Genes with Internal Repeats Require the THO Complex for Transcription

by

Vladimir Voynov

B.A., Clark University, Worcester, MA
1999

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OR PHILOSOPHY IN BIOLOGY

AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
MAY 2006
[June 2006]

Signature of Author: _____
Department of Biology

Certified by: _____
Gerald R. Fink, PhD.
Thesis Advisor

Accepted by: _____
Chairman of the Graduate Committee

Genes with Internal Repeats Require the THO Complex for Transcription

by

Vladimir Voynov

Submitted to the Department of Biology on May 26, 2006
in partial fulfillment of the requirements for the degree of

Doctor or Philosophy in Biology

ABSTRACT

A genome-wide screen for promoter-independent regulators of the *Saccharomyces cerevisiae* adhesin *FLO11* identified 37 novel regulators. Among the mutants with the strongest phenotype were the deletions of genes encoding subunits of the evolutionarily conserved THO complex.

The THO complex, which is recruited to actively transcribed genes, is required for the efficient expression of *FLO11* and other yeast genes that have long internal tandem repeats. *FLO11* transcription elongation in Tho^- mutants is hindered in the region of the tandem repeats, resulting in a loss of function. Moreover, the repeats become genetically unstable in Tho^- mutants. A *FLO11* gene without the tandem repeats is transcribed equally well in Tho^+ or Tho^- strains. The Tho^- defect in transcription is suppressed by overexpression of topoisomerase I, suggesting that the THO complex functions to rectify aberrant structures that arise during transcription.

Thesis Advisor: Gerald R. Fink, Ph.D.
Title: American Cancer Society Professor of Genetics, M.I.T.

Acknowledgements

I have been fortunate in many ways. After my first year of graduate school, I got to spend several months in the lab of Dr. Sylvia Sanders. She introduced me to yeast research and to the yeast research community. When I transitioned to the lab of Dr. Gerry Fink, I met many other great people. I have learned a lot from their knowledge, dedication, talents and consideration.

I would like to thank the following people in Gerry's lab for help with various aspects of this thesis project. Reeta Prusty and Robert Wheeler helped me design the genome-wide screen. Kevin Verstrepen and An Jansen were instrumental in the analysis of repetitive sequences. Hypoxic growth analysis was performed following Cintia Hongay's protocols, and GFP expression was assayed in a background strain prepared by Sherwin Chan. In addition, I would like to thank Robert Wheeler, Claire Tseng, and Chia Wu for comments on the manuscript.

The development of this thesis project profited from collaborations outside the lab as well. Dr. Steve Buratowski at Harvard Medical School, and his student Vanessa Runner mentored me on the Chromatin Immuno-precipitation experiments. Milan de Vries from Dr. Steve Bell's lab provided me with the large-scale transformation protocol. I would also like to thank the Lindquist lab for use of the Bioscreen machine, George Bell, Kim Walker and Sanjeev Pillai for help with Bioinfomatics, and Tom DiCesare for help with graphics.

My Thesis Committee Members, Dr. Frank Solomon, Dr. Chris Kaiser, Dr. Hazel Sive, and Dr. Fred Winston, provided me with key pointers all along the way. I am grateful to them for their time and advice. I would also like to thank Dr. Phil Sharp for a discussion on DNA:RNA hybrid formation.

I have been very fortunate to have Dr. Gerry Fink as my mentor. His enthusiasm and vision allowed me to pursue an unexpected finding, and to develop this thesis project. I am grateful to him for the opportunity to pursue an independent research project in his lab, and for his words of wisdom that will be remembered for years to come.

Finally, I would like to thank the connection of family and friends from Sofia, BG and Danielson, CT for their care and support.

Vladimir Voynov

Table of Contents

Chapter 1: Introduction

Repetitive sequences present an important feature of all genomes, as they are a source of genetic and functional variability. Certain variations are deleterious, while others can be advantageous. Regulation of repetitive sequences occurs at many levels.

This thesis describes a genome-wide screen for promoter-independent regulators of a fungal adhesin, and the finding that the THO complex is required for transcription of genes with long tandem repeats.

The screen was conducted in the yeast *Saccharomyces cerevisiae*, and was motivated by an interest in yeast cell wall genes with internal repeats. There are 29 genes with long (≥ 40nt) tandem repeats in *S. cerevisiae*, and 22 of them are cell wall genes (Fig. 1; Verstrepen et al., 2005). The internal repetitive sequence can generate genetic and functionally variability through loss or gain of repeats during replication (Verstrepen et al., 2005). This thesis shows that transcription in Tho$^-$ mutants also can induce genetic instability of the repetitive sequences within *FLO* genes.

This introduction begins with a brief overview of the regulatory and processing steps of cell surface genes. Then, a range of examples from bacteria, yeast, and vertebrates focuses on the role of repetitive sequences in replication, transcription, or recombination. A common theme of all these examples is that genetic instability of repetitive sequences can lead to aberrant pairing that impedes transcription or replication, and can alter function. Next, I present a detailed review on the THO complex. Although great advances have been made on the elucidation of its role in transcription elongation and recombination, the THO complex has been characterized exclusively with the use of artificial plasmid constructs. To date, no phenotypes have been associated with

endogenous genes that may require a functional THO complex for expression. The introduction ends with two questions that have been at the center of this work.

REGULATION OF EUKARYOTIC CELL SURFACE GENES

Most genes with long tandem repeats encode cell surface proteins (Verstrepen et al., 2005). Some repetitive sequences translate into peptides that are rich in serine, threonine, proline, and alanine amino acids. In spite of their low complexity, these regions are an important source of genetic and functional variability that facilitates adaptation and evolution. Below is a review of gene expression and cellular processing of cell surface proteins in eukaryotes. While some events are common to all genes in all eukaryotes, others reflect specific features of the coding sequence.

During transcription the core RNA polymerase II of eukaryotes associates with many additional transcription factors that regulate initiation, elongation and termination in order to produce pre-mRNA (Lee and Young, 2000). In addition, pre-mRNA undergoes 5'-end capping, splicing, and 3'-end polyadenylation before release and export from the nucleus. Genetic and biochemical studies suggest that these steps of transcription are not only highly coordinated, but more often than expected, tightly coupled (Fig. 2; Maniatis and Reed, 2002).

A central player in the coupling of transcriptional events is the C-terminal domain (CTD) of RNA polymerase II. The CTD consists of multiple repeats of the heptapeptide YSPTSPS, and serves as a platform for many interchangeable factors during transcription. This process is partly regulated by the phosphorylation states of CTD. At

initiation, CTD is phosphorylated at serine 5, and recruits the mRNA capping enzyme (Cho et al., 1997; Komarnitsky et al., 2000). During elongation, CTD is phosphorylated at serine 2 by the kinase Ctk1, favoring association with mRNA processing, elongation, and termination factors (Cho et al., 2001). Thus, tethering of various factors to CTD regulates initiation, elongation, splicing and termination (Maniatis and Reed, 2002).

Translational control of eukaryotic cell surface proteins begins when their N-terminal signal sequence is recognized by signal recognition particles to direct co-translational transport of the nascent peptide into the endoplasmic reticulum (reviewed in Walter et al., 1984; de Nobel and Lipke, 1994). The first post-translational modifications occur in the ER. For example, serine/threonine-rich repeats are a target for O-glycosylation. In yeast, the first mannose residues are added to the hydroxyl groups of the residues by protein O-mannosyltransferases (reviewed in Strahl-Bolsinger et al., 1999). Additional mannose residues are added in subsequent steps of the secretory pathway. O-glycosylation chains are much shorter (4-5 mannosyl residues) than N-glycosylation chains (50-100 units), but because of their multiplicity they can account for as much as 80% of the final molecular weight of the protein. O-glycosylation is thought to confer a rigid extended structure that permits the protein to reach further in the extracellular space. At the same time, glycoproteins add an additional layer to the cell surface (de Nobel and Lipke, 1994).

Cell surface proteins can remain attached to the cell in a number of ways. Most often, hydrophobic transmembrane domains permit protein incorporation into the phospholipid bilayer of the plasma membrane. Another set of proteins, like the mouse Thy-1 glycoprotein, attach to the plasma membrane via a glycosyl-phosphatidyl inositol

(GPI) anchor (Low and Kincade, 1985; Tse et al., 1985). Such proteins have a characteristic sequence at their C-terminus that is recognized and replaced with a GPI moiety in the ER. GPI-anchoring permits an additional level of regulation as the enzyme phospholipase C cleaves phosphatidyl-inositol. In yeast, GPI anchors can be cleaved to allow attachment of the protein C-terminus to the cell wall glucan polymer, thus facilitating protein exposure at the surface beyond the cell wall (de Nobel and Lipke, 1994).

REPETITIVE SEQUENCES AFFECT TRANSCRIPTION, REPLICATION, AND RECOMBINATION

Evidence in the literature suggests that repeated DNA sequences are prone to recombination because of aberrant DNA topology during transcription or replication. The following examples illustrate that mispairing can occur among repetitive sequences, and can generate genetic and functional variability.

Recombinogenic replication of repetitive sequences

Repetitive DNA sequences can misalign during replication by pairing out of register with another repeat. Subsequent DNA repair of these structures can lead to deletion or expansion of repetitive units in *E.coli* and yeast. In this sense, the resulting rearrangements are referred to as "encoded errors" (Lovett, 2004). The repetitive units can be trinucleotides, microsatellites (less than 60 nucleotides), or minisatellites (several hundred to several thousand nucleotides) (Kokoska et al., 1998). The model that best

accounts for the recombination of repetitive units is replication slippage, where misalignment leads to a DNA break, and is repaired by single-strand annealing. This model is discussed below in light of replication and recombination associated genes with well-defined functions.

Rearrangement of repetitive DNA sequences does not occur by homologous recombination, because it is independent of recA. In *E. coli* and other bacteria, recA catalyzes ATP-dependent strand exchange between homologous regions of DNA. For this reason recA independent events are thought to occur by other strand exchange pathways. RecA is evolutionarily conserved--- its yeast counterpart is Rad51 whereas mammals have seven recA-like genes. Several experimental systems identified recA-independent deletion between repeats. Dianov et al. used insertion of direct repeats within the *tet* marker on a plasmid, and selected for deletion recombinants by tetracycline resistance in Rec^+ and Rec^-cells (1991). Surprisingly, $recA^-$ mutant strains generate Tet^+ colonies, suggesting the presence of a recombination mechanism between tandem repeats that is different from homologous recombination. RecA-independent recombination also occurs between long tandem repeats on the chromosome (Lovett et al., 1993). The study further points out elements of crossover in the deletion products in a ruvA mutant that inhibits branch migration. This observation suggests that replicating sister strands can base pair, and offers a mechanism for expansion of repetitive sequences during replication.

Misalignment (slippage) of the repetitive sequences during replication can initiate repair that leads to loss or gain of repeats (Bzymek and Lovett, 2001). Suppressors and enhancers of RecA mutants provide clues about the mechanism of recA independent

recombination. In bacteria for example, two 3' exonucleases, DnaQ and ExoI, substantially increase recA-independent deletions, suggesting that the 3' end of the nascent strand may facilitate misalignments (Saveson and Lovett, 1997; Bzymek and Lovett, 1999). It is possible that misalignment involves the 5' ends of Okazaki fragments as well: in yeast, *rad27* mutants that fail to process Okazaki 5' ends also enhance recombination between repeats (Tishkoff et al, 1997; Kokoska et al., 1998).

Strong suppressors of recombination between repeats have been also identified. Mutations in *RAD50*, *RAD52*, and *RAD59* reduce recombination between repeats. In a system of inverted repeats of the *ade2* gene, deletion of *RAD52* reduces recombination frequency by 3,000 fold, whereas deletion of *RAD51* leads to a decrease of only 4 fold (Rattray and Symington, 1994). In a *rad51* mutant strain with the same inverted-repeat system, deletion of *RAD59* also results in over 1,000 reduction in the recombination frequency (Bai and Symington, 1996). In addition to their genetic interaction, *RAD52* and *RAD59* share significant amino acid similarity, further suggesting a common function.

Biochemical analyses indicate that Rad52 and Rad50 function in the generation and annealing of single stranded DNA. Following a double-strand break, single stranded regions can be observed on each side of the break. In *rad50* mutants, single stranded DNA is generated at a slower rate, whereas in *rad52* mutant single-stranded DNA is generated faster than in wild type (Sugawara and Haber, 1991). An in vitro system indicates that Rad52 binds single and double stranded DNA, and strongly promotes DNA annealing (Mortensen et al., 1996).

Taken together, the genetic and biochemical data support a replication slippage model: misalignment of repetitive sequences creates aberrant intermediates during replication and leads to double-strand breaks. Genes of the Rad52 epistatic group are required for making ssDNA on each end of the break, and for annealing (Bzymek and Lovett, 2001; Paques and Haber, 1999).

Recombination of repetitive sequences can provide adaptive advantages. For example if there is selection for function encoded by a gene that is flanked by repeats, replication slippage would allow the formation of an amplified tandem array that would permit higher overall expression of that gene's product (Lovett, 2004). On the other hand, homologous recombination between repeats flanking a gene could result in the loss of that gene's function (Paques and Haber, 1999).

A recent study on the recombination of repetitive sequences in the *Saccharomyces cerevisiae FLO1* gene indicates that the replication slippage model applies to naturally occurring genes with repeats (Verstrepen et al., 2005): deletion of the recA homolog *RAD51* has no effect on recombination frequency; *rad27* mutation increases frequency of recombination almost 40 fold; deletion of *RAD52*, *RAD50*, or *RAD59* significantly decreases recombination. Also consistent with the model, recombination of *FLO1* repeats generates shorter as well as longer than the original alleles.

In humans, expansion of intragenic trinucleotide repeats is associated with a number of diseases, including Huntington's disease, fragile X syndrome, myotonic dystrophy and Friedreich's ataxia. In these mammalian examples, repeats are also likely to expand by replication slippage, but the exact mechanism has not been directly

addressed. Depending on whether the repetitive region falls into an exon or into an intron, its expansion or contraction can manifest deleterious effects at the protein level (huntingin) or at the transcription level (frataxin) or both. Expansions of the trinucleotide repeat GAA•TTC within the first intron of the X25 (frataxin) gene cause Friedreich's ataxia because of a transcription elongation defect (Grabczyk and Usdin, 2000). The GAA•TTC repeats represent a purine•pyrimidine (R•Y) polymer that can form triple helices that can impede replication or transcription (Fig. 3; Mariappan, 1999). These examples illustrate that replication of repetitive sequences generates genetic instability and functional variability in a wide variety of organisms.

Supercoiling and template mispairing during transcription

Transcription is a highly coordinated process. Coincident with the regulatory steps of gene expression are two additional transcription-associated phenomena, DNA supercoiling and DNA:RNA pairing. When the RNA polymerase progresses along the DNA template, it creates positive supercoiling ahead of the complex and negative supercoiling behind it (Lui and Wang, 1987). In topoisomerase mutants, the DNA torsional stress is not restored (Brill and Sternglanz, 1988; Masse et al., 1999a), and the unpaired DNA is available for DNA:RNA base-paring (R loops) that can impede transcription elongation (Fig. 3; Drolet et al., 1994; Masse et al., 1999b). R loops may also arise from transcription through specific sequences in a less supercoiling-dependent manner. At the same time, R-looping and supercoiling seem to reinforce each other (Drolet, 2006).

One example of the interplay between transcription, supercoiling, and genetic instability of repetitive sequences emerged from a study of regulation at the rDNA locus in topoisomerase mutants. The rDNA genes encode ribosomal RNAs, and are arranged in a cluster of tandem repeats. In yeast, the 9.1 kb rDNA cistron is repeated 150-200 times on chromosome XII. Suppression of recombination at that locus requires both *TOP1* and *TOP2*: in the topoisomerase mutants *top1*Δ or *top2-1*(ts), mitotic recombination at the rDNA locus occurs at 50 to 200 fold higher frequency than in Top^+ cells (Christman et al., 1988). Moreover, transcription-dependent supercoiling occurs in yeast topoisomerase mutants (Brill and Sternglanz, 1988). Reduced topoisomerase 1 and 2 levels also lead to excision of rDNA into extrachromosomal circles in yeast (Kim and Wang, 1989). The circles can reintegrate resulting in a dynamic process that prevents loss of rDNA copy total number. The observation that rDNA transcription in topoisomerase mutants is affected only in the chromosomal copies, and not the extrachromosomal ones (Brill and Strenglanz, 1988), suggests that rDNA excision may be a way to escape supercoiling-induced reduction of transcription in topoisomerase mutants (Kim and Wang, 1989).

Transcription-associated recombination in prokaryotes

One of the first examples of transcription-associated recombination was documented in $recA^-$ *E. coli* cells (Ikeda and Matsumoto, 1979). Recombination between consecutive loci on pairs of co-infecting λ phage mutants was observed only after transcription through the loci was induced. Moreover, when the assay was conducted in a bacterial *rho* mutant that allows aberrant transcription past a termination site, the

recombinogenic regions also extended beyond the termination site. It was proposed that recombination resulted from an RNA:DNA intermediate, and the process was dubbed the Rpo pathway, since it appeared to be RNA polymerase-dependent.

Transcription-associated recombination in mammalian cells

In mammalian B cells, immunoglobulin class switching results from the transcription-induced recombination between switch regions composed of highly repetitive sequences that precede the corresponding constant-region gene. The total length of a switch region varies from 1 to 12 kb, and the length of a single repeat can be from 10 to 80 bp. These regions usually have a G-rich non-template strand. Class switch recombination occurs in a transcription-dependent manner possibly via R-loop formation. The nascent mRNA pairs with the template strand, leaving G-rich regions of the non-template DNA single stranded and prone to nicks. A B-cell specific activation-induced cytidine deaminase (AID) is suggested to deaminate deoxycytidine to deoxyuridine in single-stranded DNA, and to initiate repair and recombination (Chaudhuri and Alt, 2004).

Class-switch recombination (CSR), like VDJ recombination and somatic hypermutation, adds to the diversity of antibodies. The five classes for immunoglobulin heavy chain (IgH) constant regions (C_H) are encoded by exons in the order μ, δ, γ, ε, and α, downstream of the VDJ regions (Fig. 4). Immunoglobulins initially express Cμ but can switch to any of the other classes. Although the detailed functions of each class are not clearly understood, it seems that IgM and IgG activate the complement system, antibodies of the IgA class are found in secretions, and IgE antibodies interact with basophils and mast cells in the allergic response.

Class-switch recombination requires transcription through the corresponding heavy chain locus. This can be observed by treatment of B cells with different mitogens or lymphokines. Bacterial lipopolysaccharyde (LPS) is a B cell mitogen that can confer IgM to IgG2b switch. Lutzker et al. show that the γ2b locus is transcribed in the presence of LPS before γ2b class switching (1988). Moreover, addition of interleukin-4 inhibits the effect of LPS and alters the pattern of expression. γ2b transcripts are not detected, and there is no recombination to the IgG2b locus. In the same vein, mice that lack sequences 5' to the Sγ1 region fail to respond to interleukin-4 and cannot undergo a switch to the γ1 locus, a defect that leads to agammaglobulinemia (Jung et al., 1993).

Transcription of switch regions leads to R-loop formation. In vitro transcription of the murine switch region Sα from a plasmid relaxes supercoiling as visualized by gel electrophoresis (Reaban and Griffin, 1990). The supercoil-relaxing structure is an RNA:DNA hybrid since RNase H treatment restores supercoiling. The presence of ssDNA is manifested by linearization of the structure after treatment with mung bean nuclease. In contrast, the structure is insensitive to mung bean nuclease, if first treated with RNase H, suggesting that the two DNA strands reanneal after removal of RNA.

Switch regions, telomeres, rDNA, and minisatellites are the four cases of G-rich chromosomal domains. One interesting possibility is that G4 DNA (a stack of rings of four guanines) may have a role in gene regulation and stability. Duquette et al. define G-loops (Fig. 3) as cotranscriptionally formed structures with G4 DNA on the singled non-template strand, and a stable DNA:RNA hybrid on the other strand (2004). G-loops were visualized by electron microscopy after expression of mouse switch regions from plasmid constructs in vitro and in *E. coli* (Duquette et al., 2004). The presence of G4 and

DNA:RNA structures is supported by several experiments. Anti-digoxygenin/gold beads decorate the loop structures from digoxygenin-UTP transcription reactions. The decoration is abolished after treatment of the reaction product with RNase H and not with RNase A. In addition, treatment of the transcription reaction product with GQN1, an endonuclease that specifically cuts ssDNA 5' of G4 DNA, results in cleaved loops. Thus, transcription can also induce aberrant DNA topology.

THO/TREX complexes

The yeast THO complex consists of four tightly bound proteins: Hpr1 (90kDa), Tho2 (160kDa), Thp2 (40kDa), and Mft1 (60kDa) (Chavez et al., 2000). Two of these (Hpr1 and Tho2) are conserved from yeast to humans (Reed and Cheng, 2005). The THO complex was first identified in yeast (Chavez et al., 2000), and subsequently in human cells (Strasser et al., 2002) and *Drosophila* (Rehwinkel et al., 2004). The THO complex is thought to be part of the larger TREX complex, which is evolutionarily conserved in eukaryotes, and functions in transcription elongation and export. TREX not only includes the multi-subunit THO complex, but also the mRNA export proteins Sub2 (UAP56 in humans) and Yra1 (Aly in humans). Tex1 and proteins of the serine/arginine-rich (SR) family of splicing factors can also associate with the TREX complex (Fig. 5; Hurt et al., 2004; Custodio et al., 2004).

In yeast, chromatin immunoprecipitation experiments indicate that the THO complex is recruited to actively transcribed genes (Strasser et al., 2002; Abruzzi et al., 2004; Kim et al., 2004). A chromatin-immunoprecipitation experiment followed by RNase treatment suggests that Hpr1 binds DNA and not RNA (Abruzzi et al., 2004).

Tho2 can bind both RNA and dsDNA in vitro (Jimeno et al., 2002). Biochemical studies using natural templates have implicated the THO complex in recruiting the mRNA export proteins Sub2 and Yra1 to the mRNA (Zenklusen et al., 2002). The TREX complex can further recruit additional proteins, such as the RNA-binding proteins Gbr2 and Hrb1, which are related to the metazoan SR protein family of splicing factors.

A major distinction between the TREX complex in yeast and in mammals is that in yeast the TREX complex is recruited to active genes and travels the whole length of the gene with RNA polymerase II (Strasser et al., 2002; Abruzzi et al., 2004; Kim et al., 2004), whereas in mammals the TREX complex seems to be recruited to mRNA during splicing (Fig. 5). In mammalian cells Aly and UAP56 co-localize with SR proteins and other spliceosome components to a transcript that has been spliced, and not to the equivalent transcript with a mutated splicing site (Custodio et al., 2004). A proteomic analysis indicates that all TREX components are present in the human spliceosome (Zhou et al., 2002). Moreover, human THO components associate with spliced mRNA rather than with unspliced pre-mRNA (Masuda et al., 2005).

Identification and characterization of the THO complex began in yeast with the use of an artificial construct, *leu2-k::URA3-ADE2::leu2-k*, designed to screen for regulators of recombination between repetitive sequences in *S. cerevisiae* (Aguilera and Klein, 1990). A sequence of the *LEU2* gene flanks the *URA3* and *ADE2* markers on each side. This system can select as well as screen for recombinants of the *leu2-k* repeats that have recombined out the *URA3-ADE2* cassette: Ura^- Ade^- colonies are red and grow on FOA, but not on media that lack uracil (Fig. 4).

The first yeast THO complex subunit, *HPR1*, was identified because the *hpr1-1* mutant leads to 50-100 fold hyper-recombination in the artificial *leu2* duplication (Aguilera and Klein, 1990). Subsequent experiments with inducible promoters, ribozyme-containing ORFs and other artificial constructs demonstrated that the observed hyper-recombination phenotype of *hpr1* mutants is promoter-independent but transcription- and transcript-dependent (Chavez and Aguilera, 1997; Huertas and Aguilera, 2003). All four components of the yeast THO complex have been identified. Rescue of the temperature sensitivity and hyper-recombination phenotype of *hpr1*Δ by high-copy suppression identified the *THO2* gene (Piruat and Aguilera, 1998). The other two components, Thp2 and Mft1, were identified by co-immunoprecipitation and mass spectrometry (Chavez et al., 2000).

A number of phenotypes have been associated with Tho$^-$ mutants. In yeast, all single Tho$^-$ mutants are characterized by hyper-recombination of the artificial repeats construct, by 78 to 94% reduced mRNA levels of the *E. coli lacZ* gene when promoted by the inducible *GAL1* promoter, and by increased nuclear accumulation of poly(A)$^+$ after a 2-hr heat-shock (Chavez et al., 2000, Strasser et al., 2002). In addition, the *hpr1*Δ mutant has been associated with shorter life span possibly because of the 3-fold higher rate of recombination within the rDNA locus (Merker and Klein, 2002). This mutant also displays replication fork collapse along a *GAL1-lacZ* construct at inducible conditions, as well as two times more Rad52 foci than a Tho$^+$ strain, indicative of a transcription-dependent replication defect (Wellinger et al., 2006).

It has been suggested that the THO complex functions at the interface of mRNA processing and DNA topology during transcription. On one hand, mutations in several

transcriptional genes suppress the *hpr1*Δ temperature-sensitivity and hyper-recombination phenotypes in yeast: *SOH1*, *RPB2*, *SUA7* (Fan et al., 1996). Soh1 is a subunit of the RNA polymerase II mediator complex; Rpb2 is the second largest subunit of RNA polymerase II; Sua7 is the general transcription factor TFIIB. On the other hand, the yeast double mutant *top1*Δ *hpr1*Δ is inviable (Aguilera and Klein, 1990; Sadoff et al., 1995). Synthetic lethality interaction suggests a redundant function between Hpr1 and Top1. Moreover, Top1 and Hpr1 have a sequence similarity at the protein level (Aguilera and Klein, 1990). Thus, part of the role of the THO complex in transcription elongation may be to relieve transcription-induced supercoiling, which may be especially important for genes with sequences prone to form DNA:RNA hybrids between the template DNA and the nascent RNA (R loops; Fig. 3).

It is currently unknown whether mammalian or *Drosophila* Tho$^-$ mutants share the hyper-recombination phenotype of their yeast counterpart. In both *Drosophila* and human Tho$^-$ cells, poly(A)$^+$ accumulation has been observed (Rehwinkel et al., 2004; Guo et al., 2005). Consistent with the idea of coupling between RNA processing and genomic stability, a recent study shows that deletion of an SR protein, the alternative splicing factor ASF/SF2, promotes recombination via R-loop formation in a chicken B-cell line (Li and Manley, 2005).

All these studies imply a role for the THO/TREX complex in transcription elongation, mRNA export, and recombination. The hyper-recombination phenotype of Tho$^-$ mutants especially between repetitive sequences is suggestive of the interplay between transcription, RNA processing, and genetic stability. It is possible that the THO complex provides a stabilizing force that prevents transcription-associated DNA:RNA

hybrid formation and subsequent alterations in DNA structure that induce the recombination system.

UNANSWERED QUESTIONS

The above-mentioned examples suggest a role of repetitive sequences on replication, transcription elongation and recombination. In *Saccharomyces cerevisiae*, there are 29 genes with long (≥40 nt) tandem repeats (Verstrepen et al., 2005). It is estimated that 5% of human genes also have tandem repeats (O'Dushlaine et al., 2005). The genome-wide screen that was conducted as part of this thesis was designed to address the important question: Do genes with natural internal repeats require specific factors for function? As a reporter in the screen, we used expression of the *S. cerevisiae* cell surface adhesin *FLO11* from a heterologous constitutive promoter in order to identify exclusively post-transcriptional regulators. The screen was successful in identifying genes involved in events subsequent to the initiation of transcription. But, it took a surprising turn: the *FLO11* message was considerably reduced in a number of mutants (*mft1*, *thp2*). This result suggested regulation of *FLO11* at downstream transcriptional events.

Another puzzle that remains unresolved is: what genes require the THO complex for expression and function? The extensive research on the yeast THO complex has been carried out exclusively using artificial constructs with repeats or *GAL1* promoted bacterial *lacZ* gene, and has lead to the arguable conclusion that the THO complex is required for efficient transcription of long or GC-rich templates (Chavez et al., 2001). It has been shown that human Tho$^-$ cells also cannot transcribe *lacZ* properly, while GFP is

expressed at wild-type levels in the same experimental system (Li et al., 2005). But what are the endogenous genes that require THO complex function? Expression of several heat shock protein (*HSP*) genes requires functional THO complex (*SSA1* in yeast, Strasser et al., 2002; *HSP70* and *HSP83* in Drosophila, Rehwinkel et al., 2004), however, this observation has not been associated with a phenotype or mechanism. In *Drosophila melanogaster*, loss of THO complex function results in only minor differences in transcription profiles as revealed by whole genome arrays (Rehwinkel et al., 2004).

This thesis identifies the in vivo function of the THO complex for certain yeast genes with long tandem repeats. The initial identification of these genes resulted from a genome-wide scan for promoter-independent mutations that affect *FLO11* function (Chapter 2). *FLO11* is a representative of yeast genes that contain many internal repeats. I have shown that in the absence of the THO complex, the repetitive sequences affect transcription elongation in a number of genes with internal repeats (Chapter 3). Results from the genome-wide screen and the THO complex analysis, as well as possible future directions are further discussed in Chapter 4. Chapter 5 includes several appendices to Chapters 2 and 3.

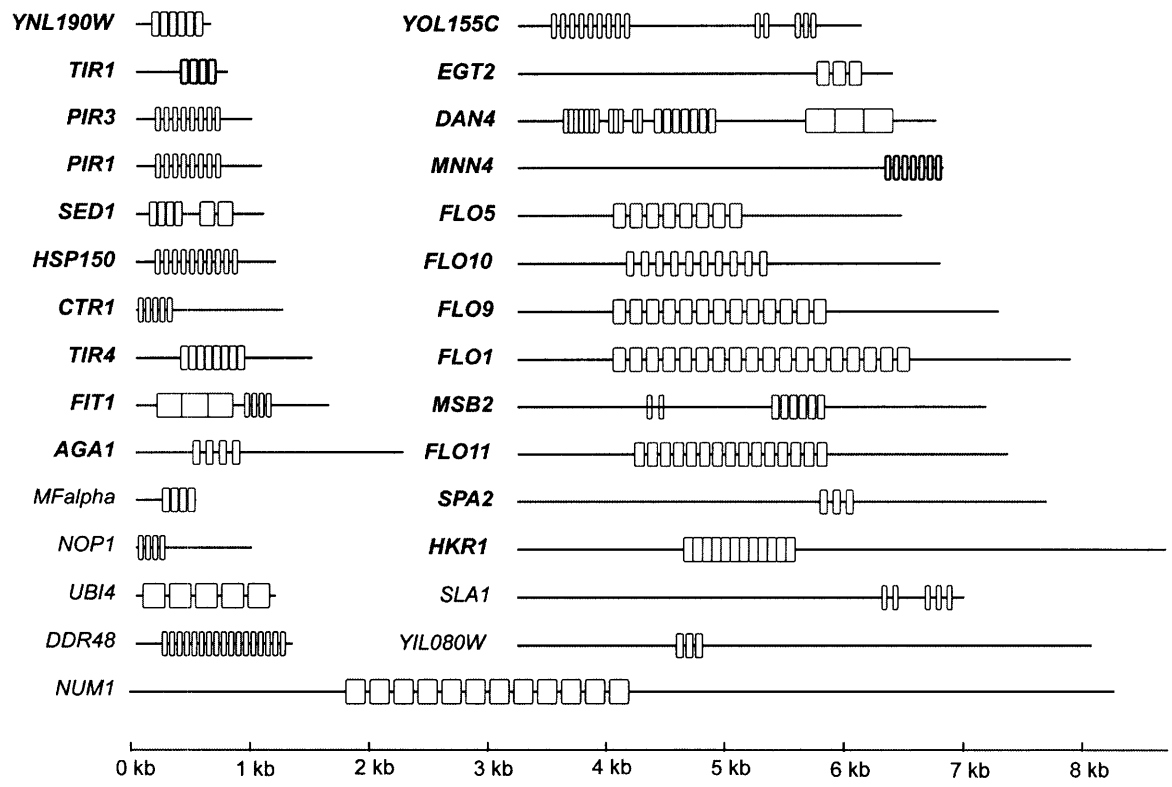


Figure 1: *S. cerevisiae* genes containing long (≥ 40nt) conserved intragenic repeats (Verstrepen et al., 2005).

Genes encoding cell surface proteins are in bold.

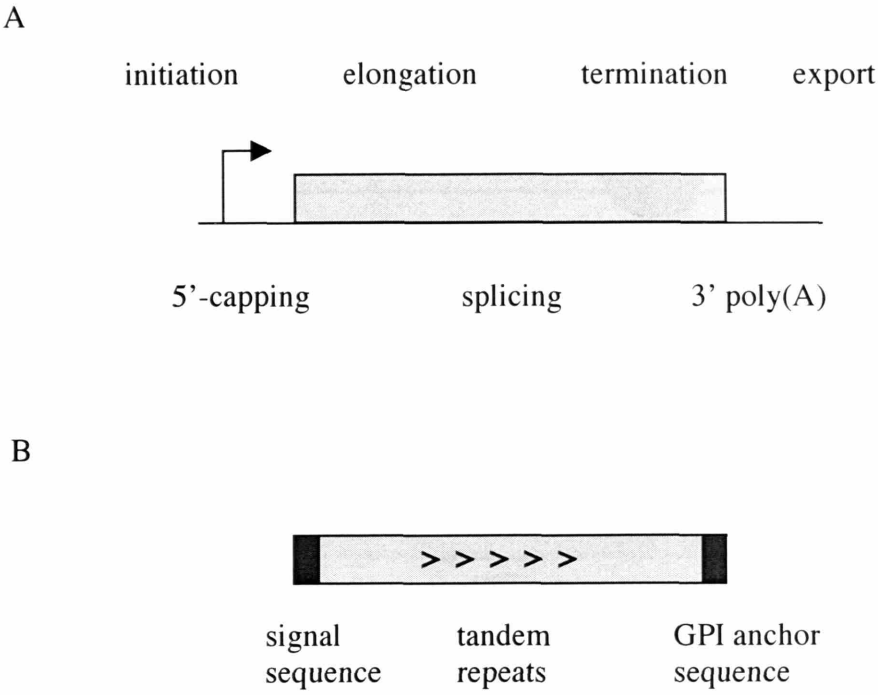


Figure 2: Regulation of eukaryotic genes that encode cell surface proteins

A. Transcription is a highly coordinated process of initiation, elongation, termination and mRNA export, and is accompanied by capping, splicing and 3' polyadenylation;

B. Intragenic motifs specify post-transcriptional regulatory steps. An N-terminal signal sequence is bound by the signal recognition particle for co-translational transport into the endoplasmic reticulum. Serine/threonine-rich tandem repeats present sites for O-glycosylation. A C-terminal sequence is replaced by a GPI anchor for attachment at the cell surface.

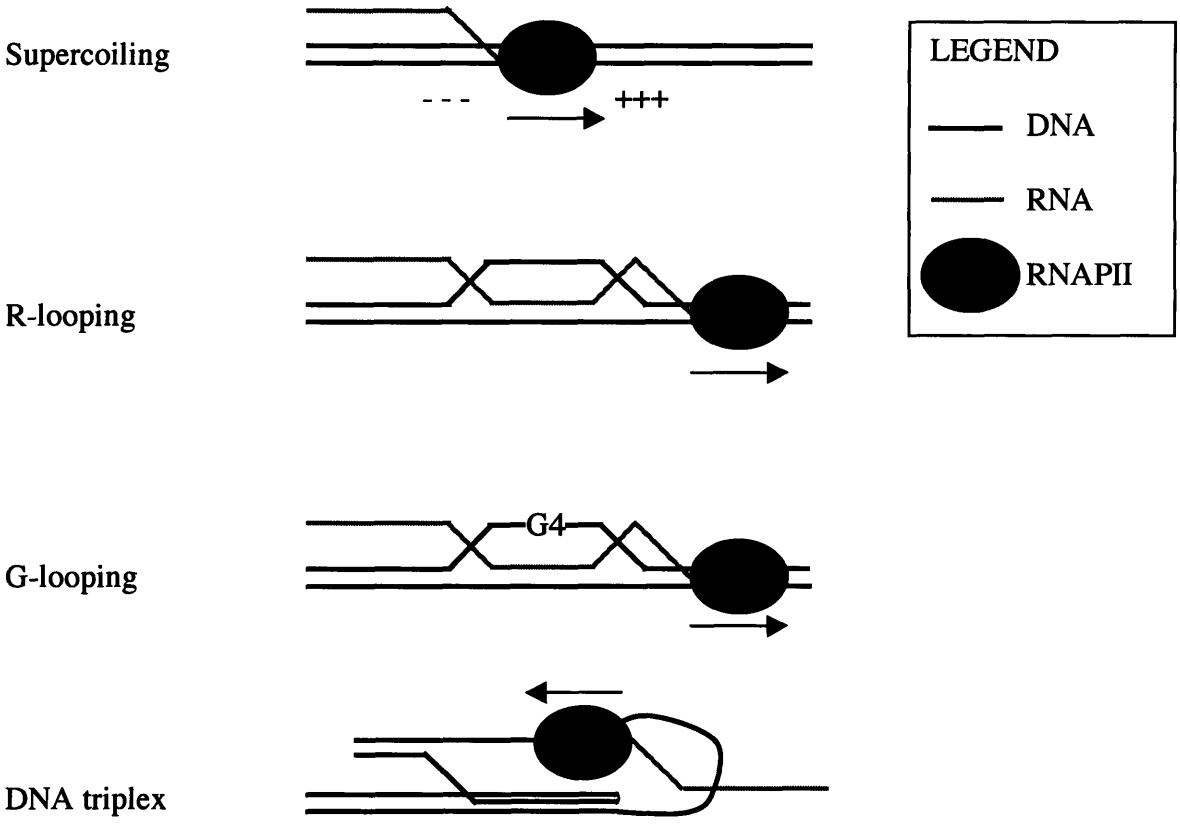


Figure 3: RNA polymerase impeding structures.

Movement of RNA polymerase along the template creates positive supercoiling ahead of the complex, and negative supercoiling behind it.

If not resolved, negative supercoiling may favor base-pairing between the template strand and the nascent RNA, resulting in DNA:RNA hybrids (R loops). The displaced region of the non-template regions is single stranded. Such structures can impede progression of subsequent RNA polymerase units.

If the single stranded fragment of the non-template strand is G-rich, it can form stable G-quadruplexes, resulting in G loops (G4).

Triplet repeats, such as the GAA•TTC repeat in the *frataxin* gene, can form a DNA triplex that stalls the polymerase during transcription.

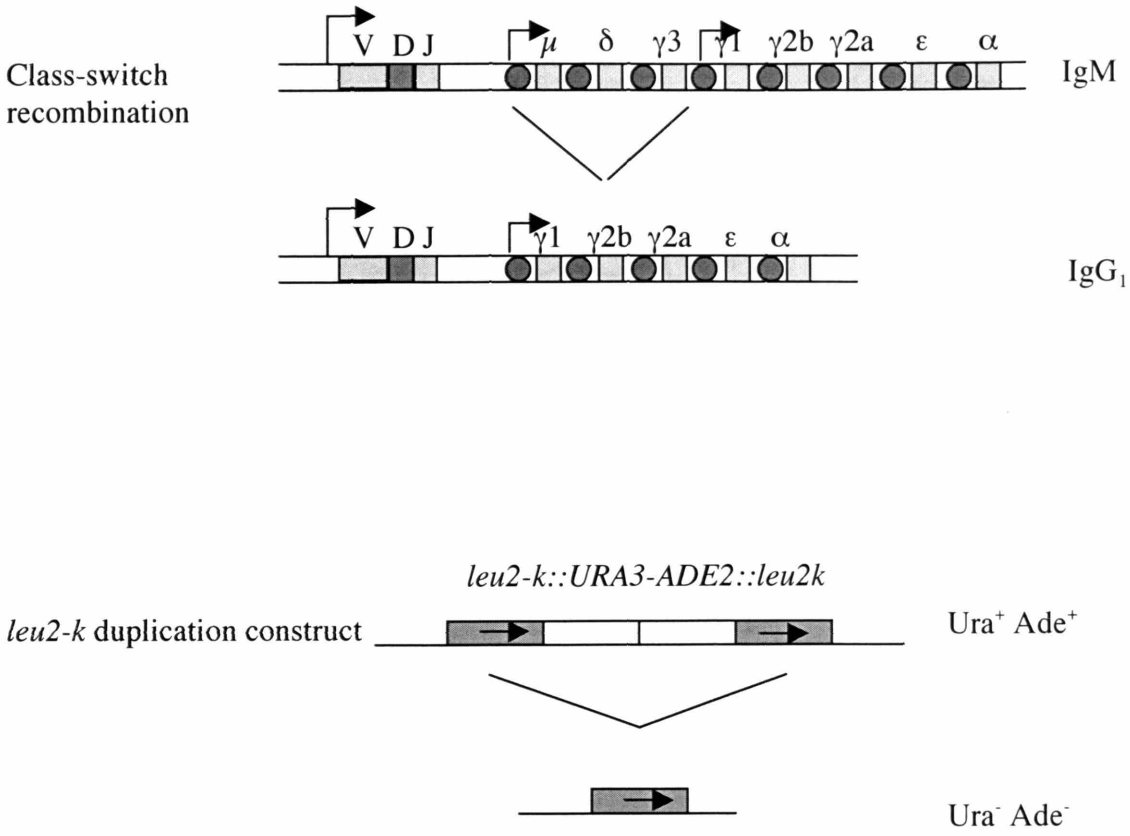


Figure 4: Transcription-dependent recombinogenic loci.

Class switch regions (circles) precede the corresponding heavy chain locus (square, μ, δ, γ3, γ1, γ2b, γ2a, ε, or α) (adapted from Chaudhuri and Alt, 2004).

The artificial construct of *LEU2* sequence flanking the markers *URA3* and *ADE2* was used to identify and analyze the THO complex (after Aguilera and Klein, 1990).

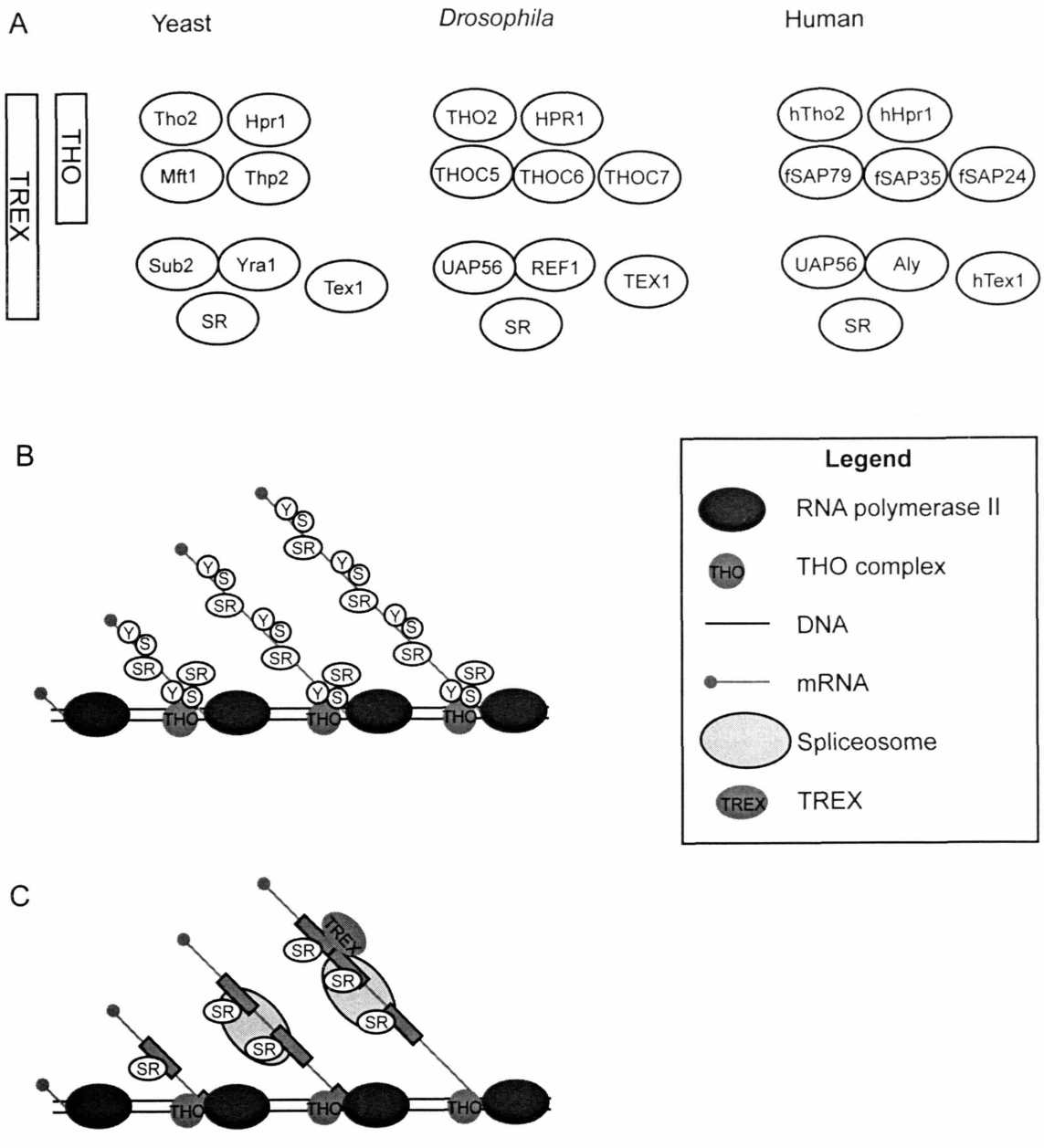


Figure 5: THO/TREX complex.

A. Subunits in *S. cerevisiae*, *D. melanogaster*, and *H. sapiens*.
B. Recruitment of the THO complex, Sub2, Yra1, and SR proteins during transcription in yeast.
C. Recruitment of THO/TREX complex during transcription and splicing in metazoans (adapted from Reed and Cheng, 2005).

REFERENCES

Abruzzi, K., Lacadie, S., and Rosbash, M. (2004). Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. EMBO J 23, 2620-2631.

Aguilera, A., and Klein, H. (1990). HPR1, a novel yeast gene that prevents intrachromosomal excision recombination, shows carboxy-terminal homology to the Saccharomyces cerevisiae TOP1 gene. Mol Cell Biol 10, 1439-1451.

Bai, Y., and Symington, L. S. (1996). A Rad52 homolog is required for RAD51-independent mitotic recombination in Saccharomyces cerevisiae. Genes Dev 10, 2025-2037.

Bentley, D. L., and Groudine, M. (1986). A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. Nature 321, 702-706.

Brill, S. J., and Sternglanz, R. (1988). Transcription-dependent DNA supercoiling in yeast DNA topoisomerase mutants. Cell 54, 403-411.

Bzymek, M., Saveson, C. J., Feschenko, V. V., and Lovett, S. T. (1999). Slipped misalignment mechanisms of deletion formation: in vivo susceptibility to nucleases. J Bacteriol 181, 477-482.

Bzymek, M., and Lovett, S. T. (2001). Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. Proc Natl Acad Sci U S A 98, 8319-8325.

Chaudhuri, J., and Alt, F. W. (2004). Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat Rev Immunol 4, 541-552.

Chavez, S., and Aguilera, A. (1997). The yeast HPR1 gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. Genes Dev 11, 3459-3470.

Chavez, S., Beilharz, T., Rondon, A., Erdjument-Bromage, H., Tempst, P., Svejstrup, J., Lithgow, T., and Aguilera, A. (2000). A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. EMBO J 19, 5824-5834.

Chavez, S., Garcia-Rubio, M., Prado, F., and Aguilera, A. (2001). Hpr1 is preferentially required for transcription of either long or G+C-rich DNA sequences in Saccharomyces cerevisiae. Mol Cell Biol 21, 7054-7064.

Christman, M. F., Dietrich, F. S., and Fink, G. R. (1988). Mitotic recombination in the rDNA of S. cerevisiae is suppressed by the combined action of DNA topoisomerases I and II. Cell 55, 413-425.

Cho, E. J., Takagi, T., Moore, C. R., and Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev 11, 3319-3326.

Cho, E. J., Kobor, M. S., Kim, M., Greenblatt, J., and Buratowski, S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. Genes Dev 15, 3319-3329.

Custodio, N., Carvalho, C., Condado, I., Antoniou, M., Blencowe, B. J., and Carmo-Fonseca, M. (2004). In vivo recruitment of exon junction complex proteins to transcription sites in mammalian cell nuclei. Rna 10, 622-633.

de Nobel, H., and Lipke, P. N. (1994). Is there a role for GPIs in yeast cell-wall assembly? Trends Cell Biol 4, 42-45.

Dianov, G. L., Kuzminov, A. V., Mazin, A. V., and Salganik, R. I. (1991). Molecular mechanisms of deletion formation in Escherichia coli plasmids. I. Deletion formation mediated by long direct repeats. Mol Gen Genet 228, 153-159.

Drolet, M., Bi, X., and Liu, L. F. (1994). Hypernegative supercoiling of the DNA template during transcription elongation in vitro. J Biol Chem 269, 2068-2074.

Drolet, M. (2006). Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. Mol Microbiol 59, 723-730.

Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., and Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes Dev 18, 1618-1629.

Fan, H. Y., Cheng, K. K., and Klein, H. L. (1996). Mutations in the RNA polymerase II transcription machinery suppress the hyperrecombination mutant hpr1 delta of Saccharomyces cerevisiae. Genetics 142, 749-759.

Fan, H. Y., Merker, R. J., and Klein, H. L. (2001). High-copy-number expression of Sub2p, a member of the RNA helicase superfamily, suppresses hpr1-mediated genomic instability. Mol Cell Biol 21, 5459-5470.

George, R., Beddoe, T., Landl, K., and Lithgow, T. (1998). The yeast nascent polypeptide-associated complex initiates protein targeting to mitochondria in vivo. Proc Natl Acad Sci U S A 95, 2296-2301.

Grabczyk, E., and Usdin, K. (2000). The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. Nucleic Acids Res 28, 2815-2822.

Guo, S., Hakimi, M. A., Baillat, D., Chen, X., Farber, M. J., Klein-Szanto, A. J., Cooch, N. S., Godwin, A. K., and Shiekhattar, R. (2005). Linking transcriptional elongation and messenger RNA export to metastatic breast cancers. Cancer Res 65, 3011-3016.

Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell 12, 711-721.

Hurt, E., Luo, M. J., Rother, S., Reed, R., and Strasser, K. (2004). Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex. Proc Natl Acad Sci U S A 101, 1858-1862.

Ikeda, H., Matsumoto, T. (1979). Transcription promotes recA-independent recombination mediated by DNA-dependent RNA polymerase in Escherichia coli. Proc Natl Acad Sci U S A 76, 4571-4575.

Jung, S., Rajewsky, K., and Radbruch, A. (1993). Shutdown of class switch recombination by deletion of a switch region control element. Science 259, 984-987.

Kim, R. A., and Wang, J. C. (1989). A subthreshold level of DNA topoisomerases leads to the excision of yeast rDNA as extrachromosomal rings. Cell 57, 975-985.

Kim, M., Ahn, S. H., Krogan, N. J., Greenblatt, J. F., and Buratowski, S. (2004). Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. Embo J 23, 354-364.

Kokoska, R. J., Stefanovic, L., Tran, H. T., Resnick, M. A., Gordenin, D. A., and Petes, T. D. (1998). Destabilization of yeast micro- and minisatellite DNA sequences by mutations affecting a nuclease involved in Okazaki fragment processing (rad27) and DNA polymerase delta (pol3-t). Mol Cell Biol 18, 2779-2788.

Komarnitsky, P., Cho, E. J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev 14, 2452-2460.

Lee, T. I., and Young, R. A. (2000). Transcription of eukaryotic protein-coding genes. Annu Rev Genet 34, 77-137.

Li, X., and Manley, J. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell 122, 365-378.

Liu, L. F., and Wang, J. C. (1987). Supercoiling of the DNA template during transcription. Proc Natl Acad Sci U S A 84, 7024-7027.

Lovett, S. T., Drapkin, P. T., Sutera, V. A., Jr., and Gluckman-Peskind, T. J. (1993). A sister-strand exchange mechanism for recA-independent deletion of repeated DNA sequences in Escherichia coli. Genetics 135, 631-642.

Lovett, S. T. (2004). Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. Mol Microbiol 52, 1243-1253.

Low, M. G., and Kincade, P. W. (1985). Phosphatidylinositol is the membrane-anchoring domain of the Thy-1 glycoprotein. Nature 318, 62-64.

Lutzker, S., Rothman, P., Pollock, R., Coffman, R., and Alt, F. W. (1988). Mitogen- and IL-4-regulated expression of germ-line Ig gamma 2b transcripts: evidence for directed heavy chain class switching. Cell 53, 177-184.

Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. Nature 416, 499-506.

Mariappan, S. V., Catasti, P., Silks, L. A., 3rd, Bradbury, E. M., and Gupta, G. (1999). The high-resolution structure of the triplex formed by the GAA/TTC triplet repeat associated with Friedreich's ataxia. J Mol Biol 285, 2035-2052.

Masse, E., and Drolet, M. (1999). Relaxation of transcription-induced negative supercoiling is an essential function of Escherichia coli DNA topoisomerase I. J Biol Chem 274, 16654-16658.

Masse, E., and Drolet, M. (1999). Escherichia coli DNA topoisomerase I inhibits R-loop formation by relaxing transcription-induced negative supercoiling. J Biol Chem 274, 16659-16664.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. Genes Dev 19, 1512-1517.

Merker, R. J., and Klein, H. L. (2002). hpr1Delta affects ribosomal DNA recombination and cell life span in Saccharomyces cerevisiae. Mol Cell Biol 22, 421-429.

Mortensen, U. H., Bendixen, C., Sunjevaric, I., and Rothstein, R. (1996). DNA strand annealing is promoted by the yeast Rad52 protein. Proc Natl Acad Sci U S A 93, 10729-10734.

O'Dushlaine, C. T., Edwards, R. J., Park, S. D., and Shields, D. C. (2005). Tandem repeat copy-number variation in protein-coding regions of human genes. Genome Biol 6, R69.

Paques, F., and Haber, J. E. (1999). Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Microbiol Mol Biol Rev 63, 349-404.

Piruat, J. I., and Aguilera, A. (1998). A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. Embo J 17, 4859-4872.

Prado, F., Piruat, J. I., and Aguilera, A. (1997). Recombination between DNA repeats in yeast hpr1delta cells is linked to transcription elongation. Embo J 16, 2826-2835.

Rattray, A. J., and Symington, L. S. (1994). Use of a chromosomal inverted repeat to demonstrate that the RAD51 and RAD52 genes of Saccharomyces cerevisiae have different roles in mitotic recombination. Genetics 138, 587-595.

Reaban, M. E., and Griffin, J. A. (1990). Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. Nature 348, 342-344.

Reed, R., and Cheng, H. (2005). TREX, SR proteins and export of mRNA. Curr Opin Cell Biol 17, 269-273.

Rehwinkel, J., Herold, A., Gari, K., Kocher, T., Rode, M., Ciccarelli, F., Wilm, M., and Izaurralde, E. (2004). Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster. Nat Struct Mol Biol 11, 558-566.

Sadoff, B., Heath-Pagliuso, S., Castano, I., Zhu, Y., Kieff, F., and Christman, M. (1995). Isolation of mutants of Saccharomyces cerevisiae requiring DNA topoisomerase I. Genetics 141, 465-479.

Saveson, C. J., and Lovett, S. T. (1997). Enhanced deletion formation by aberrant DNA replication in Escherichia coli. Genetics 146, 457-470.

Strahl-Bolsinger, S., Gentzsch, M., and Tanner, W. (1999). Protein O-mannosylation. Biochim Biophys Acta 1426, 297-307.

Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A. G., Aguilera, A., Struhl, K., Reed, R., and Hurt, E. (2002). TREX is a conserved complex coupling transcription with messenger RNA export. Nature 417, 304-308.

Sugawara, N., and Haber, J. E. (1992). Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. Mol Cell Biol 12, 563-575.

Tse, A. G., Barclay, A. N., Watts, A., and Williams, A. F. (1985). A glycophospholipid tail at the carboxyl terminus of the Thy-1 glycoprotein of neurons and thymocytes. Science 230, 1003-1008.

Verstrepen, K., Jansen, A., Lewitter, F., and Fink, G. (2005). Intragenic tandem repeats generate functional variability. Nat Genet 37, 986-990.

Walter, P., Gilmore, R., and Blobel, G. (1984). Protein translocation across the endoplasmic reticulum. Cell 38, 5-8.

Wellinger, R. E., Prado, F., and Aguilera, A. (2006). Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. Mol Cell Biol 26, 3327-3334.

Zenklusen, D., Vinciguerra, P., Wyss, J. C., and Stutz, F. (2002). Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. Mol Cell Biol 22, 8241-8253.

Zhou, Z., Licklider, L. J., Gygi, S. P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. Nature 419, 182-185.

Chapter2: A genome-wide screen for promoter-independent regulators of *FLO11*

ABSTRACT

I screened through the library of all viable yeast single-gene deletions to identify promoter-independent regulators of *FLO11*. *FLO11* is a *Saccharomyces cerevisiae* gene with many tandem repeats and is required for adhesion to solid agar. I used a construct that expresses *FLO11* from a heterologous constitutive promoter in order to avoid mutants that regulate transcription initiation, and screened for mutants with reduced adherence to agar. I identified 44 mutants with reduced adherence and these correspond to genes that fall into several functional groups: cell wall biogenesis, cytoskeleton and polarity regulation, protein sorting, mRNA processing, and chromatin remodeling. Physical and genetic interactions between the genes form a tightly linked set of regulators. Seven of these mutants have been previously shown to affect *FLO11* expression or function. Most of the mutants have more than two-fold reduced *FLO11* transcript levels, suggesting a loss of function in transcription elongation or in mRNA processing and stability. This group includes each one of the four THO complex mutants.

INTRODUCTION

In *Saccharomyces cerevisiae* there are 29 genes with long (≥40 nt) tandem repeats (Verstrepen et al., 2005). One of them, *FLO11*, encodes for a cell-surface adhesin that is required for adhesion of cells to the agar plate (Roberts and Fink, 1994; Lambrechts et al., 1996). In strains that don't express *FLO11* the cells wash off the plates. *FLO11*-related phenomena have been studied in strains of the Σ1278b background. *FLO11* is not expressed in another commonly used strain background, S288c, because of a mutation in the transcriptional activator *FLO8* (Liu et al., 1996).

The *FLO11* promoter is under a complex regulatory control. Both the MAPK and the PKA signaling cascades regulate its exceptionally large 2.5 kb promoter (Fig. 1A). Flo8 is a transcriptional activator that is absolutely required for *FLO11* expression. The effects of Nrg1 and Nrg2, which repress *FLO11* expression, are reversed by Snf1 to allow *FLO11* expression. Recently, it has been shown that the pH response regulator Rim101 also suppresses *NRG1* (Lamb and Mitchell, 2003). *FLO11* is also regulated epigenetically (Halme et al., 2004): the histone deacetylase Hda1 maintains its ON and OFF states.

FLO11 is controlled at the post-transcriptional level as well. Many of these controls occur in the secretory pathway (de Nobel and Lipke, 1994; Smits et al., 1999). The protein has several domains: an N-terminal signal sequence directs translation into the endoplasmic reticulum (ER). The central portion has an abundance of serine and threonine residues, which are targets for O-mannosylation, a process in which a short chain of 4-5 mannose residues are attached to the hydroxyl group of the serine or threonine side-chain. Glycosylation begins in the ER and continues in the Golgi

apparatus. In the end, the sugar moiety can comprise as much as 80% of the molecular weight of the glycoprotein. A C-terminal GPI-anchor sequence is replaced with a pre-assembled glycosyl-phosphatidyl-inositol (GPI) anchor. During export from the cell, the hydrophobic moiety of the GPI anchor assures attachment to the plasma membrane. Proteins like Flo11 undergo an additional step of modification where the hydrophobic part of the GPI anchor is cleaved, and the mannoprotein is covalently bound to the cell wall via attachment of the remaining portion of GPI to the β-1,6-glucan glucose-based polymer of the cell surface. Only after this multitude of processing and regulatory steps Flo11 can be functional at the cell surface.

Many of the genes with repeats encode for cell surface proteins (Verstrepen et al., 2005), and are expressed only under specific growth conditions. The yeast cell wall is composed of three major components: ß-glucan, mannoproteins, and chitin, and is about 100 nm thick. The cell wall is a very dynamic structure. In general, most of its weight comes from equal amounts of ß-glucan and mannoproteins, and small amounts (1-5%) chitin. However, the distribution, and especially the representations of different mannoproteins can change dramatically in response to growth conditions. For example, of the genes with repeats, *TIR1* and *TIR4* are induced by cold shock or hypoxia (low oxygen levels); *HSP150* by heat shock; *DAN4* by anaerobic conditions; *FLO11* by nitrogen starvation in diploid cells; *AGA1* by pheromone; *EGT2*, *FLO11*, *HSP150*, *PIR1*, *PIR3*, *SED1*, *YOL155c* by sporulation (Smits et al., 1999).

In a similar vein, yeast genes with repetitive sequences are regulated by compensatory mechanisms of the cell wall. For example, during normal aerobic growth, *CWP1* and *CWP2* (these genes do not have long tandem repeats) that encode for two

major mannoproteins are strongly expressed, while *DAN4*, *TIR1*, and *TIR4* (genes with repeats) are repressed; in anaerobic growth the regulation is reversed: CWP1 and CWP2 expression is repressed, and the DAN/TIR genes are expressed (Abramova et al., 2001). Another study compared whole-genome expression profiles in different cell wall mutants. In a *gas1*Δ mutant, characterized by strongly reduced ß-glucan levels, expression of the mannoprotein encoding genes, *SED1* and *PIR3* is upregulated (Lagorce et., 2003). These results support the idea of a compensatory mechanism at the cell wall that substitutes for reduced levels of one component with another.

In our screen for genes that are required for promoter-independent regulation of genes with internal repeats, we used *FLO11* expression from a heterologous constitutive promoter for two reasons. First, it allowed us to bypass numerous mutants that might affect transcription initiation of *FLO11*, as opposed to a processing step that may be specified by the ORF sequence. For example, the S288c library that we wanted to use in our mutant hunt is unable to express *FLO11* because of a mutation in the transcriptional activator *FLO8* (Liu et al., 1996) (Fig. 1B). Second, expression of *FLO11* from the constitutive *TEF* promoter confers strong enough adhesion in the S288c *flo8*- strain background to allow identification of non-adherent mutants (Fig. 1C).

METHODS

Yeast strains and growth conditions

Strains in two genetic backgrounds, S288c and Σ1278b, were used in these studies (Table 1). The deletion library is in the S288c background, which has a mutation in the *flo8* gene (Liu et al., 1996). As *FLO8* encodes a transcription factor required for *FLO11*

expression, the screen of the library for mutations that caused the Flo⁻ phenotype was performed with a P_{TEF}*FLO11* construct. This construct not only permits the screen of the S288c deletion library, but also reports *FLO11* promoter-independent trans-acting mutations. Each mutant is a complete deletion of the respective gene. After they were identified in the S288c screen, each was transformed into the Σ1278b 10560-23C strain and assayed for non-adherence.

For the yeast deletion library transformation, mutant strains in 96-well plates were preincubated with the *URA3/CEN* P_{TEF}*FLO11* plasmid B4126 and standard PEG/LiOAc/TE/ssDNA mix for 3 hrs at 30°C, followed by a 45-min heat shock at 42°C. Transformants were grown on SC-URA selective media, first liquid (3 days), then solid (2 days). A pool of transformants for each mutant was patched on a YPD rectangular plate, and tested for adhesion after 1 day growth at 30°C by a gentle wash under running water.

Primers and plasmid construction

Primer pairs for real-time PCR analysis were designed with Primer Express software. The plasmid B4126 was constructed by transferring a *Stu*I/*Age*I fragment that contains *FLO11* from B4050 (Guo et al., 2000) into the p416TEF *CEN-URA3* plasmid (Mumberg et al., 1995) linearized with *Eco*RI and *Xho*I.

mRNA analysis

Total RNA was isolated from 10 ml cultures grown to O.D. $_{600}$ 1.0 using hot acid phenol. DNaseI treatment was carried out for 30 min (Epicentre). Reverse transcription of 0.3 μg

RNA was performed for 30 min at 48°C with 12.5 U of MultiScribe reverse transcriptase (Applied Biosystems) and 2.5 μM random hexamers. One seventh of the cDNA product was used for real-time PCR analysis with reagents from Applied Biosystems and the ABI 7500 real-time PCR system. Probes at the 3' end of ORFs were used when available. Normalization was to *ACT1*, except when analyzing Σ1278b Tho⁻ mutants, where we noticed a slight upregulation of *ACT1* in Tho⁻ mutants compared to other controls. In those cases, normalization was to *SCR1*, a gene transcribed by RNA Polymerase III, and to the translation elongation factor *TEF*. The histograms present data from two to four independent experiments.

Northern hybridization was performed on 10 μg of RNA samples after gel electrophoresis. The blots were first hybridized with a *FLO11* probe, and then, with an *SCR1* probe.

RESULTS

A genome-wide screen identifies novel regulators of *FLO11*

The gene knockout library of *S. cerevisiae* containing all viable single gene deletions was screened to identify genes that are required for *FLO11* function. *FLO11*, a gene with many long internal tandem repeats, confers adhesion of cells to inert substrates such as agar (Roberts and Fink, 1994; Lambrechts et al., 1996). The screen utilized a P_{TEF}*FLO11* construct in which the *FLO11* gene was transcribed from the constitutive *TEF* promoter. This construct confers adherence to solid agar in S288c strains (Fig. 1C), and was used to avoid isolating mutations in genes required for transcription regulation and initiation. Each mutant of the S288c deletion library was transformed with P_{TEF}*FLO11*, and tested

for agar adhesion (Fig. 2B). This primary screen identified 79 mutants with reduced agar adhesion; of those, 44 show a phenotype in Σ1278b P_{FLO11}*FLO11* as well (Fig. 3).

FLO11 mRNA analysis in the identified mutants suggests a possible defect

I analyzed the level of *FLO11* mRNA using Northern blotting as well as real-time PCR (rtPCR) in the mutants that also had a phenotype in Σ1278b. The Northern analysis indicated the presence of one major mRNA product in wild-type and in all mutant strains. Real-time PCR offered higher accuracy in the analysis. *FLO11* transcript levels were compared against levels of *SCR1*, *TEF1*, and *ACT1*. In the analysis of strains of this background we noticed the least fluctuation between levels in wild type and mutant strains for *SCR1* and *TEF1*, so these were used for normalization of *FLO11* mRNA. In most mutants, *FLO11* message is less than 50% of wild type (Fig. 4B, chapter 5). Among those, *gas1*Δ, a cell-wall mutant, and *thp2*Δ, a transcription elongation mutant, expresses *FLO11* at less than 15% of wild type levels, whereas *vps36*Δ, a mutant of the ESCRT II complex, expresses *FLO11* at less than 5% of wild type (Fig. 4A, and 4D). Relative *FLO11* mRNA levels in all analyzed mutants are incorporated in Figure 5. The result seemed surprising at first, considering that the screen was designed with the idea to find post-transcriptional regulators of a cell surface protein. At the same time, it suggested that in the case of *FLO11*, post-initiation transcriptional events could be an important level of regulation.

Reduction of *FLO11* mRNA levels in most of the mutants requires the *FLO11* coding sequence

Reduction of *FLO11* mRNA in the identified mutants might be independent of the promoter sequence, as the *FLO11*-dependent non-adhesive phenotype was assayed both when *FLO11* is expressed from the *TEF1* promoter and from its native promoter. To determine whether the *FLO11* coding sequence was responsible, we analyzed the transcript levels in $P_{FLO11}GFP$ strains in which *GFP* replaces the *FLO11* ORF. *GFP* mRNA and *FLO11* mRNA were compared by rtPCR in the corresponding wild type and mutant strains. Although *FLO11* mRNA levels are significantly reduced in the three mutants I analyzed (*gas1*Δ, *thp2*Δ, *vps36*Δ), *GFP* expression from the *FLO11* promoter is nearly at wild-type levels in the *gas1*Δ and *thp2*Δ mutants. However, it remained at around 5% in the *vps36*Δ mutant (Fig. 4D). The same trend of expression of $P_{FLO11}GFP$ in the mutants can be visualized by the roughly equivalent GFP fluorescence in wild type, *gas1*Δ and *thp2*Δ strains, but the absence of fluorescence in the *vps36*Δ mutant (Fig. 4C). This result suggests that *FLO11* mRNA downregulation in the *gas1*Δ and *thp2*Δ mutants depends mostly upon the presence of the *FLO11* coding sequence.

The identified genes fall into several functional groups

The identified genes fall into several well-defined functional groups, and often times are within the same complex with other hits of the screen. Synthetic interactions further connect the genes in our set. Only seven of the 44 mutants we identified have been previously associated with *FLO11* function: *BEM4*, *SRV2*, *WHI3*, and the ESCRT

complex genes *VPS23, 28, 32, 36*. Sometimes genes cannot be ascribed to a single functional group, but we divided the hits from our screen in the following classes (Fig. 5).

Genes that regulate cell wall structure: GAS1, SMI1, GPI7, LDB16, YNL171C

There are three major components of the yeast cell wall: ß-glucan, mannoproteins, and chitin. Two types of ß-glucan polymers are present in the cell surface, ß-1,3-glucan and ß-1,6-glucan, reflecting the linkage pattern of the glucose subunit in the polymer. Gas1 and Smi1 are required for ß-1,3-glucan synthesis and assembly. Gas1 is a GPI-anchored protein at the plasma membrane with a ß-1,3-glucanosyl-transferase activity, and plays a major role in polymer cross-linking at the cell wall (Popolo and Vai, 1999). Smi1/Knr4 is also involved in ß-1,3-glucan biosynthesis. Gpi7 localizes to the plasma membrane and cell wall; *gpi7*Δ mutants have a larger fraction of GPI-containing proteins released in the growth medium as opposed to retained at the cell wall, suggesting a function of Gpi7 in processing GPI-anchors for proper attachment in the cell wall (Richard et al., 2002). *LDB16* (low dye binding) was identified in a screen for reduced mannophosphate (by alcian blue staining of N-linked oligosaccharides) at the yeast cell surface (Corbacho et al., 2004). Deletion of *YNL171C* is synthetic lethal with *gas1*Δ and *bem4*Δ.

Cytoskeleton and polarity: SRV2, ARP1, SEP7, CDC10, VRP1, SLA1, BEM4, SPC72

Srv2 is involved in Ras protein signaling and cytoskeleton organization. The actin-related protein Arp1 is a component of the dynactin complex, and is required for spindle orientation and nuclear migration. The septins Sep7 and Cdc10 localize at the mother-daughter bud neck and function in localization of divisional landmarks and in

cytokinesis. Vrp1 (verprolin) and Sla1 (synthetic lethal with actin) are involved in cytoskeleton organization and endocytosis. The bud-emergence protein Bem4 functions in cytoskeleton organization and Rho protein signaling. Spc72 is a component of the cytoplasmic γ-tubulin (Tub4) complex. *spc72* mutants have reduced mannophosphate at the cell surface.

Protein sorting: VPS3, -21, -23, -28, -32, -36, -65

Vps3 is required for vacuolar protein and acidification maintenance; Vps21 is a GTPase required for transport during endocytosis. *VPS65* is a dubious open reading frame of 315 nucleotides, and 75% of it overlap *SFH1*, a gene that encodes for a subunit of the RSC chromatin remodeling complex. *VPS23*, *28*, *32*, and *36* are components of the ESCRT complexes that function in protein targeting to the vacuole.

Transcription elongation and mRNA export: THO2, HPR1, MFT1, THP2, SAC3, THP1, SPT4, TRF4, DHH1, DEF1, RPB4

A number of genes with a role in transcription elongation and mRNA export are required for adhesion to agar. Tho2, Hpr1, Mft1 and Thp2 are the four subunits of the THO complex thought to be involved in transcription elongation and export. Sac3 and Thp1 form a heterodimer complex that functions in mRNA export. Spt4 can both activate and inhibit transcription elongation. Also, *spt4*Δ *mft1*Δ and *spt4*Δ *hpr1*Δ double mutants are synthetic lethal (Rondon et al., 2003). *TRF4* was identified in a synthetic lethality screen of a *top1*Δ mutant, and shown to be synthetically lethal with an *hpr1*Δ mutant as well (Sadoff et al., 1995). Trf4 is also called Pap2 for Poly (A) polymerase, indicating a

function in RNA control. Dhh1 is a cytoplasmic DExD/H-box helicase that coordinates mRNA function and decay; it localizes to cytoplasmic processing bodies, and can sequester mRNAs for degradation or for subsequent translation (Sheth and Parker, 2003). Dhh1 is also specifically required for G1/S DNA damage checkpoint recovery (Bergkessel and Reese, 2004). Def1 is an RNAPII degradation factor when RNAPII is in an elongation complex. Rpb4 is the RNA polymerase II subunit B32.

Chromatin remodeling: RSC1, RSC2, NPL6, LDB7, SIN3, TAF14, YAF9

The hits from my screen include 4 components of the RSC complex: Rsc1, Rsc2, Npl6, and Ldb7. Of those, Ldb7 has been shown to regulate mannophosphate levels at the cell surface (Corbacho et al., 2004). Sin3 forms a histone deacetylase complex with Rpd3, and has a role in chromatin silencing at rDNA and telomeres (Sun and Hampsey, 1999). Taf14 is a subunit of TFIID, TFIIF, and SWI/SNF complexes, involved in RNA polymerase II transcription initiation and in chromatin modification. Taf14 interacts with another hit from the screen, *SMI1,* for β-1,3-glucan biosynthesis, and is synthetic lethal with two other hits, *sla1*Δ and *yaf9*Δ. Yaf9 is a subunit of both the NuA4 histone H4 acetyltransferase complex and the SWR1 complex, and has homology to human leukemogenic protein AF9.

COMA complex: CTF19, MCM21

Ctf19 and Mcm21 are two components of the four-subunit COMA complex. *CTF19* was identified as a chromosome transmission fidelity mutant in a synthetic dosage lethality screen (Hyland et al., 1999); *ctf19* mutants accumulate at G2/M, and show a defect of

centromeres to bind microtubules, suggesting a function in kinetochore assembly and attachment of spindle microtubules.

Ungrouped: BFR1, CWH8, MRPL28, WHI3

Bfr1 is involved in protein metabolism. Bfr1 associates with cytoplasmic mRNP complex and with polyribosomes (Lang et al., 2001); loss of Bfr1 disrupts the interaction of Scp160 with polyribosomes. Cwh8 is a dolichyl pyrophosphate phosphatase located in the ER, and functions in N-glycosylation of proteins and in lipid biosynthesis (van Berkel et al., 1999; Fernandez et al., 2001). Mrpl28 is the mitochondrial ribosomal protein large subunit. Whi3 binds the mRNA of *CLN3*, a G1 cyclin that promotes G1/S transition; *WHI3* is required for suppressing Cln3 function in meiosis, filamentation and mating in Saccharomyces cerevisiae (Gari et al., 2001).

Five ORFs that show a non-adherent phenotype in both the primary and secondary screens are not included in the final list. Four of them are dubious ORFs that overlap known genes (in parentheses), that also have a phenotype: *YNL140C* (*THO2*), *YGR064W* (*SPT4*), *YLR358C* (*RSC2*), and *YPL017C* (*CTF19*). Deletion of a fifth gene, *PTC1/CWH47*, has a moderate non-adherent phenotype in Σ1278b but the mutant strain was lost. *PTC1* encodes for a type 2C phosphatase. It dephosphorylates Hog1 and inactivates the osmosensing MAPK pathway. The deletion mutant also exhibits a calcofluor white hypersensitivity, indicative of a cell wall defect.

Genes that are required for adhesion in the primary but not in the secondary screen

Not all mutants with a phenotype in the primary screen show a phenotype in the secondary. One group of mutants has a weak phenotype in the first screen, and no discernable phenotype in the secondary screen. The genes that correspond to these 17 deletion mutants are: *KRE6, PMT1, PMT2, AXL2, CYK2/HOF1, CLA4, SRO9, RGP1, VPS23/STP22, VPS24, STO1/GCR3, SWC2/VPS72, SWC5/AOR1, SWC6/VPS71, YDR532C, YEL059W*, and *YMR031W-A*. *KRE6* is required for β-1,6-glucan synthesis. The protein O-mannosyltransferases encoded by *PMT1* and *PMT2* localize to the endoplasmic reticulum membrane, and add the first mannose residues to the hydroxyl groups of serine and threonine side chains. Axl2 is an integral plasma membrane protein that localizes to the bud neck, and is required for axial budding. Cyk2 is also a bud-neck localized protein and is required for cytokinesis. Cla4 is a signal transduction kinase involved in septin ring assembly and cytokinesis. Sro9 participates in the organization of actin filaments, and can also bind RNA and associate with translating ribosomes. Rgp1 is a subunit of a Golgi membrane exchange factor that catalyzes nucleotide exchange on the GTPase Ypt6. Vps23 is a subunit of ESCRT I, and Vps24 of ESCRT III. Sto1 is a subunit of a nuclear mRNA cap-binding protein complex. Deletion of *STO1* suppresses the temperature-sensitivity phenotype of *hpr1*Δ mutant (Uemura et al., 1996); *STO1* also interacts genetically with the topoisomerase *TOP1*, and plasmid DNA from *sto1* mutants has excessive negatively supercoiling. Swc2, -5, -6 are components of the SWR1 complex that incorporates Htz1 into chromatin. The uncharacterized ORF *YDR532C* encodes a protein that co-localizes with spindle pole bodies. *YEL059W* and *YMR031W-A* are two dubious ORFs. *YMR031W-A* partially overlaps with the uncharacterized ORF

YMR031C, and is also in the 5'-untranslated region of the above-mentioned *CYK2/HOF1* genes.

Another group of mutants has a very strong phenotype in the P_{TEF}*FLO11* S288c system but no phenotype in Σ1278b P_{FLO11}*FLO11*. The genes corresponding to those 15 mutants are: *ISC1*, *FAB1*, *PDR17*, *BEM1/SRO1*, *SWA2/AUX1*, *SEC66*, *SEM1*, *VPS63*, *PRE9*, *RPN10*, *GPH1*, *SCP160*, *VPS69*, *SPT7*, and *HNT3*. Isc1 is a phospholipase C type enzyme. Fab1 is involved in phospholipid metabolism, and vacuole organization. *PDR17* encodes a phosphatidyl-inositol transfer protein involved in phospholipids transport and biosynthesis. Bem1 is required for bud-emergence and establishment of cell polarity. Swa2 is an auxilin-like protein involved in vesicular transport, ER organization and biogenesis. Sec66 is a subunit of the Sec63 complex required for translocation of presecretory proteins. Sem1 regulates exocytosis and pseudohyphal differentiation (Jantti et al., 1999): the gene was identified as a multicopy suppressor of exocyst mutants; deletion of *SEM1* induces filamentous growth in S288c diploid cells and enhances filamentation in Σ1278b diploids. Deletions of *VPS63* or *VPS69* lead to aberrant extracellular secretion of the vacuolar protein carboxypeptidase Y (Bonangelino et al., 2002). Pre9, the only non-essential subunit of the 20S proteasome, is required for filament maturation during pseudohyphal growth of *Saccharomyces cerevisiae* (Kohler 2003). *RPN10* encodes a subunit of the 19S regulatory particle of the 26S proteasome. Gph1 has a glycogen phosphorylase activity, and is required for the catabolism of glycogen. Scp160 is involved in chromatin silencing and chromosome segregation. Scp160 has been also identified as an mRNP component in polyribosomes that associates with specific rather than random mRNAs (Li et al., 2003); among those are the mRNAs

of *DHH1*, another hit of the screen mentioned above, and *YOL155C*, a glucosidase encoding cell surface gene with internal tandem repeats like *FLO11*. Spt7 is a component of the histone acetyltransferase SAGA complex that regulates transcription. Hnt3 is a nucleotide binding protein of the histidine triad superfamily.

Two mutants, *bdf1*Δ and *dig1*Δ, are strongly non-adherent in the primary screen, and hyper-adherent in the secondary screen. Bdf1 is a bromodomain-containing protein involved in transcription initiation. The *dig1* mutant has been previously shown to "dig into agar" as a result of hyper-invasiveness (Palecek et al., 2000) in the Σ1278b background, but I found that it has a loss of adhesion in S288c P_{TEF}*FLO11*.

The *ctk1*Δ mutant has a very strong non-adherent phenotype in S288c but we were not able to analyze the mutant in Σ1278b. Ctk1 phosphorylates serine 2 of the YSPTSPS repeat at the C-terminal domain of the Rpb1 subunit of RNA polymerase II during transcription (Cho et al., 2001). Judging by the slow growth of the *ctk1*Δ mutant in S288c, and by the slower growth of Σ1278b wild type strain compared to S288c wild type, *ctk1*Δ mutant in Σ1278b might be inviable.

DISCUSSION

The library of all viable single-gene deletions of *Saccharomyces cerevisiae* has been a great tool for genome-wide screens, and has permitted the identification of novel regulators of many biological processes and functions. The screen presented here is no exception to that. It was successful in the identification of novel regulators of *FLO11*. Many of the identified hits fall into well-defined and connected functional modules.

To name a few, the two screens in this study show that *FLO11*expression or function requires the chromatin remodeling RSC complex, the transcription elongation and export THO and Sac3-Thp1 complexes, mRNA retention (Dhh1, Whi3) or translation-stimulating polyribosome-binding (Bfr1, Scp160) proteins, kinetochore and spindle pole body proteins (Ctf19, Mcm21, Spc72), the endosomal ESCRT complexes, glycosylation (Cwh8, Pmt1, Pmt2) and GPI-anchor processing (Gpi7) proteins, phospholipid regulators (Isc1, Fab1, Pdr17), septins (Sep7, Cdc10), and glucan biosynthesis proteins (Gas1, Smi1).

Consistent with Flo11 biology, some of these proteins may affect Flo11 directly. Scp160 may be required for *FLO11* mRNA translation, since it has been shown to specifically bind the mRNA of *YOL155C*, another cell surface gene with internal repeats. The mannosyl-transferases Pmt1 and Pmt2 probably glycosylate the serine-threonine-rich repetitive region of Flo11. Cytoskeletal proteins and septins may facilitate Flo11 export. Isc1, Pdr17, Fab1 and Gpi7 might be required for GPI-anchor processing and attachment at the cell surface.

Other genes may affect *FLO11* indirectly. For example, *whi3* mutants are unable to control the G1/S cyclin Cln3, and may be fast-forwarding through the cell cycle, not allowing enough time for *FLO11* expression. On the other hand, the kinetochore and spindle pole body mutants may slow down the cell cycle at the G2/M phase and jeopardize *FLO11* functionality due to degradation.

The ESCRT complex is likely to regulate *FLO11* at two different levels. Recent studies indicate that the ESCRT complex is required for activation of Rim101 by proteolytic cleavage (Xu et al., 2004). In its turn, Rim101 is required for suppression of

Nrg1, a suppressor of the *FLO11* promoter (Lamb and Mitchell, 2003). This regulatory module may explain the lack of expression of *FLO11* as well as of *GFP* from the *FLO11* promoter in Σ1278b ESCRT$^-$ mutants. It was also recently shown that one of the ESCRTIII components, *VPS32*, is required for filament maturation during *FLO11*-dependent pseudohyphal growth (Kohler, 2003). The fact that a number of the ESCRT$^-$ mutants have a non-adherent phenotype in the primary screen where FLO11 is expressed from the constitutive TEF promoter suggests that the ESCRT complexes may have a second function in *FLO11* regulation.

FLO11 may be regulated in a feedback mechanism. Deletion of the plasma membrane gene *GAS1* reduces *FLO11* expression more than 5-fold in the Σ1278b strain background and affects *FLO11* function in both Σ1278b and S288c P_{TEF}*FLO11*. The promoter-independent effect of *gas1*Δ in S288c, and the role of Gas1 in cell-wall biosynthesis, suggests that the non-adherent phenotype may result from a defective cell wall architecture that cannot retain this GPI protein at the cell surface. The strong reduction of *FLO11* expression from its endogenous promoter in the *gas1*Δ mutant may be an indication of a feedback control—a damaged cell wall signals back to suppress the expression of a long non-essential gene requiring multiple post-transcriptional modifications. In the four other cell wall mutants pulled from the screen (*gpi7*Δ, *ynl171c*Δ, *smi1*Δ, *ldb16*Δ) *FLO11* transcript levels are also reduced at least two-fold.

Several gene deletions with phenotype fall within other ORFs. For example, the short dubious ORFs *VPS63* (327 nt), *VPS65* (315 nt) and *VPS69* (321 nt) overlap with *YPT6, SFH1,* and *SRP54*, respectively, on the other strand. *VPS63*, *-65*, and *-69* have no orthologs in other yeast species, and were given these names because of a vacuolar

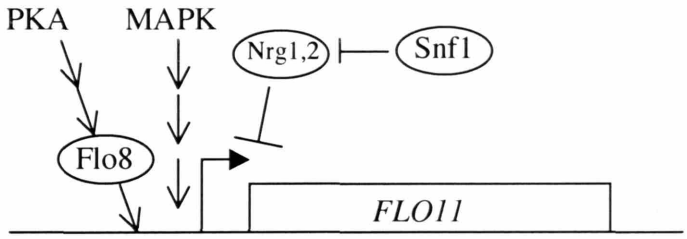
protein sorting defect of the deletion mutant (Bonangelino et al., 2002). It remains unclear whether these *VPS* genes have a function of their own or whether the observed *vps*-defect is associated with the concurrent deletion of part of the essential genes *YPT6*, *SFH1*, and *SRP54*. One experiment to address this question would be to test whether ectopic expression of the corresponding essential gene complements the observed phenotype. Another short dubious ORF *YNL171C* (369 nucleotides) is at the 3'-end of the essential gene *APC1* (5247 nt).

The known functions of the essential genes that overlap the above-mentioned dubious ORFs partially agree with the molecular biology and the other hits of this screen for promoter-independent regulators of *FLO11*. Ypt6 is a Ras-like GTPase involved in endosome to Golgi transport. The secretory pathway is required for Flo11 processing, and the screen confirmed a role of the endosomal sorting complexes (ESCRTI, II, and III) in *FLO11* regulation. Sfh1 is a member of the RSC chromatin remodeling complex, and the screen identified four other components of the RSC complex (Rsc1, Rsc2, Npl6, Ldb7) to regulate *FLO11*. Moreover, *SFH1* interacts genetically with kinetochore genes (Hsu et al., 2003), suggesting that components of the RSC and COMA complexes may regulate *FLO11* via a chromosome transmission fidelity function. Srp54 is a signal recognition particle subunit that is involved in co-translational protein targeting to the endoplasmic reticulum, a step common to all secreted and cell surface proteins. *APC1* encodes for the largest subunit of the anaphase-promoting complex, and is required for degradation of anaphase inhibitors, leading to separation of sister chromatids (Zacchariae and Nasmyth, 1999).

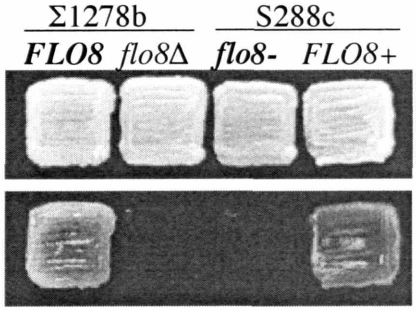
The screens in this study identified both yeast-specific and evolutionarily conserved regulators and interactions. Gas1 and Kre6 are yeast-specific proteins involved in cell wall biosynthesis. Pmt1 and Pmt2 are conserved among fungi but are absent in green plants, and present a target for antifungal agents against phytopathogenic fungi (Girrbach and Strahl, 2003). In terms of conservation, Flo11 itself resembles the human mucins with abundant serine/threonine repeats and sites for O-glycosylation. Unlike yeast Flo proteins, human mucins are either secreted or attached to the cell via a transmembrane domain instead of a GPI anchor, and contain additional domains (Lambrechts et al., 1996; Duraisamy et al., 2006). The THO complex is conserved from yeast to humans (Reed and Cheng, 2005). The RNA helicase Dhh1 that can sequester RNA in processing bodies resembles the human protooncogene p54/RCK (Bergkessel and Reese, 2004). Moreover, both the *Drosophila* and *Xenopus* orthologs of Dhh1 (Me31B and Xp54 respectively) function in development by regulating transport and translation of developmentally important mRNAs (Nakamura et al., 2001; Smillie and Sommerville, 2002).

In summary, the screen isolated 37 novel and 7 previously identified regulators of *FLO11*. Another 35 genes are required for adherence in S288c P_{TEF}*FLO11* but not in Σ1278b P_{Flo11}*FLO11*. To differentiate between promoter and strain background contributions, it may be useful to analyze the effect of these 35 mutants in a third system where the chromosomal copy of *FLO11* is placed under a constitutive or inducible promoter. The four single gene deletions of the THO complex are among the 10 mutants with the strongest non-adherent phenotype in the secondary screen. Their effect on genes with repeats is specifically analyzed in Chapter 3.

A



B



C

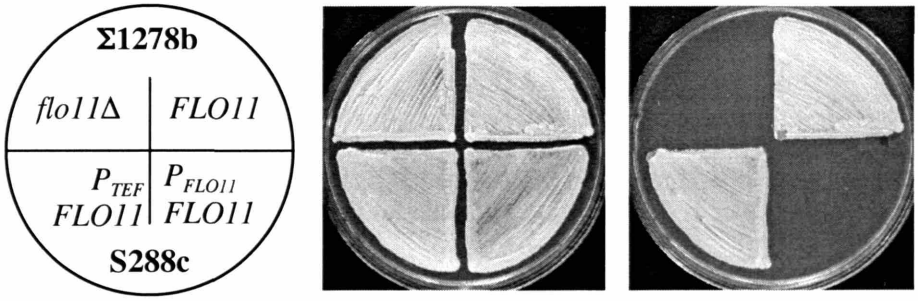


Figure 1: Transcriptional regulators of the *FLO11* promoter can be bypassed by *FLO11* expression from the constitutive *TEF* promoter.

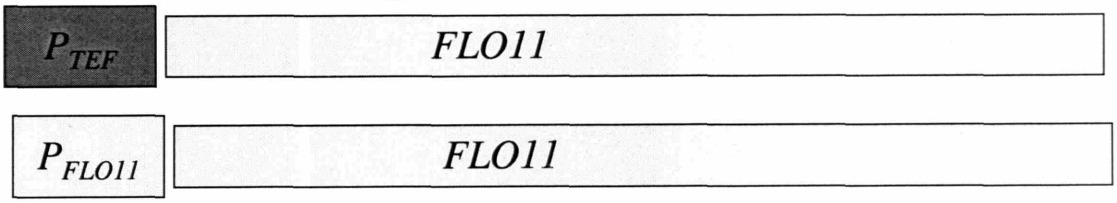
A. The MAPK and PKA signaling cascades as well as the repressors Nrg1 and Nrg2 regulate the *FLO11* promoter.

B. Strains of the S288c background have a mutation in the transcriptional activator *FLO8* and do not adhere to agar. Ectopic expression of *FLO8* wild-type allele restores adherence.

C. Expression of *FLO11* from a heterologous constitutive promoter confers adherence to agar in S288c strains.

A

promoter / background	*TEF*	*FLO11*
S288C	**PRIMARY SCREEN**	
Σ1278b		**SECONDARY SCREEN**



B

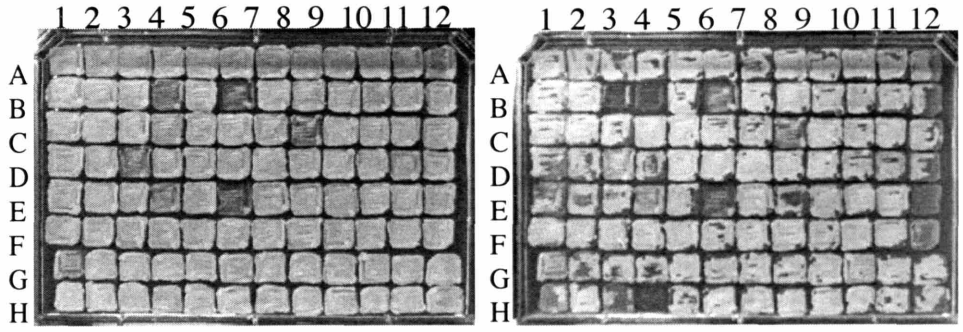


Figure 2: A screen for promoter-independent regulators of *FLO11*.

A. Experimental design.
B. Primary screen. Plate 15 of the Invitrogen *MATa* collection is shown (52 plates total). The S288c parental strain BY4741 (coordinates B3) is used as a negative control. BY4741 with P_{TEF}*FLO11* (H2) is a positive control. While most mutant strains with P_{TEF}*FLO11* adhere to agar, the *thp2*Δ mutant shows a strong non-adherent phenotype (H4).

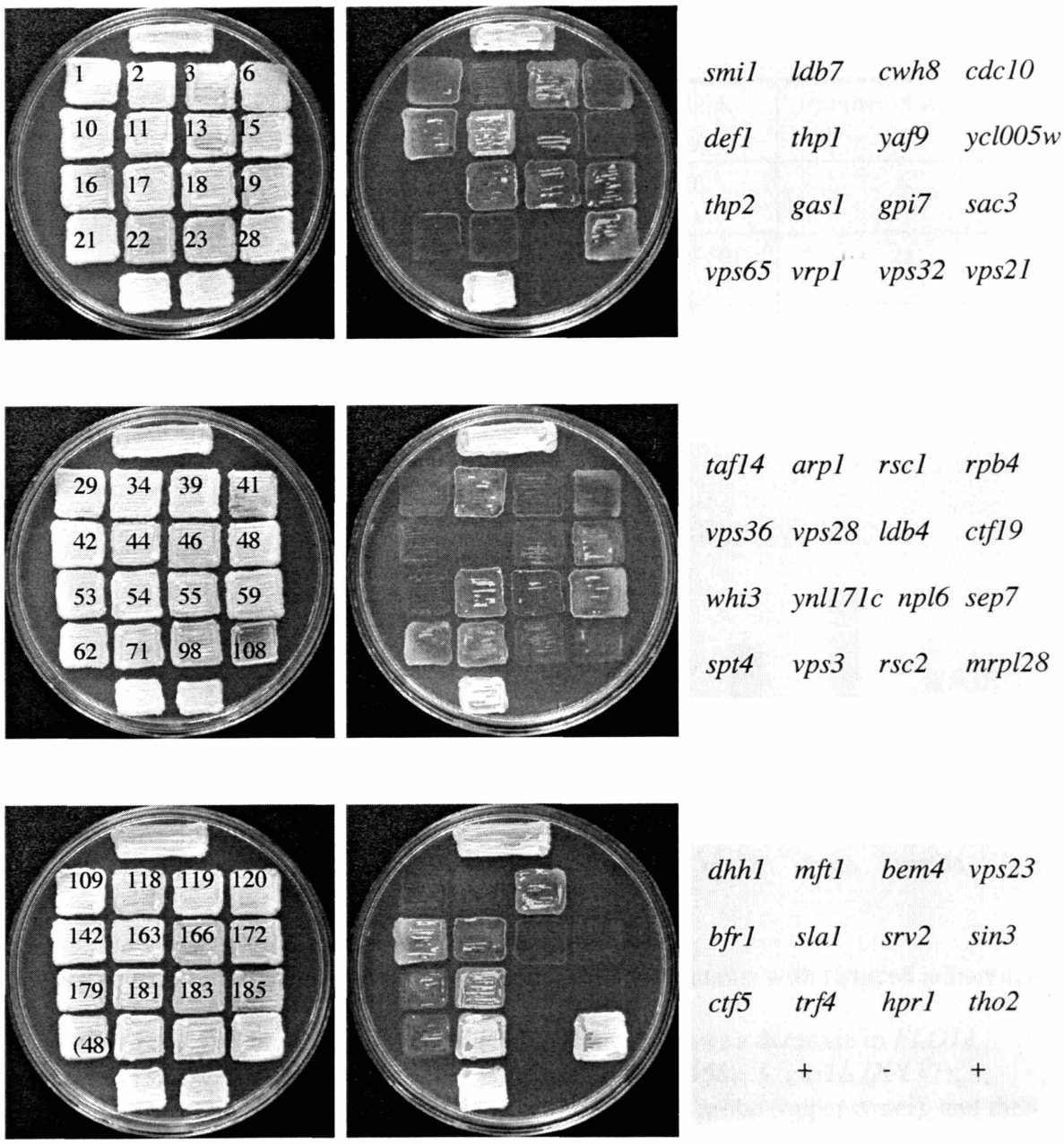
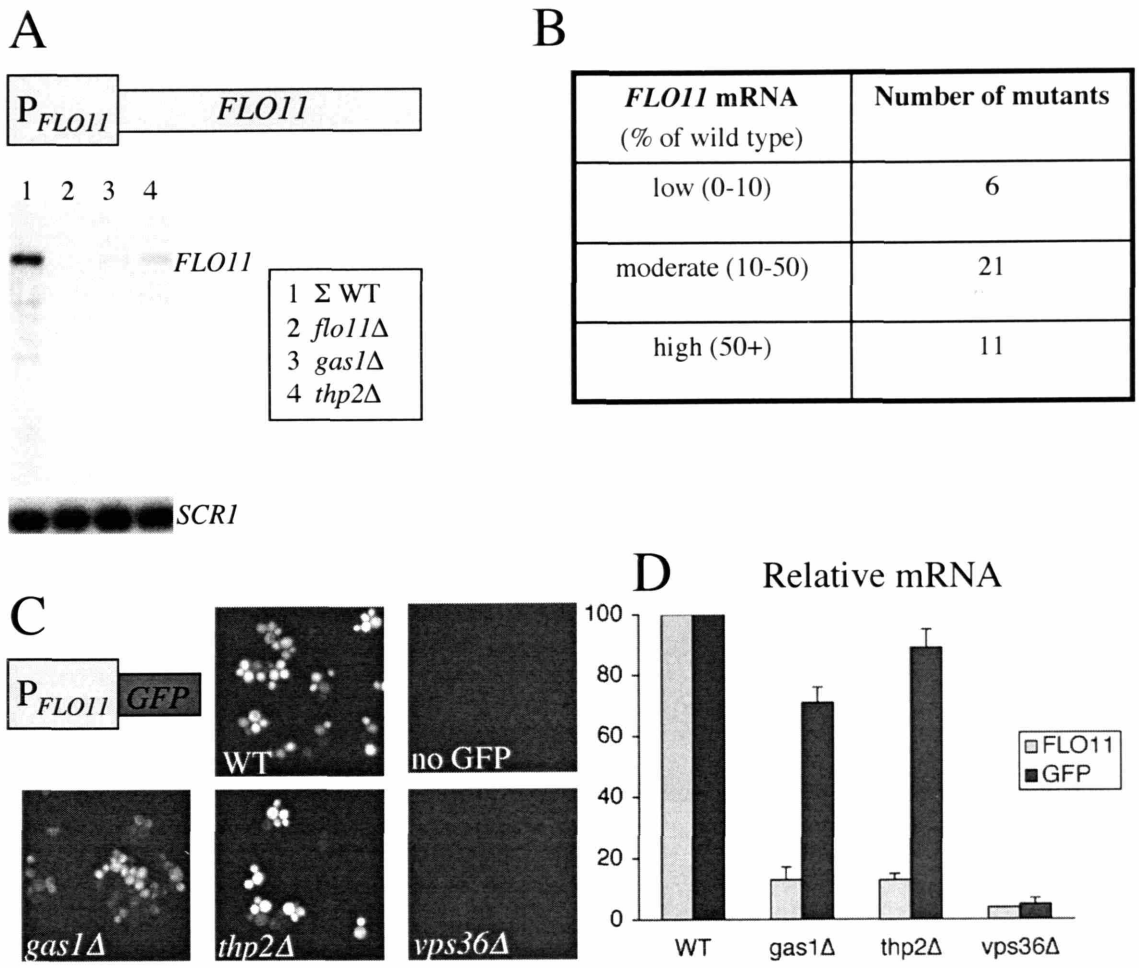


Figure 3: A secondary screen in the Σ1278b background confirms the reduced adherence phenotype of mutants identified in the primary screen.

The corresponding ORF deletions were prepared in strain 10560-23C (L7554) where *FLO11* is on the chromosome under its native promoter. The numbers on the patches indicate the corresponding XY strain. The YPD plates were washed after a 3-day incubation at 30°C. (+) refers to the wild-type Σ1278b that also appears at the top and bottom of each plate; (-) is the negative control strain *flo11*Δ (L7558) next to the positive control at the bottom of each plate. XY48 is patched twice as an additional control.



FLO11 mRNA (% of wild type)	Number of mutants
low (0-10)	6
moderate (10-50)	21
high (50+)	11

Figure 4: *FLO11* transcript and promoter analysis in the mutants with reduced adherence.

A. Northern analysis of mutants with reduced adherence shows a decrease in *FLO11* transcript: 1, Σ1278b wild-type (10560-23C); 2, *flo11Δ* (L7558); 3, *gas1Δ* (XY17); 4, *thp2Δ* (XY16). The blot was first hybridized with a *FLO11* probe (upper panel), and then with an *SCR1* probe (lower panel).

B. Cumulative data from Northern blots and real-time PCR (normalization to *SCR1* and *TEF*). Actual values of *FLO11* mRNA in specific mutants are listed in Chapter 5.

C. *GFP* expression from the *FLO11* promoter in a $P_{FLO11}GFP$ fusion construct where *GFP* replaces the *FLO11* ORF (normalizations to *SCR1* and *TEF* were nearly identical).

D. The histogram compares *FLO11* (lighter bars) and *GFP* mRNA levels (darker bars) in isogenic $P_{FLO11}FLO11$ and $P_{FLO11}GFP$ wild-type or mutant strains (real-time PCR).

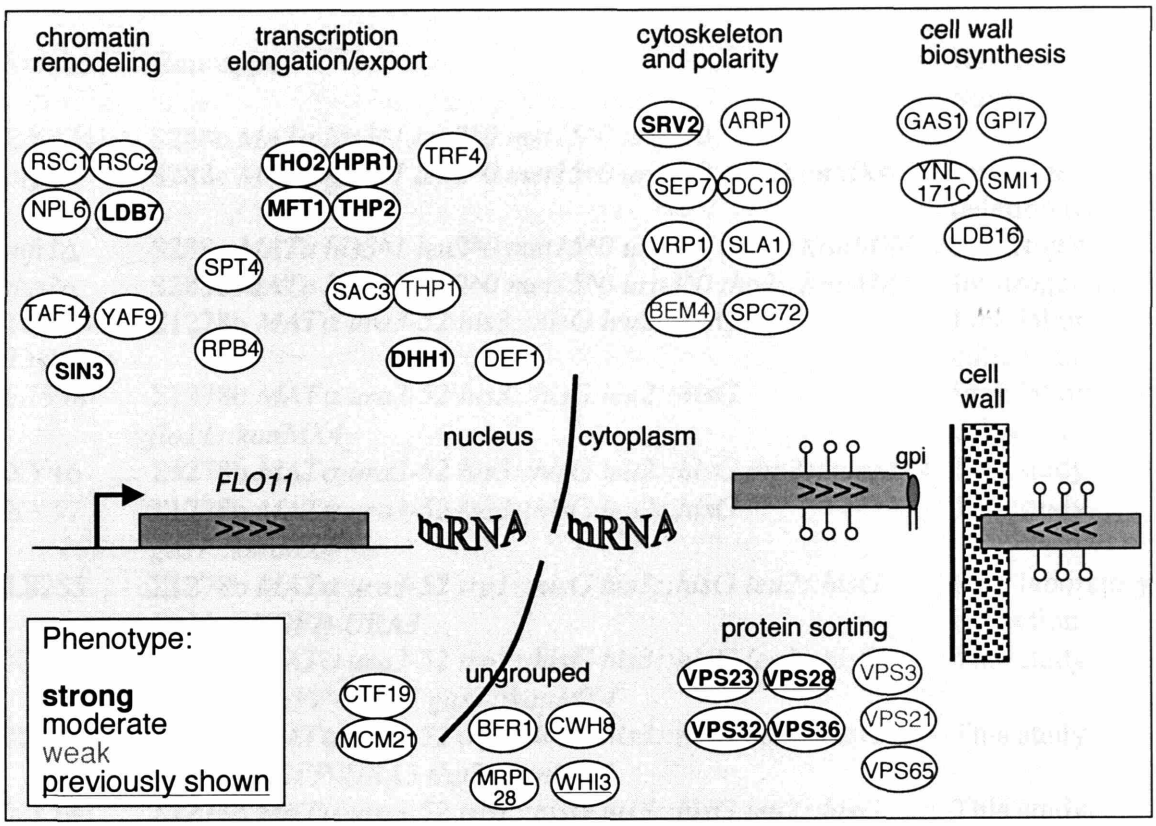


Figure 5: Summary of the results of the screen for promoter-independent regulators of *FLO11*.

Associated functions of the genes required for *FLO11* function, and suggested site of regulation.

Degree of phenotype: for strong non-adherent phenotypes gene names are in bold, for moderate in regular font, and for weak phenotype in grey font.

Underlined are genes previously shown or implicated to affect *FLO11* function.

Table 1: Yeast strains

Strain	Genotype	Reference/ Source
BY4741	S288c *MATa his3^1 leu2^0 met15^0 ura3^0*	
*orf*Δ	S288c *MATa his3^1 leu2^0 met15^0 ura3^0 orf::KanMX4*	Invitrogen deletion library
*mft1*Δ	S288c *MATa his3^1 leu2^0 met15^0 ura3^0 mft1::KanMX4*	Invitrogen *mft1*Δ
*thp2*Δ	S288c *MATa his3^1 leu2^0 met15^0 ura3^0 thp2::KanMX4*	Invitrogen *thp2*Δ
10560-23C	Σ1278b *MATα ura3-52 his3::hisG leu2::hisG*	Fink laboratory collection
L7558	Σ1278b *MATα ura3-52 his3::hisG leu2::hisG flo11::kanMX4*	Fink laboratory collection
XY16	Σ1278b *MATα ura3-52 his3::hisG leu2::hisG thp2::kanMX4*	This study
XY17	Σ1278b *MATα ura3-52 his3::hisG leu2::hisG gas1::kanMX4*	This study
L8255	Σ1278b *MATα ura3-52 trp1::hisG his3::hisG leu2::hisG flo11::yEGFP-URA3*	Fink laboratory collection
XY123	Σ1278b *MATα ura3-52 trp1::hisG his3::hisG leu2::hisG flo11::yEGFP-URA3 gas1::kanMX4*	This study
XY134	Σ1278b *MATα ura3-52 trp1::hisG his3::hisG leu2::hisG flo11::yEGFP-URA3 thp2::kanMX4*	This study
XY141	Σ1278b *MATα ura3-52 trp1::hisG his3::hisG leu2::hisG flo11::yEGFP-URA3 vps36::kanMX4*	This study

Table 2: Real-time PCR primers used in this study

ORF, position	Name	Sequence
SCR1 f 384	A47	CGGCCGGGATAGCACATA
SCR1 r 437	A48	CGCCGAAGCGATCAACTT
TEF f 349	V216	GCTGGTGGTGTCGGTGAATT
TEF r 410	V217	GCGTGTTCTCTGGTTTGACCAT
FLO11 f 3038	V290	GTTCAACCAGTCCAAGCGAAA
FLO11 r 3104	V291	GTAGTTACAGGTGTGGTAGGTGAAGTG
GFP f 24	V292	CACTGGTGTTGTCCCAATTTTG
GFP r 94	V293	CACCGGAGACAGAAAATTTGTG
ACT1 f 1038	V453	GGCTTCTTTGACTACCTTCCAACA
ACT1 r 1103	V454	GATGGACCACTTTCGTCGTATTC

REFERENCES

Abramova, N., Sertil, O., Mehta, S., and Lowry, C. (2001). Reciprocal regulation of anaerobic and aerobic cell wall mannoprotein gene expression in Saccharomyces cerevisiae. J Bacteriol 183, 2881-2887.

Aguilera, A., and Klein, H. (1990). HPR1, a novel yeast gene that prevents intrachromosomal excision recombination, shows carboxy-terminal homology to the Saccharomyces cerevisiae TOP1 gene. Mol Cell Biol 10, 1439-1451.

Bergkessel, M., and Reese, J. C. (2004). An essential role for the Saccharomyces cerevisiae DEAD-box helicase DHH1 in G1/S DNA-damage checkpoint recovery. Genetics 167, 21-33.

Bonangelino, C. J., Chavez, E. M., and Bonifacino, J. S. (2002). Genomic screen for vacuolar protein sorting genes in Saccharomyces cerevisiae. Mol Biol Cell 13, 2486-2501.

Cho, E. J., Kobor, M. S., Kim, M., Greenblatt, J., and Buratowski, S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. Genes Dev 15, 3319-3329.

Corbacho, I., Olivero, I., and Hernandez, L. (2004). Identification of low-dye-binding (ldb) mutants of Saccharomyces cerevisiae. FEMS Yeast Res 4, 437-444.

de Nobel, H., and Lipke, P. N. (1994). Is there a role for GPIs in yeast cell-wall assembly? Trends Cell Biol 4, 42-45.

Fernandez, F., Rush, J. S., Toke, D. A., Han, G. S., Quinn, J. E., Carman, G. M., Choi, J. Y., Voelker, D. R., Aebi, M., and Waechter, C. J. (2001). The CWH8 gene encodes a dolichyl pyrophosphate phosphatase with a luminally oriented active site in the endoplasmic reticulum of Saccharomyces cerevisiae. J Biol Chem 276, 41455-41464.

Gari, E., Volpe, T., Wang, H., Gallego, C., Futcher, B., and Aldea, M. (2001). Whi3 binds the mRNA of the G1 cyclin CLN3 to modulate cell fate in budding yeast. Genes Dev 15, 2803-2808.

Girrbach, V., and Strahl, S. (2003). Members of the evolutionarily conserved PMT family of protein O-mannosyltransferases form distinct protein complexes among themselves. J Biol Chem 278, 12554-12562.

Guo, B., Styles, C., Feng, Q., and Fink, G. (2000). A Saccharomyces gene family involved in invasive growth, cell-cell adhesion, and mating. Proc Natl Acad Sci U S A 97, 12158-12163.

Halme, A., Bumgarner, S., Styles, C., and Fink, G. R. (2004). Genetic and epigenetic regulation of the FLO gene family generates cell-surface variation in yeast. Cell 116, 405-415.

Hsu, J. M., Huang, J., Meluh, P. B., and Laurent, B. C. (2003). The yeast RSC chromatin-remodeling complex is required for kinetochore function in chromosome segregation. Mol Cell Biol 23, 3202-3215.

Hyland, K. M., Kingsbury, J., Koshland, D., and Hieter, P. (1999). Ctf19p: A novel kinetochore protein in Saccharomyces cerevisiae and a potential link between the kinetochore and mitotic spindle. J Cell Biol 145, 15-28.

Jantti, J., Lahdenranta, J., Olkkonen, V. M., Soderlund, H., and Keranen, S. (1999). SEM1, a homologue of the split hand/split foot malformation candidate gene Dss1, regulates exocytosis and pseudohyphal differentiation in yeast. Proc Natl Acad Sci U S A 96, 909-914.

Kohler, J. R. (2003). Mos10 (Vps60) is required for normal filament maturation in Saccharomyces cerevisiae. Molecular Microbiology 49, 1267-1285.

Kuchin, S., Vyas, V. K., and Carlson, M. (2002). Snf1 protein kinase and the repressors Nrg1 and Nrg2 regulate FLO11, haploid invasive growth, and diploid pseudohyphal differentiation. Mol Cell Biol 22, 3994-4000.

Lagorce, A., Hauser, N. C., Labourdette, D., Rodriguez, C., Martin-Yken, H., Arroyo, J., Hoheisel, J. D., and Francois, J. (2003). Genome-wide analysis of the response to cell wall mutations in the yeast Saccharomyces cerevisiae. J Biol Chem 278, 20345-20357.

Lamb, T. M., and Mitchell, A. P. (2003). The transcription factor Rim101p governs ion tolerance and cell differentiation by direct repression of the regulatory genes NRG1 and SMP1 in Saccharomyces cerevisiae. Mol Cell Biol 23, 677-686.

Lambrechts, M. G., Bauer, F. F., Marmur, J., and Pretorius, I. S. (1996). Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohyphal differentiation in yeast. Proc Natl Acad Sci U S A 93, 8419-8424.

Lang, B. D., Li, A., Black-Brewster, H. D., and Fridovich-Keil, J. L. (2001). The brefeldin A resistance protein Bfr1p is a component of polyribosome-associated mRNP complexes in yeast. Nucleic Acids Res 29, 2567-2574.

Li, A. M., Watson, A., and Fridovich-Keil, J. L. (2003). Scp160p associates with specific mRNAs in yeast. Nucleic Acids Res 31, 1830-1837.

Liu, H., Styles, C., and Fink, G. (1996). Saccharomyces cerevisiae S288C has a mutation in FLO8, a gene required for filamentous growth. Genetics 144, 967-978.

Mumberg, D., Muller, R., and Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene 156, 119-122.

Nakamura, A., Amikura, R., Hanyu, K., and Kobayashi, S. (2001). Me31B silences translation of oocyte-localizing RNAs through the formation of cytoplasmic RNP complex during Drosophila oogenesis. Development 128, 3233-3242.

Palecek, S. P., Parikh, A. S., and Kron, S. J. (2000). Genetic analysis reveals that FLO11 upregulation and cell polarization independently regulate invasive growth in Saccharomyces cerevisiae. Genetics 156, 1005-1023.

Popolo, L., and Vai, M. (1999). The Gas1 glycoprotein, a putative wall polymer cross-linker. Biochim Biophys Acta 1426, 385-400.

Prado, F., Piruat, J. I., and Aguilera, A. (1997). Recombination between DNA repeats in yeast hpr1delta cells is linked to transcription elongation. Embo J 16, 2826-2835.

Reed, R., and Cheng, H. (2005). TREX, SR proteins and export of mRNA. Curr Opin Cell Biol 17, 269-273.

Reynolds, T. B., and Fink, G. R. (2001). Bakers' yeast, a model for fungal biofilm formation. Science 291, 878-881.

Richard, M., De Groot, P., Courtin, O., Poulain, D., Klis, F., and Gaillardin, C. (2002). GPI7 affects cell-wall protein anchorage in Saccharomyces cerevisiae and Candida albicans. Microbiology 148, 2125-2133.

Roberts, R. L., and Fink, G. R. (1994). Elements of a single MAP kinase cascade in Saccharomyces cerevisiae mediate two developmental programs in the same cell type: mating and invasive growth. Genes Dev 8, 2974-2985.

Rondon, A. G., Garcia-Rubio, M., Gonzalez-Barrera, S., and Aguilera, A. (2003). Molecular evidence for a positive role of Spt4 in transcription elongation. Embo J 22, 612-620.

Sadoff, B., Heath-Pagliuso, S., Castano, I., Zhu, Y., Kieff, F., and Christman, M. (1995). Isolation of mutants of Saccharomyces cerevisiae requiring DNA topoisomerase I. Genetics 141, 465-479.

Sheth, U., and Parker, R. (2003). Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. Science 300, 805-808.

Smillie, D. A., and Sommerville, J. (2002). RNA helicase p54 (DDX6) is a shuttling protein involved in nuclear assembly of stored mRNP particles. J Cell Sci 115, 395-407.

Smits, G. J., Kapteyn, J. C., van den Ende, H., and Klis, F. M. (1999). Cell wall dynamics in yeast. Curr Opin Microbiol 2, 348-352.

Sun, Z. W., and Hampsey, M. (1999). A general requirement for the Sin3-Rpd3 histone deacetylase complex in regulating silencing in Saccharomyces cerevisiae. Genetics 152, 921-932.

Uemura, H., Pandit, S., Jigami, Y., and Sternglanz, R. (1996). Mutations in GCR3, a gene involved in the expression of glycolytic genes in Saccharomyces cerevisiae, suppress the temperature-sensitive growth of hpr1 mutants. Genetics 142, 1095-1103.

van Berkel, M. A., Rieger, M., te Heesen, S., Ram, A. F., van den Ende, H., Aebi, M., and Klis, F. M. (1999). The Saccharomyces cerevisiae CWH8 gene is required for full levels of dolichol-linked oligosaccharides in the endoplasmic reticulum and for efficient N-glycosylation. Glycobiology 9, 243-253.

Xu, W., Smith, F. J., Jr., Subaran, R., and Mitchell, A. P. (2004). Multivesicular body-ESCRT components function in pH response regulation in Saccharomyces cerevisiae and Candida albicans. Mol Biol Cell 15, 5528-5537.

Zachariae, W., and Nasmyth, K. (1999). Whose end is destruction: cell division and the anaphase-promoting complex. Genes Dev 13, 2039-2058.

Chapter 3: Genes with internal repeats require the THO complex for transcription

ABSTRACT

The evolutionarily conserved multi-subunit THO complex, which is recruited to actively transcribed genes, is required for the efficient expression of *FLO11* and other yeast genes that have long internal tandem repeats. *FLO11* transcription elongation in Tho$^-$ mutants is hindered in the region of the tandem repeats, resulting in a loss of function. Moreover, the repeats become genetically unstable in Tho$^-$ mutants. A *FLO11* gene without the tandem repeats is transcribed equally well in Tho$^+$ or Tho$^-$ strains. The Tho$^-$ defect in transcription is suppressed by overexpression of topoisomerase I, suggesting that the THO complex functions to rectify aberrant structures that arise during transcription.

INTRODUCTION

Transcription involves a highly orchestrated series of events in which the core polymerase is joined by many additional proteins that promote initiation, elongation, and termination (Lee and Young, 2000; Maniatis and Reed, 2002; Arndt and Kane, 2003). Efficient transcription is also dependent upon the configuration of the DNA template as transcription creates negative supercoils behind the polymerase and positive supercoils ahead of it (Liu and Wang, 1987; Brill and Sternglanz, 1988; Drolet, 2006). These alterations in the superhelical density could permit intragenic repetitive sequences to form structures that impede the progress of the polymerase and promote recombination (Kim and Wang, 1989; Aguilera and Klein, 1990; Prado et al., 1997). The DNA landscape may therefore influence the efficiency of transcription, and some of the elongation factors could be required to remodel the template to permit efficient transcription.

The *Saccharomyces cerevisiae* multi-subunit THO complex, which has been identified as a possible elongation component, has been associated with many aspects of RNA and DNA metabolism (Fan et al., 1996; Piruat and Aguilera, 1998; Fan et al., 2001; Jimeno et al., 2002). The complex consists of four tightly bound proteins (Hpr1, Tho2, Thp1, Mft1) (Chavez et al., 2000), two of which (Hpr1 and Tho2) are conserved from yeast to humans (Reed and Cheng, 2005). Biochemical studies using natural templates have implicated the THO complex in recruiting the mRNA export proteins Sub2 (UAP56 in humans) and Yra1 (Aly1) to the mRNA in both yeast (Zenklusen et al., 2002) and humans (Reed and Cheng, 2005). In yeast, chromatin immunoprecipitation experiments

indicate that the THO complex is recruited to actively transcribed genes (Strasser et al., 2002; Abruzzi et al., 2004; Kim et al., 2004).

The biochemical analysis of the function of the THO complex has not led to a consistent picture. Experiments using a *GAL1* promoted *E. coli lacZ* reporter construct expressed in yeast suggested that transcription elongation of the *lacZ* gene is reduced in an *hpr1*Δ mutant (Chavez and Aguilera, 1997). Further analysis using a P_{GAL}-*lacZ* system indicated that in a Tho$^-$ mutant DNA:RNA hybrids are formed in vivo between the nascent transcript and the DNA template (Huertas and Aguilera, 2003). As the transcription of GC rich *lacZ* constructs was THO-dependent, whereas that of many endogenous yeast genes was not, it was proposed that the THO complex is required for efficient transcription elongation of long and GC rich genes (Chavez et al., 2001). Studies using a long 8 kb yeast ORF whose transcription was promoted by the *GAL1* promoter show clearly that THO complex mutants affect the processivity of transcription and not the elongation rate (Mason and Struhl, 2005). Moreover, the role of the THO complex in elongation has been questioned based on the insensitivity of Tho$^-$ mutants to mycophenolic acid, a presumed inhibitor of transcription elongation (Jensen et al., 2004).

Remarkably, the genetic analysis of Tho$^-$ mutants has not resolved these puzzles, and has provided little information on native genes that require THO complex function. Mutations in any of the four genes encoding the THO complex subunits do not result in inviability at normal growth conditions, suggesting that the THO proteins are not a core component of the elongation complex. However, one class of Tho$^-$ mutants (*hpr1* –hyperrecombination) was first identified because a mutation in that gene increases the frequency of recombination between artificial tandem repeats constructed by

transformation (Aguilera and Klein, 1990). The apparent participation of the THO complex in some aspects of transcription and recombination contrasts with the absence of a phenotype in Tho$^-$ mutants that is associated with resident genes. Moreover, in *Drosophila*, loss of THO complex function results in only minor differences in transcription profiles as revealed by whole genome arrays (Rehwinkel et al., 2004).

Sequence analysis and synthetic interactions of THO genes provide insight into their function. One member of the THO complex, Hpr1, has amino acid sequence similarity to the topoisomerase Top1 (Aguilera and Klein, 1990). Top1 has been proposed to relax the negative supercoils that are generated during transcription (Liu and Wang, 1987; Brill and Sternglanz, 1988; Masse and Drolet, 1999a). The sequence similarity between Top1 and Hpr1 is likely to reflect functional redundancy with respect to DNA metabolism because *top1*Δ *hpr1*Δ double mutants are lethal (Aguilera and Klein, 1990; Sadoff et al., 1995). In addition, chromatin immunoprecipitation using tagged Hpr1 suggests that the THO proteins may be associated with DNA rather than RNA (Abruzzi et al., 2004).

In this report we show that THO function is required for the transcription of several genes containing multiple internal tandem repeats. The affected genes are not especially long and neither the genes nor the repeats are GC rich. The defect in transcription appears to be in transcription elongation, based on chromatin immunoprecipitation experiments designed to reveal RNA polymerase occupancy. Transcription is restored in Tho$^-$ mutants when the repeats are removed from the gene. As whole genome arrays comparing Tho$^+$ and Tho$^-$ strains do not reveal any general defects in transcription, these effects appear to be restricted to a subset of genes with

internal repeats. The fact that the transcriptional defects in Tho$^-$ mutants can be suppressed by overexpression of *TOP1* suggests a model in which the THO complex functions as an accessory complex that facilitates transcription past obstructive DNA configurations.

METHODS

Yeast strains and growth conditions

Strains in two genetic backgrounds, S288c and Σ1278b, were used in these studies (Table 1). The S288c *FLO8+* strain L8046 was prepared by transforming a pRS305-based *Bgl*II-cut integrating plasmid that contains a Σ1278b copy of *FLO8* (B4241) into the S288c *flo8-* strain L4242. Strains with a *FLO11* allele that lacks the repeats region, *flo11::Δrep*, were constructed in two steps. First, the *URA3* marker was amplified from a plasmid with primers V271 and V272 targeting the ends of the *FLO11* repeats region. Second, these *FLO11::URA3* strains were plated on FOA to loop out the marker.

The *LYS2+3rep* strain that has three *FLO1* repeats inserted at position 720 nt of *LYS2* was prepared in the following way. A *FLO1rep-URA3-FLO1rep* cassette was amplified from genomic DNA of strain KV133 (Verstrepen et al. 2005) with primers K428 and K429 to create overhangs for in-frame integration at *LYS2* in the strain BY4741. Transformants that were Ura$^+$ and Lys$^-$ were then streaked on FOA or SC-LYS plates to force FOA popouts, leaving behind *FLO1* repeats in *LYS2*. The *LYS2+3rep* chimera construct was confirmed by sequencing.

Strains were grown in yeast extract/peptone dextrose (YPD), unless selective media was required. Cold shock and anaerobic growth experiments were based on

previously described protocols (Abramova et al., 2001): for cold shock, cultures were grown at 30°C to O.D. $_{600}$ 1.0, and then shifted to 15°C for 90 min; strains were grown anaerobically on YPD plates supplemented with 0.5% Tween 80 and 20 μg/ml ergosterol (Sigma), and placed in an anaerobic chamber with an AnaeroPack sachet (MGC) for 3 days at 30°C. A Bioscreen apparatus (Labsystems) was used for the growth comparison of *LYS2+3rep* strains. Several reagents were used for selection or counterselection during the preparation of strains: geneticin (Gibco) at 0.2 mg/ml, hygromycin (Sigma) at 0.3 mg/ml, nourseothricin (Werner BioAgents) at 0.1 mg/ml, and 5'fluoroorotic acid (USBiological) at 1mg/ml.

The frequency of Ura$^-$ segregants of *FLO11::URA3* Tho$^+$ (XY266) or *thp2*Δ (XY454) strains was determined after growth on YPD plates for 1 day at 30°C, followed by plating on FOA and SC-URA plates at appropriate dilutions to count colony forming units.

Primers and plasmid construction

Primers are listed in Table 2. Primer pairs for real-time PCR analysis were designed with Primer Express software. The primer pairs along *FLO11* for chromatin immunoprecipitation analysis were designed to yield products of 250-300 bp. Primers for amplification of an untranscribed region on chromosome V were as previously described (Keogh and Buratowski, 2004).

mRNA analysis

Total RNA was isolated from 10 ml cultures grown to O.D. $_{600}$ 1.0 using hot acid phenol. DNaseI treatment was carried out for 30 min (Epicentre). Reverse transcription of 0.3 μg RNA was performed for 30 min at 48°C with 12.5 U of MultiScribe reverse transcriptase (Applied Biosystems) and 2.5 μM random hexamers. One seventh of the cDNA product was used for real-time PCR analysis with reagents from Applied Biosystems and the ABI 7500 real-time PCR system. Probes at the 3' end of ORFs were used when available. Normalization was to *ACT1*, except when analyzing Σ1278b Tho$^-$ mutants, where we noticed a slight upregulation of *ACT1* in Tho$^-$ mutants compared to other controls. In those cases, normalization was to *SCR1*, a gene transcribed by RNA polymerase III. The histograms present data from two to four independent experiments.

Chromatin Immunoprecipitation

Chromatin immunoprecipitations were performed as described in Keogh and Buratowski (2004). Briefly, cells were grown to O.D. $_{600}$ 0.8-1.0, fixed with formaldehyde, lysed, and sonicated. The lysates were immunoprecipitated with an anti-Rpb3 antibody (Neoclone) bound to Protein G Sepharose beads (Amersham Biosciences). Overnight incubation at 4°C was followed by four washes. The protein/DNA complexes were eluted, and the crosslinks were reversed with pronase (Calbiochem). DNA was analyzed by concurrent PCR of a *FLO11* region and an untranscribed region on chromosome V. All samples were resolved on a 6% polyacrylamide gel, and the signals were quantitated by a phosphorimager and ImageQuant software. Occupancy value for each of six regions

along *FLO11* was calculated as a ratio (IP sample/input sample) of ratios (*FLO11* specific signal/untranscribed region signal).

The ChIP assays were performed both on strains in the S288c and Σ1278b backgrounds. Although there were quantitative differences in the relative enrichment, in both backgrounds the polymerase occupancy in the Tho$^-$ strains was reduced in the 3' end of the *FLO11* strain. Better enrichment of the specific signal in the IP sample was observed for S288c than for Σ1278b strains.

Bioinformatics

GC content of DNA sequences was determined with the EMBOSS GEECEE software.

RESULTS

Reduction of *FLO11* mRNA levels in Tho$^-$ mutants requires the *FLO11* coding sequence

As was shown in Chapter 2, the non-adherent phenotype of the Tho$^-$ mutants is independent of the strain background and the promoter. When any of the THO complex subunits, *THP2*, *MFT1*, *HPR1*, and *THO2*, is separately deleted in a Σ1278b strain in which *FLO11* is under its native promoter at its resident site in the chromosome, each of the four Tho$^-$ mutants is also strongly non-adherent in this background (Fig. 1A). Thus, our screen identified the THO complex as a novel promoter-independent regulator of *FLO11*.

In Chapter 2 we found that *FLO11* expression was significantly reduced in *thp2Δ* mutant compared to a Tho$^+$ strain. This is the case for the *mft1Δ* mutants as well (Fig. 1B). The reduction appears to be independent of the promoter sequence, as the *FLO11*

levels are reduced both when *FLO11* is expressed from the *TEF1* promoter and from its native promoter. To determine whether the *FLO11* coding sequence was responsible, we analyzed the transcript levels in $P_{FLO11}GFP$ strains in which *GFP* replaces the *FLO11* ORF. *GFP* mRNA and *FLO11* mRNA were compared by rtPCR in the corresponding Tho$^+$ and *mft1Δ* strains. The level of *FLO11* mRNA is reduced about 85% in the *mft1Δ* mutant, whereas *GFP* expression from the *FLO11* promoter is nearly at wild-type levels in the Tho$^-$ mutant background (Fig. 1B). The lack of an effect of the Tho$^-$ mutants on $P_{FLO11}GFP$ can also be visualized by the roughly equivalent GFP fluorescence in Tho$^+$ and *mft1Δ* strains (Fig. 1C). This result suggests that *FLO11* mRNA downregulation in a Tho$^-$ mutant depends upon the presence of the *FLO11* coding sequence.

***FLO11* requires the THO complex for transcription elongation through the repeats**

RNA polymerase II (RNAP II) occupancy along the *FLO11* ORF was monitored in Tho$^+$ and Tho$^-$ strains by chromatin immunoprecipitation (ChIP) using an antibody to the Rpb3 subunit of the polymerase. The amount of *FLO11* DNA in the precipitate was assessed by PCR amplification. *FLO11* is an ORF of 4104 nt, the middle third of which features 15 nearly perfect tandem repeats of 1725 nt total length (Verstrepen et al., 2005). We designed 6 primer pairs along *FLO11*: one in the promoter region, two in the 5'-end proximal region, two in the 3'end region, and one in the 3' UTR (Fig. 2A). This ChIP analysis shows a gradual reduction in the level of RNAP II along *FLO11* in the *mft1Δ* *thp2Δ* mutant as compared to wild-type (Fig. 2B and C). The fact that the Tho$^-$ strain has comparable or slightly higher occupancy of RNAP II at the 5' end of the *FLO11* ORF indicates that Tho$^-$ mutants do not reduce transcription initiation of *FLO11*. At the same

time, reduced signal for the 3' end probes of *FLO11* in the *mft1Δ thp2Δ* mutant suggests lower RNAP II occupancy along the *FLO11* ORF sequence. This result indicates that the THO complex is not involved in transcription initiation but rather in transcription elongation of *FLO11*.

To examine the role of the *FLO11* repeats on transcription, we constructed a *FLO11* allele that lacks the repeat-containing region (*flo11::Δrep)* and compared the levels of *FLO11* transcription in Tho$^+$ and *mft1Δ thp2Δ* strains. *FLO11* expression is at least 65% reduced in *mft1Δ thp2Δ* compared to the wild-type strain, whereas *flo11::Δrep* expression in the mutant strain is nearly the same as that in Tho$^+$ (Fig. 2D), suggesting that the repeat region in *FLO11* is the major obstacle to transcription elongation in the Tho$^-$ mutant background.

The obstacle to transcription caused by the repeats in a Tho$^-$ mutant has a profound consequence on the genetic stability of the repeats. The stability of the repeats was measured in a *FLO11::URA3* genomic construct that contains the *URA3* gene inserted among the *FLO11* repeats (Fig. 2E). Loss of the *URA3* gene is a direct measure of the gain or loss of integral numbers of repeats (Verstrepen et al., 2005). In a Tho$^+$ strain the repeats are relatively stable, being lost at about 1.8×10^{-5}, whereas in the *thp2Δ* mutant the repeats are lost at 7.2×10^{-4} (Fig. 2E). The 40-fold higher frequency of Ura$^-$ segregants in the Tho$^-$ strain compared with Tho$^+$ suggests a greater instability of the repeats region.

Other genes with repeats require the THO complex for efficient transcription

TIR1 is a cell wall gene encoded by 765 nt, 261 of which are internal tandem repeats. The gene is required for anaerobic growth, and is induced by cold shock as well as by low oxygen levels (Abramova et al. 2001). The level of *TIR1* mRNA was measured in a Tho^+ strain and in Tho^- mutants after a 90-minute cold shock at 15°C. There is an approximately 50% reduction of *TIR1* expression in *mft1Δ* and *thp2Δ*, and more than 75% reduction in *hpr1Δ* and *tho2Δ* mutants (Fig. 3A). When grown hypoxically, *mft1Δ* and *thp2Δ* mutants show a modest growth defect, and *hpr1Δ* and *tho2Δ* a strong growth defect (Fig. 3B), consistent with the *TIR1* downregulation.

The expression of *FLO1*, another gene with long tandem repeats, is also reduced in Tho^- mutants (Fig. 3C). Flo1 is required for flocculation between yeast cells, and the reduction of *FLO1* mRNA levels is reflected in the reduced flocculation of Tho^- strains (Fig. 3D). There is no extensive sequence homology between the repeats in *FLO1* and those in *FLO11*. Several other genes with repeats (*FIT3*, *TIR4*, see Table 3) show similar dependence on the THO complex.

For comparison, we also measured the mRNA levels of several ORFs without internal repeats, of varying length and expression levels: *PMA1* (2757 nt) and *TEF1* (1377 nt), highly expressed genes, and *LYS2* (4179 nt), a gene expressed at lower levels. Expression of all three genes is unaffected in the *mft1Δ thp2Δ* mutant (Fig. 3C).

Intragenic repeats confer THO dependence

An in-frame segment containing three *FLO1* repeats (a total of 414 nt) was inserted into the *LYS2* gene to test the effect of these repeats on transcription of that gene (Fig. 4A).

LYS2 is not affected by Tho$^-$ mutants when transcribed from its cognate promoter (Fig. 3B). *LYS2+3rep* expression is 35% less in the *mft1*Δ mutant than in the Tho$^+$ strain (Fig. 4B). This difference is reflected in the growth defect of the *LYS2+3rep mft1*Δ mutant compared to the *LYS2+3rep* Tho$^+$ strain in SC-LYS (Fig. 4C and D). These data suggest that *FLO1* repeats confer THO dependence.

Overexpression of *TOP1* suppresses the Tho$^-$ defect

The partial homology between Hpr1 and Top1 (Aguilera and Klein, 1990) and the lethality of *hpr1*Δ *top1*Δ or *mft1*Δ *top1*Δ double mutants suggested an overlap between topoisomerase and THO complex function. To test this possibility, we constructed a Tho$^-$ and a Tho$^+$ strain that contained the *TOP1* gene under the highly expressed *TEF* promoter and compared these strains with Tho$^-$ and Tho$^+$ strains without the overexpression construct. *TOP1* overexpression in Tho$^-$ mutants partially restores *FLO11* mRNA levels (Fig. 5A) as well as adherence to agar (Fig. 5B). Thus, Top1 partially complements THO complex function for the efficient transcription of genes with long internal repeats.

DISCUSSION

Our data are consistent with the postulated role of the THO complex in resolving R loop DNA that may form when the transit of the RNA polymerase is impeded (Huertas and Aguilera, 2003). The model in Fig. 6 is an illustration of the proposed mechanism. A key feature of our model is that in Tho$^-$ mutants the tandemly repeated DNA found in many naturally occurring genes creates a barrier of mispaired DNA and RNA that interrupts transcription and requires the THO complex for resolution. Intragenic tandem repeats

particularly increase the chances of energetically-favorable DNA and RNA base-pairing because of the presence of identical neighboring DNA and RNA sequences. There is considerable experimental evidence that transcription elongation generates supercoiling both upstream and downstream of the moving RNA polymerase (Liu and Wang, 1987), and that R loop formation occurs if there is excessive supercoiling, which would be prevented by the action of topoisomerase 1 (Masse and Drolet, 1999b; Drolet 2006). In our model, some aspect of topoisomerase 1 function in this process can be carried out by the THO complex.

Previous studies are consistent with this model. First, several studies show that the THO complex is required for processivity and not transcription initiation (Chavez et al., 2000; Strasser et al., 2002; Kim et al., 2004; Mason and Struhl, 2005). Second, the THO complex appears to be associated with the DNA rather than the RNA during transcription (Abruzzi et al., 2004). Third, there appears to be a functional overlap between topoisomerase 1 function and THO complex function as the *top1*Δ *hpr1*Δ mutant is lethal, and Hpr1 has sequence similarity to the C-terminal region of Top1 (Aguilera and Klein, 1990). Fourth, in the absence of THO complex function, DNA:RNA hybrids have been detected during transcription (Huertas and Aguilera, 2003). Fifth, the formation of R loops has been associated with increased frequency of recombination (Huertas and Aguilera, 2003; Li and Manley, 2005) and with the absence of topoisomerase activity (Masse and Drolet, 1999b). Sixth, *top1*Δ as well as each of the THO complex mutants of yeast (*hpr1*Δ, *tho2*Δ, *thp2*Δ, *mft1*Δ) have dramatically enhanced frequencies of recombination in regions of repeated DNA (Christman et al., 1988; Aguilera and Klein, 1990; Chavez et al., 2000).

The genes affected by mutation of the THO complex have a number of similarities. The most salient attribute is that they are genes with many long tandem internal repeats. Genes with long internal tandem repeats are not a feature restricted to the yeast genome. It is estimated that 5% of human genes also have tandem repeats (O'Dushlaine et al., 2005). As we showed previously (Verstrepen et al., 2005), most of the yeast genes with internal repeats encode cell wall proteins, and the repeats are essential for cell surface interactions such as adhesion. Here we show that alleles of genes with large numbers of repeats require the THO complex for maximum expression.

Although several previous studies using recombinant constructs have suggested that the THO complex was required either for genes of high GC content or especially long genes, the yeast genes whose expression is dramatically affected do not have a high GC content (*FLO11* 46% (50% for the region of repeats), Table 3). The *FLO11* and *FLO1* genes are longer than the average yeast gene; however, transcription of yeast genes of equivalent size (*RPB1* and *LYS2*) is unaffected in Tho$^-$ mutants under standard growth conditions, and a third THO-dependent gene, *TIR1*, is only 765 nt long. Moreover, in a Tho$^+$ strain there is little difference between the expression of the long (4.1 kb) or short (2.5 kb) form of the *FLO11* gene. However, efficient transcription of the wild-type *FLO11* gene containing the repeats is dependent on a functional THO complex, whereas a *FLO11* gene without the repeats (*flo11::*Δ*rep*) is expressed at the same level in both Tho$^+$ and Tho$^-$ strains.

The presumed importance of the THO complex for maintaining the topology of the DNA template contrasts with the failure of previous studies to identify phenotypic effects of Tho$^-$ mutants on native genes. In addition, we failed to detect any dramatic

global change in the level of transcription for most genes as measured by whole genome microarrays in yeast. A similar analysis in *D. melanogaster* concluded "...the vast majority of genes are transcribed and exported independently of THO...." (Rehwinkel et al., 2004). We posit that for most genes the activity of Top1 is sufficient to prevent the topological impediments to transcription elongation. However, for genes that have repeated offensive sequences, such as the *FLO* genes, the stress on the system overwhelms the ability of Top1 to correct the defect. Under these conditions the THO complex becomes essential.

This view raises the question: is the THO complex required only for efficient transcription of genes with long tandem repeats? We think this is unlikely. First, not all genes with tandem repeats show a phenotype in the Tho$^-$ strains (Table 3). Of course, many of these genes with repeats are expressed at extremely low levels and may, like *TIR1*, only require the THO complex upon induction or some environmental stress condition that requires enhanced transcription. Second, other genes whose transcription creates stable R loops under stress conditions could also require the THO complex. It is in this sense that we posit the THO complex as a protein complex whose function is to repattern the transcription complex permitting efficient transcription elongation when transcription stalls.

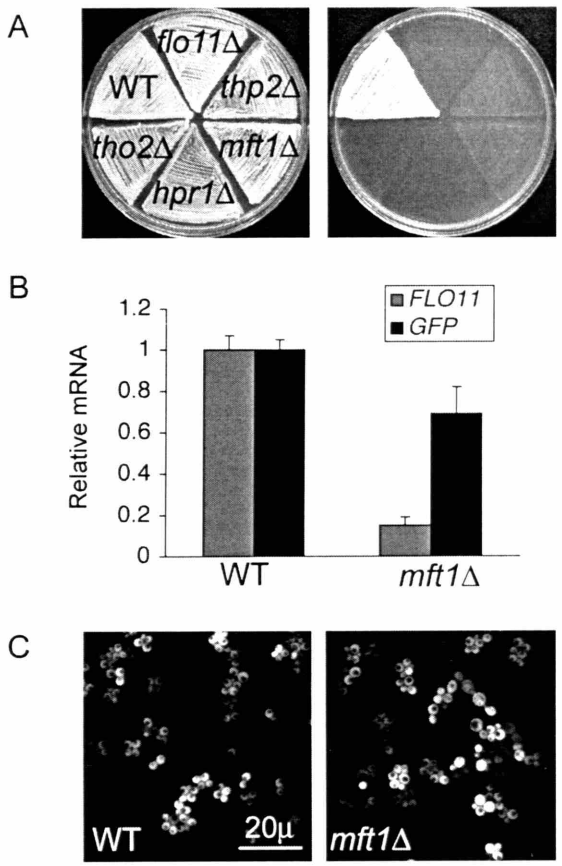
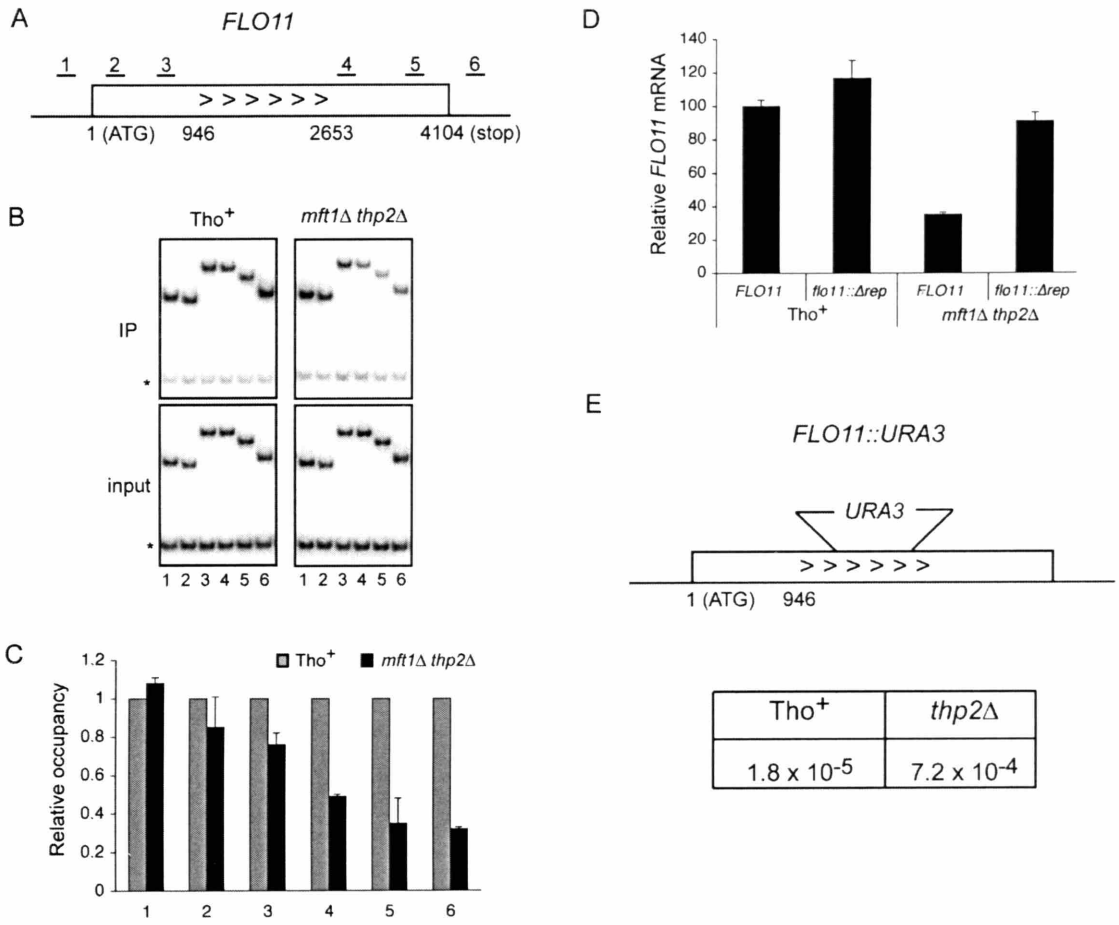


Figure1: Reduction of *FLO11* mRNA levels in Tho⁻ mutants requires the *FLO11* coding sequence

A. THO complex mutants (Tho⁻) are also defective for adherence in the Σ1278b background. Strains: wild type (10560-23C), *flo11Δ* (L7558), *thp2Δ* (XY16), *mft1Δ* (XY118), *hpr1Δ* (XY189), *tho2Δ* (XY191). On the left is a YPD plate after 3 days of incubation at 30°C, and on the right is the same plate after wash.

B. *GFP* transcription from the *FLO11* promoter in a $P_{FLO11}GFP$ fusion is unaffected by Tho⁻ mutants. The histogram compares *FLO11* (grey bars: strains as in C) and *GFP* mRNA levels (black bars: Tho⁺ (L8225), *mft1Δ* (XY136)) by real-time PCR.

C. GFP fluorescence indicates *FLO11* promoter functionality in Tho⁻ mutants. Images show GFP fluorescence of exponentially growing Tho⁺ (L8225) or *mft1Δ* (XY136) cells.



Tho+	thp2Δ
1.8×10^{-5}	7.2×10^{-4}

Figure 2: Tho⁻ mutants show a transcription elongation defect through the repeats of *FLO11*.

A. *FLO11* probes. The relative position of six DNA fragments amplified for chromatin IP analysis is shown with respect to the start codon, stop codon, and the intragenic repeats of *FLO11*. Exact positions of the primers are given in Table 2.

B. Transcription elongation of *FLO11* is defective in Tho⁻ mutants. RNA polymerase II abundance along *FLO11* was monitored by anti-Rpb3 chromatin IP using the six primer pairs in A (upper bands in each panel). The lower bands (*) correspond to a 180 bp non-transcribed region on chromosome V (Keogh and Buratowski, 2004). Strains: Tho⁺ (L8046) and *mft1Δ thp2Δ* (XY269).

C. Phosphorimager quantitation of the elongation assay in B from two independent experiments. Each of the values for probes 1-6 in the Tho⁺ strain was normalized to 1. The values for the double mutant were normalized to the corresponding wild-type probe. The raw occupancy values for the probes in the wild-type strain were typically between 3 and 8.

D. Removal of the repeats restores transcription of *FLO11* in a Tho⁻ mutant. Real-time PCR of strains: wild-type (L8046), *flo11::Δrep* (XY369), *mft1Δ thp2Δ* (XY269), *mft1Δ thp2Δ flo11::Δrep* (XY356).

E. *FLO11* repeats are less stable in a Tho⁻ mutant. The frequency of recombination was determined by measuring the frequency of Ura⁻ segregants obtained from a *FLO11::URA3* Tho⁺ (XY266) or *thp2Δ* (XY454) strain where *URA3* is flanked by *FLO11* repeats. Average measurements from four independent experiments are shown. Standard deviations were 0.6×10^{-5} for the Tho+ strain, and 8.7×10^{-5} for the *thp2Δ* mutant strain.

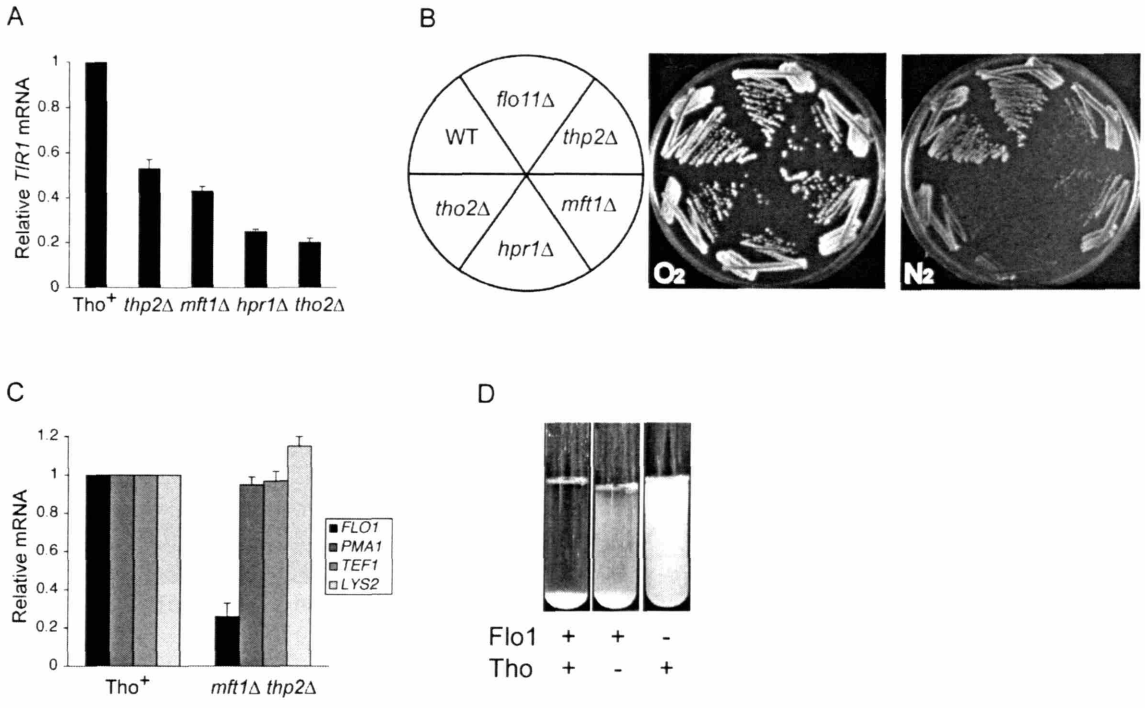


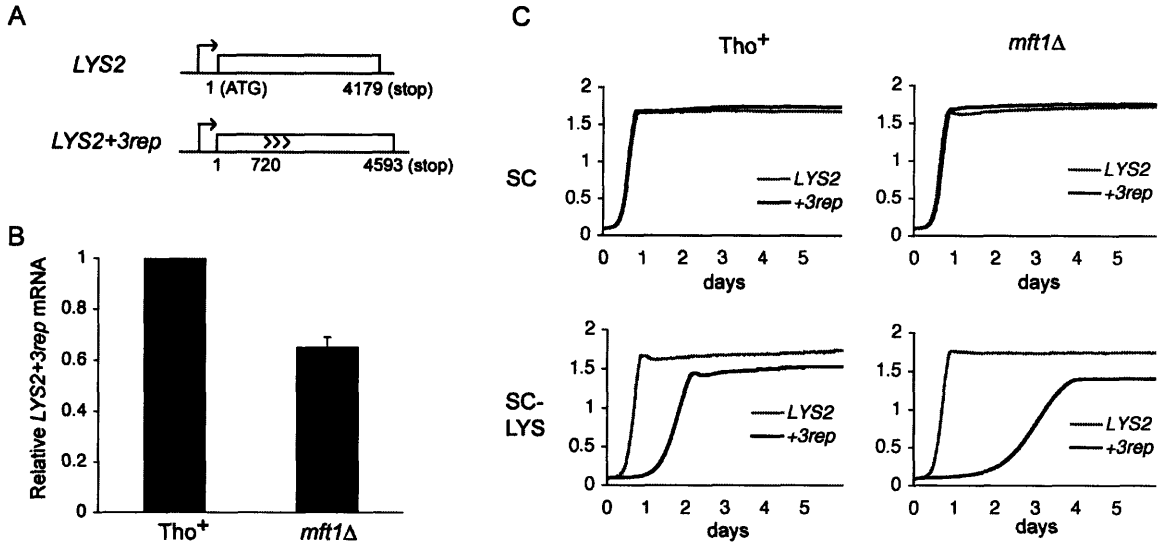
Figure 3: Expression of other genes with repeats is reduced in Tho⁻ mutants.

A. *TIR1* requires the THO complex for transcription. Expression of *TIR1* was induced by a 90-min cold-shock and measured by rt-PCR. Strains: Tho⁺ (10560-23C), *thp2Δ* (XY16), *mft1Δ* (XY118), *hpr1Δ* (XY189) and *tho2Δ* (XY191).

B. Tho⁻ mutants show an anaerobic growth defect.
Strains: wild type (10560-23C), *flo11Δ* (L7558), *thp2Δ* (XY16), *mft1Δ* (XY118), *hpr1Δ* (XY189), *tho2Δ* (XY191). The strains were streaked on YPD plates supplemented with 0.5% Tween 80 and 20 μg/ml ergosterol, and grown at 30°C aerobically or in a hypoxic chamber for 3 days.

C. *FLO1* transcription is reduced in Tho⁻ mutants. Strains: Tho⁺ (L8046), *thp2Δ mft1Δ* (XY269). Expression of three other genes, *PMA1*, *TEF1*, and *LYS2*, was measured for comparison.

D. Reduced *FLO1* expression in Tho⁻ strains results in reduction of flocculence, a trait specified by *FLO1*. Strains: Tho⁺ (L8046), *thp2Δ mft1Δ* (XY269), negative control—a strain that does not express *FLO1* (BY4741). Cells from an overnight culture were diluted to 0.1 OD/ml and grown at 30°C for 24 hrs. The test tubes were vortexed and immediately photographed.



D

Strain \ medium	SC	SC-LYS
Tho⁺ *LYS2*	1.5	1.7
mft1Δ LYS2	1.6	1.8
Tho⁺ *LYS2+3rep*	1.5	2.3
mft1Δ LYS2+3rep	1.6	5.4

Figure 4: *FLO1* repeats create THO dependence

A. *LYS2+3rep* chimera. Three *FLO1* repeats (414 nt total) were inserted at position 720 of *LYS2*.

B. Insertion of *FLO1* repeats at *LYS2* leads to reduced expression in Tho⁻ mutants. Tho⁺ (XY299) and *mft1Δ* mutant strains (XY313) carrying the *LYS2+3rep* allele were grown in SC 2% Glc to OD_{600} 1, and then shifted to SC-LYS 2% Glc for 2 hrs. Real-time PCR data from two independent experiments is shown.

C. The growth defect of strains with a *LYS2+3rep* allele is greater in Tho⁻ mutants. Strains: BY4741 (Tho⁺, *LYS2*), XY299 (Tho⁺, *LYS2+3rep*), *mft1Δ* (*LYS2*), and XY313 (*mft1Δ*, *LYS2+3rep*). The strains were grown overnight in SC 2% Glc, and diluted in a Bioscreen plate in SC 2% Glc or in SC-LYS 2% Glc in triplicate. The plate was incubated for 5.5 days with OD readings taken every 30 minutes (arbitrary optical density units on Y axis).

D. Prolonged doubling time of strains with a *LYS2+3rep* allele in Tho⁻ mutants. Overnight cultures were diluted to 0.1 OD/ml in 5ml SC or SC-LYS. OD readings were taken during the course of a day. The average doubling time (in hours) from 3 independent experiments for each strain is shown. Standard deviations were less than 10% of the corresponding average value.

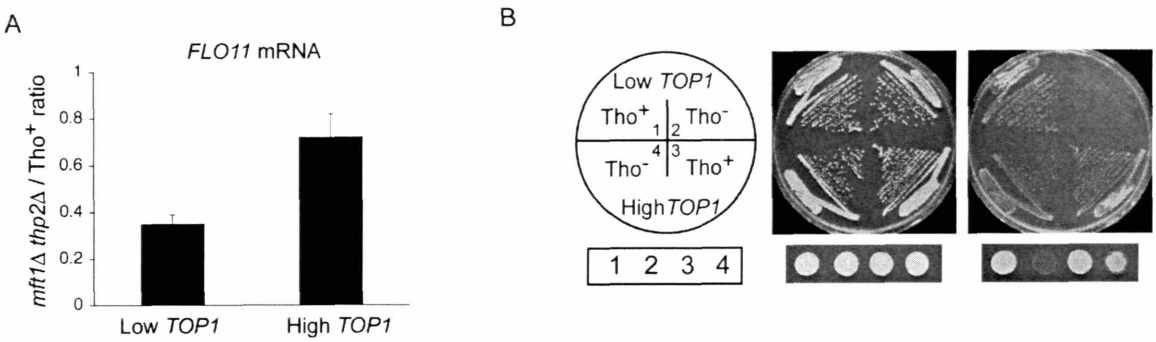


Figure 5: *TOP1* overexpression partially restores *FLO11* mRNA levels and function.

A. Expression of *TOP1* from the strong *TEF* promoter restores *FLO11* expression in Tho$^-$ mutants. Low *TOP1* is the *FLO11* mRNA ratio in Tho$^-$/Tho$^+$ strains with *TOP1* under its own promoter (Tho$^+$, L8046; *mft1Δ thp2Δ*, XY269). High *TOP1* is the *FLO11* mRNA ratio in Tho$^-$/Tho$^+$ strains with *TOP1* under the *TEF* promoter (Tho$^+$, XY426; *mft1Δ thp2Δ*, XY427). Expression of *TOP1* from the *TEF* promoter is 9 times higher than from the endogenous *TOP1* promoter (as measured by rt-pcr).

B. Overexpression of *TOP1* also partially restores adherence to agar of Tho$^-$ mutants. Strains: 1, Tho$^+$ (L8046); 2, *mft1Δ thp2Δ* (XY269); 3, Tho$^+$, high *TOP1* (XY426); 4, *mft1Δ thp2Δ*, high *TOP1* (XY427). The strains were streaked on a YPD plate, incubated for 2 days at 30°C, and photographed before and after wash. Alternatively, $2x10^6$ cells were spotted on a YPD plate, incubated for 1 day at 30°C, and washed.

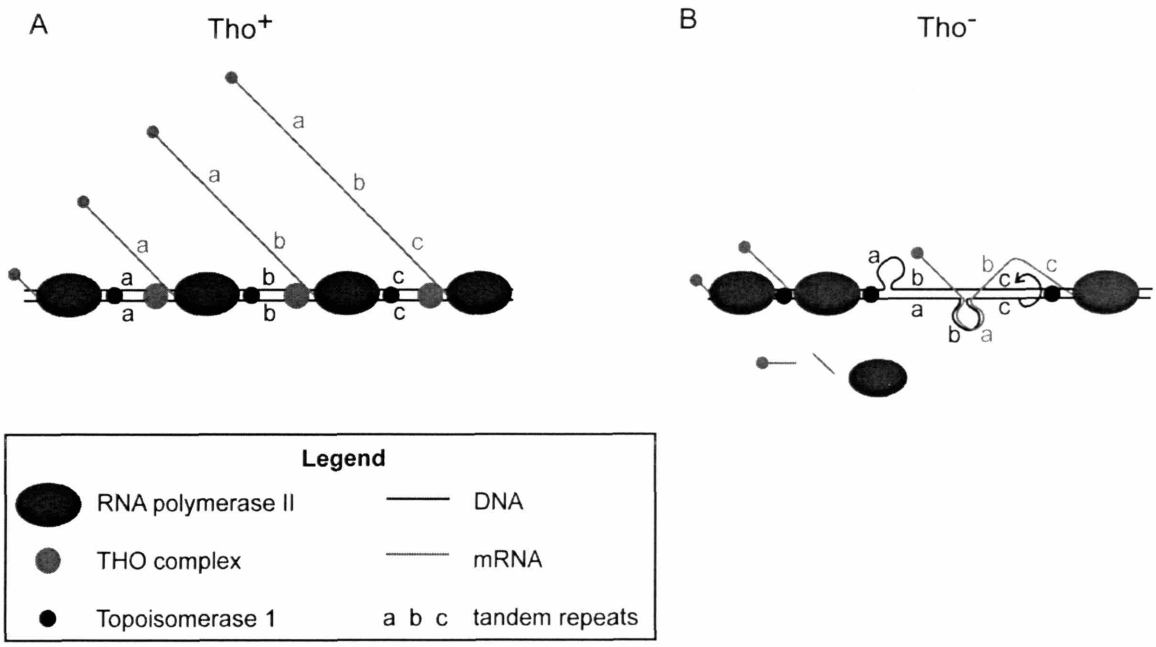


Figure 6: Transcription of intragenic repeats requires the THO complex.

A. The THO complex is required for transcription elongation of genes with internal repeats. The THO complex (grey circle) as well as Top1 (black circle) control DNA topology during transcription by RNA polymerase II (black oval) to prevent supercoiling and aberrant base paring between repetitive DNA (black) and RNA (grey) sequences (a, b, and c represent consecutive repeats on the coding and non-coding strands, and on the transcript).

B. Tandem intragenic repeats impede transcription elongation in Tho$^-$ mutants. In Tho$^-$ mutants, aberrant DNA and RNA topology facilitates DNA:RNA hybridization of repetitive sequences (one of many possible combinations of mispairing is depicted). Top1 activity is insufficient to relieve transcription-induced supercoiling, resulting in excessive negative supercoiling (arrow). Supercoiling and DNA:RNA hybrids cause the RNA polymerase II complex to stall or disengage from the DNA template.

Table 1: Yeast strains

Strain	Genotype	Reference/ Source
BY4741	S288c *MATa his3^1 leu2^0 met15^0 ura3^0*	
mft1Δ	S288c *MATa his3^1 leu2^0 met15^0 ura3^0 mft1::KanMX4*	Invitrogen *mft1Δ*
thp2Δ	S288c *MATa his3^1 leu2^0 met15^0 ura3^0 thp2::KanMX4*	Invitrogen *thp2Δ*
10560-23C	Σ1278b *MATα ura3-52 his3::hisG leu2::hisG*	Fink laboratory collection
L7558	Σ1278b *MATα ura3-52 his3::hisG leu2::hisG flo11::kanMX4*	Fink laboratory collection
XY16	Σ1278b *MATα ura3-52 his3::hisG leu2::hisG thp2::kanMX4*	This study
XY118	Σ1278b *MATα ura3-52 his3::hisG leu2::hisG mft1::kanMX4*	This study
XY189	Σ1278b *MATα ura3-52 his3::hisG leu2::hisG hpr1::kanMX4*	This study
XY191	Σ1278b *MATα ura3-52 his3::hisG leu2::hisG tho2::kanMX4*	This study
L8255	Σ1278b *MATα ura3-52 trp1::hisG his3::hisG leu2::hisG flo11::yEGFP-URA3*	Fink laboratory collection
XY136	Σ1278b *MATα ura3-52 trp1::hisG his3::hisG leu2::hisG flo11::yEGFP-URA3 mft1::kanMX4*	This study
L8046	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52*	Fink laboratory collection
XY266	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 FLO11::URA3*	This study
XY269	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 thp2::kanMX4 mft1::HygR*	This study
XY356	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 thp2::kanMX4 mft1::HygR flo11::Δrep*	This study
XY369	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 flo11::Δrep*	This study
XY299	S288c *MATa his3^1 leu2^0 met15^0 ura3^0 LYS2+3rep*	This study
XY313	S288c *MATa his3^1 leu2^0 met15^0 ura3^0 LYS2+3rep mft1::KanMX4*	This study
XY426	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 NatNT2-P_{TEF}TOP1*	This study
XY427	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 thp2::kanMX4 mft1::HygR NatNT2-P_{TEF}TOP1*	This study
XY454	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 thp2::kanMX4 FLO11::URA3*	This study

Table 2: Primers used in this study

ORF, position	Name	Sequence
Chromatin IP		
FLO11 5' f -296	V241	TGTCTTATCTGAGGAATGTCCGTG
FLO11 5' r -44	V242	ATTAGAACCACAACATGACGAGGG
FLO11 f +93	V243	CTCCGAAGGAACTAGCTGTAATTCT
FLO11 r +345	V244	TTCGTTGTAACCGTATAGTTGGACG
FLO11 f +603	V247	GACAATAATTGTGGCGGTACGAAG
FLO11 r +918	V248	AGTGCATGTCTTAGATGTGGTAGT
FLO11 f +2752	V249	ACCGAAACTACCATTGTTCCAACT
FLO11 r +3058	V250	GTTTCGCTTGGACTGGTTGAACAT
FLO11 f +3737	V253	AGTCATCTGTTGGTACTAACTCCG
FLO11 r +4037	V254	CCTTGGTAAGTACTCGAGATAGAAGG
FLO11 3' f +258	V255	ACAAGTACCGGTAGTATTGGCAC
FLO11 3' r +523	V256	CCTCTTATTCATCAAAGCCTGGTC
Chr V no-ORF f	contr f	GGCTGTCAGAATATGGGGCCGTAGTA
Chr V no-ORF r	contr r	CACCCCGAAGCTGCTTTCACAATAC
real-time PCR		
SCR1 f 384	A47	CGGCCGGGATAGCACATA
SCR1 r 437	A48	CGCCGAAGCGATCAACTT
TEF f 349	V216	GCTGGTGGTGTCGGTGAATT
TEF r 410	V217	GCGTGTTCTCTGGTTTGACCAT
FLO11 f 3038	V290	GTTCAACCAGTCCAAGCGAAA
FLO11 r 3104	V291	GTAGTTACAGGTGTGGTAGGTGAAGTG
GFP f 24	V292	CACTGGTGTTGTCCCAATTTTG
GFP r 94	V293	CACCGGAGACAGAAAATTTGTG
FLO1 f 4268	V298	TAGCTGCTGAGACGATTACCAA
FLO1 r 4348	V299	GCGTGATTAGATCTTGAAAGCGAA
LYS2 f 2081	V308	GTTGCGGTAGGGCTGATGA
LYS2 r 2150	V309	TGCGTATCTATTTCTCCTAATTCGATT
PMA1 f 2451	V363	GTCTTCTATCCCATCCTGGCAAT
PMA1 r 2532	V364	CCAACCGAATAAGGTAAACATGGT
RPB1 f 4268	V369	GTGTTTCGGAAAATGTCATTCTTG
RPB1 r 4346	V370	GACTCCTCATCGATCATCACATCA
HSP150 f 99	V379	CACTTACAGCGGTGGTGTTACC
HSP150 r 162	V380	GGAGATTGGTTGAACGGCAAT
FIT3 f 409	V387	AACACCTGGTCTCCAAGTAGTACTTCT
FIT3 r 481	V388	TGGTTGCAGTGGTTGAAGCA
TIR1 f 624	V401	CAAGACCTCTGCCATCTCTCAA
TIR1 r 690	V402	TTGCTCAGAAACAGCCTTGGT
PIR1 f 844	V403	CCACAAGCTGGTGCTATCTATGC
PIR1 r 910	V404	CACCGATGGCCAAGTTACCT
TIR4 f 1125	V451	TGACAACACTCTTGTCACCAAAGA
TIR4 r 1188	V452	TGGAACGGCAGATGTTGAAG
ACT1 f 1038	V453	GGCTTCTTTGACTACCTTCCAACA
ACT1 r 1103	V454	GATGGACCACTTTCGTCGTATTC
TOP1 f 506	V465	ATTACGATGGGAAGCCAGTAGATT
TOP1 r 568	V466	CAGCAAAGAACCCGGCTACTT
flo11::Δrep f	V271	TGCACTAAGAAGACTACTACTCCAGTACCAACCCCATCA
Strain constr.		AGCTCTACTACTGAAAGTTCTTGATTCGGTAATCTCCGA

ORF, position	Name	Sequence
flo11::Δrep r	V272	GAACAGAAGAGCTTTCAGTGCTAGAGCTGAATGGGGTT
Strain constr.		GAAGATGGAGCGGGTAATAACTGATATAA
LYS2+rep f	K428	GATAGTTTACCTGATCCAACTAAGAACTTGGGCTGGTG
Strain constr.		CGATTTCGTGGGGTGTATTCACCTAAGTCAATCTAACT
		GTACTGTCCCTGA
LYS2+rep r	K429	TGGAGTCTCCACAACACAGGTTCTCTCTGGGAAGGCTT
Strain constr.		CAGCATTGTCCTGGAAAATGTC-GATAGAGCTGGTGAT
		TTGTCCTGAA

Table 2: Primers used in this study (continued from the previous page)

Table 3: Gene expression in Tho⁻ mutants

ORF	Tho⁻/Tho⁺	length (nt)	GC%	repeats length (position)	repeats GC%
FLO11	0.15-0.35	4104	0.46	1725 (937-2661)	0.50
FLO1	0.22-0.30	4614	0.45	2391 (841-3231)	0.47
TIR1	0.21-0.51	765	0.47	261 (343-603)	0.51
FIT3	0.31-0.47	615	0.51	306 (76-381)	0.52
TIR4	0.47-0.80	1464	0.47	399 (373-771)	0.50
PIR1	0.96-1.02	1026	0.46	375 (208-582)	0.48
HSP150	0.91-1.11	1164	0.49	399 (286-684)	0.50
RPB1	1.01-1.28	5202	0.41	486 (4651-5136)	0.50
PMA1	0.91-0.96	2757	0.42	no repeats	n/a
TEF1	0.92-0.96	1377	0.44	no repeats	n/a
LYS2	1.12-1.18	4179	0.40	no repeats	n/a
lacZ	0.06-0.22*	3075	0.56*	no repeats	n/a

Cumulative real-time PCR data is presented as a ratio of expression in Tho⁻ mutants to expression in Tho⁺ strains. Length of the full ORF and of the repeats region if applicable, as well as GC content of each, are also shown for comparison.

The data for all ORFs is based on results for S288c *FLO8*⁺ strains, except for *FLO11* where the cumulative data for Σ1278b and S288c *FLO8*⁺ is shown, and for *FIT3* that was analyzed in S288c *flo8*⁻ strains.

**lacZ* data is from Chavez et al., 2000.

REFERENCES

Abramova, N., Sertil, O., Mehta, S., and Lowry, C.V. (2001). Reciprocal regulation of anaerobic and aerobic cell wall mannoprotein gene expression in Saccharomyces cerevisiae. J. Bacteriol. *183*, 2881-2887.

Abruzzi, K.C., Lacadie, S., and Rosbash, M. (2004). Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. EMBO J. *23*, 2620-2631.

Aguilera, A., and Klein, H.L. (1990). HPR1, a novel yeast gene that prevents intrachromosomal excision recombination, shows carboxy-terminal homology to the Saccharomyces cerevisiae TOP1 gene. Mol. Cell Biol. *10*, 1439-1451.

Arndt, K.M., and Kane, C.M. (2003). Running with RNA polymerase: eukaryotic transcript elongation. Trends Genet. *19*, 543-550.

Brill, S.J., and Sternglanz, R. (1988). Transcription-dependent DNA supercoiling in yeast DNA topoisomerase mutants. Cell *54*, 403-411.

Chavez, S., and Aguilera, A. (1997). The yeast HPR1 gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. Genes Dev. *11*, 3459-3470.

Chavez, S., Beilharz, T., Rondon, A.G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J.Q., Lithgow, T., and Aguilera. A. (2000). A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. EMBO J. *19*, 5824-5834.

Chavez, S., Garcia-Rubio, M., Prado, F., and Aguilera, A. (2001). Hpr1 is preferentially required for transcription of either long or G+C-rich DNA sequences in Saccharomyces cerevisiae. Mol. Cell. Biol. *21*, 7054-7064.

Christman, M.F., Dietrich, F.S., and Fink, G.R. (1988). Mitotic recombination in the rDNA of S. cerevisiae is suppressed by the combined action of DNA topoisomerases I and II. Cell *55*, 413-425.

Drolet, M. (2006). Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. Mol. Microbiol. *59*, 723-30.

Fan, H.Y., Cheng, K.K., and Klein, H.L. (1996). Mutations in the RNA polymerase II transcription machinery suppress the hyperrecombination mutant hpr1 delta of Saccharomyces cerevisiae. Genetics *142*, 749-759.

Fan, H.Y., Merker, R.J., and Klein, H.L. (2001). High-copy-number expression of Sub2p, a member of the RNA helicase superfamily, suppresses hpr1-mediated genomic instability. Mol. Cell. Biol. *21*, 5459-5470.

Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol. Cell *12*, 711-721.

Jensen, T.H., Boulay, J., Olesen, J.R., Colin, J., Weyler, M., and Libri, D. (2004). Modulation of transcription affects mRNP quality. Mol. Cell *16*, 235-244.

Jimeno, S., Rondon, A.G., Luna, R., and Aguilera, A. (2002). The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. EMBO J. *21*, 3526-3535.

Keogh, M.C., and Buratowski, S. (2004). Using chromatin immunoprecipitation to map cotranscriptional mRNA processing in Saccharomyces cerevisiae. Methods Mol. Biol. *2557*, 1-16.

Kim, M., Ahn, S.H., Krogan, N.J., Greenblatt, J.F., and Buratowski, S. (2004). Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. EMBO J. *23*, 354-364.

Kim, R.A., and Wang, J.C. (1989). A subthreshold level of DNA topoisomerases leads to the excision of yeast rDNA as extrachromosomal rings. Cell *57*, 975-985.

Lambrechts, M.G., Bauer, F.F., Marmur, J., and Pretorius, I.S. (1996). Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohyphal differentiation in yeast. Proc. Natl. Acad. Sci. USA *93*, 8419-8424.

Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annu. Rev. Genet. *34*, 77-137.

Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell *122*, 365-378.

Liu, L.F., and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. Proc. Natl. Acad. Sci. USA *84*, 7024-7027.

Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. Nature *416*, 499-506.

Masse, E., and Drolet, M. (1999a). Relaxation of transcription-induced negative supercoiling is an essential function of Escherichia coli DNA topoisomerase I. J. Biol. Chem. *274*, 16654-16658.

Masse, E., and Drolet, M. (1999b). Escherichia coli DNA topoisomerase I inhibits R-loop formation by relaxing transcription-induced negative supercoiling. J. Biol. Chem. *274*, 16659-16664.

Mason, P.B., and Struhl, K. (2005). Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. Mol. Cell *17*, 831-840.

O'Dushlaine, C.T., Edwards, R.J., Park, S.D., and Shields, D.C. (2005). Tandem repeat copy-number variation in protein-coding regions of human genes. Genome Biol. *6*, R69.

Piruat, J.I., and Aguilera, A. (1998). A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. EMBO J. *17*, 4859-4872.

Prado, F., Piruat, J.I., and Aguilera, A. (1997). Recombination between DNA repeats in yeast hpr1delta cells is linked to transcription elongation. EMBO J. *16*, 2826-2835.

Reed, R., and Cheng, H. (2005). TREX, SR proteins and export of mRNA. Curr. Opin. Cell. Biol. *17*, 269-273.

Rehwinkel, J., Herold, A., Gari, K., Kocher, T., Rode, M., Ciccarelli, F.L., Wilm, M., and Izaurralde, E. (2004). Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster. Nat. Struct. Mol. Biol. *11*, 558-566.

Roberts, R.L., and Fink, G.R. (1994). Elements of a single MAP kinase cascade in Saccharomyces cerevisiae mediate two developmental programs in the same cell type: mating and invasive growth. Genes Dev. *8*, 2974-2985.

Sadoff, B.U., Heath-Pagliuso, S., Castano, I.B., Zhu, Y., Kieff, F.S., and Christman, M.F. (1995). Isolation of mutants of Saccharomyces cerevisiae requiring DNA topoisomerase I. Genetics *141*, 465-479.

Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R., and Hurt, E. (2002). TREX is a conserved complex coupling transcription with messenger RNA export. Nature *417*, 304-308.

Verstrepen, K.J., Jansen, A., Lewitter, F., and Fink, G.R. (2005). Intragenic tandem repeats generate functional variability. Nat. Genet. *37*, 986-990.

Zenklusen, D., Vinciguerra, P., Wyss, J.C., and Stutz, F. (2002). Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. Mol. Cell. Biol. *22*, 8241-8253.

Chapter 4: Summary, Discussion and Future Directions

SUMMARY

Intragenic repetitive sequences provide a source of genetic and functional variability. Serving as "encoded errors," they can replicate into shorter or longer alleles that alter function (Lovet, 2004; Verstrepen, 2005). My results suggest that internal repeats can also create a barrier to transcription elongation that requires the THO complex for resolution.

Chapter 2 describes the initial genome-wide screen for promoter-independent regulators of *FLO11*. Flo11 is a *S. cerevisiae* cell-surface glycoprotein with multiple repeats. It confers adherence of cells to solid agar. In our screen, we used the S288C deletion library, a construct of *FLO11* under a heterologous constitutive promoter, and the agar adhesion assay. My supposition was that the screen design would identify components unique to the post-transcriptional activity of cell wall proteins— secretion, modification, cell-surface attachment. The screen was successful in identifying genes involved in events subsequent to the initiation of transcription. But, it took a surprising turn: the *FLO11* message was considerably reduced in a number of mutants. This result suggested regulation of *FLO11* at downstream transcriptional events. Among the mutants with the strongest non-adherent phenotype were the for single gene deletion mutants of the THO complex. Based on the previously implicated role of the THO complex in transcription elongation and recombination between repeats, we hypothesized that in Tho$^-$ mutants *FLO11* is non-functional because of a transcription elongation defect through the intragenic repeats.

Chapter 3 describes experiments that address this hypothesis. A chromatin immunoprecipitation experiment, designed to monitor the occupancy of the RNA

polymerase along the *FLO11* ORF shows normal polymerase levels at the 5'-end of the ORF, but reduced levels at the 3'-end in a Tho$^-$ mutant compared to a Tho$^+$ strain. This result suggests that reduced levels of *FLO11* message arise from a transcription elongation and not initiation defect in Tho$^-$ mutants. The observed defect is repeats-dependent, as removal of the repetitive sequence in *FLO11* abolishes the transcriptional defect in Tho$^-$ mutants. Moreover, Tho$^-$ mutants increase the genetic instability of the repetitive region. Overexpression of Topoisomerase I partially restores *FLO11* expression levels and function in Tho$^-$ mutants, suggesting that the elongation defect may be associated with excessive negative supercoiling in the repetitive region during transcription. Since expression of other genes with repeats is reduced in Tho$^-$ strains, we suggest that repetitive sequences in general predispose to aberrant DNA topology in Tho$^-$ mutants during transcription. In this sense, the THO complex serves to rectify aberrant structures during transcription.

DISCUSSION AND FUTURE DIRECTIONS

FLO11 sets a precedent as an endogenous gene that requires the THO complex for function

Previous studies on the THO complex have been conducted on artificial constructs with repeats, bacterial artificially overexpressed genes, or random native genes without association with phenotypes. Here we show that the repetitive regions of *FLO* genes are genetically unstable and create a transcription elongation barrier. We believe that this

applies to most yeast genes with internal repeats, although it is difficult to capture expression of many of these genes in a single experiment, as many of the yeast genes with repeats are expressed at special environmental conditions, such as anaerobic growth, cold shock, or sporulation.

One interesting question is whether our finding in yeast extends to other organisms. It is suggested that the THO/TREX complex is differentially recruited in yeast and metazoans, the main difference coming from the fact that very few yeast genes get spliced. Thus, the yeast THO complex is recruited to the transcribed region by the RNA polymerase complex, while in metazoans, TREX is recruited by the spliceosome (Reed and Cheng, 2005). Moreover, only two of the four yeast THO complex subunits, Hpr1 and Tho2, are conserved in eukaryotes. Still, the end function of the TREX complex may be more similar than anticipated in either situation. For example, *lacZ* expression is reduced in both yeast and human Tho$^-$ mutants (Li et al., 2005). In *Drosophila* as well as in yeast, several heat-shock proteins are downregulated in the mutants (Rehwinkel et al, 2004). One can speculate that human mucins (the orthologs to the yeast *FLO* genes) would also require the THO/TREX complex for function.

A recent comparative analysis of expression profiles of genes in breast tumors indicates an upregulation of hHPR1 (p84N5/hTREX84/hThoc1). This protein physically interacts with UAP56, and its depletion by dsRNA reduces cytoplasmic levels of poly(A)$^+$ RNA, consistent with TREX biology (Guo et al., 2005). Moreover, treatment of a breast cancer cell line with hTREX84-specific siRNA reduces the proliferative potential of these cells. While this study identifies hHpr1 as a prognostic marker of breast cancer, it raises the question as to expression of what genes is altered by higher levels of hHpr1.

Specifically, it would be interesting to test whether higher than normal THO complex activity would increase expression of the *MUC* genes, implicated in many types of cancer. The mucin genes are similar to the yeast *FLO* genes in that they contain multiple tandem repeats, and encode heavily glycosylated cell-surface proteins. In our analysis of *FLO1* and *FLO11* expression, we noticed a two-fold higher expression of *FLO1* in a strain with a *FLO11* allele without the repetitive sequence (*flo11::Δrep*). Changing either THO complex levels or the abundance of THO-requiring sequences may offset expression of other genes that require the THO complex.

Biochemical function of the THO complex

The biochemical function of the THO complex subunits remains unknown. Based on protein sequence similarity between Hpr1 and Top1 and synthetic lethality interaction between *top1Δ* and *hpr1Δ* (Aguilera and Klein, 1990) as well as between *top1Δ* and *mft1Δ*, one can speculate that Hpr1 and the THO complex function as a topoisomerase. In agreement with such a model, we show that overexpression of *TOP1* restores *FLO11* expression and function in an *mft1Δ thp2Δ* double mutant. A similar *TOP3* overexpression analysis may differentiate between the presence of positive supercoiling in front of the RNA polymerase or negative supercoiling behind it, since Top3 removes only negative supercoils.

At the same time, excessive supercoiling and DNA:RNA base-pairing are tightly coupled, and it is difficult to discern which event arises first. In terms of binding specificity, one study suggests that Hpr1 associates with DNA rather than RNA, based on a chromatin immunoprecipitation experiment with a tagged Hpr1 followed by RNase

treatment (Abruzzi et al., 2004). Another study shows that Tho2 can associate with both DNA and RNA in a band-shifting assay (Jimeno et al., 2002). It is possible that not all THO components function in the same way. Preventing excessive supercoiling and R looping may be a combined result of several THO components binding DNA and other components binding RNA, or recruiting other proteins to the nascent RNA.

Our model predicts formation of DNA:RNA hybrids during transcription through the *FLO11* repeats in a Tho$^-$ mutant. So far, I have attempted overexpressing *RNH1*, a yeast gene that encodes RNase H, but that does not lead to appreciable change in *FLO11* transcript levels (as measured by rt-pcr), nor does it improve the RNA polymerase processivity (assayed by chromatin IP). Another indication of aberrant pairing in the region of *FLO11* repeats would be the presence of single stranded non-template strand. Then, treatment of genomic DNA with mung bean nuclease or another enzyme with specificity to ssDNA may be used in combination with PCR on the repetitive region to compare DNA topology in a Tho$^+$ and a Tho$^-$ strain.

Other possible extensions

In another direction of research, it would be interesting to test whether Tho$^-$ mutants generate genetic and functional variation among genes with internal repeats, and furthermore, whether such an event is transcription-dependent.

Another possible extension of this thesis project is to screen the collection of yeast knock-down strains of essential genes for effect on *FLO11* function. For example, deletions of the TREX genes Sub2 and Yra1 are non-viable, and so are many of the secretory pathway mutants.

REFERENCES

Abramova, N., Sertil, O., Mehta, S., and Lowry, C. (2001). Reciprocal regulation of anaerobic and aerobic cell wall mannoprotein gene expression in Saccharomyces cerevisiae. J Bacteriol 183, 2881-2887.

Abruzzi, K., Lacadie, S., and Rosbash, M. (2004). Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. EMBO J 23, 2620-2631.

Aguilera, A., and Klein, H.L. (1990). HPR1, a novel yeast gene that prevents intrachromosomal excision recombination, shows carboxy-terminal homology to the Saccharomyces cerevisiae TOP1 gene. Mol. Cell Biol. *10*, 1439-1451.

Guo, S., Hakimi, M. A., Baillat, D., Chen, X., Farber, M. J., Klein-Szanto, A. J., Cooch, N. S., Godwin, A. K., and Shiekhattar, R. (2005). Linking transcriptional elongation and messenger RNA export to metastatic breast cancers. Cancer Res 65, 3011-3016.

Jimeno, S., Rondon, A. G., Luna, R., and Aguilera, A. (2002). The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. Embo J 21, 3526-3535.

Li, Y., Wang, X., Zhang, X., and Goodrich, D. W. (2005). Human hHpr1/p84/Thoc1 regulates transcriptional elongation and physically links RNA polymerase II and RNA processing factors. Mol Cell Biol 25, 4023-4033.

Merker, R. J., and Klein, H. L. (2002). hpr1Delta affects ribosomal DNA recombination and cell life span in Saccharomyces cerevisiae. Mol Cell Biol 22, 421-429.

Reed, R., and Cheng, H. (2005). TREX, SR proteins and export of mRNA. Curr. Opin. Cell. Biol. *17*, 269-273.

Rehwinkel, J., Herold, A., Gari, K., Kocher, T., Rode, M., Ciccarelli, F., Wilm, M., and Izaurralde, E. (2004). Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster. Nat Struct Mol Biol 11, 558-566.

Verstrepen, K., Jansen, A., Lewitter, F., and Fink, G. (2005). Intragenic tandem repeats generate functional variability. Nat Genet 37, 986-990.

Chapter 5: Appendices

Cumulative data for the non-adherent Σ1278b mutants

The genome-wide screen for promoter-independent regulators of *FLO11*, described in Chapter 2, identified 44 mutants with reduced adherence to agar in both S288c P_{TEF}*FLO11* and Σ1278b P_{FLO11}*FLO11* strain backgrounds. Most of the Σ1278b mutants were further analyzed by measuring expression levels of *FLO11* and by testing for another *FLO11*-dependent function, biofilm formation. *FLO11* mRNA levels were measured both by real-time PCR and by Northern analysis. Normalization was to *TEF1* and to *ACT1*. Biofilm formation was assayed on YPD 0.3% agar plates after 15-day incubation at room temperature as previously described (Reynolds and Fink, 2001). The results are summarized in Table 1. Adherence to solid agar of the deletion mutants of the corresponding genes is as in Chapter 2, and is included for comparison.

Pictures of the biofilm assay are included in Figure 1. The top row includes a positive control (Σ1278b wild-type, 10560-23C), a negative control (Σ1278b *flo11*Δ, L7558), and three mutants that are non-adherent phenotype in the primary screen, but are hyper-adherent or have no effect in Σ1278b. Consistent with their hyper-adherence, *bdf1*Δ and *dig1*Δ mutants form a more compact and highly wrinkled structure on the low agar plates. The Σ1278b *fab1*Δ mutant has no discernable phenotype in the agar adhesion assay but seems to have a weak defect in biofilm formation. While for most mutants the biofilm assay reinforces the findings from the adhesion assay, it also illustrates that different mutants may have disparate effects on different *FLO11*-dependent functions.

ORF	mRNA (% of WT)	agar adherence	biofilm formation
GAS1	13	+	++
GPI7	12	+	wt
SMI1	37	+	++
YNL171C	46	+	wt
LDB16	26	+	wt
SRV2	15	-	na
ARP1	73	+	wt
SEP7	62	+	++
CDC10	60	+	hyper-morphology
VRP1	25	+	-
SLA1	27	+	na
BEM4	32	++	na
SPC72	6	+	+
VPS3	13	++	+
VPS21	68	++	++
VPS23	2	-	na
VPS28	4	-	-
VPS32	3	-	-
VPS36	4	-	-
VPS65	83	+	+
THO2	na	-	na
HPR1	na	-	na
MFT1	12	-	na
THP2	13	-	+
SAC3	74	+	++
THP1	97	++	wt
SPT4	29	+	+
TRF4	na	+	na
DHH1	13	-	na
DEF1	54	+	wt
RPB4	10	+	+
RSC1	33	+	++
RSC2	21	+	na
NPL6	84	+	+
LDB7	36	-	+
SIN3	na	-	na
TAF14	74	+	++
YAF9	33	+	wt
CTF19	65	+	wt
MCM21	na	+	na
BFR1	na	+	na
CWH8	47	+	-
MRPL28	11	+	na
WHI3	14	+	+

Table 1: Summary of data for deletion mutants of the indicated genes in Σ1278b.
- strong phenotype; + moderate; ++ weak; na-not available; wt-same as wild-type.

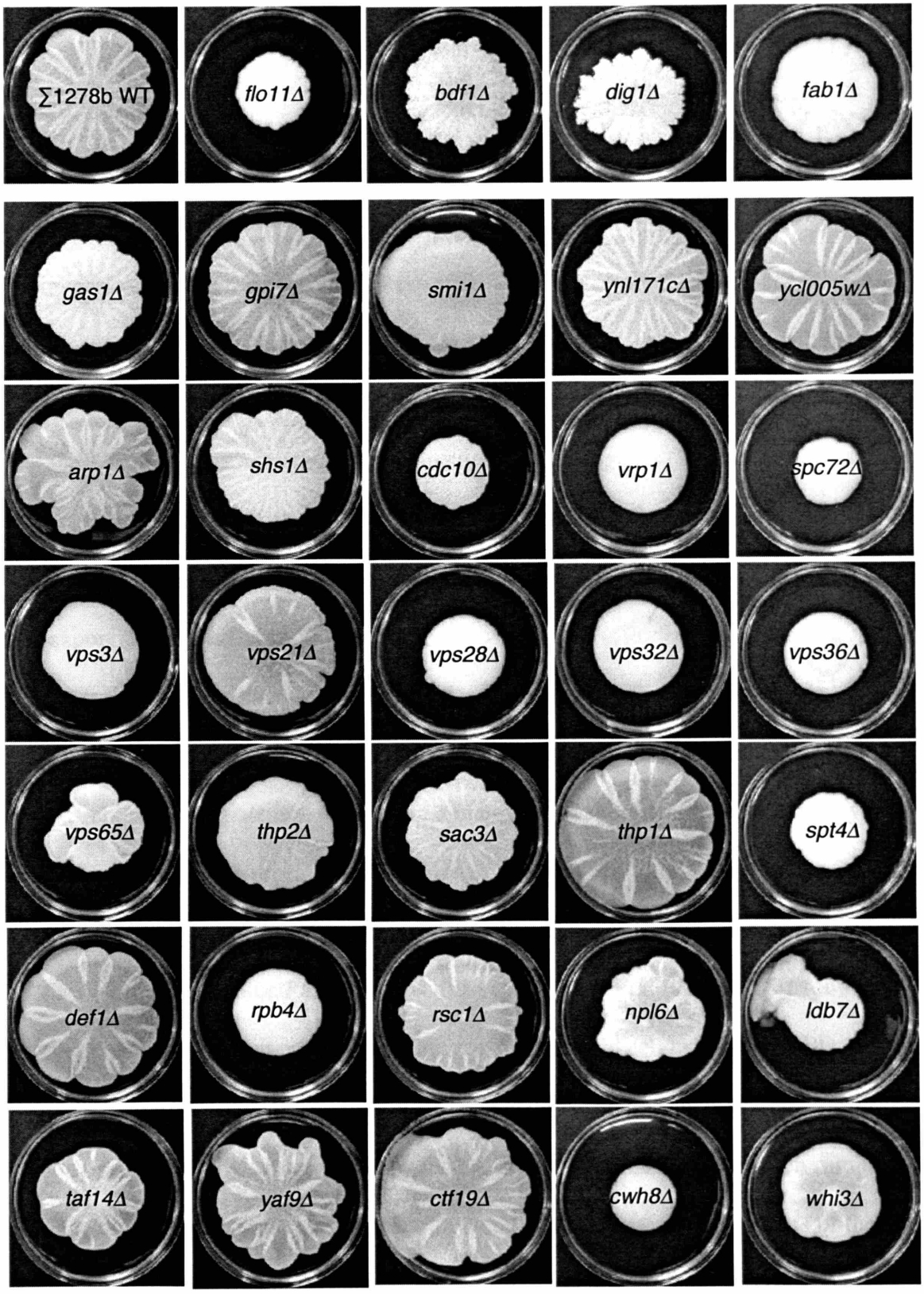


Figure 1: Biofilm assay of Σ1278b mutant strains.

Genome-wide expression profiling of Tho⁻ mutants

Ideally, we would like to know what genes require the THO complex for transcription elongation, and whether there is a functional bias towards genes with internal tandem repeats. For that, all genes would need to be transcribed at some level. A comparison between signals in an array with both 5'-end and 3'-end probes to each ORF may identify hurdles to transcription elongation in individual genes (Arndt and Kane, 2003). At this point, chromatin immuno-precipitation analysis in yeast indicates that the THO complex associates with actively transcribed genes (Strasser et al., 2002; Kim et al., 2004; Abruzzi et al., 2004), while a genome-wide expression profiling in Drosophila suggests that most genes do not require the THO complex for transcription and export (Rehwinkel et al., 2002).

It is difficult to capture expression of many genes with repeats in a single experiment because most genes with repeats require specific growth conditions for expression. Another potential problem of a microarray experiment is the specificity of the probes. Genes with repeats usually belong to families of highly homologous ORFs, such as the *FLO*, *TIR*, and *PIR* families, making it difficult to design specific probes.

We used the S288c *FLO8*⁺ strains that express both *FLO11* and *FLO1* for a whole-genome expression profiling of Tho⁻ mutants. Liquid cell cultures in YPD media were grown for 5 hours after dilution in a shaker at 30°C, and for an additional hour without shaking. RNA was prepared by the hot acid phenol method. For each sample, 15 μg total RNA was used to produce biotinylated cRNA. Fifteen micrograms of the biotinylated cRNA was then hybridized to a Yeast genome S98 array (Affymetrix). The

detailed protocols for amplification, hybridization, washing, scanning, and data analysis are available at http://www.wi.mit.edu/CMT/protocols.html. Scan-to-scan variations in intensity were corrected by applying a scaling factor to each experiment so that the overall intensity of all experiments was 100. All scaling factors were between 2.1 and 2.2. The data was floored at 26 for the control experiment. To evaluate the relative change in transcript level under different experimental conditions, a ratio was calculated by dividing the signal intensity in each experiment by that of a control experiment. Only genes with at least 3 –fold reduced expression in each of the two mutant strains (XY200 and XY204) with respect to the Tho^+ strain (L8046) were considered, resulting in 43 targets out of 2960 genes, that is 1.5% (Table 2).

Most of the reduced genes encode stress-induced or hypoxic growth-specific proteins. This set of genes may be expressed because of the flocculent nature of the strains, creating a hypoxic microenvironment. It remains unclear, whether reduced expression of these genes in Tho^- mutants is a direct result of compromised THO complex function or an indirect effect of reduced flocculation. Expression of the flocculation-specific gene *FLO1* is reduced by 55% according to the microarray data, and 70% according to real-time PCR data. Thus, the effective number of genes requiring THO complex for expression may be even less than 43 in this experiment. Among the group of most reduced genes are 3 genes that encode cell-surface serine/threonine-rich mannoproteins: *DAN1*, *TIR4*, and *SPI1*.

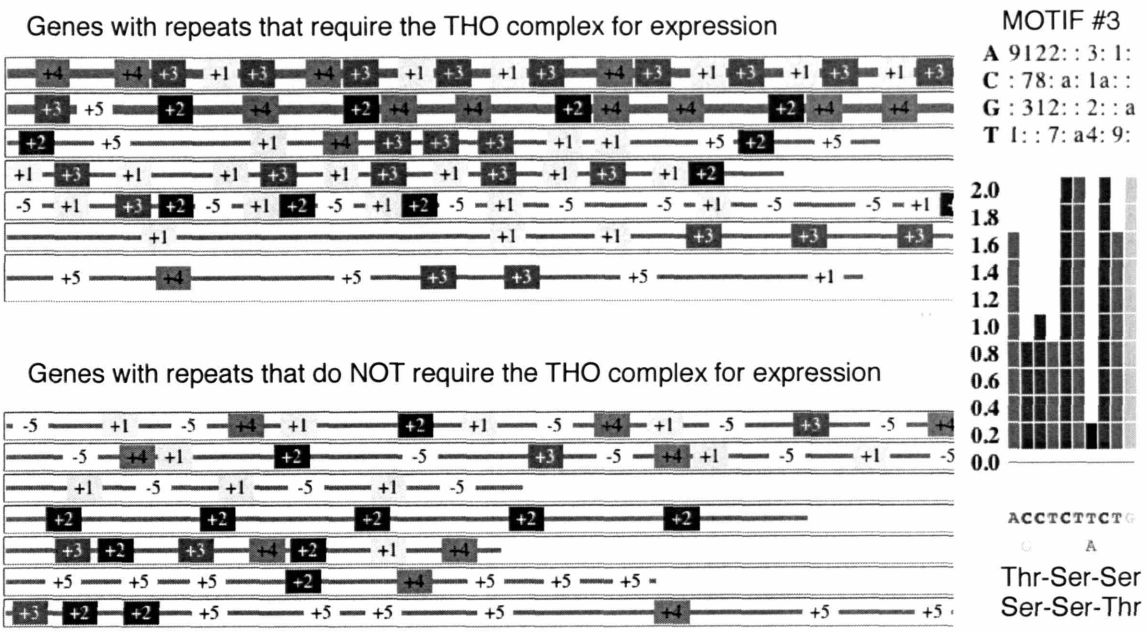
DAN1	YJR150C	cell wall mannoprotein with similarity to Tir proteins
	YER067C-A	dubious
	YNL194C	integral membrane protein at the cell cortex
	YER067W	uncharacterized
	YNR034W-A	uncharacterized
FMP33	YJL161W	found in mitochondrial proteome
HSP30	YCR021C	plasma membrane protein induced by heat shock, ethanol treatment, and entry into stationary phase
ALD3	YMR169C	Aldehyde Dehydrogenase (NAD(P)+)
GAC1	YOR178C	regulatory subunit for Glc7p
	YLR252W	dubious
RTN2	YDL204W	reticulon-like protein of unknown function
CYC7	YEL039C	Cytochrome c isoform 2, expressed under hypoxic conditions
TIR4	YOR009W	cell wall Ser/Ala-rich mannoprotein of the Srp1/Tip1 family
SOL4	YGR248W	6-phosphogluconolactonase
PIG2	YIL045W	putative type-1 protein phosphatasetargeting subunit
CTT1	YGR088W	cytosolic catalase T that removes superoxide radicals
TSL1	YML100W	regulatory subunit of trehalose-6-phosphate synthase/phosphatase complex
SYM1	YLR251W	protein required for ethanol metabolism, induced by heat shock
TSA2	YDR453C	stress inducible cytoplasmic thioredoxin peroxidase
TMA10	YLR327C	protein of unknown function that associates with ribosomes
	YFR017C	uncharacterized
FMP16	YDR070C	found in mitochondrial proteome
HSP78	YDR258C	oligomeric mitochondrial matrix chaperone
FMP12	YHL021C	found in mitochondrial proteome
GPG1	YGL121C	proposed gamma subunit of the heterotrimeric G protein
GLC3	YEL011W	glycogen branching enzyme, involved in glycogen accumulation
	YJL144W	cytoplasmic hydrophilin
HSP42	YDR171W	small cytosolic stress-induced chaperone
UIP4	YPL186C	protein of unknown function that interacts with Ulp1
GSY1	YFR015C	glycogen synthase with similarity to GSY2
	YHR087W	protein involved in RNA metabolism
SDP1	YIL113W	stress-inducible dual-specificity MAPK phosphatase
GPX1	YKL026C	phospholipids hydroperoxide glutathione peroxidase
PIC2	YER053C	mitochondrial phosphate carrier
YAK1	YJL141C	serine-threonine protein kinase, part of a glucose sensing system
	YPL247C	uncharacterized
ALD3	YMR169C	cytoplasmid aldehyde dehydrogenase
DCS2	YOR173W	non-essential protein containing a histidine triad motif
OM45	YIL136W	major constituent of the mitochondrial outer membrane
SPI1	YER150W	GPI-anchored Ser/Thr rich cell wall protein of unknown function
	YPL230W	uncharacterized
GSY2	YLR258W	glycogen synthase
TPS2	YDR074W	phosphatase subunit of trehalose-6-phosphate synthase/ phosphatase complex

Table 2: List of genes (from most reduced to least reduced expression) with at least 3-fold reduced expression in Tho$^-$ mutants (two independent experiments).

Codon bias in repeats of genes that require the THO complex

Since microarray experiments cannot provide a comprehensive expression analysis of genes with internal repeats, a variety of strains and growth conditions were used to monitor expression of particular genes in Tho^+ and Tho^- strains. *FLO11* was analyzed in Σ1278b and S288c *FLO8+* strains, *FLO1* in S288c *FLO8+* strains, *TIR1* and *TIR4* in Σ1278b as well as in S288c strains after a cold-shock induction. The rest of the genes in Table 3 were analyzed in S288c strains.

The data suggests that some genes with repeats require the THO complex for expression while others do not (Table 3). To test whether sequence motifs may account for this divide, a search for short (5-11nt) motifs was simultaneously performed on all 14 repetitive regions with the MEME software (http://meme.sdsc.edu). This analysis was carried out by Sanjeev Pillai from the Bioinformatics group at Whitehead Institute. Motif #3 is over-represented in the genes that require the THO complex for expression (Figure 2). As the sequence of this motif encodes for serine and threonine residues, codon bias may account for the observed THO-dependence of this group of genes.



TIR1

MAYTKIALFAAIAALASAQTQDQINELNVILNDVKSHLQEYISLASDSSSGFSLSSMPAG
VLDIGMALASATDDSYTTLYSEVDFAGVSKMLTMVPWYSSRLEPALKSLNGD**ASSSAAPS**
SSAAPTSSAAPSSSAAPTSSAASSSSEAKSSSAAPSSSEAKSSSAAPSSSEAKSSSAAPS
SSEAKSSSAAPSSTEAKITSAAPSSTGAKTSAISQITDGQIQATKAVSEQTENGAAKAFV
GMGAGVVAAAAMLL

PIR1

MQYKKSLVASALVATSLAAYAPKDPWSTLTPSATYKGGITDYSSTFGIAVEPIA**TTASSK**
AKRAAAISQIGDGQIQATTKTTAAAVSQIGDGQIQATTKTKAAAVSQIGDGQIQATTKTT
SAKTTAAAVSQIGDGQIQATTKTKAAAVSQIGDGQIQATTKTTAAAVSQIGDGQIQATTK
TTAAAVSQIGDGQIQATTNTTVAPVSQITDGQIQATTLTSATIIPSPAPAPITNGTDPVT
AETCKSSGTLEMNLKGGILTDGKGRIGSIVANRQFQFDGPPPQAGAIYAAGWSITPEGNL
AIGDQDTFYQCLSGNFYNLYDEHIGTQCNAVHLQAIDLLNC

Figure 2: MEME motif search among 14 repetitive regions in 12 genes (Table 3), 6 requiring the THO complex for expression, and 6 not requiring the THO complex. Motif #3 (+ indicates the coding strand) is over-represented in the first group of genes. The 10 nt sequence of this motif translates into serine and threonine. *TIR1* is the fourth gene in the THO-requiring group, and *PIR1* is the first gene in the THO-independent group. The full amino acid sequences of *TIR1* and *PIR1* are given for comparison (the repetitive regions are in bold).

ORF	Tho⁻/Tho⁺	length (nt)	GC%	repeats length (position)	repeats GC%
FLO11	0.15-0.35	4104	0.46	1725 (937-2661)	0.50
FLO1	0.22-0.30	4614	0.45	2391 (841-3231)	0.47
FIT3	0.31-0.47	615	0.51	306 (76-381)	0.52
TIR1	0.21-0.51	765	0.47	261 (343-603)	0.51
TIR4	0.47-0.80	1464	0.47	399 (373-771)	0.50
YOL155C	0.62	2904	0.47	459 (298-756)	0.52
				324 (2275-2598)	0.45
PIR1	0.96-1.02	1026	0.46	375 (208-582)	0.48
HSP150	0.91-1.11	1164	0.49	399 (286-684)	0.50
PIR3	1.1	978	0.46	188 (298-486)	0.48
YNL190W	1.22	615	0.45	312 (145-456)	0.45
SED1	0.8	1017	0.47	174 (217-390)	0.48
				234 (532-765)	0.45
RPB1	1.01-1.28	5202	0.41	486 (4651-5136)	0.50

Table 3: cumulative real-time PCR data on expression of genes with repeats. This table contains a few additional ORFs in comparison to the corresponding table in Chapter 3. *YOL155C* and *SED1* contain two repetitive regions.

10560-23C	Σ1278b *MATα ura3-52 his3::hisG leu2::hisG*	Fink laboratory collection
L7558	Σ1278b *MATα ura3-52 his3::hisG leu2::hisG flo11::kanMX4*	Fink laboratory collection
L8046	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52*	Fink laboratory collection
XY200	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 thp2::kanMX4*	This study
XY204	S288c *MATa flo8::LEU2::FLO8+ his3^200 leu2^1 ura3-52 mft1::kanMX4*	This study

Table 4: Yeast strains

REFERENCES

Abruzzi, K.C., Lacadie, S., and Rosbash, M. (2004). Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. EMBO J. *23*, 2620-2631.

Arndt, K.M., and Kane, C.M. (2003). Running with RNA polymerase: eukaryotic transcript elongation. Trends Genet. *19*, 543-550.

Kim, M., Ahn, S.H., Krogan, N.J., Greenblatt, J.F., and Buratowski, S. (2004). Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. EMBO J. *23*, 354-364.

Rehwinkel, J., Herold, A., Gari, K., Kocher, T., Rode, M., Ciccarelli, F.L., Wilm, M., and Izaurralde, E. (2004). Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster. Nat. Struct. Mol. Biol. *11*, 558-566.

Reynolds, T. B., and Fink, G. R. (2001). Bakers' yeast, a model for fungal biofilm formation. Science 291, 878-881.

Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R., and Hurt, E. (2002). TREX is a conserved complex coupling transcription with messenger RNA export. Nature *417*, 304-308.