDG606-*I343K*	*URA3*/*CEN*	This study
pBDG1534	pGTy1*his3-AI*	*TRP1*/*CEN*	(26)
pBJM24	pBDG1534-*I201T*	*TRP1*/*CEN*	This study
pBJM93	pGp22-V5	*URA3*/*CEN*	This study
pRS414	Empty	*TRP1*/*CEN*	(97)
pBDG1606	pGTy1PR- *his3-AI*	*TRP1*/*CEN*	This study
pGTy917	pGTy2-917	*URA3*/*2μ*	(100)
pJC384	Ty2-917	*URA3*/*2μ*	(45)
pBDG1565	pGAL-Yes2:1038-1616	*URA3*/*2μ*	(26)

<u>Table S4.4   CNC^R pGTy1*his3-AI* mobility.</u> Ty1-less *S. paradoxus* strains repopulated with 1 (YEM514) or 38 (YEM515) copies of Ty1-H3 and containing wild type or CNC^R pGTy1*his3-AI* were analyzed for Ty1*his3-AI* mobility as described in the Materials and Methods. *Ty1his3-AI* mobility is calculated by dividing the number of colonies that form on SC-Ura-His by number of colonies on SC-Ura, and the standard deviation (SD) from 4 replicates is shown. Frequencies from individual experiments are shown, but the entire analysis was repeated at least three times with comparable results.

Experiment	Ty1	pGTy1*his3AI*	pGTy1his3-AI mobility X 10^{-5} (SD)
A	1	WT	3500 (610)
	38	WT	1.8 (0.5)
	38	Gag N183D	420 (70)
	38	Gag A273V	440 (96)
	38	Gag I201T	620 (110)
	38	Gag K186Q	760 (120)
B	1	WT	2100 (300)
	38	WT	3.1 (0.4)
	1	Gag N183D	1900 (250)
	1	Gag A273V	2200 (340)
	1	Gag I201T	2100 (250)
	1	Gag K186Q	2500 (530)

C	1	WT	990 (320)
	38	WT	1.3 (0.3)
	38	Gag P173L	85 (12)
	38	Gag D180N	45 (8.6)
	38	Gag T218A	53 (12)
	38	Gag T218I	43 (12)
	38	Gag E287T	43 (16)
D	1	WT	2300 (270)
	38	WT	2 (0.3)
	1	Gag P173L	320 (90)
	1	Gag D180N	2700 (360)
	1	Gag T218A	3100 (350)
	1	Gag T218I	2700 (320)
	1	Gag E287T	2700 (300)
E	1	WT	1900 (300)
	38	WT	1.2 (0.3)
	38	Gag T55A Q351L	22 (3.9)
	38	Gag K250E	14 (3.1)
F	1	WT	1000 (270)
	38	WT	1.5 (0.4)
	1	Gag T55A Q351L	1000 (170)
	1	Gag K250E	5.6 (1.3)

G	1	WT	1900 (370)
	38	WT	1 (0.2)
	38	Gag V336I	13 (2.9)
	38	Gag Q350R S395L	18 (2.7)
	38	Gag M174I R362G	51 (17)
	38	Gag I189V S411G	30 (8.8)
H	1	WT	2200 (290)
	38	WT	0.5 (0.1)
	1	Gag V336I	3200 (370)
	1	Gag Q350R S395L	1800 (350)
	1	Gag M174I R362G	1900 (150)
	1	Gag I189V S411G	2900 (520)
I	1	WT	1400 (200)
	38	WT	1.8 (0.3)
	1	RT D518G/V519A	6.1 (1.1)
	38	RT D518G/V519A	20 (2.9)
I	1	WT	2400 (330)
	38	WT	1 (0.2)
	1	I201T/A273V	190 (19)
	38	I201T/A273V	85 (36)

Table S4.5 Ty2-917*his3-AI* mobility. A Ty1-less *S. paradoxus* strain with Ty2-917*his3-AI* (DG3664) was transformed with the following multicopy vectors (2μ): (A) empty (pRS416), pGTy2-917 (pGTy2-917) or pTy2-917 (pJC384) (B) empty (pYES2) or p22 (pBDG1565). Cells were grown at 22˚ C for two days in SC media containing glucose (A) or galactose (B). Numbers represent Ty2*his3-AI* mobility events per cell and standard deviations are provided in parentheses. Mobility assays were repeated at least three times and representative results are shown.

	Vector (2μ)	Ty2-917*his3-AI* mobility x 10^{-6} (SD)	Fold change
A	empty	6.5 (1.0)	1
	pGTy2-917	5.5 (1.6)	↓1.2
	Ty2-917	9.5 (1.5)	↑1.5
B	empty	9.0 (2.7)	1
	p22	4.9 (2.7)	↓1.8

Table S4.6 PR/p4 is not required for CNC.

Vector (2μ)	Ty1 *his3-AI* mobility x 10^{-6} (SD)	Fold decrease
empty	490 (140)	1
pGPOLΔ	3.3 (1.5)	150
pGPOLΔ*d1*	3.6 (0.5)	140
pYES2-p45	3.0 (1.7)	163

CHAPTER 5

CONCLUSIONS

The role of Ty1 Gag in the nuclear export of Ty1 mRNA

In Chapter 2, Gag-p45 was demonstrated to be the sole Ty1 protein required for the formation of Ty1 retrosomes. To study the role of Gag-p45 in Ty1 mRNA trafficking, a galactose-inducible Ty1 element carrying a frameshift mutation that eliminated the expression of Ty1 Gag and Gag-Pol proteins was expressed in Ty1-less *S. paradoxus* (Figure 2.1). The Ty1fs frameshift mutation occurs immediately after the *GAG* ATG and results in a premature stop codon approximately 128 nt downstream. Although this Ty1fs RNA transcript was shown to be a substrate for degradation via nonsense-mediated decay (NMD) (Figure 2.3), enough of the full length transcript was present to study its localization in the absence of Ty1 proteins. Expression of Ty1fs RNA revealed the lack of cytoplasmic Ty1 RNA foci and an accumulation of Ty1fs RNA in the nucleus. The lack of cytoplasmic Ty1fs RNA is likely due to degradation by cytoplasmic complexes, including the NMD machinery, Xrn1 in P bodies, and the cytoplasmic exosome (Figure 2.3). Importantly, the nuclear localization of Ty1fs RNA was not influenced by the NMD machinery (Figure 2.4). When Gag was provided *in trans*, Ty1fs RNA levels increased and localized to retrosomes in the cytoplasm. The discovery that Ty1fs RNA remains in the nucleus in the absence of Gag is fascinating, raising the

possibility that Gag is required for efficient nuclear export of Ty1 RNA. In certain retroviruses such as RSV, Gag traffics into the nucleus and binds unspliced retroviral RNA (1). The RSV Gag-RNA complex is exported from the nucleus and this nuclear interaction between RSV Gag and RNA may mark the genomic RNA for packaging (2). The export of RSV Gag-RNA complexes is dependent on the conserved CRM1 nuclear export pathway, while the majority of RSV RNA transcripts exit the nucleus by association with Tip-associating protein/Nuclear RNA export factor 1 (TAP/NXF1) (2, 3). The yeast ortholog of TAP is Mex67, which also exports Ty1 RNA from the nucleus (4). In wild type or *mex67-5* mutants, Ty1 Gag was not visualized in the nucleus (Figure 1.6-7). Perhaps export of Ty1 Gag/RNA complexes in the nucleus requires the CRM1 export pathway. A common method for inhibiting CRM1 export is by treating cells with leptomycin B (LMB). To confirm or rule out Ty1 Gag nuclear trafficking, it will be important to treat yeast cells with this inhibitor and visualize Gag localization. Yeast are not naturally sensitive to LMB, but there are strains available that carry a LMB-sensitive allele of *CRM1* that can be used to test for Crm1-dependent nuclear export of Ty1 Gag (5). Another interesting question to explore is the contribution of the *GAL1* promoter to the localization of Ty1fs RNA, as it is well established that promoter sequences influence stability of mRNA transcripts and can dictate downstream localization in the cell, at least under stress conditions (6-8). Ty1fs RNA dynamics could be examined by following localization of the frame-shifted transcript expressed from its endogenous promoter, although this could be difficult due to the low level of Ty1fs RNA present in cells (9).

The importance of Gag and SRP-bound Ty1 RNA translation complexes for Ty1 RNA stability and retrosome formation

Since publication of the work presented in Chapter 2, it has been proposed that Ty1 retrosome formation occurs at SRP-associated Ty1 RNA translation complexes bound to SRP receptors at the ER (10). Stability of Gag is dependent on ER translocation, as depletion of SRP factors reduces Gag stability. Additionally, hypomorphic mutations in SRP-related genes or treatment with tunicamycin both decrease the level of SRP-mediated translocation of Ty1 Gag to the ER and increase the abundance of retrosomes. Retrosome abundance returns to normal if the depletion of SRP factors is suppressed with a low concentration of cyclohexamide, which slows translation elongation and presumably allows more time for SRP docking. In order for SRP-associated Ty1 RNA translation complexes to nucleate retrosomes, it is hypothesized that Gag would undergo retrograde transport via an unknown mechanism from the ER to the cytoplasm, where it would bind translating Ty1 mRNA. This would mean that the same pool of Ty1 mRNA is used for translation and for packaging of the gRNA into VLPs. Additional support for this model includes the observation that Ty1 retrosomes are disrupted by the global block in translation initiation upon glucose starvation (11).

Doh and coworkers identified cellular conditions, such as the depletion of SRP factors that result in low levels of Ty1 Gag protein with no effect on Ty1 RNA stability, which is different than the conclusion that Ty1 Gag is required for Ty1 RNA stability presented in Chapter 2 (10). Doh *et al.* suggested that decreased stability of Ty1fs RNA may be a result of its inability to undergo a sufficient amount of translation, rather than

due to the absence of Gag protein (10). However, it is very likely that Ty1fs RNA is translated since it is a target of NMD, which degrades RNA via a translation-dependent mechanism (12). It remains to be determined if the level of Ty1fs RNA translation is insufficient for stability of Ty1 RNA. Additionally, if SRP-docking of the translation complex is required for Ty1 RNA stability, then a SRP-localization signal likely within the N-terminus of Gag is required (10). Since the frameshift mutation in Ty1fs RNA alters the sequence of the peptide being translated, this would likely block SRP docking. These hypotheses regarding Ty1 RNA stability could be addressed by producing a full length Ty1 mRNA that encodes most of the N-terminus of Gag prior to reaching an introduced premature stop codon. This would allow a higher level of translation of Ty1 mRNA and allow for SRP-docking of Ty1 RNA translation complexes, considering that the Gag N-terminus might be required for SRP binding. Also, the nucleic acid chaperone activity of Gag is located in the C-terminus, not in the N-terminus, so this truncated protein will likely not be involved in binding Ty1 RNA (13, 14). Interestingly, Malagon and Jensen performed a similar experiment in a different context (15). While full length Ty1 mRNA was not studied, a *GAG* mRNA carrying various deletion or frameshift mutations producing C-terminally truncated Gag proteins were tested for RNA levels and the ability to form retrosomes. These *GAG* mRNAs, which did not encode full length Gag, produced lower levels of RNA in cells compared to the wild type *GAG* transcript and did not form retrosomes. Their study supports the idea that full length Gag-p45, rather than SRP-docked Ty1 translation complexes, is most important for Ty1 RNA stability. Regardless, it is clear from our work that Gag-p45 provided from a separate *GAG* transcript stabilizes full length Ty1fs RNA (Figure 2.7). Both SRP-

docking of Ty1 translation complexes and subsequent Gag binding may be important for stabilizing Ty1 RNA and promoting retrosome formation; however, further investigation will be required to understand how these different RNA-binding events contribute to retrosomes.

<u>Understanding CNC in the natural context</u>

In Chapters 3 and 4, a mechanism for CNC in yeast was proposed that involves several posttranslational effects upon VLP assembly and function mediated by a truncated Gag protein (p22) produced from a subgenomic Ty1 RNA (Figure 5.1). Epitope tagged p22 expressed via a *GAL1* promoter interacted with Gag (Figures 3.11 and 4.S2). This interaction mediates the following disruptions to Ty1 replication: a block in retrosome formation (Figure 3.10), a block in normal VLP assembly (Figure 3.8-9), and inhibition of VLP maturation (Figures 3.8 and 4.6). Most of these biochemical effects are observed during the simultaneous overexpression of pGTy1 and/or p22, and thus are likely exaggerated compared to the *in vivo* scenario, where Ty1 gene product concentrations are much lower. It is clear that the more p22 produced in the system, the greater the inhibition of Ty1 retrotransposition, suggesting that a ratio of Gag:p22 might be critical for CNC. However, it is important to understand properties of CNC in a natural condition where Ty1 products are not overexpressed. Considering a cell that carries a single versus multiple Ty1 copies in the genome, it seems reasonable that all Ty1 products, including Ty1 mRNA, Ty1i RNA, Gag, and p22, would increase coordinately. Is the Gag:p22 ratio important *in vivo* and why does p22 appear more potent at higher Ty1 copy number?

To answer these questions, it will be important to identify the Gag substrate (monomeric or oligomeric) that binds p22 with the highest affinity. In Chapters 3 and 4, it was demonstrated that a tagged version of p22 can coimmunoprecipitate Gag expressed from the endogenous 32 Ty1 elements in S288C (Figure 3.11) or from cells that express Gag from a pGTy1 element (Figure 4. S2). These strains are quite different in their ability to form VLPs. Endogenous VLPs are not visualized in S288C, while cells expressing pGTy1 readily form VLPs (16). Thus, it is reasonable to propose that p22 can interact with Gag in assembly stages prior to assembly of a complete VLP, although we cannot rule out that p22 may also interact with fully assembled VLPs. If p22 interacts with Gag at a stage prior to VLP formation, then it could be interacting with monomeric Gag, oligomeric Gag, or higher order assembly intermediates. By treating cells with the chemical cross-linking reagent formaldehyde, interacting proteins are "locked" in place and monomeric and low order oligomeric complexes can be separated on low percentage polyacrylamide gels under denaturing conditions. Protein cross-linking has been used with HIV-1 to determine Gag multimerization state in various cellular compartments (17). Utilizing this approach with Ty1 Gag has successfully revealed monomeric or low order oligomeric Gag present in the cell, but there is no evidence so far that p22-V5 interacts with these structures (JMT, unpublished data). High-resolution cell fractionation analyses may be necessary to identify the specific Gag complexes that interact with p22. Taken together, the data supports the idea that p22 interacts with higher order assembly intermediates of Gag or partially assembled VLPs. In addition, this would draw similarities to the capsid-binding restriction factors Fv1 and Trim5α,

which do not exhibit binding to monomeric CA, but rather to a polymerized CA lattice (18, 19).

Interactions between p22 and a Gag assembly intermediate may explain why p22 has differential effects on Ty1 mobility in the presence of varying Ty1 copy number (Figure 5.2). In cells with low Ty1 copy number (CNC⁻), the concentration of Ty1 mRNA and Gag proteins is also low. It is not known whether CNC⁻ cells support Ty1 retrosome formation due to the fact that the concentration of Ty1 products is below the detection limit for FISH/IF microscopy. However, CNC⁻ cells support high levels of Ty1 mobility (9, 20), indicating that retrotransposition-competent VLPs form in these cells, although this has not been confirmed by TEM. This CNC⁻ environment, including the local or global concentrations of Ty1 products and host cofactors involved in VLP assembly and function, is ideal for Ty1 retrotransposition. In contrast, for cells with a high Ty1 copy number (CNC⁺), the concentration of Ty1 mRNA and Gag is much higher and retrosomes form in the cytoplasm (16). Transposition in the CNC⁺ laboratory strain S288C is very low and VLPs are not visible by TEM (16, 21). It is well established that retroviral Gag proteins undergo concentration-dependent assembly (22-25). If Ty1 Gag assembly is concentration-dependent, then there will be more assembly occurring in CNC⁺ cells, although not complete as judged by the lack of morphologically distinct particles (16). It is logical that if p22 binds Ty1 assembly intermediates, then CNC⁺ cells provide an environment with many more available binding sites (Figure 5.2). Thus, an increase in copy number results in a Ty1 Gag concentration that surpasses a threshold for concentration-dependent Gag assembly. The more assembly intermediates present, the more binding sites that are available for p22. This hypothesis could be tested in

several ways. First, it has been shown that replacing the nucleocapsid domain of HIV-1 Gag with a self-dimerizing leucine zipper can increase Gag assembly, even in the absence of nucleic acids (26). If a leucine zipper is introduced to the C-terminus of Ty1 Gag to enhance assembly of Gag in CNC$^-$ cells, would CNC$^-$ cells now be more sensitive to p22? Second, studies to biochemically characterize Ty1 assembly intermediates (size exclusion chromatography and/or ultracentrifugation of cellular extracts) would be instrumental to identifying the propensity for assembly in low versus high copy strains. Strains repopulated with the Ty1 separation of function mutation A1123G (Figure 3.2; Table 3.2) will also be important to include in these types of analyses to define Ty1 product concentration and efficiency of assembly in the absence of p22 in high copy strains. Third, another avenue to consider is the increasingly advanced microscopy techniques that are being developed to analyze the stoichiometry and diffusion of molecules in live cells, as was recently used to demonstrate the presence of oligomeric HIV-1 Gag in the cytoplasm (27). These methods may reveal different Gag interactions, assembly intermediates, and potential substrates for p22 binding as Ty1 copy number increases.

References

1.   Garbitt-Hirst R, Kenney S, Parent L. Genetic evidence for a connection between Rous sarcoma virus gag nuclear trafficking and genomic RNA packaging. J Virol 2009;83(13):6790-6797.

2.   Scheifele LZ, Garbitt RA, Rhoads JD, Parent LJ. Nuclear entry and CRM1-dependent nuclear export of the Rous sarcoma virus Gag polyprotein. Proc Natl Acad Sci U S A 2002;99(6):3944-3949.

3.   Gruter P, Tabernero C, von Kobbe C, Schmitt C, Saavedra C, Bachi A, Wilm M, Felber BK, Izaurralde E. TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. Mol Cell 1998;1(5):649-659.

4.   Malagon F, Jensen T. The T-body: a new cytoplasmic RNA granule in Saccharomyces cerevisiae. Mol Cell Biol 2008.

5.   Neville M, Rosbash M. The NES-Crm1p export pathway is not a major mRNA export route in Saccharomyces cerevisiae. EMBO J 1999;18(13):3746-3756.

6.   Bregman A, Avraham-Kelbert M, Barkai O, Duek L, Guterman A, Choder M. Promoter elements regulate cytoplasmic mRNA decay. Cell 2011;147(7):1473-1483.

7.   Trcek T, Larson DR, Moldon A, Query CC, Singer RH. Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. Cell 2011;147(7):1484-1497.

8.   Zid BM, O'Shea EK. Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast. Nature 2014;514(7520):117-121.

9.   Garfinkel D, Nyswaner K, Wang J, Cho J. Post-transcriptional cosuppression of Ty1 retrotransposition. Genetics 2003;165(1):83-99.

10.   Doh JH, Lutz S, Curcio MJ. Co-translational localization of an LTR-retrotransposon RNA to the endoplasmic reticulum nucleates virus-like particle assembly sites. PLoS Genet 2014;10(3):e1004219.

11.   Ashe MP, De Long SK, Sachs AB. Glucose depletion rapidly inhibits translation initiation in yeast. Mol Biol Cell 2000;11(3):833-848.

12.   Gonzalez CI, Bhattacharya A, Wang W, Peltz SW. Nonsense-mediated mRNA decay in Saccharomyces cerevisiae. Gene 2001;274(1-2):15-25.

13.   Cristofari G, Ficheux, D., Darlix, J.L. The Gag-like Protein of the Yeast Ty1 Retrotransposon Contains a Nucleic Acid Chaperone Domain Analogous to Retroviral Nucleocapsid Proteins. Journal of Biological Chemistry 2000;275:19210-19217.

14.   Nishida Y, Pachulska-Wieczorek K, Blaszczyk L, Saha A, Gumna J, Garfinkel DJ, Purzycka KJ. Ty1 retrovirus-like element Gag contains overlapping restriction factor and nucleic acid chaperone functions. Nucleic Acids Res 2015;43(15):7414-7431.

15.   Malagon F, Jensen TH. T-body formation precedes virus-like particle maturation in S. cerevisiae. RNA Biol 2011;8(2):184-189.

16.   Checkley MA, Nagashima K, Lockett SJ, Nyswaner KM, Garfinkel DJ. P-body components are required for Ty1 retrotransposition during assembly of retrotransposition-competent virus-like particles. Molecular and Cellular Biology 2010;30(2):382-398.

17.   Kutluay SB, Bieniasz PD. Analysis of the initiating events in HIV-1 particle assembly and genome packaging. PLoS pathogens 2010;6(11):e1001200.

18.   Hilditch L, Matadeen R, Goldstone DC, Rosenthal PB, Taylor IA, Stoye JP. Ordered assembly of murine leukemia virus capsid protein on lipid nanotubes directs specific binding by the restriction factor, Fv1. Proc Natl Acad Sci U S A 2011;108(14):5771-5776.

19.   Stremlau M, Perron M, Lee M, Li Y, Song B, Javanbakht H, Diaz-Griffero F, Anderson DJ, Sundquist WI, Sodroski J. Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. Proc Natl Acad Sci U S A 2006;103(14):5514-5519.

20.   Matsuda E, Garfinkel DJ. Posttranslational interference of Ty1 retrotransposition by antisense RNAs. Proc Natl Acad Sci USA 2009;106(37):15657-15662.

21.   Curcio MJ, Garfinkel DJ. Single-step selection for Ty1 element retrotransposition. Proc Natl Acad Sci U S A 1991;88(3):936-940.

22.   Campbell S, Rein A. In vitro assembly properties of human immunodeficiency virus type 1 Gag protein lacking the p6 domain. J Virol 1999;73(3):2270-2279.

23.   Campbell S, Vogt VM. In vitro assembly of virus-like particles with Rous sarcoma virus Gag deletion mutants: identification of the p10 domain as a morphological determinant in the formation of spherical particles. J Virol 1997;71(6):4425-4435.

24.   Lingappa JR, Reed JC, Tanaka M, Chutiraka K, Robinson BA. How HIV-1 Gag assembles in cells: Putting together pieces of the puzzle. Virus research 2014;193:89-107.

25.   Lingappa JR, Thielen BK. Assembly of immature HIV-1 capsids using a cell-free system. Methods Mol Biol 2009;485:185-195.

26.   Zhang Y, Qian H, Love Z, Barklis E. Analysis of the assembly function of the human immunodeficiency virus type 1 gag protein nucleocapsid domain. J Virol 1998;72(3):1782-1789.

27.	Hendrix J, Baumgartel V, Schrimpf W, Ivanchenko S, Digman MA, Gratton E, Krausslich HG, Muller B, Lamb DC. Live-cell observation of cytosolic HIV-1 assembly onset reveals RNA-interacting Gag oligomers. J Cell Biol 2015;210(4):629-646.

28.	Garfinkel DJ, Tucker JM, Saha A, Nishida Y, Pachulska-Wieczorek K, Blaszczyk L, Purzycka KJ. A self-encoded capsid derivative restricts Ty1 retrotransposition in Saccharomyces. Curr Genet 2015.
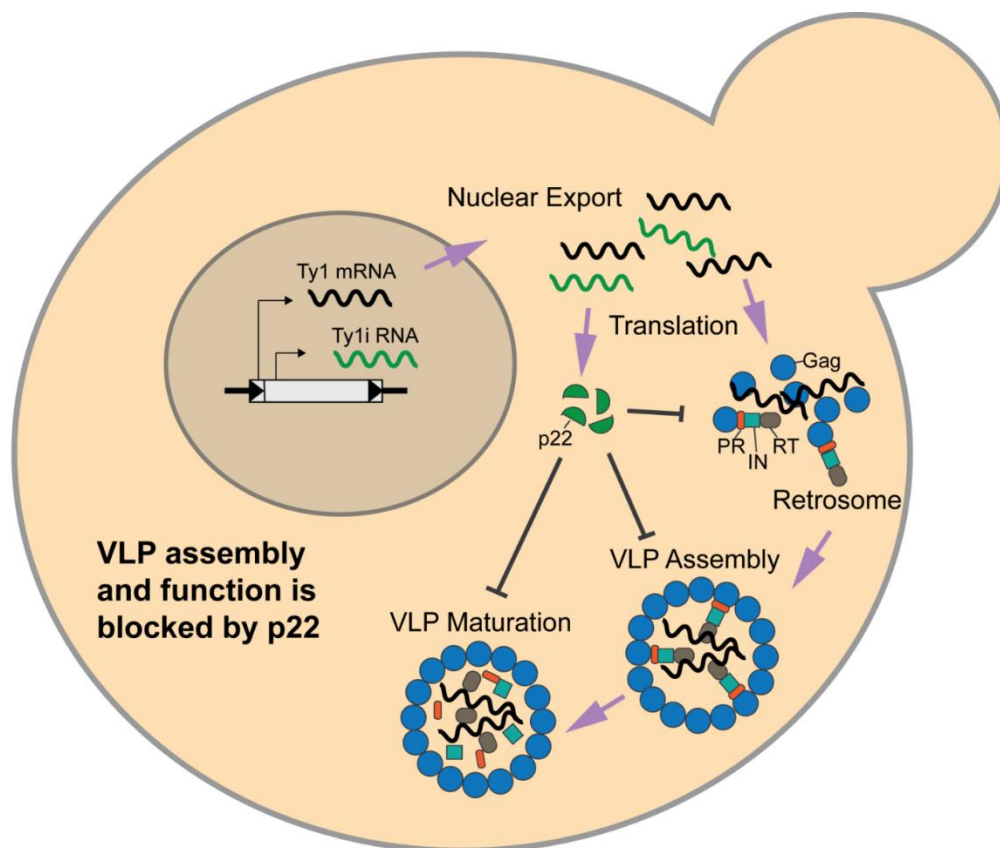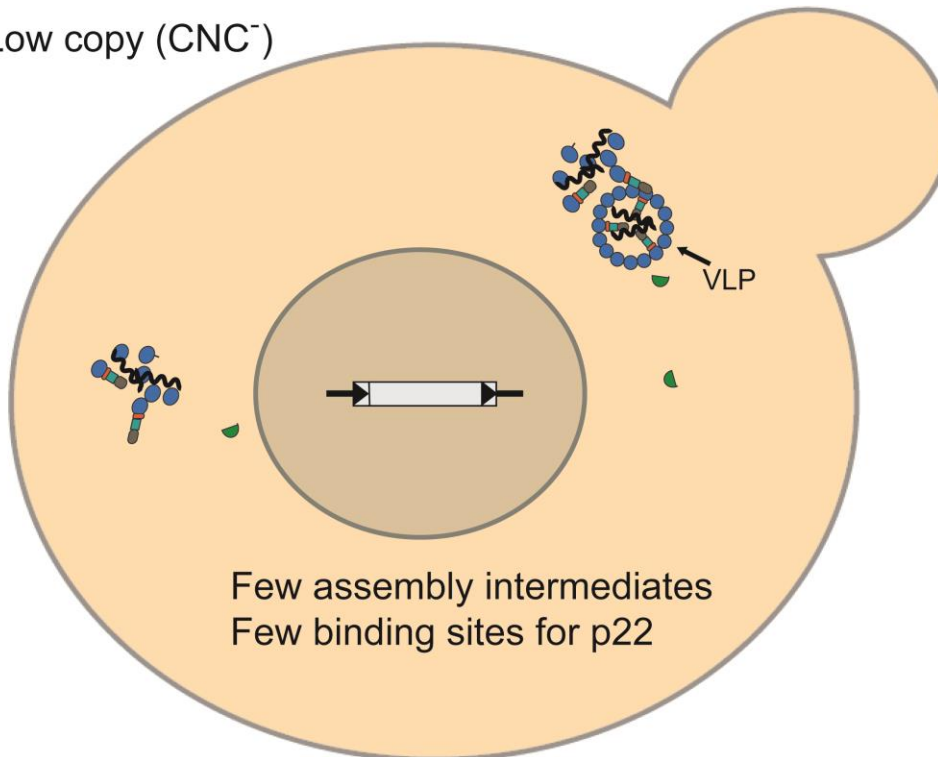
<u>Figures</u>

Figure 5.1. Ty1 VLP functions inhibited by p22. An abbreviated Ty1 replication cycle is shown highlighting steps affected by p22 (green semi-circles). When p22 and Ty1 are co-expressed, p22 colocalizes with Gag and disrupts retrosome formation. p22 associates with Ty1 VLPs and alters assembly and maturation. Used with permission from (28).

Low copy (CNC$^-$)

VLP

Few assembly intermediates
Few binding sites for p22

High copy (CNC$^+$)

assembly intermediates

Many assembly intermediates
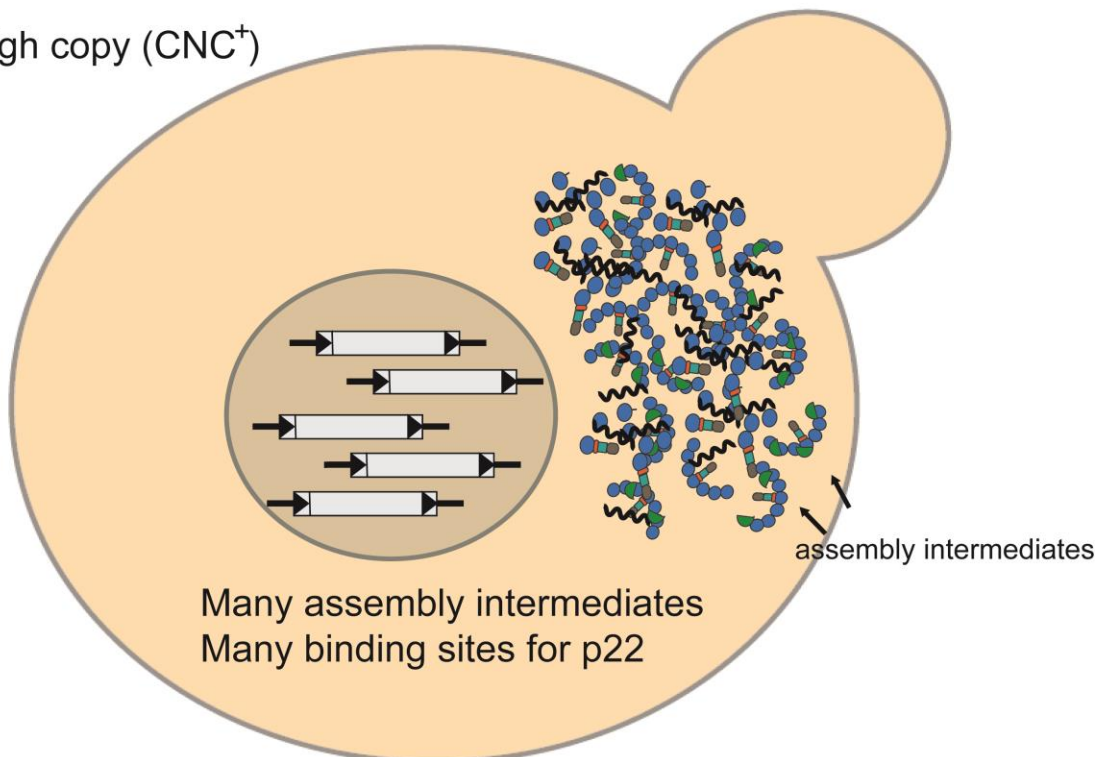Many binding sites for p22

Figure 5.2 A model for CNC: p22 binds preferentially to assembly intermediates. In CNC⁻ cells (top), there are few assembly intermediates, as VLP assembly is highly efficient. In CNC⁺ cells (bottom), there are many assembly intermediates due to concentration-dependent assembly of Gag in retrosomes. If p22 (green semi-circles) preferentially binds assembly intermediates, then there are more binding sites for p22 in CNC⁺ cells.