UCSF
UC San Francisco Electronic Theses and Dissertations

Title
CDK Regulation of Replication Proteins: Mcm2-7 and DNA Polymerase Alpha Primase

Permalink
https://escholarship.org/uc/item/0678c85d

Author
Liku, Muluye Ejigu

Publication Date
2008-09-10

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

CDK Regulation of Replication Proteins:
Mcm2-7 and DNA Polymerase Alpha Primase

by

Muluye E Liku

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Dedicated to

Teyeku Demessew and Liku Ejigu

Acknowledgments

Firstly, I would like to express gratitude to my thesis advisor, Joachim Li for a being a great example of what it means to be committed to teaching. This is evident in the respect and admiration expressed towards Joachim by students and faculty alike. His strict adherence to systematic and thorough execution of thought and experiments, especially in the face of failure are admirable and what lead to success in science. I hope to implement these tools of success in my future aspirations.

I would also like to give special thanks to my superb thesis committee members Carol Gross, David Morgan, and David Toczyski. Thanks for helping guide the ship of my graduate training. Your many scientific and personal advices have been extremely helpful and I shall never forget your kindness. I want to give special thanks to Carol who made the fateful initial contact (like explorers of old) at my undergraduate campus and peaked my interest in science. If it was not for her I would not be pursuing science at UCSF.

Two other heroes I want to express my gratitude towards are Erin O'Shea and Peter Walter. When I first came to UCSF I was not sure that I wanted to do biological science but the first day I met Erin it was evident that I had found a new passion and curiosity. My momentum was redirected towards biological sciences. My great thanks to Erin for expressing so much confidence and support. Peter Walter has always been a source of great wisdom about life especially as it relates to science. Many thanks for the numerous advices. I have learned a great deal from your examples.

I would like to extend a special thanks to the many past and present members of the Li lab, who have put up with me and have been kind enough to share their time and understanding of the world. Many things have I learned.
Many thanks to unnamed UCSF mentors, colleagues and friends for the years of encouragement.

My heartful gratitude to Mrs. Ota. As my first teacher she was able to bridge a great gulf between us when I first came to the US and I didn't speak a word of English. I couldn't have asked for a better teacher to beckon me to come out and explore the world that awaited me. I am forever grateful.

Last but definitely not least my million thanks to my family for the many years of support as I embarked on many journeys to attain lofty goals.

Chapter 2 of this dissertation are reprints of material as it appears in Liku ME, Nguyen VQ, Rosales AW, Irie K, Li JJ (2005). CDK phosphorylation of a novel NLS-NES module distributed between two subunits of the Mcm2-7 complex prevents chromosomal rereplication. Mol Biol Cell *16(10)*, 5026-39. Chapter 3 of this dissertation are reprints of material as it appears in Moses AM, Liku ME, Li JJ, Durbin R (2007). Regulatory

evolution in proteins by turnover and lineage-specific changes of cyclin-dependent kinase consensus sites. Proc Natl Acad Sci USA *104(45):* 17713-8.

CDK Regulated Localization of Replication Proteins

Muluye E. Liku

Abstract

The precise inheritance of genetic material requires that eukaryotic cells initiate DNA replication in a highly regulated manner. To prevent any inappropriate re-replication of the genome, initiation at each of the hundreds to thousands of replication origins must occur only once per cell cycle. In *Saccharomyces cerevisiae*, cyclin dependent kinases have a critical role in inhibiting reinitiation, and must do so without interfering with their role in promoting initiation. Previously, we had shown that CDKs inhibit multiple proteins involved in an early step of replication initiation, i.e. assembly of the pre-replicative complex (pre-RC), so as to prevent re-initiation of DNA replication. Specifically, we had demonstrated that making ORC, Cdc6, and Mcm2-7 refractory to CDK inhibition promotes limited reinitiation and re-replication from some origins. Despite this finding, our understanding of the mechanisms used by CDKs to prevent re-initiation are incomplete.

This thesis presents two different studies that expand our understanding of these mechanisms. In the first part, we acquired a more in depth understanding of how CDKs inhibit Mcm2-7 activity by promoting their nuclear export. We identified key transport regulatory modules on Mcm2 and Mcm3 that control the import and export of the Mcm2-7 complex and showed that phosphorylation of the Mcm3 module alters the balance between the two events so as to favor export. We showed that this Mcm3 module and its

CDK regulation evolved quite recently in a budding yeast lineage that includes *Saccharomyces cerevisiae* and closely related yeast. A more detailed analysis of CDK phosphorylation consensus sites on Mcm3 and other pre-RC components showed that although the precise number and position of sites is not conserved, the clustering of these sites in specific regions is.

Our second study on the CDK inhibition of reinitiation was motivated by the limited re-replication that we observed when we deregulated ORC, Cdc6, and Mcm2-7. This result suggests that CDKs target additional replication components to prevent re-initiation, and we investigated the possibility that DNA polymerase alpha primase, specifically its Pol1 and Pol12 subunits, might be one of these components. We showed that the CDK phosphorylation of Pol1 and Pol12, which appears to be conserved through humans, are unlikely to be activating events that trigger replication initiation, but are in fact delayed till after replication is mostly complete, consistent with a late inhibitory role. We also showed that CDK phosphorylation of Pol1 is required to promote the relocalization of Pol alpha primase from nucleoplasm to nuclear periphery in G2/M phases. Because of these suggestive hints, although we could not establish a role for Pol 1 or Pol12 phosphorylation in the block to re-replication, we suspect that such a role may be uncovered with further investigation.

Table of Contents

		page
CHAPTER 1	Introduction	1
CHAPTER 2	CDK phosphorylation of a novel NLS-NES module distributed between two subunits of the Mcm2-7 complex prevents chromosomal rereplication	17
CHAPTER 3	Regulatory evolution in proteins by turnover and lineage specific changes of cyclin dependent kinase consensus sites	100
CHAPTER 4	CDK regulation of Polymerase Alpha Primase complex	144
CHAPTER 5	Conclusion	205

List of Tables

page

CHAPTER 2

Table 1 Base substitutions in the different mutant mcm3 alleles 90-91

Table 2 Plasmids used in Chapter 2 92-93

Table 3 Integrating tagging plasmids used in Chapter 2 94

Table 4 pGAL expression plasmids used in Chapter 2 95

Table 5 Strains used in Chapter 2 96-99

CHAPTER 3

Table 1 The 55 characterized CDK sites used in Fig. 2*b* and SI Fig. 4*a* used in Chapter 3 138-43

CHAPTER 4

Table 1 Plasmid used in Chapter 4 200

Table 2 Strains used in Chapter 4 201-3

Table 3 chromosomal segmental duplications observed in Chapter 4 204

List of Figures

page

CHAPTER 2

Figure 1	68
Figure 2	69
Figure 3	70
Figure 4	71
Figure 5	72
Figure 6	73
Figure 7	74
Figure 8	75
Figure 9	76
Figure 10	77

Supplementary Figure 1	88
Supplementary Figure 2	89

CHAPTER 3

Figure 1	127
Figure 2	128
Figure 3	129
SI Figure 4	132
SI Figure 5	133-34
SI Figure 6	135-37

CHAPTER 4

Figure 1	184-87
Figure 2	189

Figure 3	191
Figure 4	193-95
Figure 5	197
Figure 6	199

Chapter 1

Introduction

Introduction

Reproduction of life at the cellular level involves the transmission of the genetic material. The dogma states that cells want the transmission to occur without much change to the genome and in an efficient manner. To achieve this faithful transmission the cell engages in two major events: genomic DNA replication and segregation of the resulting two copies of genomic DNA. This effort is divided into the cell cycle, characterized by cells cycling through G1, S, G2 and M phases. Much of the discussion in this thesis will be focused on the process of DNA replication and in particular its regulation, which occurs around S phase.

The cell devotes a considerable repertoire of proteins to ensure DNA replication occurs accurately and efficiently. Since this process consumes so much resource the cell has evolved regulatory mechanisms to ensure that the cell is well prepared for proper DNA replication before actually initiating replication. Thus regulation of DNA replication occurs at the step of initiation. Many of the studies of this process were facilitated by the arrangement of DNA replication into two mutually exclusive stages: pre-replicative complex (pre-RC) formation and initiation along with elongation of replication. Each of these steps is regulated to ensure that the replication occurs faithfully. Consequently, it was possible to assign function to replication proteins based on their activities at these stages by determining execution points of activity.

DNA replication initiates at specific positions on the chromosomes. To allow for efficient replication of the genome, eukaryotic cells have hundreds to thousands of initiation sites. There are two factors that are essential for defining an initiation site: the DNA sequence (*cis* element) and the protein complexes that bind to it (*trans* factors). The only eukaryotic organism demonstrated to have defined DNA sequences in which DNA replication initiation has preferred sites of initiation is *Saccharomyces cerevisiae*. In other eukaryotes origin selection is thought to be for the most part guided by chromatin structure and not specific DNA sequences with a few exceptions (Stephen P Bell & Dutta 2002). The *Saccharomyces cerevisiae* origin was defined as a 100-150bp DNA sequence containing three to four 10-15bp stretch demonstrated to be sufficient to initiate DNA replication (Stephen P Bell & Dutta 2002).

There are numerous *trans* factors that are critical for defining a DNA replication initiation site. The first set of protein complexes that form at origins are the pre-Replicative Complex (pre-RC) (J F Diffley 1996; Kelly & Brown 2000; B Stillman 1996). The assembly of the pre-RC is essential for initiation at these origins. The potentiation to initiate DNA replication of these origins conferred by pre-RC assembly is termed "licensing" (Blow & Hodgson 2002). The pre-RC formation requires a pioneer factor, Origin Recognition Complex (ORC) to bind to origins (S P Bell & B Stillman 1992). In *Saccharomyces cerevisiae* the defined DNA sequence element recognized by ORC is called an autonomously replicating sequence (ARS) or generally as an origin. There are approximately ~336-420 ARS elements bound by ORC distributed among the 16 chromosomes of S. cerevisiae (Wyrick et al. 2001). The binding of ORC to origins

allows for the recruitment of Cdc6 and Cdt1 to each origin (Stephen P Bell & Dutta 2002). These two components are required to load the Mcm2-7 complex, the presumed replicative helicase. Once these pre-RC components are assembled the origin is "licensed" and is competent for the process of replication initiation (Stephen P Bell & Dutta 2002). The transition from pre-RC to initiation complex involves several events but is controlled by cyclin dependent kinases (CDKs) and Dbf4 dependent Kinase (DDK) activities (Arias & Walter 2007).

The cyclin dependent kinases in *Saccharomyces cerevisiae* are composed of a group of nine regulatory cyclin subunits who independently associate with the catalytic subunit, *CDC28* (the CDK) (D O Morgan 1997). The activity of the CDK depends on its association with the cyclins. Although the CDK subunit of the complex is present throughout the cell cycle the cyclins are regulated both at the transcriptional and post translational level. This regulation is critical to promote the cell cycle. Similarly, the DDK is composed of a cyclin like subunit, *DBF4*, and the catalytic subunit *CDC7*. The activity of Cdc7 depends on its association with Dbf4. Like the cyclins, Dbf4 transcription and its protein level are cell cycle regulated (RA Sclafani 2000).

CDK activity promotes replication initiation

The classical work by Rao and Johnson simply and elegantly demonstrated, fusing hela cells at different stages of the cell cycle, that replication occurs once and only once per cell cycle(Rao & Johnson 1970). In addition their work demonstrates that some event in G1 causes DNA to have potential to replicate and that some factors from S-phase promoted DNA replication (Rao & Johnson 1970). Subsequent work had implicated the CDK as being essential for DNA replication initiation (Blow & Nurse 1990; L H Hartwell et al. 1974; Schwob & Nasmyth 1993) and (Reviewed in Kelly & Brown 2000). However, the essential targets of this CDK were unknown until the discovery of *SLD2* by Araki's group (Masumoto et al. 2002). Subsequently SLD3 was also identified as a second essential CDK substrate to promote replication initiation (Zegerman & John F X Diffley 2007; Seiji Tanaka et al. 2007). The phosphorylation of these proteins, along with presence and activity of other factors, facilitates the loading of the replicative DNA polymerases in order to initiate replication (Arias & Walter 2007).

DDK activity promotes replication initiation

The initial work done to implicate an essential role for Cdc7 in DNA replication was made three decades ago (Leland H Hartwell et al. 1973; L H Hartwell 1973; R A Sclafani

& Jackson 1994; Patterson et al. 1986). At this stage no group has identified the essential target(s) of the DDK required to initiate DNA replication. However, a number of studies suggest it may be the Mcm2-7 complex. First, the Dbf4/Cdc7 requirement can be bypassed by a *mcm5-bob1* mutation (an amino acid P83L). This structural change in Mcm5 causes an inefficient yet active Mcm2-7 complex to promote DNA replication (Hardy et al. 1997; Hoang et al. 2007; Robert A Sclafani, Tecklenburg & Pierce 2002). Additional genetic evidence linking the Mcms and DDK is the isolation a *DBF4* allele that is an allele-specific suppressor of *mcm2-1* and subsequently a physical interaction from yeast extracts was demonstrated (Lei et al. 1997). Secondly a number of groups have shown that Mcm2, 4, and 6 are phosphorylated by Dbf4/Cdc7 *in vitro* and possibly *in vivo* (Hisao Masai et al. 2006; Lei et al. 1997; Montagnoli et al. 2006; Sheu & Bruce Stillman 2006; Cho et al. 2006; H Masai et al. 2000; Yuki Komamura-Kohno et al. 2006). In addition some groups report that Dbf4/Cdc7 phosphorylation of the Mcm proteins at serines and threonines is directed by adjacent acidic residues and in some cases these acidic residues are generated by prior CDK phosphorylation at serines and threonines of the consensus sequence (H Masai et al. 2000; Montagnoli et al. 2006; Cho et al. 2006; Yuki Komamura-Kohno et al. 2006).

CDK prevents DNA re-replication

In addition to promoting DNA replication initiation the CDK activity is also required to prevent re-replication by blocking pre-RC assembly but exactly how it achieved this regulation was not clear (Hayles et al. 1994; Dahmann, J F Diffley & Nasmyth 1995;

Piatti et al. 1996). More specifically, the protein targets of the kinase were not identified in these earlier studies. Nonetheless it was appreciated that in the G1/S transition upon activation of CDK and initiation of replication the pre-RC was inhibited from reassembling. Subsequent work showed a number of pre-RC components as targets of the kinase and that their regulation was critical to prevent DNA re-replication (Nguyen, Co & Li 2001). The model that emerged from this work and others is that DNA replication is arranged into two mutually exclusive stages to prevent re-replication. The first stage is the assembly of pre-RCs in G1 when there is no CDK activity. In the second stage CDK activity rises and promotes replication initiation but concomitantly prevents pre-RC assembly. The second irreversible step, triggering of initiation prevents a state of licensing that occurs in G1. Other organisms employ different mechanisms and not all use CDK to prevent reinitiation. This thesis will be focused on this control as we understand it in *Saccharomyces cerevisiae*; for review of other organisms see (Arias & Walter 2007).

Cdc6 regulation

CDK regulates Cdc6 activity through multiple modes (Honey & Futcher 2007). One method of regulation is at the transcriptional level. Through phosphorylation of Swi5 it inactivates transcription and through inhibitory phosphorylation of Whi5 indirectly promotes transcriptional activation (Costanzo et al. 2004; de Bruin et al. 2004; Piatti, Lengauer & Nasmyth 1995). Although not established there was some evidence that CDK may also regulate transcription of Cdc6 by phosphorylating a transcription co-

factor, Swi6 (Sidorova, Mikesell & Breeden 1995). CDK also regulates the stability of Cdc6 by direct phosphorylation. Upon phosphorylation, Cdc6 is ubiquitinated by the SCF and is subsequently proteolyzed (Drury, Perkins & J F Diffley 2000). Another level of regulation is through Clb2/Cdc28 binding of Cdc6 at the N-terminal segment and is believed to inhibit Cdc6 by sequestration. This binding requires prior phosphorylation of CDK sites at the N-terminus by the S-phase CDK Clb5/Cdc28 (Mimura et al. 2004).

Orc1-6 regulation

The CDK regulation of ORC was the least characterized of the three mechanisms the CDK uses to prevent a second round of DNA replication. Nonetheless it was clear that the CDK targets at least two subunits of the ORC complex, *ORC2* and *ORC6*. Both are phosphorylated *in vitro* and by Clb5/Cdc28 with high specificity relative to Clb2/Cdc28 (Ubersax et al. 2003; Loog & David O Morgan 2005). In addition it was shown that this phosphorylation occurred in vivo and was critical to prevent re-initiation of DNA replication (Nguyen, Co & Li 2001). An additional level of regulation was identified in *ORC6* that was important in preventing re-replication. The interaction occurred between an RXL motif on ORC6 and Clb5/Cdc28 complex (Wilmes et al. 2004).

Mcm2-7 regulation

The Mcm2-7 complex is considered the replicative DNA helicase. It has been shown to have helicase activity in vitro as a subcomplex and full complex (Y Ishimi 1997; Bochman & Schwacha 2008). In addition the complex is found at replication forks and is required for initiation and elongation of DNA replication (O. M. Aparicio, Weinstein & S P Bell 1997; Labib, Tercero & J F Diffley 2000). The CDK regulation of Mcm2-7 complex is less clear.

There is some evidence suggesting that phosphorylation of human Mcm4 inhibits helicase activity of Mcm4,6,7 complex (Y Ishimi & Y Komamura-Kohno 2001). In mammalian cells, phosphorylation of Mcm4 at specific sites during the cell cycle inhibits helicase activity (Yuki Komamura-Kohno et al. 2006). The regulation of Mcm2-7 complex has been better analyzed in *Saccharomyces cerevisiae*. Earlier investigation had identified a cell cycle regulated localization phenotype of the Mcms in which the Mcm2-7 accumulated in the nucleus in late mitosis and persisted until the activation of the CDKs in late G1(Nguyen et al. 2000; Labib, J F Diffley & Kearsey 1999). Upon high CDK activity the Mcm2-7 are gradually excluded from the nucleus. The gradual exclusion was due to tethering of the Mcm2-7 to DNA in S-phase (Nguyen et al. 2000). Thus suggesting the chromatin bound pool is protected from the CDK regulated export but the soluble pool is subject to even G1 phase CDK activity (Labib, J F Diffley & Kearsey 1999). These two works suggested that the CDK dependent nuclear exclusions of the Mcm2-7 complex were important for preventing reinitiation. However, it was not until the seminal paper from our lab that showed nuclear exclusion of the Mcm2-7 complex was critical to prevent re-initiation of DNA replication (Nguyen, Co & Li 2001).

Nonetheless none of these works elucidate the exact mechanism of CDK regulation of the Mcm2-7 complex localization. Our effort to understand this regulation is the focus of chapter two. Briefly we identify NLS signals on *MCM2* and *MCM3* that are critical for the nuclear localization of the complex. Furthermore, we identify an NES on *MCM3* adjacent to the NLS. Moreover, we find that we can recapitulate the cell cycle regulated localization of the complex by fusing the NLS and NES signals from *MCM2* and *MCM3* to a heterologous protein, GFP. This allowed us to examine the cdk regulation of the localization without disrupting normal cellular role of the complex. Lastly we extend our analysis of NLS-NES transport module to the endogenous proteins and determine that CDK phosphorylation of Mcm3 promotes the nuclear exclusion of the Mcm2-7 complex. However, we note that this is not essential for complete export of the complex. Suggesting that there maybe additional CDK target sites that are important for nuclear exclusion of the Mcm2-7 complex.

Evolution of CDK regulation of the Mcm2-7 localization

In Chapter 3 we show that CDK phosphorylation site position and number are not conserved in a number of CDK substrates. However, work in the field has demonstrated that the homologue/ortholog in different species is regulated similarly despite the changes in position and number of CDK phosphorylation sites. The pre-RC protein ORC1 provided the best example of a protein whose regulation by phosphorylation was maintained despite the turnover of phosphorylation sites. We propose that the regulation is conserved through conservation of the clustering of phosphorylation sites. In addition,

although *Saccharomyces cerevisiae* regulates localization of Mcm2-7, other more divergent organisms do not regulate the nuclear-cytoplasmic localization. This raised the question of whether there was evolution of CDK regulation of Mcm3. We wanted to investigate the mechanism of evolution of this regulation. Investigation of this mechanism of evolution led to the finding that the Mcm3 CDK sites flanking the NLS were gained in a lineage leading to *Saccharomyces cerevisiae*. Interestingly, the gain in CDK regulation coevolved a nuclear export signal (NES).

CDK regulation of Polymerase alpha primase (Pol alpha primase)

In Chapter 4 we present some intriguing finding on potential CDK regulation of Pol alpha primase. The Polymerase alpha primase (Pol alpha primase) complex is composed of four essential subunits *POL1, POL12, PRI1* and *PRI2*. The Pol alpha primase complex's primer synthesis activity is essential for initiation and elongation of DNA replication. This essential role and cell cycle regulated activities prompted us to investigate a possible regulation of Pol alpha primase by CDK to prevent reinitiation. The removal of CDK phosphorylation sites on POL1 (*pol1-cdk13A*) abrogates the cell cycle regulated nuclear periphery localization. Moreover, introduction of the *pol1-cdk13A* allele in strains carrying alleles of *orc2-cdk6A orc6-cdk4A* and *MCM7-2NLS* (OM) that disrupts CDK regulation had increased frequency of segmental duplications and chromosomal

aneuploidy. When an ectopic copy of *pGAL-delntcdc6* was introduced the resultant strains (OMCP1) didn't undergo more reinitiation than strains that had just the OMC perturbations. The emergence of the genomic instabilities associated with strains carrying mutations in Pol alpha primase suggests toxicity linked to these mutations perhaps a potential low undetectable level of reinitiation.

References:

Aparicio, O.M., Weinstein, D.M. & Bell, S.P., 1997. Components and dynamics of DNA replication complexes in S. cerevisiae: redistribution of MCM proteins and Cdc45p during S phase. *Cell*, 91(1), 59-69.

Arias, E.E. & Walter, J.C., 2007. Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. *Genes & Development*, 21(5), 497-518.

Bell, S.P. & Stillman, B., 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature*, 357(6374), 128-34.

Bell, S.P. & Dutta, A., 2002. DNA replication in eukaryotic cells. *Annual Review of Biochemistry*, 71, 333-74.

Blow, J.J. & Nurse, P., 1990. A cdc2-like protein is involved in the initiation of DNA replication in Xenopus egg extracts. *Cell*, 62(5), 855-62.

Blow, J.J. & Hodgson, B., 2002. Replication licensing--defining the proliferative state? *Trends in Cell Biology*, 12(2), 72-8.

Bochman, M.L. & Schwacha, A., 2008. The Mcm2-7 complex has in vitro helicase activity. *Molecular Cell*, 31(2), 287-93.

de Bruin, R.A.M. et al., 2004. Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. *Cell*, 117(7), 887-98.

Cho, W. et al., 2006. CDC7 kinase phosphorylates serine residues adjacent to acidic amino acids in the minichromosome maintenance 2 protein. *Proceedings of the National Academy of Sciences of the United States of America*, 103(31), 11521-6.

Costanzo, M. et al., 2004. CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast. *Cell*, 117(7), 899-913.

Dahmann, C., Diffley, J.F. & Nasmyth, K.A., 1995. S-phase-promoting cyclin-dependent kinases prevent re-replication by inhibiting the transition of replication origins to a pre-replicative state. *Current Biology: CB*, 5(11), 1257-69.

Diffley, J.F., 1996. Once and only once upon a time: specifying and regulating origins of DNA replication in eukaryotic cells. *Genes & Development*, 10(22), 2819-30.

Drury, L.S., Perkins, G. & Diffley, J.F., 2000. The cyclin-dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle. *Current Biology: CB*, 10(5), 231-40.

Hardy, C.F. et al., 1997. mcm5/cdc46-bob1 bypasses the requirement for the S phase activator Cdc7p. *Proceedings of the National Academy of Sciences of the United States of America*, 94(7), 3151-5.

Hartwell, L.H., 1973. Three additional genes required for deoxyribonucleic acid synthesis in Saccharomyces cerevisiae. *Journal of Bacteriology*, 115(3), 966-74.

Hartwell, L.H. et al., 1974. Genetic control of the cell division cycle in yeast. *Science (New York, N.Y.)*, 183(120), 46-51.

Hartwell, L.H. et al., 1973. Genetic Control of the Cell Division Cycle in Yeast: V. Genetic Analysis of cdc Mutants. *Genetics*, 74(2), 267-286.

Hayles, J. et al., 1994. Temporal order of S phase and mitosis in fission yeast is determined by the state of the p34cdc2-mitotic B cyclin complex. *Cell*, 78(5), 813-22.

Hoang, M.L. et al., 2007. Structural changes in Mcm5 protein bypass Cdc7-Dbf4 function and reduce replication origin efficiency in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, 27(21), 7594-602.

Honey, S. & Futcher, B., 2007. Roles of the CDK phosphorylation sites of yeast Cdc6 in chromatin binding and rereplication. *Molecular Biology of the Cell*, 18(4), 1324-36.

Ishimi, Y., 1997. A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex. *The Journal of Biological Chemistry*, 272(39), 24508-13.

Ishimi, Y. & Komamura-Kohno, Y., 2001. Phosphorylation of Mcm4 at specific sites by cyclin-dependent kinase leads to loss of Mcm4,6,7 helicase activity. *The Journal of Biological Chemistry*, 276(37), 34428-33.

Kelly, T.J. & Brown, G.W., 2000. Regulation of chromosome replication. *Annual Review of Biochemistry*, 69, 829-80.

Komamura-Kohno, Y. et al., 2006. Site-specific phosphorylation of MCM4 during the cell cycle in mammalian cells. *The FEBS Journal*, 273(6), 1224-39.

Labib, K., Diffley, J.F. & Kearsey, S.E., 1999. G1-phase and B-type cyclins exclude the DNA-replication factor Mcm4 from the nucleus. *Nature Cell Biology*, 1(7), 415-22.

Labib, K., Tercero, J.A. & Diffley, J.F., 2000. Uninterrupted MCM2-7 function required for DNA replication fork progression. *Science (New York, N.Y.)*, 288(5471), 1643-7.

Lei, M. et al., 1997. Mcm2 is a target of regulation by Cdc7-Dbf4 during the initiation of DNA synthesis. *Genes & Development*, 11(24), 3365-74.

Loog, M. & Morgan, D.O., 2005. Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. *Nature*, 434(7029), 104-8.

Masai, H. et al., 2000. Human Cdc7-related kinase complex. In vitro phosphorylation of MCM by concerted actions of Cdks and Cdc7 and that of a criticial threonine residue of Cdc7 bY Cdks. *The Journal of Biological Chemistry*, 275(37), 29042-52.

Masai, H. et al., 2006. Phosphorylation of MCM4 by Cdc7 kinase facilitates its interaction with Cdc45 on the chromatin. *The Journal of Biological Chemistry*, 281(51), 39249-61.

Masumoto, H. et al., 2002. S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. *Nature*, 415(6872), 651-5.

Mimura, S. et al., 2004. Phosphorylation-dependent binding of mitotic cyclins to Cdc6 contributes to DNA replication control. *Nature*, 431(7012), 1118-23.

Montagnoli, A. et al., 2006. Identification of Mcm2 phosphorylation sites by S-phase-regulating kinases. *The Journal of Biological Chemistry*, 281(15), 10281-90.

Morgan, D.O., 1997. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annual Review of Cell and Developmental Biology*, 13, 261-91.

Nguyen, V.Q. et al., 2000. Clb/Cdc28 kinases promote nuclear export of the replication initiator proteins Mcm2-7. *Current Biology: CB*, 10(4), 195-205.

Nguyen, V.Q., Co, C. & Li, J.J., 2001. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature*, 411(6841), 1068-73.

Patterson, M. et al., 1986. Molecular characterization of cell cycle gene CDC7 from Saccharomyces cerevisiae. *Molecular and Cellular Biology*, 6(5), 1590-8.

Piatti, S. et al., 1996. Activation of S-phase-promoting CDKs in late G1 defines a "point of no return" after which Cdc6 synthesis cannot promote DNA replication in yeast. *Genes & Development*, 10(12), 1516-31.

Piatti, S., Lengauer, C. & Nasmyth, K., 1995. Cdc6 is an unstable protein whose de novo synthesis in G1 is important for the onset of S phase and for preventing a 'reductional' anaphase in the budding yeast Saccharomyces cerevisiae. *The EMBO Journal*, 14(15), 3788-99.

Rao, P.N. & Johnson, R.T., 1970. Mammalian cell fusion: studies on the regulation of DNA synthesis and mitosis. *Nature*, 225(5228), 159-64.

Schwob, E. & Nasmyth, K., 1993. CLB5 and CLB6, a new pair of B cyclins involved in DNA replication in Saccharomyces cerevisiae. *Genes & Development*, 7(7A), 1160-75.

Sclafani, R.A., 2000. Cdc7p-Dbf4p becomes famous in the cell cycle. *Journal of Cell Science*, 113 (Pt 12), 2111-7.

Sclafani, R.A. & Jackson, A.L., 1994. Cdc7 protein kinase for DNA metabolism comes of age. *Molecular Microbiology*, 11(5), 805-10.

Sclafani, R.A., Tecklenburg, M. & Pierce, A., 2002. The mcm5-bob1 bypass of Cdc7p/Dbf4p in DNA replication depends on both Cdk1-independent and Cdk1-dependent steps in Saccharomyces cerevisiae. *Genetics*, 161(1), 47-57.

Sheu, Y. & Stillman, B., 2006. Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. *Molecular Cell*, 24(1), 101-13.

Sidorova, J.M., Mikesell, G.E. & Breeden, L.L., 1995. Cell cycle-regulated phosphorylation of Swi6 controls its nuclear localization. *Molecular Biology of the Cell*, 6(12), 1641-58.

Stillman, B., 1996. Cell cycle control of DNA replication. *Science (New York, N.Y.)*, 274(5293), 1659-64.

Tanaka, S. et al., 2007. CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature*, 445(7125), 328-32.

Ubersax, J.A. et al., 2003. Targets of the cyclin-dependent kinase Cdk1. *Nature*, 425(6960), 859-64.

Wilmes, G.M. et al., 2004. Interaction of the S-phase cyclin Clb5 with an "RXL" docking sequence in the initiator protein Orc6 provides an origin-localized replication control switch. *Genes & Development*, 18(9), 981-91.

Wyrick, J.J. et al., 2001. Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. *Science (New York, N.Y.)*, 294(5550), 2357-60.

Zegerman, P. & Diffley, J.F.X., 2007. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature*, 445(7125), 281-5.

Chapter 2

CDK phosphorylation of a novel NLS-NES module distributed between two subunits of the Mcm2-7 complex prevents chromosomal rereplication

ABSTRACT

Cyclin-dependent kinases (CDKs) use multiple mechanisms to block reassembly of pre-replicative complexes (pre-RCs) at replication origin to prevent inappropriate re-replication. In *Saccharomyces cerevisiae*, one of these mechanisms promotes the net nuclear export of a pre-RC component, the Mcm2-7 complex, during S, G2 and M phases. Here we identify two partial nuclear localization signals (NLSs) on Mcm2 and Mcm3 that are each necessary, but not sufficient, for nuclear localization of the Mcm2-7 complex. When brought together in *cis*, however, the two partial signals constitute a potent NLS, sufficient for robust nuclear localization when fused to an otherwise cytoplasmic protein. We also identify a Crm1-dependent nuclear export signal (NES) adjacent to the Mcm3 NLS. Remarkably, the Mcm2-Mcm3 NLS and the Mcm3 NES are sufficient to form a transport module that recapitulates the cell cycle-regulated localization of the entire Mcm2-7 complex. Moreover, we show that CDK regulation promotes net export by phosphorylation of the Mcm3 portion of this module and that nuclear export of the Mcm2-7 complex is sufficient to disrupt replication initiation. We speculate that the distribution of partial transport signals among distinct subunits of a complex may enhance the specificity of protein localization and raise the possibility that previously undetected distributed transport signals are used by other multiprotein complexes.

INTRODUCTION

The faithful transmission of genetic information during cell division requires that complete duplication of the genome during S phase strictly alternate with accurate segregation of the duplicated genome during M phase. Eukaryotic cells ensure that their genome is duplicated precisely once per cell cycle by enforcing a single round of replication initiation at each of the hundreds to thousands of replication origins scattered throughout their genome. We and others have shown that re-initiation in the budding yeast *Saccharomyces cerevisiae* causes a rapid and serious insult to the genome, triggering a DNA damage response and cell cycle arrest (Archambault *et al*., 2005; Green and Li, 2005). In other metazoans, rereplication also induces checkpoint responses and, in some cases, leads to apoptosis (Mihaylov *et al*., 2002; Melixetian *et al*., 2004; Zhu *et al*., 2004). Thus, restricting DNA replication initiation to a single round per cell cycle is critical for genome integrity and cell survival.

Cyclin-dependent kinases play a critical role in the cell cycle regulation of replication initiation by controlling both the activation and formation of the pre-replicative complex (pre-RC), a critical intermediate in the initiation reaction (reviewed in Bell and Dutta, 2002; Diffley, 2004). Assembly of the pre-RC in G1 phase, when CDK activity is low, makes origins competent for replication initiation later in the cell cycle when CDK activity is induced. The pre-RC is assembled when the origin recognition complex (ORC) binds origins and recruits Cdc6 and Cdt1 to help load the putative replicative helicase, the heterohexameric Mcm2-7 complex. Activation of the

pre-RC by CDKs and Cdc7-Dbf4 kinase then leads to recruitment of additional replication proteins and the triggering of initiation. This activation is accompanied by disassembly of the pre-RC: Cdc6 and Cdt1 dissociate from the origin and the Mcm2-7 complex is thought to move with the replication fork as part of the replisome.

In addition to triggering initiation, CDKs can inhibit reinitiation by blocking reassembly of the pre-RC (Diffley, 2004). This inhibitory function has been investigated most extensively in *S. cerevisiae*, where at least three inhibitory targets of the CDK Cdc28 have been identified: ORC, Cdc6, and the Mcm2-7 complex (Nguyen *et al.*, 2001). The mechanism by which this inhibition occurs is best understood for Cdc6, which is regulated by Cdc28 both transcriptionally and post-translationally (Moll *et al.*, 1991; Piatti *et al.*, 1995; Drury *et al.*, 1997; Mimura *et al.*, 2004). Cdc28 kinase inhibits ORC function by phosphorylation of Orc2 and Orc6 (Nguyen *et al.*, 2001) and binding of the S-phase cyclin Clb5 to Orc6 (Wilmes *et al.*, 2004). In other eukaryotes, CDK phosphorylation has also been shown to promote the ubiquitin mediated degradation of Cdt1 (Liu *et al.*, 2004; Sugimoto *et al.*, 2004; Thomer *et al.*, 2004).

We and others have shown that Cdc28 promotes the nuclear exclusion of the Mcm2-7 complex (Labib *et al.*, 1999; Nguyen *et al.*, 2000). In G1 phase, when Cdc28 kinase levels are low, the Mcm2-7 complex accumulates in the nucleus independent of its loading onto origins. After Cdc28 kinase becomes active in late G1 phase, the Mcm2-7 complex experiences net nuclear export until it is excluded from the nucleus. Little is understood about how Cdc28 controls localization of the Mcm2-7 complex. For

example, it is not known whether Cdc28 promotes Mcm2-7 nuclear export by directly phosphorylating the Mcm complex, by targeting some of the many replication proteins known to interact with the complex, or by modulating the activity of the transport machinery.

Most transport of proteins across the nuclear envelope is mediated by a family of nucleocytoplasmic transport receptors, which shuttle proteins through the nuclear pore in a unidirectional manner (reviewed in Weis, 2003). Nuclear import is mediated by the binding of nuclear import receptors to nuclear localization signals (NLSs) and nuclear export is mediated by the binding of nuclear export receptors to nuclear export signals (NESs). Proteins that shuttle between the nucleus and cytoplasm, like the Mcm2-7 complex, often contain both NLSs and NESs, with the relative rates of nuclear import and export specified by these signals determining the steady state localization of a protein. Hence, mechanisms that directly or indirectly affect the activity of NLSs and NESs can control the nucleocytoplasmic localization a protein (reviewed in Jans *et al*., 2000). Proteins that do not contain nuclear transport signals can also be nuclear localized by associating with a protein that does contain these signals. In fact, studies that have dissected the transport signals responsible for localization of multiprotein complexes have uncovered numerous examples of a single subunit providing the transport signal for the entire complex (Maridor *et al*., 1993; Pereira *et al*., 1998; Leslie *et al*., 2004; Subramaniam and Johnson, 2004; Wendler *et al*., 2004).

In this study, we identify two partial nuclear localization signals on two distinct subunits of the Mcm2-7 complex, Mcm2 and Mcm3. These elements are each necessary for nuclear localization of the entire Mcm2-7 complex, but unlike canonical NLSs, neither is sufficient for this localization. Together, however, they exhibit strong import activity when fused to a heterologous protein. We have also identified an NES adjacent to the Mcm3 NLS. All three signals effectively constitute a transport module that, when fused to a heterologous protein, can recapitulate the cell-cycle regulated localization of the Mcm2-7 complex. We show that CDK phosphorylation of the Mcm3 portion of this transport module promotes the net nuclear export of both a heterologous fusion protein and the Mcm2-7 complex. We also demonstrate that this CDK directed export is sufficient to disrupt the initiation of DNA replication, establishing that this regulation contributes significantly to the control of replication initiation. Finally, we suggest that distribution of transport modules among distinct subunits of a complex may couple protein localization to complex assembly and could be used by other multiprotein complexes.

MATERIALS AND METHODS

Yeast strains, media, and growth

S. cerevisiae strains used in this study were derivatives of YJL310 (Detweiler and Li, 1998; Nguyen *et al*., 2000) or W303. YEP medium and synthetic complete medium (Guthrie and Fink, 1991) were supplemented with 2% dextrose (YEPD; SDC), or 2%

raffinose (YEPR; SRafC). The *GAL10* promoter (*pGAL*) was induced by addition of 2% galactose unless otherwise indicated.

For *in vivo* labeling of Mcm3, low phosphate YEPD was prepared as follows. 10g yeast extract and 20g bacto peptone were dissolved in 1 liter of water and 10 ml 1M $MgSO_4$ and 10ml concentrated NH_4OH were added. A cloudy precipitate, formed during 30 min of stirring, was removed by filtration through Whatman # 1 filter. The pH was adjusted to 5.8 with concentrated HCl and the media autoclaved then supplemented with 10ml ADE/TRP (each 5mg/ml).

Plasmids and Strains

Plasmids and strains used in this study are described in detail in supplementary information.

Cell growth, arrest, and release

To arrest cells, α-factor was used at a final concentration of 50 ng/ml (all strains were *bar1*, see Figures 2, 3, 4, 6, and 10), and nocodazole (NOC) was used at a final concentration of 15 μg/ml (see Figures 2, 3, 4, 6, and 8). For Figures 2, cell cycle blocks were relieved by filtering the cells, washing them three times with an equal volume of resuspension medium prewarmed to the appropriate temperature, then resuspending them in the appropriate medium. For Figure 4B and Figure 6, Cells were released from the α-factor arrest by addition of Pronase at 100μg/mL. To inactivate Crm1-T539C, Leptomycin B (a gift of Minoru Yoshida (Nishi *et al*., 1994; Kudo *et al*., 1999)) was added to a final concentration of 100ng/ml.

Immunoblot analysis

Immunoblot analysis was performed as described (Nguyen *et al*., 2001). Blots were probed with c-Myc polyclonal antibody at a 1:200 dilution (sc-789, Santa Cruz Biotechnology).

Flow Cytometry Analysis

Samples were fixed in Ethanol and processed as described in (Green and Li, 2005).

Fluorescence microscopy

For fluorescence microscopy of live cells for Figures 2, 3, and 10, cells were rapidly washed with PBS (137mM NaCl, 2.7mM KCl, 43mM Na_2HPO_4, 14mM KH_2PO_4 at pH 7.4) and visualized using a Leica DMLB fluorescence microscope with a 100X PL

Fluotar oil immersion objective. Images were acquired with an Optronics DEI-750 CCD camera using the Scion Image Software program. Live cell GFP fluorescence microscopy for Figures 4, 5, 6B, 8, and 9 were performed as follows: cells (OD_{600}≈0.2) were washed with 1mL SDC, pelleted and resuspended in 40μL of SDC and visualized within 8 minutes of sampling using 60X PlanApo oil objective on an Olympus BX60 microscope with EndowBandpass GFP filter cube (Chroma, Rockingham VT Cat.#41017). For Figure 6A time point samples were fixed in 100% Ethanol. Samples were pelleted and washed in PBS with 50ng/mL DAPI, visualized using 60X PlanApo oil objective on an Olympus BX60 microscope with Endowbandpass GFP filter cube. Images were acquired with Openlab 3.1.7 software (Improvision) driving a Hamamatsu ORCA ER CCD (exposure time was on the order of 300ms). Figure panels were assembled using Openlab 3.1.7 and Illustrator 10.0.

In vitro phosphorylation of Mcm3

One ng of purified Cdc28-His_6 and 10ng purified Clb2-MBP (a gift of Jeff Ubersax, Morgan lab, UC San Francisco, CA) mixed in 2μl of storage buffer (300mM NaCl, 25mM Hepes-KOH pH 7.4, 10% glycerol) for 5 min on ice, were added to 23μl kinase buffer (50mM Hepes-KOH pH 7.4, 2mM $MgCl_2$, 1mM DTT) containing 0.1mM ATP, 2μCi [γ-^{32}P] ATP (Amersham), and 1μg purified GST-Mcm3, 1μg purified GST-Mcm3-cdk-5A, or 1μg purified GST-Mcm3-cdk-7A, and incubated at 25^{0}C for 15 min. The reaction was stopped by adding 10μl 4X SDS sample buffer (250mM Tris-HCl pH 6.8, 2.8M β-mercaptoethanol, 10% SDS, 0.06% bromophenol blue and 40% glycerol)

and boiling for 5 min. The reaction products were resolved on an 8% SDS-PAGE, stained with coomassie, then dried and subjected to autoradiography.

In vivo phosphate labeling of Mcm3

YJL4313, YJL4315, and YJL4324 expressing Myc-Mcm3 were grown in low phosphate YEPD to OD_{600} of 0.5-0.8 at 30^0C. 1 OD unit was spun down, resuspended in 1ml low phosphate YEPD containing 1mCi orthophosphate (Amersham, PBS.13A ^{32}P-orthophosphate, acid free), and incubated for 60 min at 30° C. The labeled cells were pelleted in a 1.5 ml screw cap tube and placed on ice. In parallel, 15 OD units of exponentially growing YJL2160 expressing untagged Mcm3 were pelleted, resuspended in 1 ml dH_20, and added to the labeled cell pellet. After thorough mixing, the labeled and unlabeled cells were repelleted and frozen in liquid nitrogen. Frozen cell pellets were resuspended in 350μl lysis buffer (25mM Hepes pH 8.0, 150mM NaCl, 0.1% NP-40, 1mM Na_3VO_4, 1mM EDTA pH 8.0, 80mM ß-Glycerol Phosphate, and 50mM NaF) containing 1mM phenylmethylsulfonyl fluoride (PMSF), 1μg/ml leupeptin, 10μg/ml aprotinin, 2mM benzamidine, and 1μg/ml pepstatin . Cells were lysed by bead beating in a Mini Bead Beater (Biospec Products Cat# 693) for two 1 min pulses. After pelleting the cell debris, the supernatant was clarified by centrifugation for 10 min at 20,800g at 4^0C, then incubated with 1μl of c-Myc monoclonal antibody (9E10, MMS-150R, Covance Inc.) prebound to 10 μl of magnetic beads (Dynabead Protein G, 100.03, Dynal Biotech Inc.). After 2 hr incubation the magnetic beads were washed and resuspended in 25μl of 2X SDS sample buffer. The sample was loaded on a 4%-15% SDS-PAGE gradient gel (BioRad) and the dried gel developed on a phosphorimager.

RESULTS

Mcm2 and Mcm3 each contain a sequence required for nuclear localization of their respective proteins

As a first step to understanding the CDK regulation of Mcm2-7 nucleocytoplasmic localization we first sought to identify the nuclear localization signals (NLSs) and nuclear export signals (NESs) responsible for nucleocytoplasmic transport of the Mcm proteins. Our previous work, suggesting that Mcm proteins colocalize as a complex (Nguyen *et al.*, 2000), raised the possibility that NLSs or NESs on one or more of the Mcm proteins could be responsible for nuclear or cytoplasmic localization of the entire complex. We scanned the amino acid sequences of the six Mcm proteins (Mcm2 – Mcm7) for the two NLS sequence motifs that are recognized and bound by the import receptor adapter, importin alpha (reviewed in Jans *et al.*, 2000). One of these motifs, represented by the SV40 NLS (PKKKRKV), contains a single cluster of highly basic residues, and the other motif, represented by the nucleoplasmin NLS (KRPAATKKAGQAKKKKL) contains two smaller clusters of basic residues separated by 7-22 amino acids. Four good matches to these motifs were identified on Mcm2 (residues 5-9, RRRRR; residues 150-155, RRRRRR), Mcm3 (residues 766-772, PKKRQRV), and Mcm7 (residues 199-219, RR-13aa-RRYRKK).

To determine whether any of these sequences were required for nuclear transport of the Mcm proteins, we initially examined whether these sequences were essential for cell viability. We reasoned that a sequence required for nuclear localization of any of the Mcm proteins would also be essential for viability, since the Mcm proteins must perform their essential replication function in the nucleus. We generated mutant *mcm* genes with alanines substituting for multiple basic residues in the identified sequences and attempted to replace the endogenous genes with these mutant genes by two-step gene replacement. Haploid strains expressing mutant Mcm proteins with alanine substitutions on Mcm2 (at amino acid residues 150-155) or on Mcm7 (at residues R214, R216, and K217) were viable and exhibited growth rates indistinguishable from wild-type strains. Thus, these sequences are not required for nuclear localization of Mcm proteins. In contrast, we could not isolate haploid strains expressing mutant Mcm proteins containing alanine substitutions in Mcm2 (at residues 5-9) or Mcm3 (at residues 766-772) (Figure 1A; *mcm2-nls* and *mcm3-nls*) (Supplementary Figure 1). Fusing sequences encoding two tandem copies of the SV40 NLS onto the mutant *mcm2-nls* or *mcm3-nls* genes did allow isolation of gene replacement strains expressing these NLS-tagged mutant genes. Tetrad analysis confirmed that mutations in residues 5-9 of Mcm2 or 766-772 of Mcm3 resulted in inviability and that fusion of the SV40 NLS to these mutated Mcm proteins restored viability. Together these results demonstrate that Mcm2 residues 5-9 and Mcm3 residues 766-772 sequences are essential for viability and that their essential role may be to direct the nuclear localization of their respective proteins.

To directly examine the role of residues 5-9 of Mcm2 and 766-772 of Mcm3 in the localization of their respective protethese ins, we used fluorescence microscopy to examine the subcellular distribution of mutant Mcm2 and Mcm3 fused to GFP during the G1 phase of the cell cycle (Figure 2A and 2B). We previously reported that wild-type Mcm2, 3, 4, 6 and 7 tagged with GFP are fully functional when expressed as the sole copy of their respective Mcm proteins, indicating that these fusion proteins can complex with other Mcms (Nguyen *et al*., 2000). In Figure 2A and 2B, the fusion proteins were expressed in addition to the endogenous wild-type *MCM* genes. The latter supported the viability of these cells but did not interfere with the fluorescence analysis of the GFP fusion proteins. As expected, GFP fusions to wild-type Mcm2 or Mcm3 accumulated in the nucleus during G1 phase. In contrast, Mcm2-nls-GFP and GFP-Mcm3-nls were distributed throughout the cytoplasm. This mislocalization could be rescued by fusing two tandem copies of the SV40 NLS to the mutant fusion proteins; both Mcm2-nls-GFP-$SVNLS_2$ and $SVNLS_2$-GFP-Mcm3-nls were constitutively nuclear. These results directly demonstrate that residues 5-9 of Mcm2 and residues 766-772 of Mcm3 are required for the nuclear localization of their respective proteins. Hence, we refer to these residues as the Mcm2 NLS and Mcm3 NLS, respectively. Our results corroborate previously published results in *S. cerevisiae*, implicating the Mcm3 residues in nuclear localization of Mcm3 (Young *et al*., 1997), and in S. *pombe*, implicating N-terminal Mcm2 residues in nuclear localization of Mcm2 (Pasion and Forsburg, 1999).

The Mcm2 and Mcm3 NLSs are each required for localization of the Mcm2-7 complex

We have previously shown that fusion of two tandem SV40 NLSs to any Mcm subunit promotes the constitutive nuclear localization of both that subunit and each of the other Mcm subunits (Nguyen *et al*., 2000). Hence, we suspected the lethality arising from mutation of the Mcm2 or Mcm3 NLSs could be rescued by fusing the SV40 NLS to other Mcm subunits. To examine this possibility, we reattempted two-step gene replacement of *MCM2* and *MCM3* with *mcm2-nls* and *mcm3-nls*, respectively, in haploid strains containing two tandem copies of the SV40 NLS fused to other Mcm proteins (Supplementary Figure 1). We could successfully replace *MCM2* with *mcm2-nls* when the SV40 NLSs were fused to Mcm3, Mcm4, Mcm5, or Mcm6 and could successfully replace *MCM3* with *mcm3-nls* when the SV40 NLSs were fused to Mcm2, Mcm4, Mcm5, or Mcm6. Furthermore, the mutant Mcm2-GFP or GFP-Mcm3 in these strains displayed constitutive nuclear localization (data not shown). These results indicate that mislocalization of the mutant Mcm2-GFP or GFP-Mcm3 can be rescued by ensuring the nuclear localization of one other Mcm subunit. These results further suggest that mutating the Mcm2 or Mcm3 NLS disrupts the nuclear localization of all other Mcm subunits; otherwise an Mcm subunit that could retain its nuclear localization would have rescued the NLS mutations without requiring fusion to the SV40 NLS.

To directly examine whether the Mcm2 or Mcm3 NLS is required for the nuclear localization of other Mcm subunits, we performed a set of experiments exemplified by the one shown in Figure 2C. In this experiment, we examined the effect of mutating the NLS in Mcm2 on the nuclear import of Mcm7-GFP during the transition from G2/M to G1 phase. Because the NLS mutation is lethal we complemented the mutant *mcm2-nls*

gene with an *MCM2* gene encoding a conditionally degraded version of the Mcm2 protein (*mcm2-td*). Mcm2-td is targeted for ubiquitin-mediated degradation by both raising the temperature to 37°C and shifting cells into galactose containing media to induce the E3 ubiquitin ligase Ubr1; under these restrictive conditions there is no detectable Mcm2-td after 60 minutes (Labib *et al*., 2000). In this conditionally complemented strain we could examine how the NLS-defective Mcm2 protein directs the localization of Mcm7-GFP after Mcm2-td is degraded. In parallel, we generated control strains where either the wild-type *MCM2* gene or the suppressed mutant gene *mcm2-nls-GFP-SVNLS*$_2$ was introduced into the *mcm2-td* parent strain.

Experimental cells expressing Mcm7-GFP, Mcm2-nls, and Mcm2-td under permissive conditions for Mcm2-td (rich medium containing raffinose at 25° C) were arrested in metaphase with nocodazole. Once arrested, we induced degradation of Mcm2-td by shifting the cells to restrictive conditions (adding galactose at 37°C). After 30 min, the cells were released from the nocodazole arrest into an α-factor G1 arrest, still under restrictive conditions. Mcm proteins normally enter the nucleus during this G2/M to G1 phase transition, but Mcm7-GFP failed to accumulate in the nucleus of these cells (Figure 2C). In contrast, Mcm7-GFP strongly accumulated in the nucleus of the control strains expressing either Mcm2 or Mcm2-nls-SVNLS$_2$ (Figure 2C). Moreover, when cells were maintained at permissive conditions for the Mcm2-td proteins, Mcm7-GFP accumulated in the nucleus in all three strains (data not shown). These results indicate that the Mcm2 NLS is required for the nuclear localization of Mcm7 in G1 phase. Similar experiments with Mcm3 (Figure 2D) demonstrate that the Mcm3 NLS is also required for

nuclear localization of Mcm7. Finally, by repeating these experiments with GFP fused to other Mcm subunits, we have been able to show that the Mcm2 NLS is required for nuclear localization of Mcm3 and Mcm4, and the Mcm3 NLS is required for nuclear localization of Mcm2 (data not shown). Taken together, these results suggest that the Mcm2 and Mcm3 NLSs are required, not just for nuclear localization of their respective proteins, but of the entire Mcm2-7 complex.

<u>Together the Mcm2 and Mcm3 NLS are sufficient for strong nuclear localization activity</u>

The classical definition of an NLS is a sequence that is both necessary and sufficient for directing the nuclear localization of proteins. To determine whether the Mcm2 and Mcm3 NLSs are sufficient to direct the nuclear localization of proteins, we fused them to three tandem copies of GFP (GFP_3). With a combined molecular mass of 82kD, these tandem GFPs are larger than the 60kD upper limit for proteins to diffuse through nuclear pores and therefore must be actively transported to enter and exit the nucleus (reviewed in Weis, 2003). These fusion proteins were placed under the control of the regulatable *GAL10* promoter. To ensure that the observed localization was not inherited from a previous stage of the cell cycle, yeast cells containing these fusion constructs were first arrested in G1 phase or G2/M phase before the fusion proteins were induced by galactose.

When the SV40 NLS was fused to GFP_3, strong nuclear localization was observed. In contrast, sequences containing the Mcm2 NLS (amino acids 1-17) or the

Mcm3 NLS (amino acids 760-789), and hereafter referred to as NLS2 and NLS3, respectively, showed only a very weak ability to localize GFP_3 to the nucleus at either stage of the cell cycle (Figure 3). Hence, individually, these two sequences are insufficient to confer robust nuclear localization on a heterologous protein. This conclusion is consistent with the observation that neither NLS is sufficient to direct the Mcm2-7 complex into the nucleus in the absence of the other. Together, however, the NLS2 and NLS3 strongly directed GFP_3 into the nucleus in both G1 and G2/M phases (Figure 3). Thus, the weak Mcm2 and Mcm3 NLSs can functionally act as a single strong NLS. Using a larger segment spanning the Mcm3 NLS (amino acids 746-789), which increases the spacing between the Mcm2 and Mcm3 NLSs (to 30 amino acids), also resulted in strong composite NLS activity..

Our findings differ from the previous study in *S. cerevisiae* that identified the Mcm3 NLS (Young *et al.*, 1997). In that study, the sequence was reported to be not only necessary for nuclear localization of Mcm3, but also sufficient for nuclear localization of a heterologous protein. That conclusion, however, was based on a slightly different segment spanning the Mcm3 NLS (amino acids 755-781 versus our segment of amino acids 760-789) examined in combination with 50 amino acids of the Leu2 protein. When we fused that Mcm3 segment without the Leu2 segment to our tandem GFP reporter, we still observed poor NLS activity relative to the SV40 NLS or the combined NLS2-NLS3 (data not shown). Thus, our examination of an isolated Mcm3 NLS indicates that this NLS only has weak activity.

Mcm3 contains a Crm1-dependent NES which cooperates with the Mcm2 and Mcm3 NLSs to form a cell-cycle regulated transport module

We have previously shown that the Mcm2-7 complex undergoes net nuclear export when cells activate Cdc28 kinase activity (Nguyen *et al*., 2000). Because each subunit is too large to diffuse through the nuclear pore, we suspected the complex contains nuclear export signals on one or more subunits. The most recognizable NES motif identified to date is a leucine-rich motif that recruits the nucleocytoplasmic export receptor Crm1 (Fornerod *et al*., 1997; Fukuda *et al*., 1997; Ossareh-Nazari *et al*., 1997; Stade *et al*., 1997) and is exemplified by the NESs of the HIV REV and PKIα (LQLPPLERLTL and LALKLAGLDI respectively),(Fischer *et al*., 1995; Wen *et al*., 1995). One such motif (LQRRLQLGL; aa 834-842) is present in a 67 amino acid segment (aa 790-856) immediately C-terminal of the Mcm3 NLS segment. To test the export activity of this potential NES, we added it to a GFP_3 reporter construct containing a composite Mcm2-Mcm3 NLS. In this new construct (Figure 1B) the adjacent Mcm3 NLS and NES are derived from one contiguous 111 amino acid segment of Mcm3 (aa 746-856).

When the resulting fusion protein was induced in exponentially growing cells, it displayed cell cycle regulated localization that was reminiscent of the CDK regulation of Mcm2-7 localization (Labib *et al*., 1999; Nguyen *et al*., 2000). The protein was distributed throughout the cytoplasm of budded cells, which contain active Cdc28 kinase, and was predominantly nuclear in unbudded G1 phase cells, which contain little or no

active Cdc28 kinase. Moreover, the fusion protein was strongly nuclear in G1 cells arrested with α-factor, and was distributed throughout the cytoplasm in G2/M cells arrested with nocodazole (Figure 4A-two left panels). The cytoplasmic localization in nocodazole-arrested cells was dependent on the leucine-rich motif (Figure 1B and 4A).

To demonstrate that this cytoplasmic localization was due to net nuclear export during passage through the cell cycle, we introduced the GFP_3 fusion construct into cells containing a leptomycin B-sensitive allele of *CRM1*, *crm1-T539C* (Neville and Rosbash, 1999). Cells expressing the fusion protein were released from a G1 arrest into a G2/M arrest, either in the presence or absence of 100ng/ml leptomycin B. By the time both cultures had completed S phase (60 min) the GFP_3 fusion protein was distributed throughout the cytoplasm in the absence of leptomycin B, but remained strongly nuclear in its presence (Figure 4B). Similarly, addition of leptomycin B to exponentially growing cells expressing the fusion protein resulted in constitutive nuclear localization of the protein (data not shown). Thus, the redistribution of this fusion protein from nucleus to cytoplasm is indeed due to nuclear export, and is dependent on the Crm1 export receptor as well as the leucine-rich motif. We henceforth refer to the 67 amino acid segment downstream of the Mcm3 NLS as the Mcm3 NES or NES3. Importantly, the Mcm2 NLS and the contiguous Mcm3 NLS and NES behave as a minimal transport module (NLS2-NLS3NES3) that recapitulates the cell cycle regulated localization of the entire Mcm2-7 complex.

The Mcm3 NES promotes the nuclear export of Mcm subunits

We next examined whether the Mcm3 NES promotes nuclear export in the context of the full Mcm2-7 complex. To do this, we generated a haploid strain in which the wild-type endogenous *MCM3* gene was replaced by a mutant *GFP-Mcm3-nes* gene, which contains alanine substitutions at the leucine residues of the Mcm3 leucine-rich motif (AQRRAQAGA). The ability to generate such a strain indicates the Mcm3 leucine-rich motif does not perform an essential function. As a control, we generated a *GFP-MCM3* strain, which expresses a wild-type fusion protein. Both experimental and control strains divided at identical rates and contained a similar distribution of cells throughout the cell cycle by both budding indices and flow cytometry (data not shown). However, the mutant GFP-Mcm3-nes could be detected in virtually all uninucleate budded cells, whereas GFP-Mcm3 could only be detected in 40% of small budded cells and 4% of uninucleate large budded cells (Figure 5A). Similar results were observed if GFP was fused to Mcm7 instead of Mcm3 (Figure 5B), suggesting that the Mcm3 NES also promotes the nuclear export of multiple Mcm subunits.

To confirm this role for the Mcm3 NES, we compared the distribution of GFP-Mcm3 with mutant GFP-Mcm3-nes protein in cells synchronously released from an α-factor arrest into a nocodazole arrest (Figure 6A). Both flow cytometry and budding indices confirmed that cell cycle progression was not affected by mutation of the Mcm3 NES. As expected, at the beginning of the time course (Figure 6A, 0 min), when all cells were unbudded G1 cells, the GFP fusion proteins of both experimental and control strains

were strongly nuclear. However, 60 min after release from α-factor, when most cells were small-budded and in late S phase, there was a dramatic difference. In the *GFP-MCM3* strain, few of the small budded cells (5/100 cells) retained detectable nuclear levels of the fusion protein, whereas in the *GFP-mcm3-nes* strain, residual nuclear accumulation of the fusion protein could be detected in almost all small budded cells (122/123). Similar, but less striking differences were observed at 50 and 70 min after release from α-factor arrest (data not shown). These data indicate that Mcm3 NES promotes the nuclear export of Mcm3. Eventually, as cells remained at the nocodazole arrest and became large-budded (Figure 6A, 100 min), nuclear accumulation of the GFP fusion proteins gradually became undetectable (0/71 for GFP-Mcm3 and 1/54 for GFP-Mcm3-nes), indicating that the NES is not absolutely required for the net nuclear export of Mcm3. Nonetheless, the NES is essential for the timely export of Mcm3, and without this timely export, Mcm3 is not effectively cleared from the nuclei of cycling cells (Figure 5A). A very similar delay in nuclear export was observed in a *GFP-Mcm3 crm1-T539C* leptomycin B sensitive strain, if leptomycin B was added upon release from alpha factor arrest (data not shown). These results suggest that the Mcm3 NES functions through Crm1 in the full Mcm2-7 complex as it does in the GFP_3 fusion protein.

Nuclear export of Mcm7-GFP is also delayed in a *mcm3-nes* strain relative to an *MCM3* strain (Figure 6B). 50 min after release from α-factor arrest, 42% (56/134) of the small budded *MCM3* cells retained barely detectable nuclear accumulation of Mcm7-GFP, whereas almost 90% (94/110) of the small budded *mcm3-nes* cells retained residual nuclear accumulation of Mcm7-GFP. Similar but smaller differences could be seen at 40

and 60 min after release from α-factor (data not shown). By 70 min, however, most cells were large-budded, and almost no wild-type (0/77) or mutant (3/172) large budded cells exhibited detectable nuclear accumulation. Again, the *MCM7-GFP crm1-T539C* strain showed a similar delay in Mcm7 nuclear export (data not shown). Thus, although there appears to be another partially redundant export signal(s) for the Mcm2-7 complex, we conclude that the Mcm3 NES functions as a Crm1-dependent export signal for at least two subunits of the complex. Also, because the export defect in the *crm1-T539C* mutant phenocopies the export defect of the *mcm3-nes* mutant, it appears that the partially redundant export signal(s) may function through a different export receptor besides Crm1.

Mcm3 is a substrate of Cdc28 kinase

We and others have previously shown that Cdc28 kinase activity promotes the net nuclear export of Mcm proteins (Labib *et al.*, 1999; Nguyen *et al.*, 2000), raising the possibility that this regulation is through direct phosphorylation of the Mcm2-7 complex. A scan of the amino acid sequences of all six Mcm proteins for the full consensus CDK phosphorylation site (S/T)-P-X-(K/R) (Nigg, 1993) only identified two sites on Mcm4 and five sites on Mcm3. The sites on Mcm3 were of particular interest because they are all located within the Mcm3 portion of the NLS-NES transport module (Figure 1A). Four sites flank the basic region of the Mcm3 NLS and the fifth site is adjacent to the leucine-rich motif of the Mcm3 NES. Two additional sites that satisfy a more degenerate CDK phosphorylation site consensus ((S/T)-P) are positioned between the basic region

and leucine-rich motif. For the experiments discussed below, we generated mutations in these sites that substitute alanine for the phosphoacceptor serine or threonine of these consensus sites.

Figure 7A shows that recombinant Clb2-Cdc28 kinase can phosphorylate purified GST-Mcm3 *in vitro*, confirming previous reports of Mcm3 phosphorylation by purified Clb2-Cdc28 and Clb5-Cdc28 kinases (Ubersax *et al*., 2003; Loog and Morgan, 2005). Importantly, GST-Mcm3-cdk5A and GST-Mcm3-cdk7A, which contain mutations in the five full CDK consensus sites and all seven potential CDK sites, respectively (Figure 1A), were both poorly phosphorylated. We also examined Mcm3 phosphorylation *in vivo* by metabolically labeling cells with 32P orthophosphate and observed that Mcm3 displayed significantly more phosphorylation than, Mcm3-cdk5A, and Mcm3-cdk7A (Figure 7B). Together these results suggest that the Mcm3 portion of the NLS-NES transport module is a target of Cdc28 kinase *in vitro* and *in vivo*.

<u>The Mcm3 CDK consensus phosphorylation sites regulate the transport activity of the NLS-NES module.</u>

We next asked whether Cdc28 phosphorylation of the Mcm3 NLS3NES3 segment regulates the activity of the NLS-NES transport module. We examined the effect of mutating the Mcm3 consensus CDK phosphorylation sites on the localization of the GFP_3 fusion protein containing the transport module. We introduced alanine substitutions in all five full consensus CDK sites or just the four sites flanking the Mcm3 NLS (Figure 1B).

Representative cells from exponentially growing cultures expressing the wild-type or mutant fusion proteins are shown in Figure 8A. As described earlier, the wild-type fusion protein displayed a range of subcellular distributions from nuclear to cytoplasmic depending on the cell cycle position of individual cells. In contrast, both mutant proteins were constitutively nuclear. These results show that Cdc28 promotes the net nuclear export of the GFP_3 fusion protein through phosphorylation of the NLS-NES transport module. They suggest that phosphorylation acts as a switch that flips the activity of the NLS-NES transport module from directing net nuclear import to directing net nuclear export.

Phosphorylation of the NLS-NES transport module promotes the net nuclear export of the Mcm2-7 complex

To determine whether Cdc28 regulation of the NLS-NES transport module contributes to the cell cycle regulated export of the entire Mcm2-7 complex, we investigated the effect of mutating the CDK consensus sites of the transport module in the endogenous *MCM3* gene. We first examined three strains where the wild-type *MCM3* gene was replaced by *GFP-mcm3-cdk4A*, *GFP-mcm3-cdk5A*, or *GFP-mcm3-cdk7A* (Figure 1A). At a metaphase arrest imposed by nocodazole, all three strains displayed partial nuclear retention of their mutant GFP-Mcm3 (Figure 8B), in contrast to wild-type GFP-Mcm3, which showed no such retention (Figure 8C, GFP-Mcm3 and (Nguyen *et al.*, 2000)). This inability to fully export the mutant GFP-Mcm3 proteins was also observed in exponentially growing cells. Figure 9A (first two panels and accompanying bar

graphs) shows a quantitative analysis of the nuclear localization of GFP-Mcm3-cdk4A and GFP-Mcm3 in unbudded, small budded, and uninucleate large budded cells. GFP-Mcm3-cdk4A persists in the nuclei of small budded and uninucleate large budded cells while GFP-Mcm3 is disappearing. These observations show that the CDK consensus sites in the Mcm3 NLS-NES module are required for the efficient cytoplasmic localization of Mcm3.

To examine the effect of the CDK consensus site mutations on the net nuclear export of other Mcm proteins, *MCM3* was replaced by *mcm3-cdk5A* in *MCM2-GFP*, *MCM4-GFP*, and *MCM7-GFP* strains. At a nocodazole arrest, partial nuclear retention of the GFP fusion protein was observed in all three *mcm3-cdk5A* strains, in contrast to the full cytoplasmic distribution observed in the congenic *MCM3* strains (Figure 8C). Similarly, in exponentially growing cells, the Mcm-GFP fusion proteins in these strains were never fully cleared from the nucleus. Partial nuclear retention of these three Mcm-GFP proteins was also observed in both nocodazole arrested and exponentially growing cells when *mcm3-cdk5A* was replaced by *mcm3-cdk4A* strains (Figure 9B and data not shown). These results suggest that the CDK consensus sites in the Mcm3 portion of the NLS-NES transport module are required for efficient nuclear export of each Mcm protein. We conclude that Cdc28 phosphorylation of NLS-NES module promotes the net nuclear export of the Mcm2-7 complex.

Because some nuclear export of Mcm proteins was still observed in *mcm3-cdk4A*, *mcm3-cdk5A*, and *mcm3-cdk7A* strains, it appears that phosphorylation of the Mcm3

NLS-NES module is not the sole mechanism by which Cdc28 promotes the net nuclear export of Mcm2-7. Whatever the additional mechanism, it presumably requires one or more NES(s) in the Mcm2-7 complex. To determine whether the Mcm3 NES contributes to this residual export we replaced the *MCM3* ORF with *mcm3-cdk4A-nes* in *GFP-MCM3* and *MCM7-GFP* strains. The mutant Mcm3 expressed in these strains contains alanine substitutions in both the four CDK consensus sites flanking the NLS and the leucine-rich repeat of the NES (Figure 1A). The GFP-Mcm fusion proteins in these strains were strongly nuclear throughout the cell cycle, indicating that the combination of CDK consensus site and NES mutations in Mcm3 could completely abrogate the net nuclear export of Mcm3 and Mcm7 (Figure 9A and 9B, last panels and graph). These results provide further evidence of the importance of the NLS-NES transport module in the regulation of Mcm protein localization.

Phosphomimic mutation of Mcm3 promotes the net nuclear export of Mcm proteins and impairs replication initiation

We have shown that phosphoryation of the CDK consensus sites in the NLS-NES transport module is necessary for the efficient net nuclear export of the Mcm2-7 complex after G1 phase. To examine whether phosphorylation of these CDK consensus sites is sufficient to promote this export during G1 phase, we mutated the phosphoacceptor residues of all five full consensus sites to aspartic acid or glutamic acid to mimic constitutive phosphorylation of these sites (Figure 1A). This mutant *mcm3* allele, *mcm3-*

cdk5ED, was substituted for the wild type endogenous *MCM3* gene by two–step gene replacement in *MCM2-GFP*, *GFP-MCM3*, *MCM4-GFP*, or *MCM7-GFP* strains. As described earlier (Figure 6A, and (Labib *et al*., 1999; Nguyen *et al*., 2000)), these GFP fusion proteins normally concentrate in the nucleus during G1 phase. In the *mcm3-cdk5ED* mutant background, in contrast, much of these proteins were redistributed to the cytoplasm at a G1 phase arrest (Figure 10A). Some residual nuclear localization was still observed, although this is not surprising given that (1) phosphomimic mutations only partially resemble phosphorylated residues, (2) CDK phosphorylation of sites beyond the CDK consensus sites on Mcm3 may be necessary for full exclusion of the Mcm2-7 complex, and (3) the loading of Mcm2-7 onto chromatin makes these proteins refractory to cytoplasmic redistribution. Despite these limitations, the phosphomimic mutations on Mcm3 were sufficient to promote significant, albeit incomplete, export of the Mcm2-7 complex in G1 phase.

Such incomplete export may account for the ability to isolate *mcm3-cdk5ED* strains, as complete exclusion of Mcm2-7 from the nucleus during G1 phase would presumably be lethal. Nonetheless, the phosphomimic mutations clearly compromised the cell cycles of these cells. While exponentially growing liquid cultures of wild-type *GFP-MCM3* control strains doubled every 90 min, *GFP-mcm3-cdk5ED* mutant strains doubled every 140-150 min. Analysis of microcolonies derived from individually plated *GFP-mcm3-cdk5ED* cells not only confirmed that a significant percent of these cells divide much slower than wild-type *GFP-MCM3* cells, but also established that 10-12% of the mutant cells generated fully arrested microcolonies after less than 3-4 divisions

(Supplementary Figure 2A). Hence, although a mutant cell lineage could be propagated, considerable inviability was experienced every generation. Flow cytometry and budding indices of exponentially growing *GFP-mcm3-cdk5ED* cells showed that they spend a greater proportion of their cell cycle in G2/M phase relative to *GFP-MCM3* cells (Supplementary Figure 2B). The simplest interpretation of these observations is that poor nuclear accumulation of Mcm2-7 during G1 phase compromises the initiation of DNA replication, making it difficult to complete a full S phase in a timely manner and triggering a checkpoint delay or arrest at G2/M phase. Many known replication initiation mutants, including the originally isolated *mcm* mutants, display a similar accumulation of G2/M cells (Gibson *et al*., 1990; Hennessy *et al*., 1991; Foiani *et al*., 1994; Merchant *et al*., 1997; Jacobson *et al*., 2001).

To determine whether replication initiation is indeed disrupted in *GFP-mcm3-cdk5ED* mutant, we examined the rate of plasmid loss during multiple generations of nonselective growth (Figure 10B). Failure to initiate DNA replication on a plasmid will enhance its intrinsic loss rate. In control *MCM3* and *GFP-MCM3* strains, the plasmid YCp50 was lost at a rate of 3.0% and 4.4% per generation, respectively. In the *GFP-mcm3-cdk5ED* strain this rate was increased to 18.5%. Increased plasmid loss rates due to defective initiation can often be suppressed by increasing the number of origins on the plasmid. The plasmid pJW1112, which contains an additional seven tandem copies of the *H4ARS* origin inserted into Ycp50, specifically reduced the plasmid loss rate in the *GFP-mcm3-cdk5ED* strain (data not shown), indicating that the elevated plasmid loss seen in the *GFP-mcm3-cdk5ED* strain is due to defective replication initiation.

This defect in initiation could be due to mislocalization of the Mcm2-7 complex or to localization-independent effects of these mutations on the initiation function of the complex. To distinguish between these possibilities, we fused the *mcm3-cdk5ED* to two tandem copies of the SV40 NLS (*SVNLS$_2$-mcm3-cdk5ED*). This NLS fusion restored strong nuclear localization of Mcm2-GFP, GFP-Mcm3, Mcm4-GFP, and Mcm7-GFP in G1 phase (Figure 10A) and restored liquid culture doubling times and microcolony expansion to wild-type levels (data not shown). In addition, this fusion restored plasmid loss rates in the *SVNLS$_2$-GFP-mcm3-cdk5ED* strain to wild-type levels (Figure 10B). We conclude that the replication initiation defect observed in the *GFP-mcm3-cdk5ED* is primarily due to mislocalization of Mcm proteins in G1 phase. Considering the incomplete extent of this mislocalization arising from the phosphomimic mutations, these results suggest that bona fide ectopic phosphorylation of the Mcm3 portion of the NLS-NES module in G1 phase would severely impair replication initiation in S phase and that the normal CDK regulation of Mcm2-7 localization contributes significantly to restraining reinitiation of DNA replication during and after S phase.

DISCUSSION

The Mcm2-7 complex is localized by a transport module distributed over more than one subunit

In our effort to understand CDK regulation of Mcm2-7 localization, we have mapped transport signals that promote the nuclear import and nuclear export of this complex. We show that three transport signals distributed across two subunits effectively cooperate to form a regulatory transport module that plays a key role in controlling the localization of the Mcm2-7 complex. Importantly, the isolated transport module fused to a heterologous protein is sufficient to recapitulate the cell cycle regulated localization of the entire complex. Two of the transport signals in this module are weak NLSs on Mcm2 and Mcm3. Individually, each is required but not sufficient for robust nuclear accumulation of the Mcm2-7 complex and the heterologous protein. The third signal, positioned next to the Mcm3 NLS, is a Crm1 dependent NES containing a leucine-rich motif.

Previous analyses of nuclear localization of protein complexes have identified single subunits that are responsible for nuclear localization of an entire complex (Maridor *et al*., 1993; Pereira *et al*., 1998; Leslie *et al*., 2004; Subramaniam and Johnson, 2004; Wendler *et al*., 2004). To our knowledge, our analysis of Mcm2-7 nuclear localization in budding yeast provides the first documentation of a complex whose nuclear localization requires transport signals on more than one subunit. Distributing multiple required transport signals over distinct subunits provides one way to make complex formation a prerequisite for nuclear accumulation. Imposing such a prerequisite could be particularly important in preventing individual subunits or subcomplexes with unrestrained activity from accumulating in the nucleus, where they might threaten the integrity of the genome. The Mcm4-6-7 subcomplex, for example, exhibits helicase activity *in vitro* that is

suppressed by other Mcm subunits in the full complex (Ishimi *et al*., 1998; Lee and Hurwitz, 2001). Preventing such an active subcomplex from accumulating in the nucleus unless it first associates with Mcm2 and Mcm3 may safeguard against inappropriate or uncontrolled unwinding of genomic DNA. Thus, using transport modules distributed over multiple subunits of a multiprotein complex could increase the specificity with which these subunits are localized by coupling their accumulation in either nucleus or cytoplasm to incorporation into a complex.

The use of weak transport signals distributed over more than one polypeptide potentially challenges the identification of transport signals directing the localization of multiprotein complexes. Although the weak NLSs identified in this work were somewhat evident based on their basic sequence composition, other weak or partial transport signals (for either import or export) may not be readily identified due to lack of sequence homology to known canonical signals. This raises the possibility that unrecognized distributed localization signals may program the localization of other multiprotein complexes. We also note that in addition to the transport module that we have identified on the Mcm2-7 complex, other NLSs or NESs may contribute to the localization of the complex; indeed our data suggests that other NESs that do not function through the Crm1 export receptor may work in parallel with the Mcm3 NES.

Future analysis of the mechanism by which the Mcm2 and Mcm3 NLS cooperate to promote nuclear localization of the Mcm2-7 complex should provide insight into how large multiprotein complexes are transported across the nuclear pore. Particularly

interesting is the question of whether the NLSs act as a single bipartite NLS that strongly binds a single nuclear transport receptor, or whether they independently and weakly bind separate nuclear transport receptors. In the latter case, two transport receptors might still transport the entire complex as one unit across the nuclear pore, or they might transport separate subcomplexes that can only be retained and accumulate in the nucleus through mutual interaction. Whatever the precise mechanism, the end result, as suggested by this study, is to couple complex formation to nuclear accumulation.

Complex formation can also be coupled to nucleocytoplasmic localization if transport signals are masked or unmasked by protein interactions or conformational changes associated with complex formation (Reviewed in Kaffman and O'Shea, 1999; Jans *et al*., 2000). Such a mechanism has been proposed for nuclear localization of Mcm2-7 in both *S. pombe* (Pasion and Forsburg, 1999) and *S. cerevisiae* (Labib *et al*., 2001) based on the interdependence among Mcm proteins for maintaining their nuclear localization. When the complex is nuclear (throughout the cell cycle in *S. pombe* and during G1 phase in *S. cerevisiae*), conditionally disrupting one Mcm subunit results in nuclear export of the remaining Mcm subunits. It has been suggested that complex formation may expose inaccessible NLSs or mask competing NESs on Mcm subunits such that nuclear import is favored over nuclear export (Pasion and Forsburg, 1999). Although direct evidence for this model is lacking, we cannot rule out the possibility that such a mechanism works in parallel with the distributed transport module reported here.

The NLS-NES transport module is a regulatory target of CDKs

In addition to providing important transport signals for localization of the Mcm2-7 complex, we have shown that the distributed NLS-NES transport module provides a key target for CDK regulation of this localization. The unphosphorylated module promotes net nuclear import and the phosphorylated module promotes net nuclear export. Future experiments will be needed to address the precise mechanism by which CDK phosphorylation alters the activity of this module. In other examples of CDK regulated protein localization, CDK phosphorylation has been implicated in modulating NLS and/or NES activity (Moll *et al*., 1991; Sidorova *et al*., 1995; Deng *et al*., 2004; Geymonat *et al*., 2004; Harreman *et al*., 2004). A classic and well defined example is provided by the related CDK, Pho80-Pho85, whose phosphorylation of the Pho4 NES and NLS promotes nuclear export and inhibits nuclear import, respectively (Komeili and O'Shea, 1999). Based on these precedents, phosphorylation of the Mcm3 transport module could down regulate its NLS activity, upregulate its NES activity, or both. Furthermore, in the context of the full Mcm2-7 complex, phosphorylation of the Mcm3 transport module could induce conformational changes that alter the activity or accessibility of other transport signals in the complex.

Some residual nuclear export of Mcm subunits persists following alanine substitutions in the Mcm3 CDK consensus sites, suggesting that CDKs may target more than the Mcm3 transport module to promote nuclear export. The only other full CDKconsensus phosphorylation sites ((S/T)-X-P-(K/R)) in the Mcm2-7 complex are at the N-terminus of Mcm4. However, alanine substitution at these sites alone or in

combination with mutation of the Mcm3 CDK sites did not have any striking effect on Mcm2-7 localization (data not shown). Hence, further work will be necessary to refine our understanding of Cdc28 regulation of Mcm2-7 localization.

Nuclear export of Mcm2-7 provides a significant contribution to replication control

The cell cycle regulated nuclear export of replication initiation proteins provides an appealing mechanism for preventing rereplication within a single cell cycle. Regulated localization of replication proteins has also been observed in metazoans, where Mcm proteins are constitutively nuclear. In both mammalian and Xenopus cells, ectopically expressed Cdc6 is exported from the nucleus in a CDK dependent manner (Saha *et al*., 1998; Jiang *et al*., 1999; Petersen *et al*., 1999; Pelizon *et al*., 2000; Delmolino *et al*., 2001). Although interpretation of these studies is complicated by evidence that a portion of the endogenous Cdc6 population remains bound to nuclear structures throughout the cell cycle (Mendez and Stillman, 2000; Alexandrow and Hamlin, 2004), exporting the unbound population may still be important in the block to rereplication. Demonstrating such a role, however, has been difficult, in part because multiple layers of rereplication control make most individual mechanisms dispensable. In budding yeast, simultaneous disruption of multiple mechanisms has established a role for regulated Mcm localization in the block to rereplication (Nguyen *et al*., 2001; Mimura *et al*., 2004; Wilmes *et al*., 2004), but the limited rereplication that is induced makes it difficult to assess the relative significance of each individual mechanism.

In this study, we show that phosphomimic mutations of the CDK consensus sites at the Mcm3 NLS-NES segment promote the premature net export of Mcm2-7 from G1 nuclei. Even though this net export does not result in complete nuclear exclusion of the complex, it impairs the initiation of DNA replication and delays cell division. Presumably, the more effective nuclear exclusion mediated by *bona fide* CDK phosphorylation is sufficient to block reinitiation of DNA replication from most, if not all, origins. Thus, our data provides the first evidence that CDK regulation of the localization of a replication initiation factor may play a significant role in the inhibition of rereplication in eukaryotes. Analogous regulation of other replication initiation proteins in other eukaryotes may have a similarly important impact on the block to rereplication.

ACKNOWLEDGEMENTS

We thank Anita Sil, David Toczyski, David Morgan, Hiten Madhani, Erin O'Shea, and members of the Li laboratory for critical reading of the manuscript. We thank Minoru Yoshida for the generous gift of Leptomycin B and Karsten Weis for the crm1 strains and plasmids. This work was supported by grants to J. Li from the ACS (RPG-99-169-01-CCG), NIH (RO1 GM59704), and Brook Byers Award. MEL was support by a Ford Foundation Predoctoral Fellowship and NIH predoctoral fellow (1 F31 CA110268-01). VQN was funded by an NIH Training Grant (T32 CA09270) and a UCSF Chancellor's Fellowship. AWR was funded by NIGMS grant (1 R25 GM56847).

REFERENCES

Alexandrow, M.G., and Hamlin, J.L. (2004). Cdc6 chromatin affinity is unaffected by serine-54 phosphorylation, S-phase progression, and overexpression of cyclin A. Mol Cell Biol *24*, 1614-1627.

Archambault, V., Ikui, A.E., Drapkin, B.J., and Cross, F.R. (2005). Disruption of mechanisms that prevent rereplication triggers a DNA damage response. Mol Cell Biol *25*, 6707-6721.

Bell, S.P., and Dutta, A. (2002). DNA replication in eukaryotic cells. Annu Rev Biochem *71*, 333-374.

Delmolino, L.M., Saha, P., and Dutta, A. (2001). Multiple mechanisms regulate subcellular localization of human CDC6. J Biol Chem *276*, 26947-26954.

Deng, W., Lin, B.Y., Jin, G., Wheeler, C.G., Ma, T., Harper, J.W., Broker, T.R., and Chow, L.T. (2004). Cyclin/CDK regulates the nucleocytoplasmic localization of the human papillomavirus E1 DNA helicase. J Virol *78*, 13954-13965.

Detweiler, C.S., and Li, J.J. (1998). Ectopic induction of Clb2 in early G1 phase is sufficient to block prereplicative complex formation in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *95*, 2384-2389.

Diffley, J.F. (2004). Regulation of early events in chromosome replication. Curr Biol *14*, R778-786.

Drury, L.S., Perkins, G., and Diffley, J.F. (1997). The Cdc4/34/53 pathway targets Cdc6p for proteolysis in budding yeast. Embo J *16*, 5966-5976.

Fischer, U., Huber, J., Boelens, W.C., Mattaj, I.W., and Luhrmann, R. (1995). The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. Cell *82*, 475-483.

Foiani, M., Marini, F., Gamba, D., Lucchini, G., and Plevani, P. (1994). The B subunit of the DNA polymerase alpha-primase complex in Saccharomyces cerevisiae executes an essential function at the initial stage of DNA replication. Mol Cell Biol *14*, 923-933.

Fornerod, M., Ohno, M., Yoshida, M., and Mattaj, I.W. (1997). CRM1 is an export receptor for leucine-rich nuclear export signals. Cell *90*, 1051-1060.

Fukuda, M., Asano, S., Nakamura, T., Adachi, M., Yoshida, M., Yanagida, M., and Nishida, E. (1997). CRM1 is responsible for intracellular transport mediated by the nuclear export signal. Nature *390*, 308-311.

Geymonat, M., Spanos, A., Wells, G.P., Smerdon, S.J., and Sedgwick, S.G. (2004). Clb6/Cdc28 and Cdc14 regulate phosphorylation status and cellular localization of Swi6. Mol Cell Biol *24*, 2277-2285.

Gibson, S.I., Surosky, R.T., and Tye, B.K. (1990). The phenotype of the minichromosome maintenance mutant mcm3 is characteristic of mutants defective in DNA replication. Mol Cell Biol *10*, 5707-5720.

Green, B.M., and Li, J.J. (2005). Loss of Rereplication Control in Saccharomyces cerevisiae Results in Extensive DNA Damage. Mol Biol Cell *16*, 421-432.

Harreman, M.T., Kline, T.M., Milford, H.G., Harben, M.B., Hodel, A.E., and Corbett, A.H. (2004). Regulation of nuclear import by phosphorylation adjacent to nuclear localization signals. J Biol Chem *279*, 20613-20621.

Hennessy, K.M., Lee, A., Chen, E., and Botstein, D. (1991). A group of interacting yeast DNA replication genes. Genes and Development *5*, 958-969.

Ishimi, Y., Komamura, Y., You, Z., and Kimura, H. (1998). Biochemical function of mouse minichromosome maintenance 2 protein. J Biol Chem *273*, 8369-8375.

Jacobson, M.D., Munoz, C.X., Knox, K.S., Williams, B.E., Lu, L.L., Cross, F.R., and Vallen, E.A. (2001). Mutations in SID2, a novel gene in Saccharomyces cerevisiae, cause synthetic lethality with sic1 deletion and may cause a defect during S phase. Genetics *159*, 17-33.

Jans, D.A., Xiao, C.Y., and Lam, M.H. (2000). Nuclear targeting signal recognition: a key control point in nuclear transport? Bioessays *22*, 532-544.

Jiang, W., Wells, N.J., and Hunter, T. (1999). Multistep regulation of DNA replication by Cdk phosphorylation of HsCdc6. Proc Natl Acad Sci U S A *96*, 6193-6198.

Kaffman, A., and O'Shea, E.K. (1999). Regulation of nuclear localization: a key to a door. Annu Rev Cell Dev Biol *15*, 291-339.

Komeili, A., and O'Shea, E.K. (1999). Roles of phosphorylation sites in regulating activity of the transcription factor Pho4. Science *284*, 977-980.

Kudo, N., Matsumori, N., Taoka, H., Fujiwara, D., Schreiner, E.P., Wolff, B., Yoshida, M., and Horinouchi, S. (1999). Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region. Proc Natl Acad Sci U S A *96*, 9112-9117.

Labib, K., Diffley, J.F.X., and Kearsey, S.E. (1999). G1-phase and B-type cyclins exclude the DNA-replication factor Mcm4 from the nucleus. Nature Cell Biol. *1*, 415-422.

Labib, K., Kearsey, S.E., and Diffley, J.F. (2001). MCM2-7 proteins are essential components of prereplicative complexes that accumulate cooperatively in the nucleus during G1-phase and are required to establish, but not maintain, the S-phase checkpoint. Mol Biol Cell *12*, 3658-3667.

Labib, K., Tercero, J.A., and Diffley, J.F. (2000). Uninterrupted MCM2-7 function required for DNA replication fork progression. Science *288*, 1643-1647.

Lee, J.K., and Hurwitz, J. (2001). Processive DNA helicase activity of the minichromosome maintenance proteins 4, 6, and 7 complex requires forked DNA structures. Proc Natl Acad Sci U S A *98*, 54-59.

Leslie, D.M., Zhang, W., Timney, B.L., Chait, B.T., Rout, M.P., Wozniak, R.W., and Aitchison, J.D. (2004). Characterization of karyopherin cargoes reveals unique mechanisms of Kap121p-mediated nuclear import. Mol Cell Biol *24*, 8487-8503.

Liu, E., Li, X., Yan, F., Zhao, Q., and Wu, X. (2004). Cyclin-dependent kinases phosphorylate human Cdt1 and induce its degradation. J Biol Chem *279*, 17283-17288.

Loog, M., and Morgan, D.O. (2005). Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. Nature in press.

Maridor, G., Gallant, P., Golsteyn, R., and Nigg, E.A. (1993). Nuclear localization of vertebrate cyclin A correlates with its ability to form complexes with cdk catalytic subunits. J Cell Sci *106 (Pt 2)*, 535-544.

Melixetian, M., Ballabeni, A., Masiero, L., Gasparini, P., Zamponi, R., Bartek, J., Lukas, J., and Helin, K. (2004). Loss of Geminin induces rereplication in the presence of functional p53. J Cell Biol *165*, 473-482.

Mendez, J., and Stillman, B. (2000). Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. Mol Cell Biol *20*, 8602-8612.

Merchant, A.M., Kawasaki, Y., Chen, Y., Lei, M., and Tye, B.K. (1997). A lesion in the DNA replication initiation factor Mcm10 induces pausing of elongation forks through chromosomal replication origins in Saccharomyces cerevisiae. Mol Cell Biol *17*, 3261-3271.

Mihaylov, I.S., Kondo, T., Jones, L., Ryzhikov, S., Tanaka, J., Zheng, J., Higa, L.A., Minamino, N., Cooley, L., and Zhang, H. (2002). Control of DNA replication and chromosome ploidy by geminin and cyclin A. Mol Cell Biol *22*, 1868-1880.

Mimura, S., Seki, T., Tanaka, S., and Diffley, J.F. (2004). Phosphorylation-dependent binding of mitotic cyclins to Cdc6 contributes to DNA replication control. Nature *431*, 1118-1123.

Moll, T., Tebb, G., Surana, U., Robitsch, H., and Nasmyth, K. (1991). The role of phosphorylation and the CDC28 protein kinase in cell cycle- regulated nuclear import of the S. cerevisiae transcription factor SWI5. Cell *66*, 743-758.

Neville, M., and Rosbash, M. (1999). The NES-Crm1p export pathway is not a major mRNA export route in Saccharomyces cerevisiae. Embo J *18*, 3746-3756.

Nguyen, V.Q., Co, C., Irie, K., and Li, J.J. (2000). Clb/Cdc28 kinases promote nuclear export of the replication initiator proteins Mcm2-7. Curr Biol *10*, 195-205.

Nguyen, V.Q., Co, C., and Li, J.J. (2001). Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. Nature *411*, 1068-1073.

Nigg, E.A. (1993). Cellular substrates of p34(cdc2) and its companion cyclin-dependent kinases. Trends Cell Biol *3*, 296-301.

Nishi, K., Yoshida, M., Fujiwara, D., Nishikawa, M., Horinouchi, S., and Beppu, T. (1994). Leptomycin B targets a regulatory cascade of crm1, a fission yeast nuclear protein, involved in control of higher order chromosome structure and gene expression. J Biol Chem *269*, 6320-6324.

Ossareh-Nazari, B., Bachelerie, F., and Dargemont, C. (1997). Evidence for a role of CRM1 in signal-mediated nuclear protein export. Science *278*, 141-144.

Pasion, S.G., and Forsburg, S.L. (1999). Nuclear localization of Schizosaccharomyces pombe Mcm2/Cdc19p requires MCM complex assembly. Mol Biol Cell *10*, 4043-4057.

Pelizon, C., Madine, M.A., Romanowski, P., and Laskey, R.A. (2000). Unphosphorylatable mutants of Cdc6 disrupt its nuclear export but still support DNA replication once per cell cycle. Genes Dev *14*, 2526-2533.

Pereira, G., Knop, M., and Schiebel, E. (1998). Spc98p directs the yeast gamma-tubulin complex into the nucleus and is subject to cell cycle-dependent phosphorylation on the nuclear side of the spindle pole body. Mol Biol Cell *9*, 775-793.

Petersen, B.O., Lukas, J., Sorensen, C.S., Bartek, J., and Helin, K. (1999). Phosphorylation of mammalian CDC6 by cyclin A/CDK2 regulates its subcellular localization. Embo J *18*, 396-410.

Piatti, S., Lengauer, C., and Nasmyth, K. (1995). Cdc6 is an unstable protein whose de novo synthesis in G1 is important for the onset of S phase and for preventing a

'reductional' anaphase in the budding yeast Saccharomyces cerevisiae. Embo J *14*, 3788-3799.

Saha, P., Chen, J., Thome, K.C., Lawlis, S.J., Hou, Z.H., Hendricks, M., Parvin, J.D., and Dutta, A. (1998). Human CDC6/Cdc18 associates with Orc1 and cyclin-cdk and is selectively eliminated from the nucleus at the onset of S phase. Mol Cell Biol *18*, 2758-2767.

Sidorova, J.M., Mikesell, G.E., and Breeden, L.L. (1995). Cell cycle-regulated phosphorylation of Swi6 controls its nuclear localization. Mol Biol Cell *6*, 1641-1658.

Stade, K., Ford, C.S., Guthrie, C., and Weis, K. (1997). Exportin 1 (Crm1p) is an essential nuclear export factor. Cell *90*, 1041-1050.

Subramaniam, P.S., and Johnson, H.M. (2004). The IFNAR1 subunit of the type I IFN receptor complex contains a functional nuclear localization sequence. FEBS Lett *578*, 207-210.

Sugimoto, N., Tatsumi, Y., Tsurumi, T., Matsukage, A., Kiyono, T., Nishitani, H., and Fujita, M. (2004). Cdt1 phosphorylation by cyclin A-dependent kinases negatively regulates its function without affecting geminin binding. J Biol Chem *279*, 19691-19697.

Thomer, M., May, N.R., Aggarwal, B.D., Kwok, G., and Calvi, B.R. (2004). Drosophila double-parked is sufficient to induce re-replication during development and is regulated by cyclin E/CDK2. Development *131*, 4807-4818.

Ubersax, J.A., Woodbury, E.L., Quang, P.N., Paraz, M., Blethrow, J.D., Shah, K., Shokat, K.M., and Morgan, D.O. (2003). Targets of the cyclin-dependent kinase Cdk1. Nature *425*, 859-864.

Weis, K. (2003). Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. Cell *112*, 441-451.

Wen, W., Meinkoth, J.L., Tsien, R.Y., and Taylor, S.S. (1995). Identification of a signal for rapid export of proteins from the nucleus. Cell *82*, 463-473.

Wendler, P., Lehmann, A., Janek, K., Baumgart, S., and Enenkel, C. (2004). The bipartite nuclear localization sequence of Rpn2 is required for nuclear import of proteasomal base complexes via karyopherin alphabeta and proteasome functions. J Biol Chem *279*, 37751-37762.

Wilmes, G.M., Archambault, V., Austin, R.J., Jacobson, M.D., Bell, S.P., and Cross, F.R. (2004). Interaction of the S-phase cyclin Clb5 with an "RXL" docking sequence in the initiator protein Orc6 provides an origin-localized replication control switch. Genes Dev *18*, 981-991.

Young, M.R., Suzuki, K., Yan, H., Gibson, S., and Tye, B.K. (1997). Nuclear accumulation of Saccharomyces cerevisiae Mcm3 is dependent on its nuclear localization sequence. Genes Cells *2*, 631-643.

Zhu, W., Chen, Y., and Dutta, A. (2004). Rereplication by depletion of geminin is seen regardless of p53 status and activates a G2/M checkpoint. Mol Cell Biol *24*, 7140-7150.

FIGURE LEGENDS

Figure 1. Wild-type and mutant Mcm2 and Mcm3 nucleocytoplasmic transport signals.
A. Key amino acid sequences of wild-type and mutant Mcm2 NLS and Mcm3 NLS-NES transport module. Putative NLSs and leucines in leucine-rich motif are in bold. Consensus CDK phosphorylation sites are underlined with putative phosphoacceptor residues in gray. Amino acids not spelled out are indicated in parentheses. Amino acid substitutions in the mutant Mcm proteins used in this study are indicated below the wild-type sequences.
B. A schematic of the MCM transport module containing the Mcm2 NLS and the Mcm3 NLS-NES. Below the schematic are the WT and corresponding mutant alleles of this construct that are discussed in the text.

Figure 2. The Mcm2 and Mcm3 NLSs are each required for nuclear localization of the Mcm2-7 complex.
A. The Mcm2 NLS is required for nuclear localization of Mcm2-GFP. YJL3265 (*MCM2::{MCM2-GFP}*), YJL1231 (*MCM2::{mcm2-nls-GFP}*), and YJL1228 (*MCM2::{mcm2-nls-GFP-SVNLS$_2$})* were arrested in G1 phase with α-factor then examined by fluorescence microscopy.
B. The Mcm3 NLS is required for nuclear localization of GFP-Mcm3. YJL2669 (*MCM3::{GFP-MCM3}*), YJL2675 (*MCM3::{GFP-mcm3-nls}*), and YJL2665 (*MCM3::{SVNLS$_2$-GFP-mcm3-nls}*) were arrested in G1 phase with α-factor then examined by fluorescence microscopy.

C. The Mcm2 NLS is required for nuclear localization of Mcm7-GFP. Cultures of YJL3765 (*MCM7-GFP mcm2-td MCM2*), YJL3840 (*MCM7-GFP mcm2-td mcm2-nls*), and YJL3799 (*MCM7-GFP mcm2-td mcm2-nls-SVNLS*$_2$) growing exponentially at 23°C were arrested in G2/M phase by addition of nocodazole for 3 hr. Galactose was added and cultures were shifted to 37°C for 30 min to degrade Mcm2-td. Cells were then released from G2/M phase into a G1 phase block by shifting them to fresh medium containing α-factor for 2 hr (still in the presence of galactose and at 37°C). Cells examined by fluorescence microscopy are shown just before the G2/M phase release (NOC arrest) and at the G1-phase block (α-factor arrest).

D. The Mcm3 NLS is required for nuclear localization of Mcm7-GFP. Cultures of YJL3464 (*MCM7-GFP mcm3-td MCM3*), YJL3469 (*MCM7-GFP mcm3-td mcm3-nls*), and YJL3474 (*MCM7-GFP mcm3-td SVNLS*$_2$*-mcm3-nls*) were subjected to the same experimental protocol described for Figure 2C.

Figure 3. Together the Mcm2 and Mcm3 NLSs are sufficient to direct the nuclear localization of a heterologous protein. Overnight cultures of YJL310 containing *URA3*-marked centromeric plasmids pAR109 (*pGAL-SV40NLS-GFP*$_3$), pAR110 (*pGAL-NLS2-GFP*$_3$), pAR101 (*pGAL-NLS3-GFP*$_3$), pAR113 (*pGAL-NLS2-NLS3-GFP*$_3$), pAR126 (*pGAL-nls2-NLS3-GFP*$_3$), or pAR127 (*pGAL-NLS2-nls3-GFP*$_3$) and growing in SRaf-Ura medium were shifted to YEPRaf medium for 90 min before splitting each culture in two and adding α-factor to one half and nocodazole to the other. One hour later, as the cultures were approaching a complete arrest, galactose was added to induce synthesis of

the GFP_3 fusion proteins. After 2 hr of induction, cells were examined by fluorescence microscopy.

Figure 4. Mcm3 contains an NES, which in combination with the Mcm2 NLS and Mcm3 NLS, directs the cell cycle regulated localization of GFP_3.

A. The cytoplasmic localization of NLS2-NLS3NES3-GFP_3 in G2/M phase is dependent on the leucine-rich motif of the Mcm3 NES. Cultures of YJL4662 (*crm1Δ trp1::{pGAL-NLS2-NLS3NES3-GFP$_3$),TRP1} [crm1-T539C]*) and YJL4860 (*crm1Δ trp1::{pGAL-NLS2-NLS3nes3-GFP$_3$),TRP1} [crm1-T539C]*) constitutively expressing their GFP_3 fusion proteins in YEPGal medium were arrested for 90 min in either G1 phase with α-factor or G2/M phase with nocodazole before being examined by fluorescence microscopy.

B. NLS2-NLS3NES3-GFP_3 is exported from the nucleus by a Crm1 dependent mechanism. YJL4662 constitutively expressing NLS2-NLS3NES3-GFP_3 in YEPGal medium was arrested in G1 phase with α-factor. At time 0, cells were released from the arrest in the presence or absence of 100ng/mL leptomycin B (LMB). Samples were collected at the indicated times for flow cytometry and fluorescence microscopy.

Figure 5. The Mcm3 NES is required for efficient nuclear export of Mcm3 and Mcm7 in cycling cells.

A. Mutation of the Mcm3 NES increases the population of cells containing nuclear GFP-Mcm3 during exponential growth. Exponentially growing cultures of YJL2162 (*GFP-MCM3*) (top) and YJL2741 (*GFP-mcm3-nes*) (bottom) were

examined by fluorescence and DIC microscopy. Representative fluorescent fields are shown. Cells were categorized as unbudded (UB), small budded (SB) or uninucleate large budded (LB) based on their DIC image and DAPI fluorescence. Binucleate large budded cells, which comprise approximately half of all large budded cells, are unseparated post mitotic cells (many in G1 phase) and were thus not included in the analysis. Each of these categories was further subclassified into cells with or without detectable nuclear GFP fluorescence above cytoplasmic levels as exemplified by the pictures of individual cells. Bar graphs show the percent of cells with nuclear or nonnuclear GFP fluorescence and the total number of cells counted (in parentheses) for each bud stage.

B. Mutation of the Mcm3 NES increases the population of cells containing nuclear Mcm7-GFP during exponential growth. The same experiment and analysis described in Figure 5A was performed for YJL1979 (*MCM3 MCM7-GFP*) and YJL5439 (*mcm3-nes MCM7-GFP*).

Figure 6. The Mcm3 NES is required for the timely nuclear export of Mcm3 and Mcm7.

A. Mutation of the Mcm3 NES delays the nuclear export of GFP-Mcm3. YJL2162 (*GFP-MCM3*) and YJL2741 (*GFP-mcm3-nes*) cells were arrested in G1 phase with alpha factor and at time 0 synchronously released into a G2/M phase arrest with nocodazole. At the indicated times, samples were taken for flow cytometry and fluorescence microscopy. (Left) Representative images from 0, 60, and 100 min. (Right) Flow cytometry profiles.

B. Mutation of the Mcm3 NES delays the nuclear export of Mcm7-GFP. YJL1979 (*MCM3 MCM7-GFP*) and YJL5439 (*mcm3-nes MCM7-GFP*) were treated and analyzed as described in C, except microscopic images from 0, 50, and 70 min time point are shown.

Figure 7. *In vitro* and *in vivo* phosphorylation of Mcm3 is dependent on the consensus CDK phosphorylation sites in the Mcm3 NLS3NES3 module.

A. GST-Mcm3 is phosphorylated by Cdc28-Clb2 kinase *in vitro*. *In vitro* kinase reactions were performed with purified Cdc28-Clb2 kinase mixed with purified GST-Mcm3, GST-Mcm3-cdk5A, or GST-Mcm3-cdk7A. Reaction products electrophoresed on SDS-PAGE were subjected to autoradiography (top) and coomassie staining (bottom).

B. The CDK consensus sites in the Mcm3 NLS3NES3 are required for *in vivo* phosphorylation of Mcm3. YJL4110 (*MCM3*), YJL4313 (*Myc_6-MCM3*), YJL4324 (*Myc_6-mcm3-cdk5A*), and YJL4315 (*Myc_6-mcm3-cdk7A)* were metabolically labeled with ^{32}P orthophosphate for 1 hr before lysis and immunoprecipitation with 9E10 anti-Myc monoclonal antibody. Immunoprecipitates were electrophoresed on SDS-PAGE and either subjected to autoradiography (top) or immunoblotted with rabbit anti-Myc polyclonal antibodies (bottom).

Figure 8. The Mcm3 consensus CDK phosphorylation sites are required for nuclear exclusion of the Mcm2-7 complex.

A. Net nuclear export of NLS2-NLS3NES3-GFP_3 is dependent on the consensus CDK phosphorylation sites in NLS3NES3. YJL4662 (*crm1Δ trp1::{pGAL-NLS2-NLS3NES3-*

GFP$_3$),TRP1} [crm1-T539C]), YJL5750 (*trp1::{pGAL-NLS2-NLS3NES3-cdk5A-GFP$_3$),TRP1}*) and YJL5753 (*trp1::{pGAL-NLS2-NLS3NES3-cdk4A-GFP$_3$),TRP1}*) growing exponentially and constitutively expressing their GFP$_3$ fusion proteins in YEPGal medium were examined by fluorescence microscopy.

B. Mcm3 CDK consensus sites are required for nuclear exclusion of Mcm3. Fluorescent microscopy of nocodazole arrested YJL2714 (*GFP-mcm3-cdk4A*), YJL2720 (*GFP-mcm3-cdk5A*), and YJL2314(*GFP-mcm3-cdk7A*).

C . Mcm3 Cdk consensus sites are required for nuclear exclusion of other Mcm subunits. Flouorescence microscopy of nocodazole arrested YJL2033 (*mcm2::{MCM2-GFP}),* YJL4165 (*mcm2::{MCM2-GFP} mcm3-cdk5A*), YJL2160 (*GFP-MCM3*), YJL2720 (*GFP-mcm3-cdk5A*), YJL2037 (*mcm4::{MCM4-GFP}*), YJL4169 (*mcm4::{MCM4-GFP}, mcm3-cdk5A*), YJL2217 (*mcm7::{MCM7-GFP}*), and YJL4167 (*mcm7::{MCM7-GFP}, mcm3-cdk5A*).

Figure 9. Together the Mcm3 NES and the consensus CDK sites in Mcm3 NLS are required for nuclear export of GFP-Mcm3 and Mcm7-GFP.

A. GFP-Mcm3 is strongly nuclear throughout the cell cycle when both the leucine-rich motif of the Mcm3 NES and the four CDK consensus sites flanking the Mcm3 NLS are mutated. Exponentially growing cultures of YJL2162 (*GFP-MCM3*), YJL2714 (*GFP-mcm3-cdk4A*), and YJL5216 (*GFP-mcm3-cdk4A-nes*) were examined by fluorescence and DIC microscopy. Cells were categorized as unbudded (UB), small budded (SB) or uninucleate large budded (LB) based on their DIC image and DAPI fluorescence. Binucleate large budded cells, which comprise approximately half of all large budded

cells, are unseparated post mitotic cells (many in G1 phase) and were thus not included in the analysis. Each bud category was further subclassified into cells with or without detectable nuclear GFP fluorescence above cytoplasmic levels. Bar graphs show the percent of cells with nuclear or nonnuclear GFP fluorescence and the total number of cells counted (in parentheses) for each bud category.

B. Mcm7-GFP is strongly nuclear throughout the cell cycle when both the leucine-rich motif of the Mcm3 NES and the four CDK consensus sites flanking the Mcm3 NLS are mutated. Exponentially growing cultures of YJL1979 (*MCM3 MCM7-GFP*), YJL5691 (*mcm3-cdk4A mcm7::{MCM7-GFP}*), and YJL5221 (*mcm3-cdk4A-nes MCM7-GFP*) were examined by fluorescence and DIC microscopy as described in Figure 9A.

Figure 10. Phosphomimic mutations of the Mcm3 CDK consensus sites promote net nuclear export of Mcm proteins in G1 phase and this mislocalization impairs replication initiation.

A. (First Row) YJL2160 (*GFP-MCM3*), YJL1265 (*GFP-mcm3-cdk5ED*), YJL1260 (*$SVNLS_2$-GFP-mcm3-cdk5ED*) . (Second Row) YJL2033 (*mcm2::{MCM2-GFP}),* YJL4162 (*mcm2::{MCM2-GFP} mcm3-cdk5ED*), YJL4094 (*mcm2::{MCM2-GFP} $SVNLS_2$-mcm3-cdk5ED*). (Third Row) YJL2037 (*mcm4::{MCM4-GFP}*), YJL4103 (*mcm4::{MCM4-GFP} mcm3-cdk5ED*), YJL4098 (*mcm4::{MCM4-GFP} $SVNLS_2$-mcm3-cdk5ED*). (Fourth Row) YJL2217 (mcm7::{*MCM7-GFP}*), YJL4108 (*mcm7::{MCM7-GFP} mcm3-cdk5ED*), YJL4096 (*mcm7::{MCM7-GFP} $SVNLS_2$-mcm3-cdk5ED*). All strains were arrested in G1 phase with α-factor for 90 min (>95% unbudded) before being examined by fluorescence microscopy.

B. **Mcm mislocalization due to the *mcm3-cdk5ED* mutation impairs replication initiation. Plasmid loss rates per were measured over 12-20 generations for plasmids YCp50 (1 ARS) and pJW1112 (8 ARSs) in YJL310 (*MCM3*), YJL2160 (*GFP-MCM3*), YJL1265 (*GFP-mcm3-cdk5ED*), and YJL1259 (*$SVNLS_2$-GFP-mcm3-cdk5ED*). Histogram shows average loss rate per generation and standard error for four independent isolates of each plasmid-yeast pair. Complete failure to replicate a plasmid would result in a theoretical loss rate of 50% per generation.**

Figure 1

A

NLS
1 MSDNRRRRREEDDSDSE 17
MCM2 WT
mcm2-nls AAAAA

Leucine rich motif of NES
NLS
761 848
MCM3 WT SPQKSPKKRQKVRQPASNSGSPIKSTPRR(8aa)TPSS(22aa)SPLPADEEAELQRRLQLGLRVSPRR
mcm3-nls AAA A
mcm3-nes A A A A
mcm3-cdk4A A A A A
mcm3-cdk5A A A A A A
mcm3-cdk7A A A A A A A A
mcm3-cdk4A-nes A A A A A A A A
mcm3-cdk5ED D D D E D

B

NLS2 NLS3 NES3 GFP_3

NLS2-NLS3NES3 MSDNRRRRREE(23aa)SPQKSPKKRQKVRQPASNSGSPIKSTPRR(41aa)EAELQRRLQLGLRVSPRR(8aa)
NLS2-NLS3nes3 A A A A
NLS2-NLS3NES3-cdk4A A A A A
NLS2-NLS3NES3-cdk5A A A A A A

Figure 2

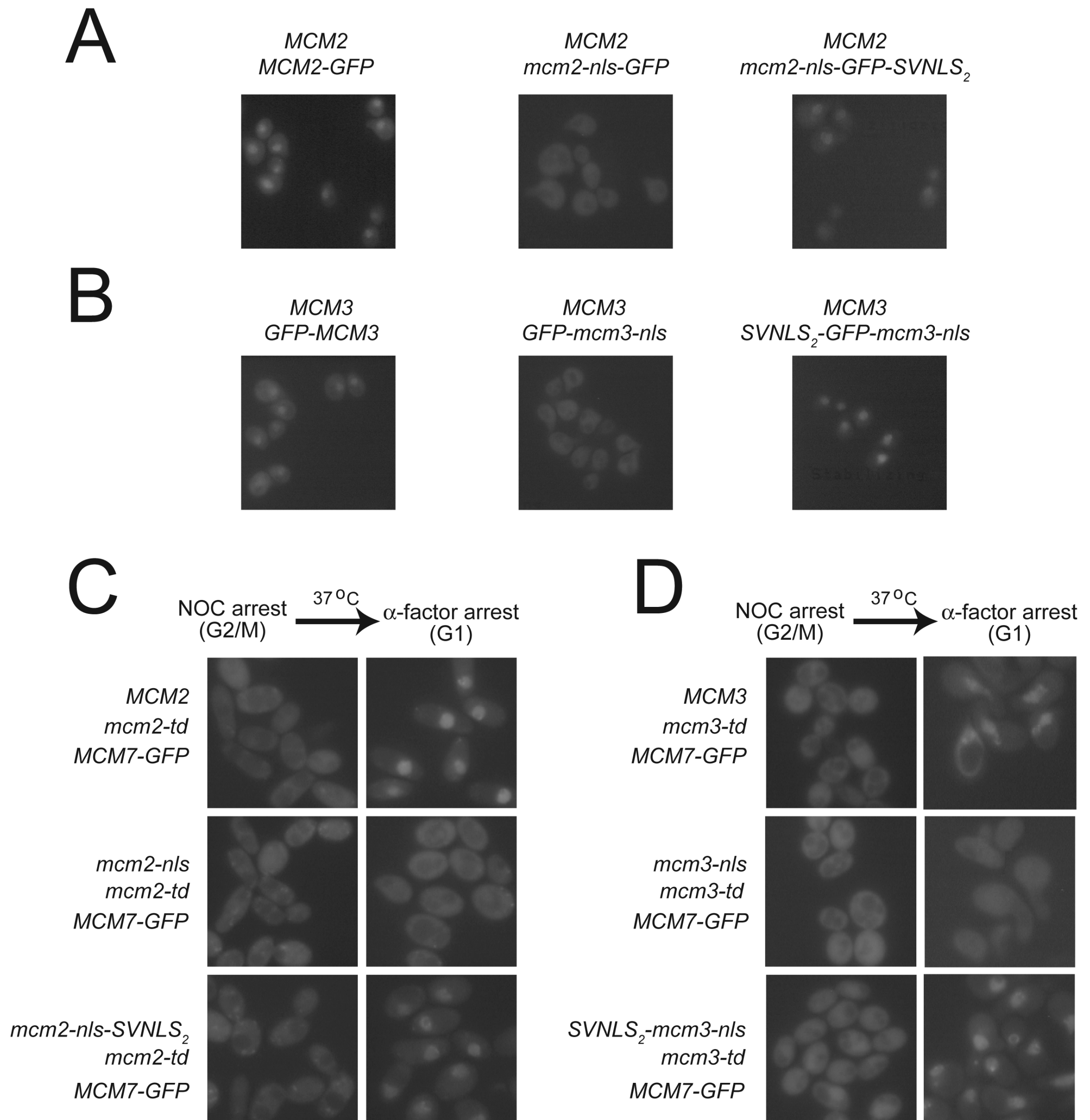


Figure 3

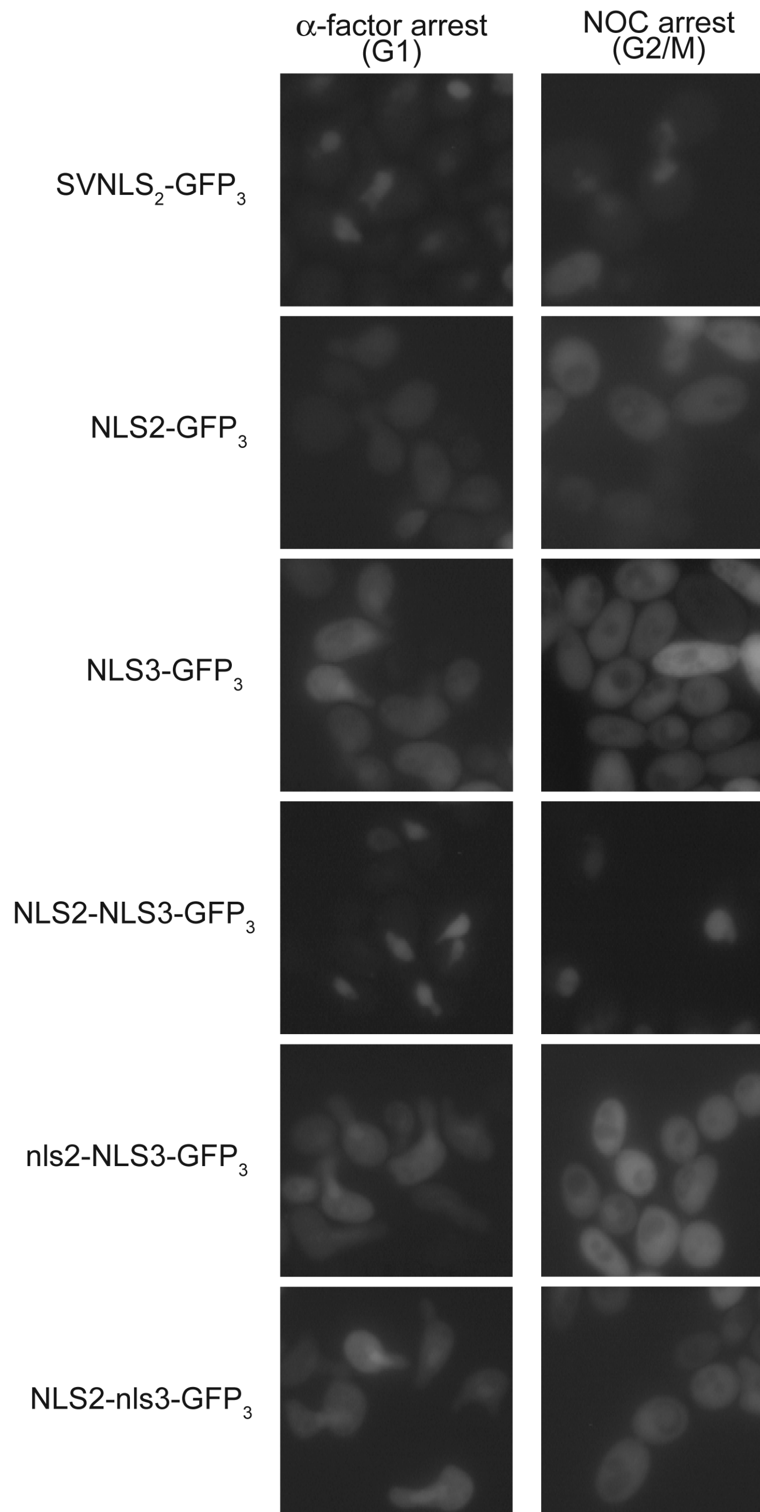
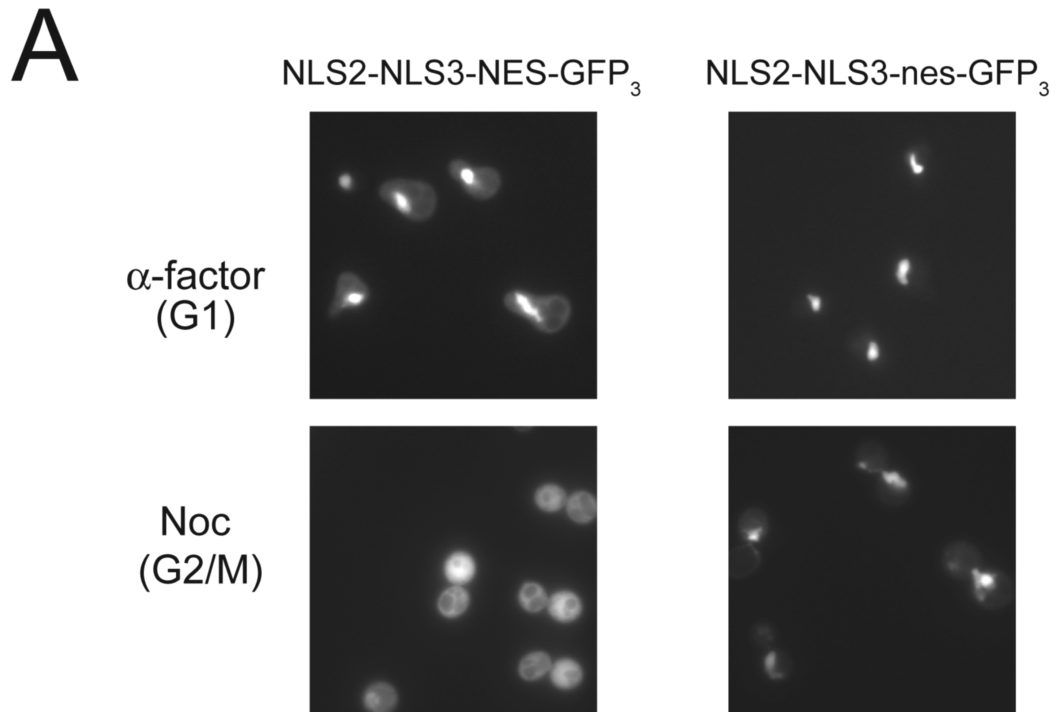


Figure 4



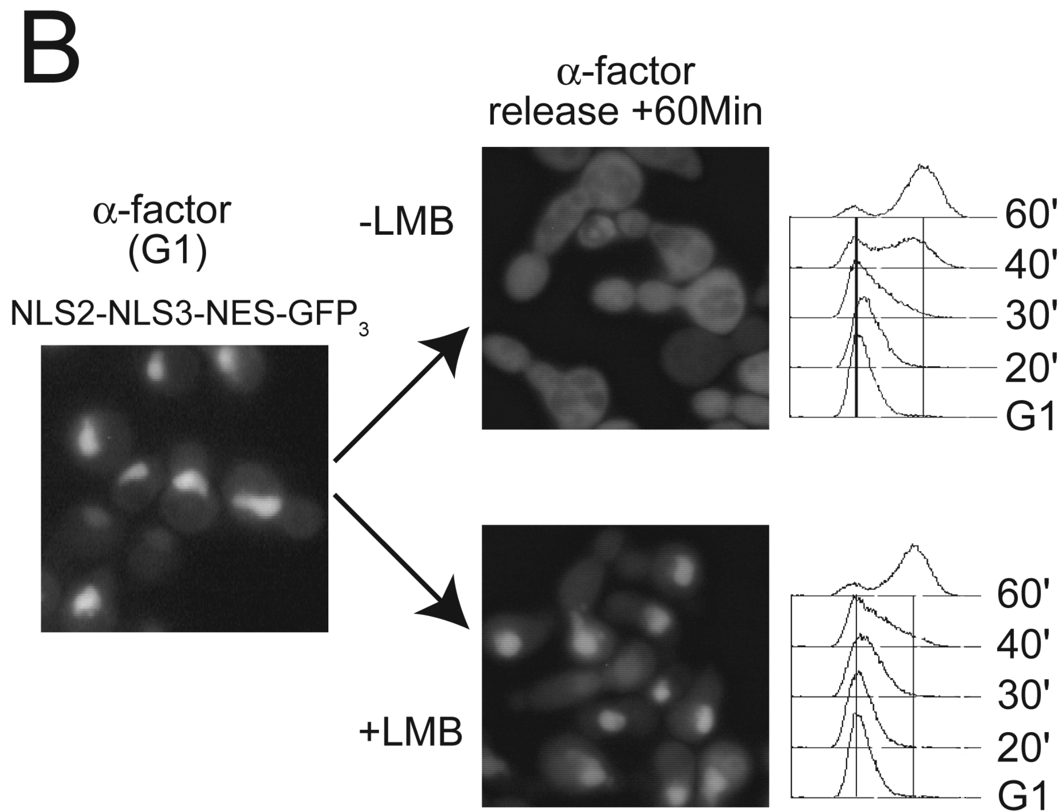


Figure 5

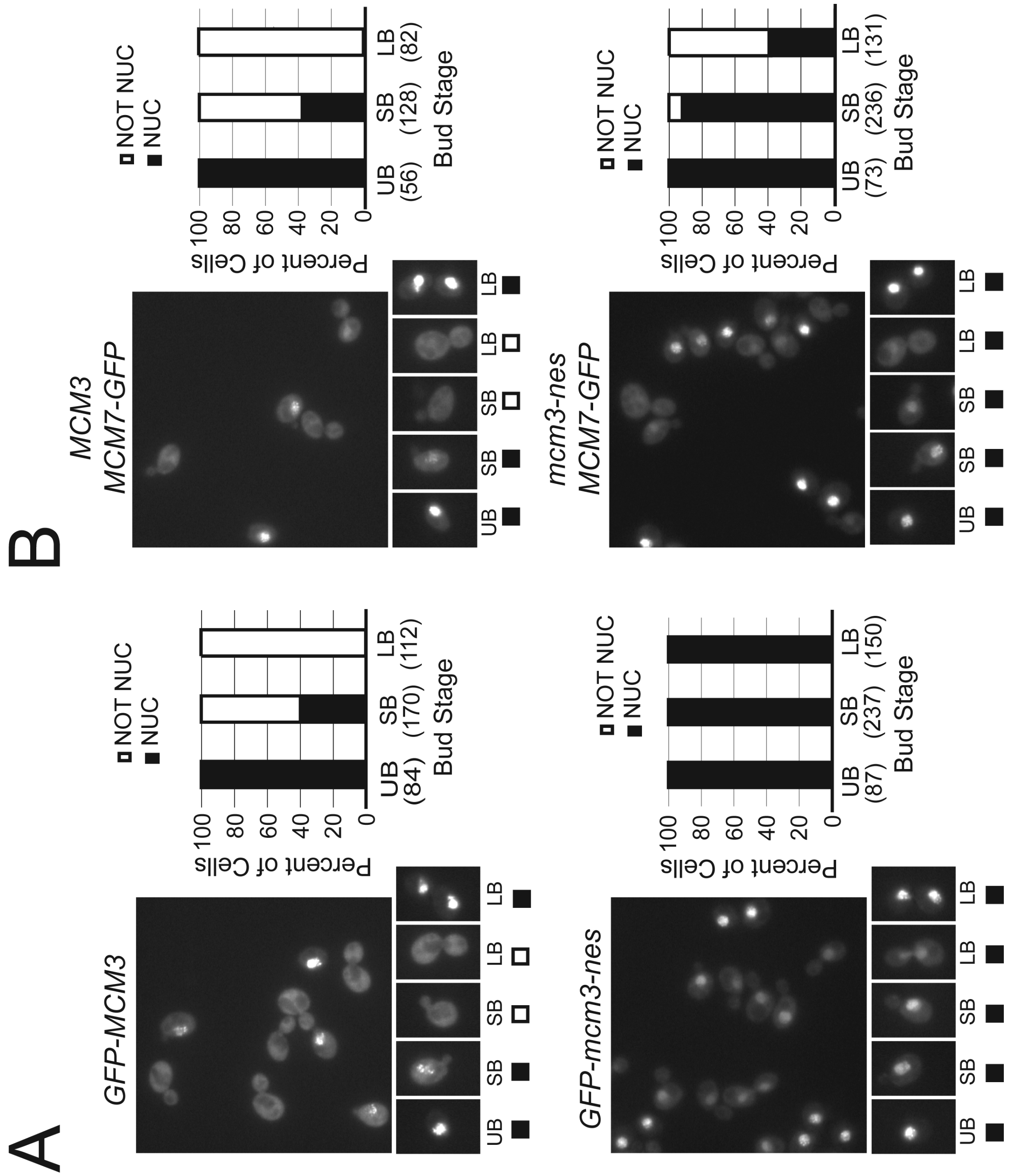
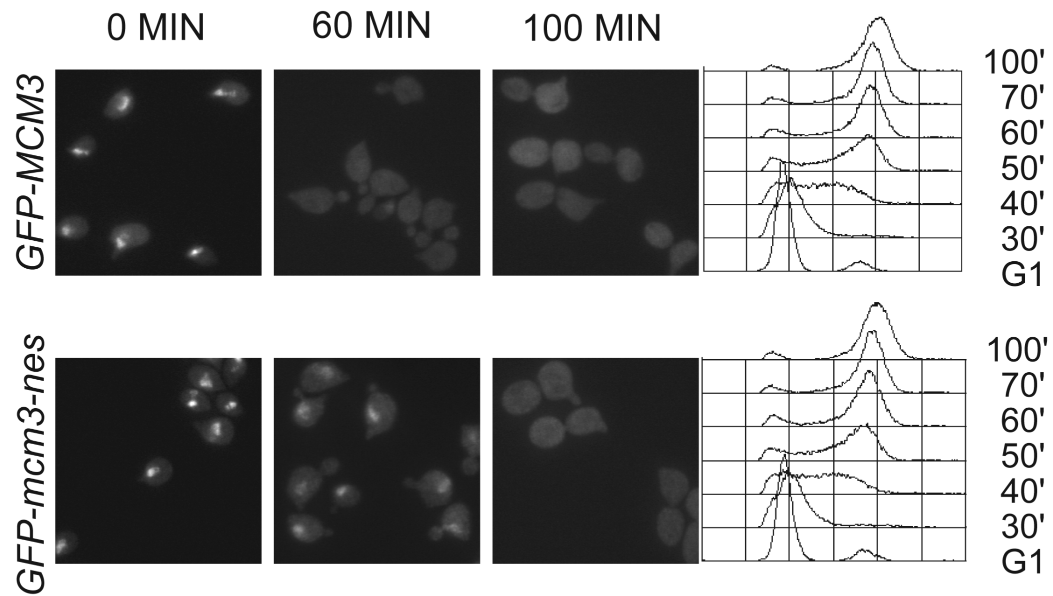


Figure 6

A



B

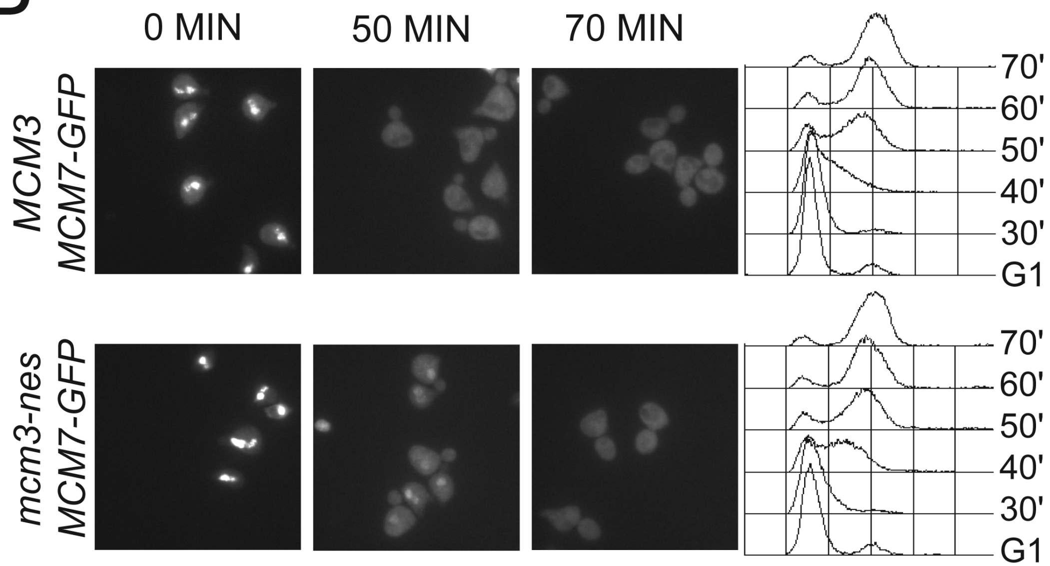


Figure 7

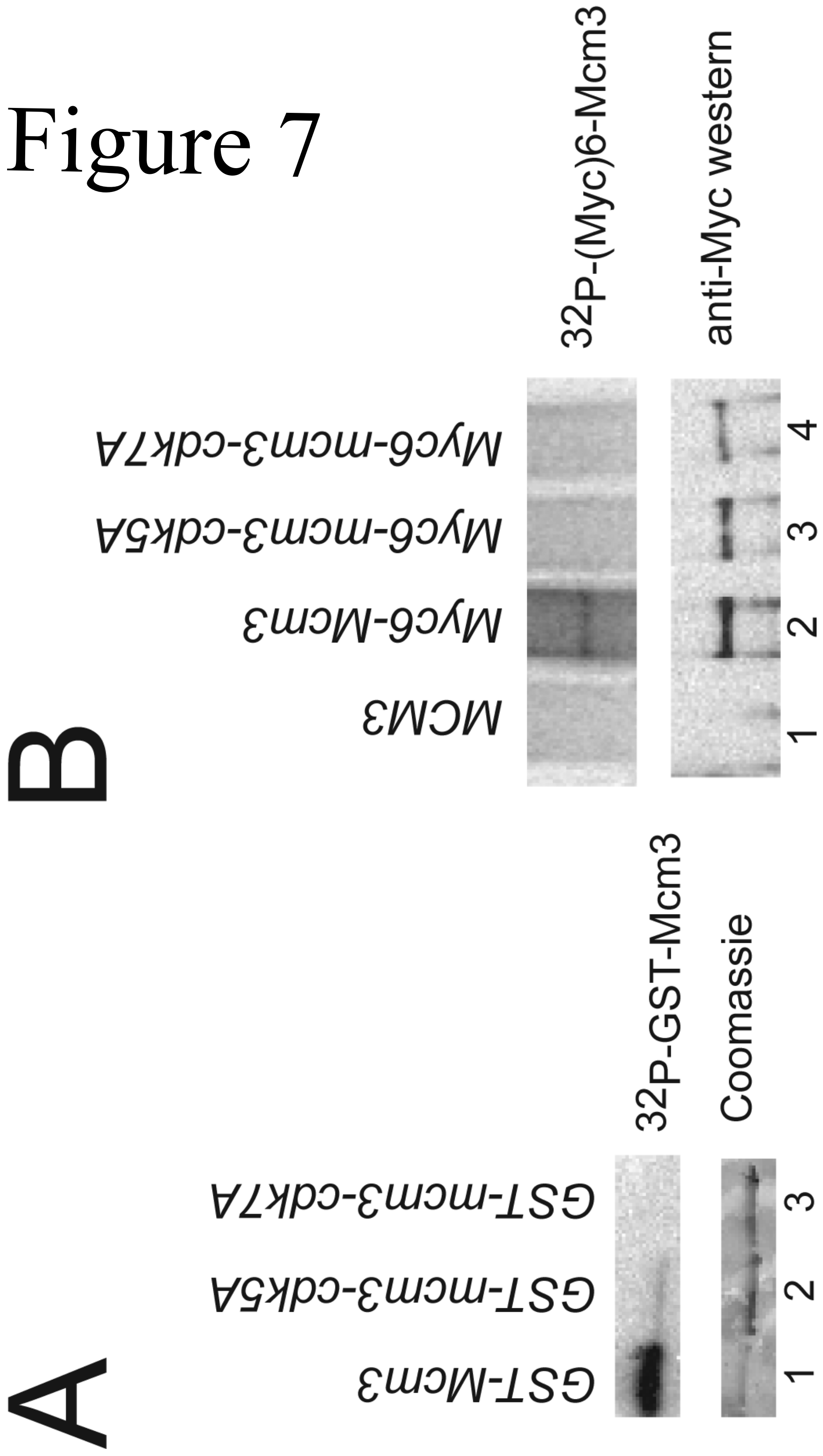


Figure 8

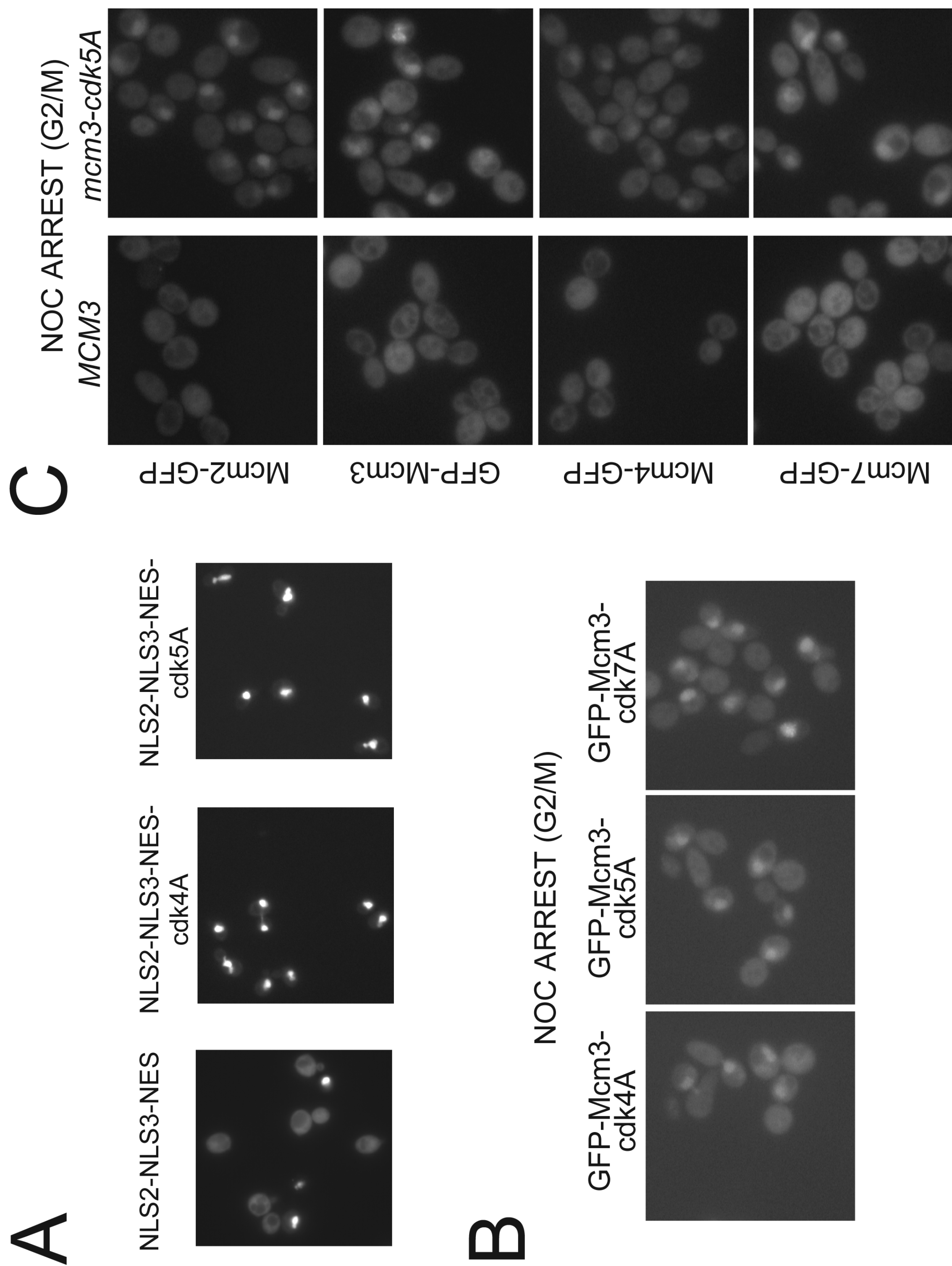


Figure 9

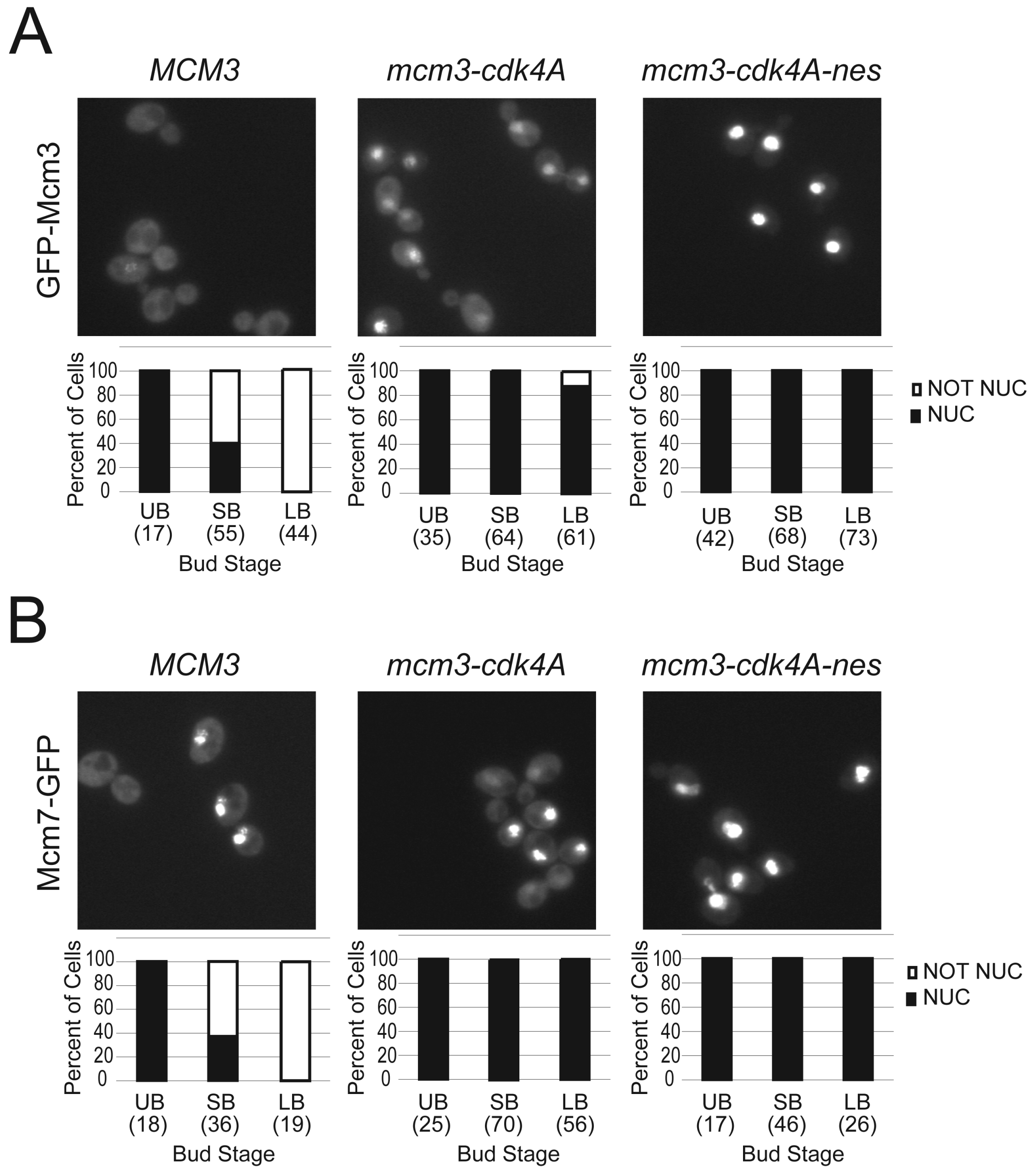
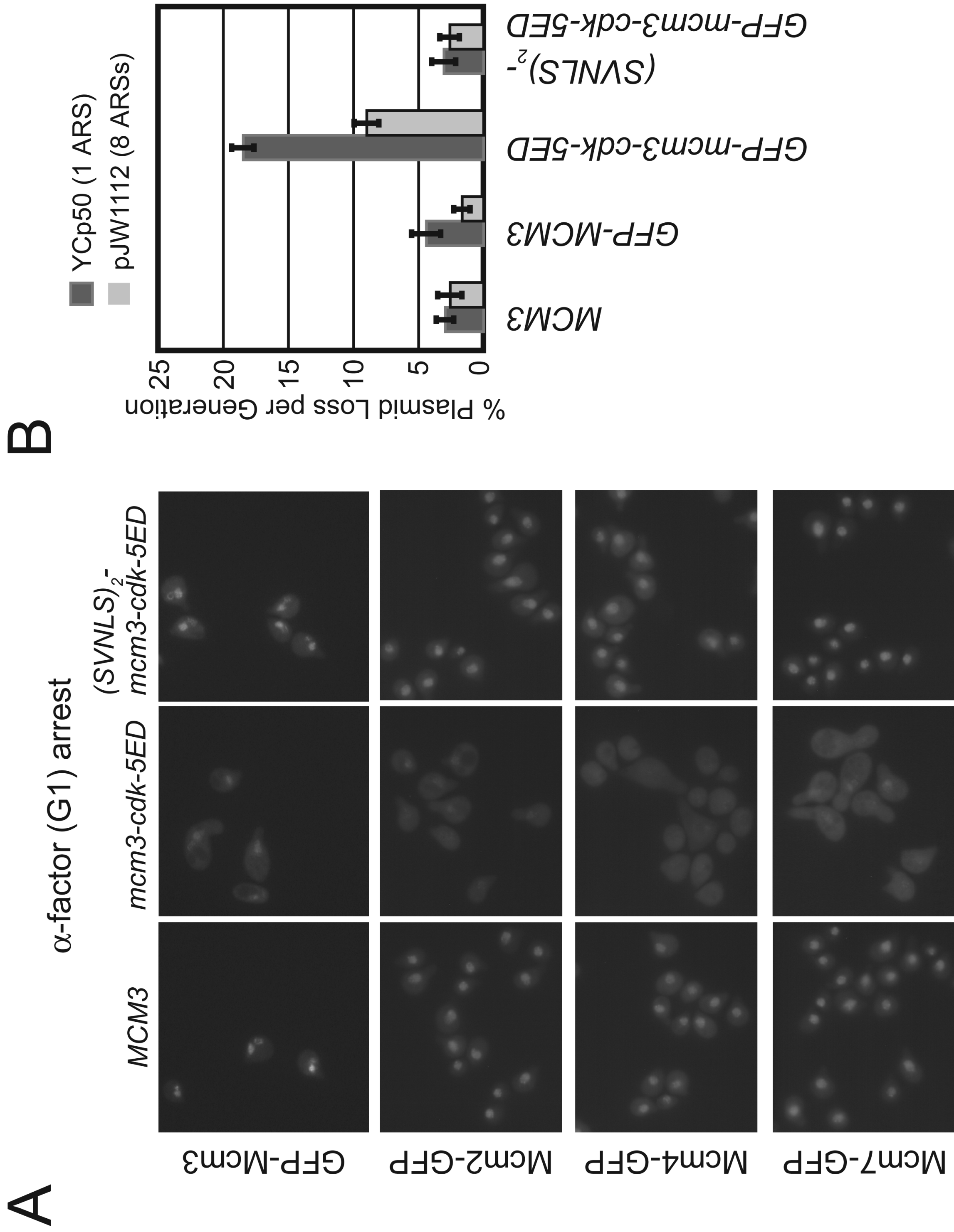


Figure 10



SUPPLEMENTARY MATERIALS AND METHODS

Integrating Replacement Plasmids for MCM Genes

All MCM3 integrating replacement plasmids for two-step gene replacement are derivatives of pKI1296 (Nguyen *et al*., 2000), a pRS306-based plasmid (Sikorski and Hieter, 1989) which contains the *MCM3* gene, an *SgrAI* - *Not*I linker (5'-GCACCGGTGGCGGCCGC-3') inserted between the first and second codons of the *MCM3* ORF, and a GFP cassette encoding the S65T V163A variant (Nguyen *et al*., 2000) inserted in the *Not*I site. The immediate parent of the gene replacement plasmids used in this study is pKI1359, which was generated from pKI1296 by introducing seven silent base substitutions between the *Cla*I and *Bsu36*I sites of the *MCM3* ORF (see Table 1). The resulting *MCM3-silent* allele has five added restriction sites (*Bgl*II, *Nhe*I, *BssH*II, *Pst*I and *Avr*II) and one removed site (*Bam*HI).

Additional base substitutions (see Table 1) were introduced into pKI1359 to mutate various motifs between the *Cla*I and *Bsu36*I sites of the *MCM3* ORF, generating the following *mcm3* mutant alleles: (1) *nls* – Mcm3 NLS basic motif mutated from PKKRQRV to PAAAQAV and marked by addition of a *Pvu*II site (2) *nes* –Mcm3 NES leucine-rich motif mutated from LQRRLRLGL to AQRRARAGA and marked by addition of an *Apa*I site; (3) *cdk4A* –phosphoacceptor residues of the four full consensus CDK sites flanking the NLS mutated to alanine and marked from N- to C- terminus, respectively, by addition of *Apa*I (replacing *Bgl*II site in pKI1359), *Nar*1, *Sac*I, and *Ava*I sites; (4) *cdk5A* –cdk4A mutations plus phosphoacceptor residue of the full consensus CDK site adjacent to the leucine rich motif mutated to alanine and marked by addition of

*Nar*I site; (5) *cdk7A* –cdk5A mutations plus TP and SP sites between the basic and leucine rich motif mutated to alanine and marked respectively by addition of *Fnu4H*I and *Hae*III restriction sites (6) *cdk5ED* – phosphoacceptor residues of the five full consensus CDK sites to aspartic acid (from serine) or glutamic acid (from threonine) and marked from N- to C-terminus, respectively, by addition of *BamH*I (replacing *Bgl*II site in pKI1359), *PpuM*I, *BstE*II, *Dde*I, and *Sau3A* sites. Plasmids containing these *mcm3* alleles are described in Table 2. In plasmids lacking *GFP*, the *GFP* cassette was excised by *Not*I digestion and recircularization. In plasmids containing *SVNLS*$_2$, an *SgrA*I cassette encoding two tandem copies of the SV40 NLS, was inserted into the *SgrA*I site. The sequence of the *SVNLS*$_2$ cassette including flanking *SgrA*I restriction sites is:

5'-CACCGGTGAGATCTGGTCCTCCAAAAAAGAAGAGAAAGGTTGAAGGTGGA TCTGGTCCTCCAAAAAAGAAGAGAAAGGTTGAAGGTGGATCTAGACCACCG GTG-3'. In plasmids containing *Myc*$_6$, two *Not*I cassettes each encoding three copies of the Myc epitope replaced the *GFP* cassette in the *Not*I site. The sequence of the triple Myc cassette including flanking NotI restriction sites is:

5'-GCGGCCGCGCAGAAGAACAAAAGTTGATTTCTGAAGAAGACTTGAACGGC GAACAAAA G TTGATTTCTGAAGAAGACTTGAACGGATCCGAACAAAAGTT GATTTCTGAAGAAGACTTG AACTGCGGCCGC-3'. In pVN165, which contains the *mcm3-td* allele expressed under the control of the endogenous *MCM3* promoter, an *SgrA*I cassette encoding the ts-degron module (without any promoter sequences) was inserted in the *SgrA*I site. This ts-degron cassette was PCR amplified from pPW58 (Dohmen *et al*., 1994) with oligonucleotides

OJL282 5'-CAAGGCACCGGTGCAGATTTTCGTCAAGACTTTGACCG-3' and

OJL283 5'-CAGACCACCGGTGCGGTACCGTCTTTCTTCTCG-3'.

All but two *MCM2* integrating replacement plasmids for two-step gene replacement (see Table 2) were derived from pKI1151 (Nguyen *et al.*, 2000), a pRS306-based plasmid (Sikorski and Hieter, 1989) that contains the *MCM2* gene, a *Not*I-*SgrA*I linker (5'-GGCGGCCGCGCACCGGTG-3') inserted between the last and stop codons of the *MCM2* ORF, and a GFP-cassette (encoding the S65T V163A variant) inserted at the *Not*I site. pKI1500 is identical to pKI1151 except that nucleotides 13 to 26 of the *MCM2* ORF are mutated from 5'-AGAAGACGTAGACG-3', which encodes RRRRR of the Mcm2 NLS (*NLS2*), to 5'-GCCGCTGCCGCAGC-3', which encodes AAAAA of a mutant Mcm2 NLS (*nls2*). pKI1495 is identical to pKI1500 except that the $SVNLS_2$ cassette is inserted in the *SgrA*I site. pKI1145, pAR146, and pAR145 are identical to pKI1151, pKI1500, and pKI1495, respectively, except for the absence of the *GFP* cassette in the *Not*I site. The two *MCM2* integrating replacement plasmids that were not derived from pKI1151 are pAR138 and pKI1365. pKI1365 was derived from pJL973 (Nguyen *et al.*, 2000), which contains the *MCM2* gene cloned into pRS306, by insertion of an *SgrA*I-*Not*I linker 5'-GCACCGGTGGGCGGCCGC-3' between the first and second codon of the *MCM2* ORF. To generate pAR138, the *SgrA*I ts-degron cassette described above for pVN165 was inserted into the *SgrA*I site of pKI1365 at the 5' end of the MCM2 ORF. All integrating replacement plasmids were introduced into yeast strains by two-step gene replacement as described in Table 5.

Integrating Tagging Plasmids for MCM Genes

Integrating tagging plasmids (Table 3) for introducing two copies of the SV40 NLS ($SVNLS_2$) at the N-terminus or Mcm3 (pRS1412) or Mcm6 (pKI1408), or the C terminus of Mcm2 (pKI1410), Mcm4 (pKI1404), or Mcm5 (pKI1406), were generated as follows. The *NgoM*IV-*Aat*II fragments from previously described pKI1218, pKI1226, pKI1341, pKI1188 and pKI1323 (Nguyen *et al*., 2000) were inserted into the *NgoM*IV-*Aat*II vector fragment of pRS304 (Sikorski and Hieter, 1989) to construct pKI1404, pKI1406, pKI1408, pKI1410 and pKI1412, respectively.

Integrating tagging plasmids (Table 3) for tagging Mcm2 (pKI1183), Mcm4 (pKI1212), or Mcm7 (pKI1198) at their C-terminus with GFP (the S65T V163A variant) were previously described (Nguyen *et al*., 2000). All integrating tagging plasmids were introduced into yeast strains by one-step loop-in integration as described in Table 5.

pGAL Expression Plasmids

All GFP reporter plasmids (Table 4) were derived from pKI1518, which contains the *GAL10* promoter and sequences encoding three tandem copies of GFP (the F64L S65T variant (Nguyen *et al*., 2000) inserted in a modified multiple cloning site of pRS316, a *URA3* marked *CEN-ARS* vector (Sikorski and Hieter, 1989). A cassette containing the *GAL10* promoter flanked by *Kpn*I sites (5'-GGTACCCGGTGC-3' linker added to 5' end of both strands) was PCR amplified from pPH44 (Phil Hieter, Michael Smith Laboratories University of British Columbia Vancouver,BC, Canada) and inserted in the *Kpn*I site at one end of the pRS316 multiple cloning site oriented such that the *GAL10* promoter transcribes in the direction of the multiple cloning site. The sequence of the promoter is a variant of the *GAL10* sequence listed in the Saccharomyces Genome

Database (SGD) from nucleotide –668 to +3 (where position +1 is the first nucleotide of the *GAL10* ORF). On the transcribed strand of the *GAL10* cassette is an insertion of 5'-ATATGTATATGG-3' between positions –83 and –82 of the SGD sequence and base substitutions at the following SGD positions: G-239A, G-253T, A-417C, C-532G, C-537T, T-575C and G-602T (SGD base on left; cassette base on right). An *EcoR*I – *BamH*I fragment containing the tandem GFP sequences (*GFP*$_3$) from EB0757 (Kaffman *et al*., 1998) was inserted between the *EcoR*I and *BamH*I sites of the multiple cloning site. Finally, the portion of the multiple cloning site in pRS316 from *Sal*I to *EcoR*I was replaced with the sequence

5'-GTCGACTCCGGAATCGATAAGCTTCCTGAGGGAGCACCGGTGAACCCGGG TGAATTC-3', resulting in the restriction sites from *GAL10* promoter to *GFP*$_3$ reading *Xho*I-*Sal*I-*BspE*I-*Cla*I-*Hind*III-*Bsu36*I-*SgrA*I-*Sma*I-*EcoR*I. pJL1551, a *TRP1* integrating version of pKI1518, was created by replacing the entire multiple cloning site of pRS304 (Sikorski and Hieter, 1989) with that of pKI1518. pAR101, 110, 113, 126, 127 (Table 4) contain mutant and wild-type sequence cassettes from *MCM2* and *MCM3* inserted into pKI1518. pML104, 105, 116, and pMT103 (Table 4) are similarly related to pJL1551. The cassettes and the restriction sites used to insert them are described below using a nucleotide numbering system where +1 is the first position of the relevant ORF. The *NLS2* cassette is composed of nucleotide 1-51 from wild-type *MCM2* precisely inserted between *Sal*I and *BspE*I. The *nls2* cassette is composed of nucleotide 1-51 from the *mcm2-nls* allele (described above) precisely inserted between *Sal*I and *BspE*I. The *NLS3* cassette is composed of nt 2278-2367 from the *MCM3-silent* allele (Table 1) precisely inserted between *Cla*I and *Hind*III. The *nls3* is composed of nt 2278-2367 of the *mcm3-*

nls allele (Table 1) precisely inserted between *Cla*I and *Hind*III. The *NLS3NES3, NLS3nes3*, *NLS3NES3-cdk4A*, *NLS3NES3-cdk5A*, and cassettes are identical, respectively, to the *Cla*I to *Bsu36*I fragments from *MCM3-silent*, *mcm3-nes*, *mcm3-cdk4A*, and *mcm3-cdk5A* (described in Table 1) and were inserted into the *Cla*I and *Bsu36*I sites of pKI1551. Finally, the *SVNLS* cassette in pAR109 is composed of the sequence 5'-CCTCCTAAGAAAAAGAGAAAGGTCGAAGACCCCAAGGAC-3' (encodes PPKKKRKVEDPK) precisely inserted between *Cla*I and *Hind*III. All sequences between the *GAL10* promoter and the *GFP*$_3$ were confirmed by sequencing.

High copy plasmids expressing wild-type (pVN185) and mutant (pVN186 and pVN187) Mcm3 proteins under the control of the *pGAL* promoter were constructed as follows. To construct pVN185, we first removed restriction sites from the pRS426 multicloning site by digesting with *Ecl136*II and *Cla*1, filling in the recessed ends with Klenow, and recircularizing the blunt-ended fragments. The 1177bp *Stu*I to *Sal*I fragment from the resulting vector clone was then replaced by a *Stu*I to *Sal*I fragment from pEG(KT)-MCM3 (Lei *et al*., 1996) to generate pVN185. pVN186 and pVN187 were constructed by inserting the 2094bp *Bgl*II to *Asp718* fragment from pKI1358 and pKI1401 respectively into the *Bgl*II to *Asp718* vector fragment of pVN185. pKI1458 contains *UBR1* under the expression of the *GAL1* promoter in the *TRP1*-marked integrating vector pJL804, which was constructed by inserting the *Not*I-*Xho*I *GAL1* promoter fragment from pJL806 (Nguyen *et al*., 2001) into the *Not*I-*Xho*I vector fragment of pRS304 (Sikorski and Hieter, 1989). An N-terminal *BamH*I-*Spe*I fragment, PCR amplified from pUBR1 (A. Varshavsky, California Institute of Technology, Pasadena, California) using the oligonucleotides

OJL855 5’-CTCTGGATCCATGTCCGTTGCTGATGATGATTTAGG-3’and OJL856 5’-GCTTCTTCATCTCCACAGTCAC-3’, and a C-terminal *Spe*I-*Pst*I fragment from pUBR1 were cloned into the *BamH*I and *Pst*I restriction sites of pJL804 in a three-way ligation. pML104, 105, 116, pMT103, and pKI1458 were introduced into yeast strains by one step loop-in integration at the *trp1-289* locus as described in Table 5. pAR101, 109, 110, 113, 126, 127 were transformed into YJL4110 and maintained under selection for the *URA3* marker.

REFERENCES

Detweiler, C.S., et al. (1998). Ectopic induction of Clb2 in early G1 phase is sufficient to block prereplicative complex formation in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *95*, 2384-2389.

Dohmen, R.J., et al. (1994). Heat-inducible degron: a method for constructing temperature-sensitive mutants. Science *263*, 1273-1276.

Jaspersen, S.L., et al. (1998). A late mitotic regulatory network controlling cyclin destruction in Saccharomyces cerevisiae. Mol Biol Cell *9*, 2803-2817.

Kaffman, A., et al. (1998). Phosphorylation regulates association of the transcription factor Pho4 with its import receptor Pse1/Kap121. Genes Dev *12*, 2673-2683.

Lei, M., et al. (1996). Physical interactions among Mcm proteins and effects of Mcm dosage on DNA replication in Saccharomyces cerevisiae. Mol Cell Biol *16*, 5081-5090.

Nguyen, V.Q., et al. (2000). Clb/Cdc28 kinases promote nuclear export of the replication initiator proteins Mcm2-7. Curr Biol *10*, 195-205.

Nguyen, V.Q., et al. (2001). Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. Nature *411*, 1068-1073.

Sikorski, R.S., et al. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics *122*, 19-27.

Supplementary Figure 1.

Lethal mutation of either Mcm2 or Mcm3 NLSs can be rescued by fusing tandem SV40 NLSs to a single Mcm protein in the Mcm2-7 complex.

A. The *URA3* marked plasmid pKI1500 (*mcm2-nls-GFP*) is integrated in the following strains by homologous recombination 5' of the *MCM2* ORF: YJL1231 (*MCM2::{mcm2-nls-GFP, URA3}*), YJL3074 (*MCM2::{mcm2-nls-GFP, URA3} SVNLS$_2$-MCM3*), YJL3070 (*MCM2::{mcm2-nls-GFP, URA3} MCM4-SVNLS$_2$*), YJL3072 (*MCM2::{mcm2-nls-GFP, URA3} MCM5-SVNLS$_2$*), and YJL3076 (*MCM2::{mcm2-nls, URA3} MCM6-SVNLS$_2$*). pKI1495 (*mcm2-nls-GFP-SVNLS$_2$*) is similarly integrated in YJL1228 (*MCM2::{SVNLS$_2$-mcm2-nls, URA3}*). These strains contain both a WT and mutant copy of the *MCM2* locus in direct repeat, and homologous recombination between these loci would result in excision of the integrated plasmid. The strains were plated on 5-FOA to select for this excision, which, depending on the position of the recombination event, would leave a single *MCM2* locus containing either the WT or mutant *MCM2* NLS sequence (see diagram). PCR and restriction analysis was used to distinguish between these two outcomes. The number of excisants containing the mutation and the total number of excisants screened are tabulated.

B. The *URA3* marked plasmid pKI1362 (*GFP-mcm3-nls*) is integrated by homologous recombination 3' of the *MCM3* ORF in the following strains: YJL2675 (*MCM3::{GFP-mcm3-nls, URA3}*), YJL3445 (*MCM3::{GFP-mcm3-nls, URA3} MCM2-SVNLS$_2$*), YJL3438 (*MCM3::{GFP-mcm3-nls, URA3} MCM4-SVNLS$_2$*), YJL3441 (*MCM3::{GFP-mcm3-nls, URA3} MCM5-SVNLS$_2$*), and YJL3444 (*MCM3::{GFP-mcm3-nls, URA3} MCM6-SVNLS$_2$*). pKI1356 (*SVNLS$_2$-GFP-mcm3-nls*) is similarly integrated in YJL2665

(*MCM3::{SVNLS$_2$-GFP-mcm3-nls, URA3}*). These strains were plated on 5-FOA to select for plasmid excision and the excisants were analyzed as described in Supplementary Figure 1A

Supplementary Figure 2.

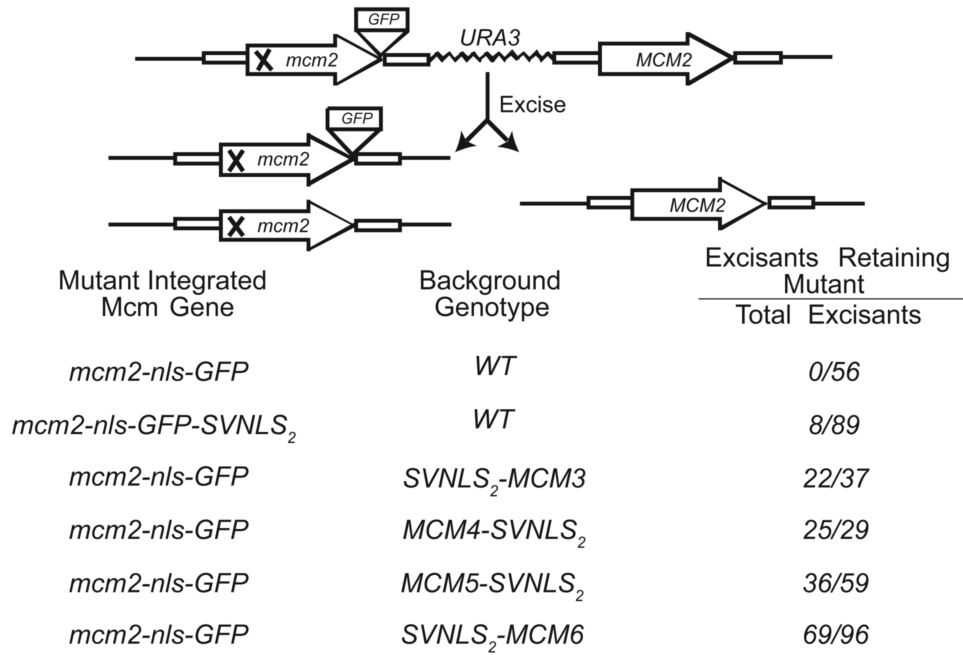
The phosphomimic mutations of the Mcm3 CDK consensus sites can slow or arrest the cell cycle and induce an accumulation of G2/M cells.

A. YJL1265 (*GFP-mcm3-cdk5ED*) and YJL2160 (*GFP-MCM3*) cells growing exponentially in YEPD medium were sonicated and plated on YEPD plates at 23°C. The number of cell lobes in each of 200 colonies was counted at the indicated times.

B. YJL1265 (*GFP-mcm3-cdk5ED*) and YJL2160 (*GFP-MCM3*) cells growing exponentially in YEPD medium were fixed in ethanol for FACS analysis and in formaldehyde for budding indices.

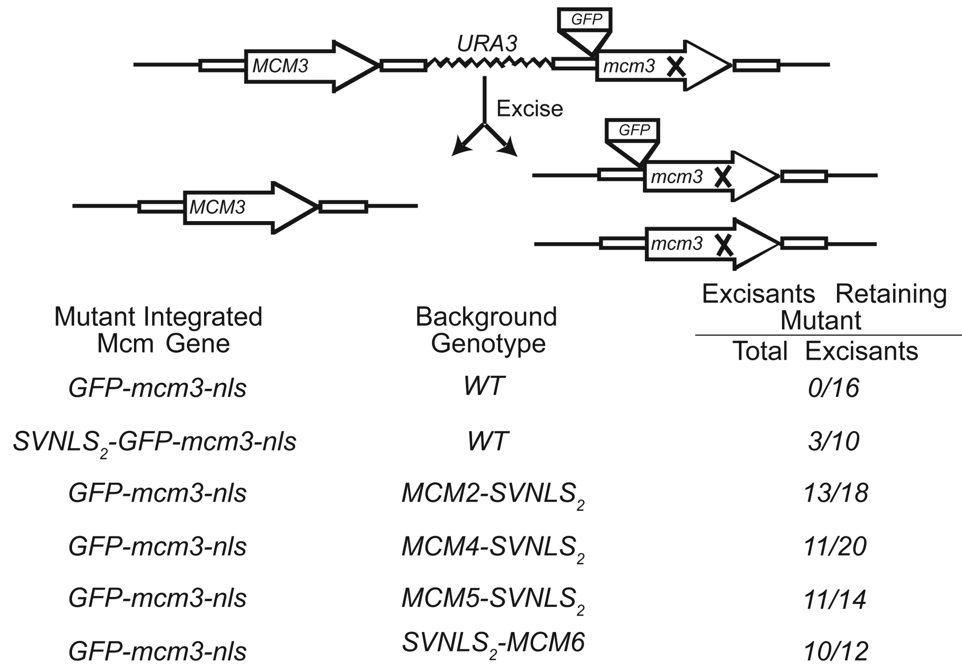
SupFig 1

A



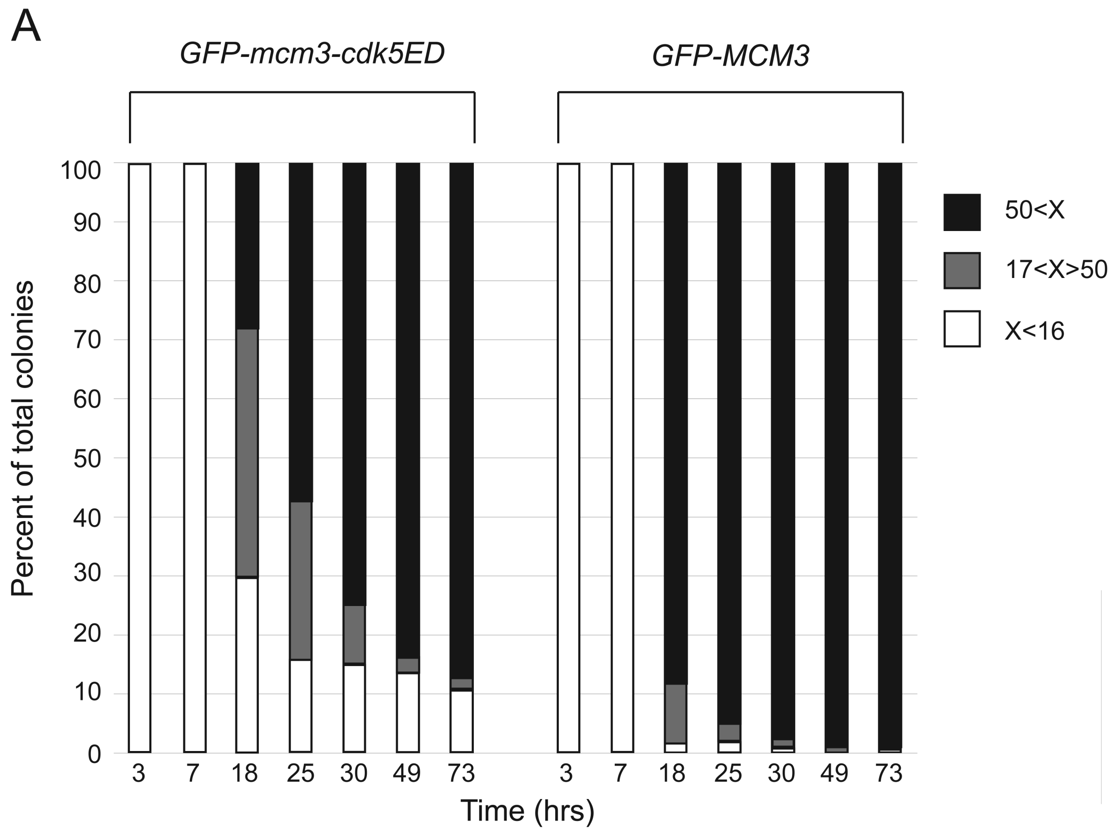
Mutant Integrated Mcm Gene	Background Genotype	Excisants Retaining Mutant / Total Excisants
mcm2-nls-GFP	*WT*	*0/56*
mcm2-nls-GFP-SVNLS$_2$	*WT*	*8/89*
mcm2-nls-GFP	*SVNLS$_2$-MCM3*	*22/37*
mcm2-nls-GFP	*MCM4-SVNLS$_2$*	*25/29*
mcm2-nls-GFP	*MCM5-SVNLS$_2$*	*36/59*
mcm2-nls-GFP	*SVNLS$_2$-MCM6*	*69/96*

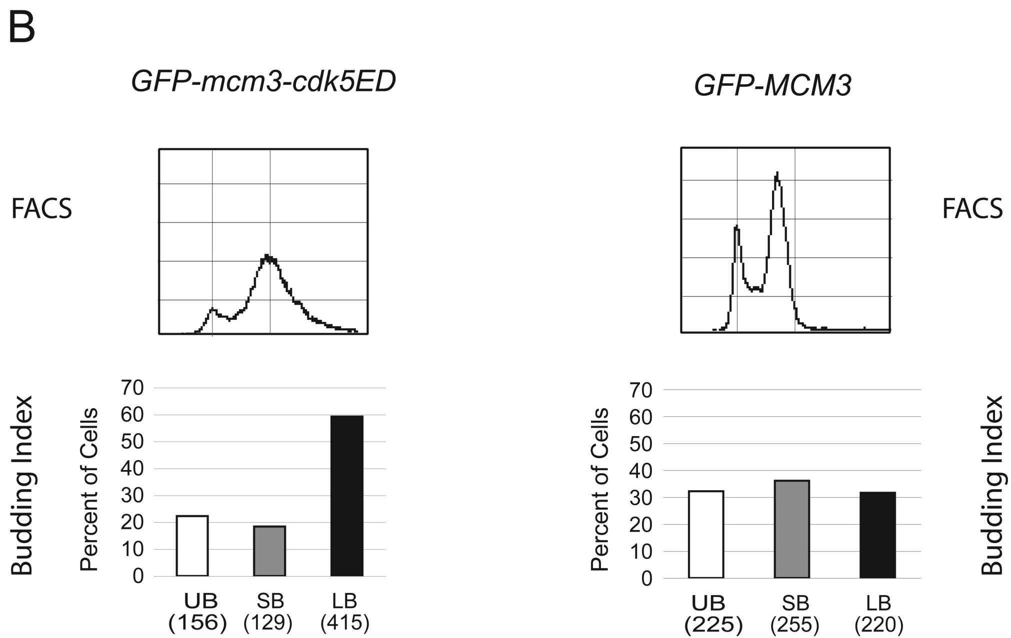
B



Mutant Integrated Mcm Gene	Background Genotype	Excisants Retaining Mutant / Total Excisants
GFP-mcm3-nls	*WT*	*0/16*
SVNLS$_2$-GFP-mcm3-nls	*WT*	*3/10*
GFP-mcm3-nls	*MCM2-SVNLS$_2$*	*13/18*
GFP-mcm3-nls	*MCM4-SVNLS$_2$*	*11/20*
GFP-mcm3-nls	*MCM5-SVNLS$_2$*	*11/14*
GFP-mcm3-nls	*SVNLS$_2$-MCM6*	*10/12*

SupFig2





Supplementary Table 1. Base substitutions in *mcm3* alleles						
silent	*nls*	*nes*	*cdk4A*	*cdk5A*	*cdk7A*	*cdk5ED*
G2280a	G2298a	T2500g	a2280G	a2280G	a2280G	a2280G
A2328t	A2299g	T2501c	T2281g	T2281G	T2281g	T2281g
A2340t	A2300c	A2502c	T2283c	T2283C	T2283c	C2282a
A2406g	A2301t	C2512g	A2292g	A2292g	A2292g	A2292g
G2484a	A2302g	T2513c	T2293g	T2293g	T2293g	T2293g
A2538t	A2303c	T2514c	A2295g	A2295g	A2295g	C2294a
A2541g	A2304c	2518G	G2298a	G2298a	G2298a	A2295c
	A2305g	T2524c	t2340A	t2340A	t2340A	G2398a
	G2306c	G2526	T2341g	T2341g	T2341g	T2341g
	A2307t		C2343t	C2343t	C2343t	C2342a
	A2311g		A2356g	A2356g	A2356g	A2356g
	G2312c		T2358c	T2358c	T2358c	C2357a
	A2313t		A2361g	A2361g	A2361g	T2358g
				T2533g	A2392g	G2532t
				T2535g	G2394t	T2533g
					T2470g	C2534a
					A2472c	
					T2533g	
					T2535g	

Base position is listed relative to the first nucleotide of the ORF (+1) with starting nucleotide on left and substituted nucleotide on right. In most cases the starting nucleotide is wild type (upper case) and substituted nucleotide is mutant (lower cases),

Supplementary Table 1. Base substitutions in *mcm3* alleles (cont'd)

but in a few cases a mutated nucleotide is reverted back to wild-type. Base subsitution for all alleles aside from the *silent* allele were introduced into the *silent* allele.

Supplementary Table 2. Integrating replacement plasmids

Plasmids	*mcm* Allele	Source
pAR128	*SVNLS*$_2$*-mcm3-nls*	This Study
pAR138	*mcm2-td*	This Study
pAR145	*mcm2-nls-SVNLS*$_2$	This Study
pAR146	*mcm2-nls*	This Study
pKI1145	*MCM2*	This Study
pKI1151	*MCM2-GFP*	(Nguyen *et al.*, 2000)
pKI1296	*GFP-MCM3*	This Study
pKI1304	*Myc*$_6$*-MCM3*	This Study
pKI1356	*SVNLS*$_2$*-GFP-mcm3-nls*	This Study
pKI1358	*GFP-mcm3-cdk7A*	This Study
pKI1359	*GFP-MCM3-silent*	This Study
pKI1362	*GFP-mcm3-nls*	This Study
pKI1363	*GFP-mcm3-nes*	This Study
pKI1365	*MCM2*	This Study
pKI1392	*MCM3-silent*	This Study
pKI1394	*mcm3-nls*	This Study
pKI1400	*GFP-mcm3-cdk4A*	This Study
pKI1401	*GFP-mcm3-cdk5A*	This Study
pKI1495	*mcm2-nls-GFP-SVNLS*$_2$	This Study

Supplementary Table 2. Integrating replacement plasmids cont'd

Plasmids	*mcm* Allele	Source
pKI1500	*mcm2-nls-GFP*	This Study
pKI1517	*SVNLS*$_2$*-GFP-mcm3-cdk5ED*	This Study
pKI1525	*GFP-mcm3-cdk5ED*	This Study
pML106	*GFP-mcm3-cdk4A-nes*	This Study
pVN165	*Myc*$_6$*-mcm3-td*	This Study
pVN181	*SVNLS*$_2$*-mcm3-cdk5ED*	This Study
pVN195	*Myc*$_6$*-mcm3-cdk5A*	This Study
pVN196	*Myc*$_6$*-mcm3-cdk7A*	This Study

The *GFP* allele used in Table 2 plasmids contains the S65T and V163A mutations. The vector backbone is pRS306.

Supplementary Table 3. Integrating tagging plasmids

Plasmids	MCM Allele	Vector Backbone	Source
pKI1183	*CtermMCM2-GFP*	pRS306	(Nguyen *et al.*, 2000)
pKI1198	*CtermMCM7-GFP*	pRS306	(Nguyen *et al.*, 2000)
pKI1212	*CtermMCM4-GFP*	pRS306	(Nguyen *et al.*, 2000)
pKI1404	*CtermMCM4-SVNLS$_2$*	pRS304	This Study
pKI1406	*CtermMCM5-SVNLS$_2$*	pRS304	This Study
pKI1408	*SVNLS$_2$-NtermMCM6*	pRS304	This Study
pKI1410	*CtermMCM2-SVNLS$_2$*	pRS304	This Study
pKI1412	*SVNLS$_2$-NtermMCM3*	pRS304	This Study

The *GFP* allele used in Table 3 plasmids contains the S65T and V163A mutations

Supplementary Table 4. pGAL expression plasmids

Plasmids	Key Features	Vector Backbone	Source
pAR101	*pGAL10-NLS3-*GFP_3	pRS316	This Study
pAR109	*pGAL10-SVNLS-*GFP_3	pRS316	This Study
pAR110	*pGAL10-NLS2-*GFP_3	pRS316	This Study
pAR113	*pGAL10-NLS2-NLS3-*GFP_3	pRS316	This Study
pAR126	*pGAL10-nls2-NLS3-*GFP_3	pRS316	This Study
pAR127	*pGAL10-NLS2-nls3-*GFP_3	pRS316	This Study
pML104	*pGAL10-NLS2-NLS3NES3-*GFP_3	pRS304	This Study
pML105	*pGAL10-NLS2-NLS3nes3-*GFP_3	pRS304	This Study
pML116	*pGAL10-NLS2-NLS3NES3-cdk4A-*GFP_3	pRS304	This Study
pMT103	*pGAL10-NLS2-NLS3NES3-cdk5A-*GFP_3	pRS304	This Study
pVN185	*pGAL-GST-MCM3*	pRS426	This Study
pVN186	*pGAL-GST-(mcm3-cdk7A)*	pRS426	This Study
pVN187	*pGAL-GST-(mcm3-cdk5A)*	pRS426	This Study
pKI1458	*pGAL1-UBR1*	pRS304	This Study
pKI1518	*pGAL10*	pRS316	This Study
pKI1551	*pGAL10*	pRS304	This Study

The *GFP* allele used in Table 4 plasmids contains the F64L and S65T mutations.

Supplementary Table 5. Yeast Strains

Strain	Genotype	Source
YJL305	*MATα leu2-3,112 ura3-52, trp1-289*	Congenic to YJL310 (Ray Deshaies RDY487)
YJL309	*MATα leu2-3,112 ura3-52, trp1-289 pep4::TRP1*	Congenic to YJL310 (Ray Deshaies RDY493)
YJL310	*MATa leu2-3,112 ura3-52 trp1-289 bar1::LEU2*	(Detweiler and Li, 1998)
YJL312	*MATa leu2-3,112 ura3-52 trp1-289 bar1::LEU2 pep4::TRP1*	Congenic to YJL310 (Ray Deshaies)
YJL1228 *	*MCM2::{mcm2-nls-GFP-SVNLS$_2$, URA3}*	pKI1495 into YJL310[c]
YJL1231 *	*MCM2::{mcm2-nls-GFP, URA3}*	pKI1500 into YJL310[c]
YJL1259 *	*SVNLS$_2$-GFP-mcm3-cdk5ED*	pKI1517 into YJL310[a]
YJL1260 *	*SVNLS$_2$-GFP-mcm3-cdk5ED*	Sister Isolate to YJL1259
YJL1263 *	*mcm3-cdk5ED*	pKI1525 into YJL310[a]
YJL1265 *	*GFP-mcm3-cdk5ED*	pKI1525 into YJL310[a]
YJL1337 *	*Myc$_6$-mcm3-td MCM7-GFPtrp1-289::{pGAL-UBR1, TRP1}*	pVN165[a] and pKI1458[c] into YJL1977
YJL1977 *	*MCM7-GFP*	(Nguyen *et al*., 2000)
YJL1979 *	*MCM7-GFP*	Sister isolate to YJL1977
YJL2033 *	*mcm2::{MCM2-GFP, URA3}*	pKI1183 into YJL310[b]
YJL2037 *	*mcm4::{MCM4-GFP, URA3}*	pKI1212 into YJL310[b]
YJL2160 *	*GFP-MCM3*	pKI1296 into YJL310[a]
YJL2162 *	*GFP-MCM3*	Sister Isolate to YJL2160
YJL2217 *	*mcm7::{MCM7-GFP, URA3}*	pKI1198 into YJL310[b]
YJL2314 *	*GFP-mcm3-cdk7A*	pKI1358 into YJL310[a]

Supplementary Table 5. Yeast Strains (cont'd)

Strain	Genotype	Source
YJL2665 *	*MCM3::{SVNLS$_2$-GFP-mcm3-nls, URA3}*	pKI1356 into YJL310[c]
YJL2669 *	*MCM3::{GFP-MCM3-silent, URA3}*	pKI1359 into YJL310[c]
YJL2675 *	*MCM3::{GFP-mcm3-nls, URA3}*	pKI1362 into YJL310[c]
YJL2714 *	*GFP-mcm3-cdk4A*	pKI1400 into YJL310[a]
YJL2717 *	*mcm3-cdk4A*	pKI1400 into YJL310[d]
YJL2720 *	*GFP-mcm3-cdk5A*	pKI1401 into YJL310[a]
YJL2724 *	*mcm3-cdk5A*	pKI1401 into YJL310[d]
YJL2741 *	*GFP-mcm3-nes*	pKI1363 into YJL310[a]
YJL3070 *	*MCM2::{mcm2-nls-GFP, URA3} mcm4::{MCM4-SVNLS$_2$, TRP1}*	pKI1404 into YJL1231[b]
YJL3072 *	*MCM2::{mcm2-nls-GFP, URA3} mcm5::{MCM5-SVNLS$_2$, TRP1}*	pKI1406 into YJL1231[b]
YJL3074 *	*MCM2::{mcm2-nls-GFP, URA3} mcm3::{SVNLS$_2$-MCM3, TRP1}*	pKI1412 into YJL1231[b]
YJL3076 *	*MCM2::{mcm2-nls-GFP, URA3} mcm6::{MCM6-SVNLS$_2$, TRP1}*	pKI1408 into YJL1231[b]
YJL3265 *	*MCM2::{MCM2-GFP, URA3}*	pKI1151 into YJL310[c]
YJL3438 *	*MCM3::{GFP-mcm3-nls, URA3} mcm4::{MCM4-SVNLS$_2$, TRP1}*	pKI1404 into YJL2675[b]
YJL3441 *	*MCM3::{GFP-mcm3-nls, URA3} mcm5::{MCM5-SVNLS$_2$, TRP1}*	pKI1406 into YJL2675[b]
YJL3444 *	*MCM3::{GFP-mcm3-nls, URA3} mcm6::{MCM6-SVNLS$_2$, TRP1}*	pKI1408 into YJL2675[b]
YJL3445 *	*MCM3::{GFP-mcm3-nls, URA3} mcm2::{MCM2-SVNLS$_2$, TRP1}*	pKI1410 into YJL2675[b]
YJL3464 *	*Myc$_6$-mcm3-td::{MCM3-silent, URA3} MCM7-GFPtrp1-289::{pGAL-UBR1, TRP1}*	pKI1392 into YJL1337[c]
YJL3469 *	*Myc$_6$-mcm3-td::{mcm3-nls, URA3} MCM7-GFPtrp1-289::{pGAL-UBR1, TRP1}}*	pKI1394 into YJL1337[c]
YJL3474 *	*Myc$_6$-mcm3-td::{SVNLS$_2$-mcm3-nls, URA3} MCM7-GFPtrp1-289::{pGAL-UBR1, TRP1}*	pAR128 into YJL1337[c]
YJL3644 *	*mcm2-td MCM7-GFPtrp1-289::{pGAL-UBR1, TRP1}*	pAR138[a] and pKI1458[c] into YJL1977

Supplementary Table 5. Yeast Strains (cont'd)

Strain	Genotype	Source
YJL3765 *	*mcm2-td ura3-52::{MCM2, URA3) MCM7-GFPtrp1-289::{pGAL-UBR1, TRP1}*	pKI1145 into YJL3644[c]
YJL3799 *	*mcm2-td ura3-52::{mcm2-nls-SVNLS$_2$, URA3) MCM7-GFPtrp1-289::{pGAL-UBR1, TRP1}*	pAR145 into YJL3644[c]
YJL3840 *	*mcm2-td ura3-52::{mcm2-nls, URA3} MCM7-GFPtrp1-289::{pGAL-UBR1, TRP1*	pAR146 into YJL3644[c]
YJL4045	*SVNLS$_2$-mcm3-cdk5ED*	pVN181 into YJL310[a]
YJL4094 *	*SVNLS$_2$-mcm3-cdk5ED mcm2::{MCM2-GFP, URA3}*	pKI1183; into YJL4045[b]
YJL4096 *	*SVNLS$_2$-mcm3-cdk5ED mcm7::{MCM7-GFP, URA3}*	pKI1198 into YJL4045[b]
YJL4098 *	*SVNLS$_2$-mcm3-cdk5ED mcm4::{MCM4-GFP, URA3}*	pKI1212 into YJL4045[b]
YJL4103 *	*mcm3-cdk5ED mcm4::{MCM4-GFP, URA3}*	pKI1212 into YJL1263[b]
YJL4108 *	*mcm3-cdk5ED mcm7::{MCM7-GFP, URA3}*	pKI1198 into YJL1263[b]
YJL4110	*MATa ade2-1 trp1-1 can1-100 leu2-3,112 his3-11,15 ura3-1 bar1::hisG psi+*	congenic to W303 (Jaspersen *et al*., 1998)
YJL4162 *	*mcm3-cdk5ED mcm2::{MCM2-GFP, URA3}*	pKI1183 into YJL1263[b]
YJL4165 *	*mcm3-cdk5A mcm2::{MCM2-GFP, URA3}*	pKI1183 into YJL2724[b]
YJL4167 *	*mcm3-cdk5A mcm7::{MCM7-GFP, URA3}*	pKI1198 into YJL2724[b]
YJL4169 *	*mcm3-cdk5A mcm4::{MCM4-GFP, URA3}*	pKI1212 into YJL2724[b]
YJL4313	*MATa Myc$_6$-MCM3 ade2-1 trp1-1 can1-100 leu2-3,112 his3-11,15 ura3-1 bar1::hisG psi+*	pKI1304 into YJL4110[a]
YJL4315	*MATa Myc$_6$-mcm3-cdk7A ade2-1 trp1-1 can1-100 leu2-3,112 his3-11,15 ura3-1 bar1::hisG psi+*	pVN196 into YJL4110[a]
YJL4324	*MATa Myc$_6$-mcm3-cdk5A ade2-1 trp1-1 can1-100 leu2-3,112 his3-11,15 ura3-1 bar1::hisG psi+*	PVN195 into YJL4110[a]
YJL4367	*MATα crm1::LEU2 trp1-1 ade2–1 ura3–1 his3–11,15 leu2–3,112 can1–100 [crm1-lmb, HIS3]*	Karsten Weis (KWY175)
YJL4578	*MATa crm1::LEU2 trp1-1 ade2–1 ura3–1 his3–11,15 leu2–3,112 can1–100 [crm1-lmb, HIS3]*	Mating type switch of YJL4367
YJL4662	*MATa crm1::LEU2 trp1-1::{pGAL10-NLS2-NLS3NES3-GFP$_3$, TRP1} ade2–1 ura3–1 his3–11,15 leu2–3,112 can1–100 [crm1-lmb, HIS3]*	pML104 into YJL4578[c]

Supplementary Table 5. Yeast Strains (cont'd)

Strain	Genotype	Source
YJL4860	*MATa crm1::LEU2 trp1-1::{pGAL10-NLS2-NLS3nes3-*GFP_3*, TRP1} ade2–1 ura3–1 his3–11,15 leu2–3,112 can1–100 [crm1-lmb, HIS3]*	pML105 into YJL4578[c]
YJL5216 *	*GFP-mcm3-cdk4A-nes*	pML106 into YJL310[a]
YJL5221 *	*mcm3-cdk4A-nes MCM7-GFP*	pML106 into YJL1979[d]
YJL5439 *	*mcm3-nes MCM7-GFP*	pKI1363 into YJL1979[a]
YJL5691 *	*mcm3-cdk4A mcm7::{MCM7-GFP, URA3}*	pKI1198 into YJL2717[b]
YJL5750 *	*trp1-289::{pGAL10-NLS2-NLS3NES3-cdk5A-*GFP_3*, TRP1}*	pMT103 in YJL310[c]
YJL5753 *	*trp1-289::{pGAL10-NLS2-NLS3NES3-cdk4A-*GFP_3*, TRP1}*	pML116 in YJL310[c]

*These strains are congenic to YJL310 whose full genotype is *MAT*a *ura3-52 leu2-3,112 trp1-289 bar1::LEU2*.

(a)Strain derived by two-step gene replacement using indicated plasmid

(b)Strain derived by one-step integrative tagging using indicated plasmids

(c)Strain derived by one-step integration using indicated plasmid

Supplementary Table 5. Yeast Strains (cont'd)

(d)Strain derived by two-step gene replacement such that the *GFP* portion of the plasmid was removed from the chromosome during the second step of plasmid excision.

CHAPTER 3

Regulatory evolution in proteins by turnover and lineage specific changes of cyclin dependent kinase consensus sites

Abstract

Evolutionary change in gene regulation is a key mechanism underlying the genetic component of organismal diversity. Here, we study evolution of regulation at the posttranslational level by examining the evolution of cyclin-dependent kinase (CDK) consensus phosphorylation sites in the protein subunits of the pre-replicative complex (RC). The pre-RC, an assembly of proteins formed during an early stage of DNA replication, is believed to be regulated by CDKs throughout the animals and fungi. Interestingly, although orthologous pre-RC components often contain clusters of CDK consensus sites, the positions and numbers of sites do not seem conserved. By analyzing protein sequences from both distantly and closely related species, we confirm that consensus sites can turn over rapidly even when the local cluster of sites is preserved, consistent with the notion that precise positioning of phosphorylation events is not required for regulation. We also identify evolutionary changes in the clusters of sites and further examine one replication protein, Mcm3, where a cluster of consensus sites near a nucleocytoplasmic transport signal is confined to a specific lineage. We show that the presence or absence of the cluster of sites in different species is associated with differential regulation of the transport signal. These findings suggest that the CDK regulation of MCM nuclear localization was acquired in the lineage leading to Saccharomyces cerevisiae after the divergence with Candida albicans. Our results begin to explore the dynamics of regulatory evolution at the posttranslational level and show interesting similarities to recent observations of regulatory evolution at the level of transcription.

Introduction

The contribution of regulatory evolution to biological diversity is increasingly well appreciated (1–4). The identification of changes in transcriptional regulatory proteins (5, 6) and, more frequently, the cis-elements they recognize in noncoding DNA (reviewed in ref. 7), has provided mechanistic insight into the evolution of gene regulation.

Genes are regulated at multiple levels, however. In eukaryotes, posttranslational regulation of protein activity by phosphorylation is of particular importance (8). Although little is known in general about the evolution of this type of regulation, comparative studies of posttranslational modification sites in phosphorylase (9, 10) and fructose 1-6-bisphosphatase (11) revealed that they were not conserved between homologues.

Recent studies have applied computational approaches to databases of protein sequences to perform comparative studies on larger scales. For example, targets of protein kinase A were predicted based on conservation of consensus sites between Candida albicans and Saccharomyces cerevisiae (12). Another study examined regulation of cell-cycle proteins in four species and proposed coevolution between posttranslational regulation by phosphorylation and transcriptional regulation (13).

Phosphoregulation plays a critical role in cell-cycle control (14–16). For example, it has been found in several species that after the initiation of DNA replication, to ensure that a single round of DNA replication occurs in each eukaryotic cell cycle, a subset of the

DNA replication machinery (the pre-RC) is directly inhibited by cyclin-dependent kinase (CDK) (17, 18).

Here, we examine the evolution of regulation of the pre-RC by CDKs. Several features of this system make it attractive for evolutionary analysis. First, the pre-RC proteins are found in single copy in many animals and fungi (17), so it is relatively easy to identify their orthologs in most species. Also, human CDKs have been shown to rescue yeast CDK mutations (19, 20), suggesting little change in the functional capabilities of the kinase. Finally, CDK is a proline-directed serine/threonine kinase (21) with a well defined consensus site S/T-P-X-R/K (where X is any amino acid). Evolutionary loss of the critical S/T or P is likely to preclude phosphorylation by CDK in that species.

In some cases, the specific consensus sites likely to be phosphorylated by CDK in vivo have been determined through a combination of experimental methods; we refer to these sites as “characterized.” In addition, CDK target proteins often contain multiple CDK consensus sites closely spaced in their primary amino acid sequence; we refer to these as “clusters.” Previous studies have noted that, even when clusters of characterized sites are found in orthologous pre-RC components, the individual consensus sites are not always conserved in position or number (22, 23). We refer to this as “turnover” of sites and suggest that it is consistent with regulation through mechanisms that impose loose constraints on spacing and number of phosphorylation sites (ref. 24; see Discussion).

Our analysis of evolutionary changes in CDK consensus sites in pre-RC proteins reveals examples of both turnover of characterized sites in preserved clusters and lineage-specific changes in the clusters of sites. We suggest that the CDK regulation of nuclear localization of the pre-RC component Mcm3 (25) was gained on the lineage leading to S. cerevisiae after the divergence from C. albicans, and we provide experimental support for this model.

Results

Signatures of CDK Regulation in Pre-RC Proteins.

To get a broad sense of the conservation of CDK regulation, we obtained sequences and orthologs for pre-RC proteins (see Methods) from 21 species with complete genome sequences publicly available, selected so that their phylogenetic positions were informative amongst the animals and fungi. For each protein, we identified experimentally verified CDK targets where consensus sites had been characterized ("P"s in Fig. 1 and refs. 22, 23, and 25–38) and also calculated the SLR statistic (see Methods), which measures the overrepresentation and spatial clustering of strong (S/T-P-X-R/K, where X is any amino acid) and weak (S/T-P) CDK consensus matches (Fig. 1), which we have shown to be predictive of CDK regulation (39). Because the pre-RC is expected to be regulated by CDKs in all these species, a simple expectation is that the same proteins would be targets in all species. Indeed, we find proteins that have high values of SLR across many (Orc1, Mcm4) or all (Cdc6) of the species examined, suggesting that

regulation has been preserved since a common ancestor. However, other proteins (Orc2) show less consistent patterns, while some (Orc6, Mcm2, Mcm3) show lineage-specific patterns. In these cases, the changes in statistical signal could be due to either bona fide changes in regulation or incorrect classification by our statistical method, but in at least one of these cases, we see a functional difference corresponding to the statistical difference (see below).

Turnover of Functional CDK Consensus Sites.

Even when regulation appears conserved, as has been noted in previous studies (22, 23), we found that the numbers and positions of CDK consensus sites were not always conserved. A striking example of this is the linker region of ORC1, which contains a strong cluster of CDK consensus matches in all of the animals and most of the fungi (Figs. 1 and 2 a). Sites in this region are phosphorylated by CDK in Drosophila (23) and are involved in CDK-regulated localization and degradation of ORC1 in mammalian cells (38, 40, 41). Despite the persistence of the cluster over long evolutionary distances, examination of the numbers and positions of individual CDK consensus sites (Fig. 2 a) reveals rapidly changing organization.

It is possible that this apparent turnover of sites is due simply to difficulties in comparing highly diverged amino acid sequences, or that consensus matches in clusters do not all represent functional sites and are not constrained. To rule these out, we examined the evolution of experimentally characterized consensus sites. We consider a consensus site characterized if there is some in vivo (including cell culture) evidence of phosphorylation

and/or function in a CDK-regulated process ["P" in Fig. 1 and supporting information (SI) Table 1]. We examined these sites in alignments of orthologs from closely related species (see Methods), where most residues are unchanged and we have high-confidence in multiple alignments (84%, 74%, and 64% identical for yeast, mammals, and Drosophilae, respectively).

We found in each clade that characterized consensus sites accumulated on average fewer substitutions than the flanking residues (rates were 20%, 60%, and 27% of flanking regions for yeast, mammals, and Drosophila, respectively; SI Fig. 4a). Interestingly, despite this evidence for constraint, we also found that of the 55 experimentally characterized CDK consensus sites, 9 had substitutions in the critical S/T or P of the CDK consensus in these closely related species (not conserved, Fig. 2 b). These include a previously reported nonconserved CDK site in the N terminus of mammalian CDC6 (42). We also noted five sites that changed between strong and weak consensus matches in these alignments (Fig. 2 b). Thus, microevolutionary changes in functional CDK sites provide a potential mechanism for the changes in number and positions of consensus sites observed over long evolutionary distances.

The linker region of mammalian ORC1 (boxed region in Fig. 2 a and ref. 40) provides an extreme example of this evolutionary turnover (Fig. 2 c). Of three strong and one weak characterized consensus sites (ref. 38 and Fig. 2 a and c iii, iv, vi, and viii), only one is conserved over the mammals (Fig. 2 c iv), although it is additionally modulated by alternative splicing in mouse (43). Furthermore, one of these sites appeared within the

divergence of the primates (Fig. 2 c viii). In addition to these changes in characterized sites, we noted a region containing human-specific losses of consensus matches and a human polymorphism appears to disrupt an ancestral consensus match (Fig. 2 c x).

To test for constraint in the linker region of ORC1 more formally, we computed the ratio of nonsynonymous to synonymous substitution rates (dn/ds) (see Methods) for this region, and found it to be 0.98. Consistent with the rapid changes in the consensus sites, we could not reject the hypothesis of no constraint (dn/ds = 1, $P > 0.91$, see Methods).

Such weak constraint and rapid turnover of consensus sites in the mammals is surprising given that that cluster of sites in this region of ORC1 appears to have been retained since the common ancestor of the animals (Figs. 1 and 2 a). We therefore sought to detect constraint on the cluster of sites. We reconstructed the sequence of the ancestral ORC1 linker region (see Methods) and found it to have more consensus sites and stronger clustering than the extant human sequence (5 strong, 11 weak, SLR = 9.89 vs. 3 strong, 9 weak, SLR = 4.50). We then simulated the evolution of the ancestral sequence, using a general protein model (see Methods). Constraint on the cluster of sites should lead to greater and less variable values of SLR, so we compared a composite statistic (the difference between the mean and standard deviation of the SLR over the seven species) in the real mammalian sequences to the simulations, and found it to be significantly greater (Fig. 2 e, $P < 0.005$, $n = 5,000$). These simulation results support the model that the cluster of CDK sites in ORC1 evolves under purifying selection, even though little constraint is apparent at the amino acid level.

Lineage Specific Regulatory Evolution.

In contrast to cases like ORC1 where cluster of CDK consensus sites is largely conserved, other pre-RC proteins show considerable variation in SLR across species (Fig. 1), despite the importance of proper regulation (17, 18). For example, Mcm3 is a CDK target in S. cerevisiae (35), but CDK regulation has not been reported in S. pombe or human.

Consistent with the hypothesis of lineage-specific regulation of MCM3, we find a dramatic statistical change in the clustered CDK consensus sites on the lineage leading to S. cerevisiae (Fig. 1). This change is due to a cluster of consensus sites in the C-terminal region of S. cerevisiae Mcm3 (Sc-Mcm3-CTR) that was found to be critical for the CDK-mediated shuttling of the MCM complex in and out of the nucleus in that species (25, 44, 45). Indeed, mutation of the CDK consensus sites in the Sc-Mcm3-CTR abolished its ability to confer regulated nuclear localization to a GFP reporter construct (25). Interestingly, in contrast to S. cerevisiae, the MCM protein complex is constitutively nuclear in S. pombe and human (46). We therefore decided to test whether the changes in CDK consensus sites were associated with lineage-specific changes in regulation.

We first sought to rule out that the changes in CDK consensus sites could be explained by statistical fluctuations. To do so, we obtained Mcm3 orthologs from six additional fungi to improve resolution within the Ascomycetes (see Methods). We then used maximum

parsimony to reconstruct the ancestral organization of these sequences and infer the gains and losses of CDK consensus sites along each branch (Fig. 3 a; see Methods). For the strong consensus, we inferred 13 gains in the clade containing S. cerevisiae, significantly greater than the 5.37 expected if gains were randomly distributed proportional to the evolutionary distance on each branch ($P = 0.0037$, see Methods). For the weak consensus, we inferred 13 gains in this clade, which also greater than the 10.36 expected but is not statistically significant ($P = 0.24$). These data show that gains of strong consensus matches are nonrandomly distributed along the tree and suggest that the CDK-regulated shuttling of MCMs in and out of the nucleus in S. cerevisiae is due at least in part to changes in CDK consensus sites that occurred after the divergence from C. albicans (Fig. 3 a).

This model predicts that the region homologous to the Sc-Mcm3-CTR from species outside this clade would not confer regulated localization to a GFP reporter construct. We therefore inserted the homologous region of C. albicans Mcm3 into such a construct (Fig. 3 b and c, see Methods) and tested its localization in S. cerevisiae in cells arrested in G1 (by alpha factor) or G2 (by nocodazole). Although the S. cerevisiae construct showed nuclear localization in the G1 but not the G2 arrest (Fig. 3 d, compare iv with viii), the C. albicans construct was constitutively nuclear (Fig. 3 d, compare i with v), confirming a functional difference in this region of the protein between these species. To further resolve the evolutionary events that lead to regulated localization of MCMs in S. cerevisiae, we performed similar experiments, using the C-terminal region of Mcm3 from Candida glabrata (Fig. 3 c, ii and vi) and Kluveromyces lactis (Fig. 3 c, iii and vii) and

found that these showed regulated nuclear localization, consistent with the origin of this regulation in the ancestor of the S. cerevisiae clade.

Taken together, our experiments and sequence analysis of the Mcm3 C terminus are consistent with the model that the functional CDK consensus sites that regulate nuclear localization arose in the common ancestor of the S. cerevisiae clade after divergence from C. albicans. It is important to note that the Sc-Mcm3-CTR contains other regulatory sequences (ref. 25 and SI Fig. 5a), including a Crm1-dependent export signal, which also appeared at that time, and a basic nuclear localization signal, which is shared by all ascomycetes and may be important for the observed nuclear localization of the MCMs in S. pombe (46). Consistent with this model, we identify a basic nuclear localization signal, but no leucine rich export signals in the homologous region of the C. albicans protein (SI Fig. 5 b and c). To rule out the possibility that there were cryptic export signals or CDK sites further downstream of the region we defined as homologous to the Sc-Mcm3-CTR, we also performed all of the experiments, using the entire C terminus from each species and found similar results (data not shown).

Discussion

Inhibition of the pre-RC by CDKs to prevent rereplication is an ancient feature of the eukaryotic cell cycle (17). Our results suggest that, even though this regulatory logic is preserved, its mechanistic implementation can evolve rapidly.

For example, we found that, on average, 16% (11–21% ± SE) of characterized CDK consensus sites in pre-RC components in budding yeast, human, and Drosophila are not conserved in alignments of closely related species. In ORC1, the presence of polymorphisms in the human population suggests that the resculpting of regulatory regions continues.

Traditional models of phosphoregulation invoke allosterically driven conformational changes as a consequence of phosphorylation, which presumably require modification at precise positions in the protein structure. More recent analyses of phosphoregulation suggest alternative regulatory paradigms involving multiple phosphorylation sites that do not need to be conserved (24). Clusters of multiple phosphorylation sites can modulate interactions (25, 47–49) or provide specific dynamic properties (50–52) and these mechanisms may not depend on the specific locations or numbers of sites (24).

Consistent with this model, we found statistical evidence for constraint at the level of the cluster of consensus sites in the linker region of ORC1, despite weak constraint at the amino acid level. In clusters, when new consensus sites appear via point mutations, constraints on the ancestral sites may be relaxed, allowing them to accumulate destructive substitutions. Interestingly, this stabilizing selection model was first proposed for transcriptional enhancer elements in DNA, where, despite little similarity in primary sequence, orthologous enhancers could drive similar expression patterns by preserving clusters of transcription factor binding sites (53, 54).

In addition to turnover of consensus sites in conserved clusters, we found cases of entire clusters that are not conserved over evolution. We observed lineage-specific accumulation of consensus sites in the C terminus of S. cerevisiae Mcm3, which we showed was associated with functional differences in localization of a reporter construct (Fig. 3). We also note that the C-terminal cluster of consensus sites in yeast Cdc6 (29) shows a similar pattern, appearing even more recently (SI Fig. 5d). Because CDK inhibits the pre-RC through multiple regulatory mechanisms (35), we suggest that new mechanisms may evolve without drastic negative consequences. Thus, a possible explanation for these lineage-specific changes is “regulatory network turnover” (55), in which interactions are gained and lost in the context of a preserved regulatory logic.

Finally, we note that the accretion of regulatory motifs in the Mcm3 C terminus is analogous to the evolutionary gain of transcription factor binding sites in enhancers (56). In extending this model to phosphorylation sites, we suggest that the cooption of a new target into an existing regulatory network by acquisition of motifs for preexisting, trans-acting factors is a general mechanistic basis for evolutionary increases in regulatory specificity and, perhaps, organismal complexity.

Methods

Proteins, Orthologs, and Clustering of CDK Sites.

For the animal and yeast genomes used in Fig. 1 i and iii, protein sequences and ortholog assignments were obtained from the TreeFam database (57) and Yeast Gene Order Browser (58), respectively. To assign orthologs for the species not included in these databases (Fig. 1 ii), we obtained amino acid sequences from J. Stajich (University of California, Berkeley, CA; http://fungal.genome.duke.edu). We then aligned the fungal and animal orthologs (from TreeFam or Yeast Gene Order Browser), using T-Coffee software (59), created profile-hidden Markov models, and searched the additional genomes for matches to these profiles, using the HMMer package (http://hmmer.janelia.org, using the -forward option). We took the top hit as the ortholog in each case, except for CDC6, where the top hit was the same as the top hit for ORC1 in some of the fungi, so we took the second hit. Where a protein was present in multiple copies in a species (e.g., CDC7 in S. pombe), we excluded that protein for that species from further analyses (gray box in Fig. 1). If the HMMer e-value was >0.001 or the protein was truncated relative to other orthologs, we deemed the ortholog low confidence (gray box in Fig. 1).

For each protein in each species, we computed SLR, a log likelihood ratio statistic, which measures clustering and enrichment of motifs in a sequence. Briefly, this statistic compares the likelihood of the observed motifs and their spacing under a model that

includes clusters to that under the genomic background frequency or a model, including clusters of weak sites only (for details, see ref. 39). We computed the background frequencies of these motifs in each of the genomes studied. We reported the analysis shown in Fig. 1 by using other statistical measures and found similar results (SI Fig. 6).

Alignments of Closely Related Species.

We obtained ortholog assignments and protein sequences for each of the characterized CDK targets from budding yeast in S. paradoxis, S. mikatae, and S. bayanus from SGD (60), from human in mammals from TreeFam or from Drosophila from 12 Drosophilae (V. Iyer, D. Pollard, and M. Eisen, personal communication). These were aligned with T-Coffee, and truncated orthologs were removed, except in the case of mammalian CDT1, where only the N-terminal region was available. Alignments of all of the characterized sites are available as SI Dataset 1.

To compute the dn/ds, we obtained coding DNA sequences and inserted the gaps from the protein alignments into these. For the linker region of ORC1 (which we took to be amino acids 196–470 in the human sequence), we used paml (61) to compute maximum-likelihood branch lengths with either an unknown dn/ds or dn/ds fixed at 1, assuming the phylogeny (((human,chimp),macaque),(mouse,rat)),dog,cow). We compared two times the difference in likelihoods to a $\chi 2$ distribution with one degree of freedom. Human SNPs and alternative mouse transcripts for ORC1 were obtained from Ensembl (version

41; ref. 62). We note that dn/ds for the clusters of CDK sites were higher on average than the whole proteins, with ORC1 showing the highest value (data not shown).

Simulations of Orc1 Evolution.

To obtain the distribution of the difference of the mean and standard deviation of SLR for the ORC1 linker used the following procedure. We extracted the amino acid alignment and used paml (61), using the mammalian phylogeny described above to obtain the maximum-likelihood estimates for the branch lengths (in amino acid substitutions per site) and to reconstruct the ancestral sequence. We then used the ROSE sequence evolution software (63) to simulate (with default parameters for protein evolution) along the estimated tree starting from the ancestral sequence. Finally, we computed the average and standard deviation of the SLR in the simulated sequences for the extant species.

Reconstruction of Ancestral Mcm3 CDK Matches.

Because we wanted to reconstruct the ancestral organization of CDK matches in Mcm3 over longer evolutionary distances where we were no longer confident in the alignment of individual residues, we devised the following parsimony method. First, we obtained protein predictions for six additional Ascomycete genomes (http://fungal.genome.duke.edu/), assigned orthologs as above, made a multiple alignment of the protein sequences, using T-Coffee, and used paml to obtain maximum-likelihood estimates of the branch lengths for the tree topology shown in Fig. 3 A. We

then searched the aligned sequences for matches to the CDK consensus and created an "alignment" of CDK consensus matches by treating any CDK match within five amino acid residues as another in a different species as "aligned." For Mcm3, this yielded 31 aligned "columns," where there was a match to either the strong or weak CDK consensus in at least one species. Based on this, we used the "classical parsimony" algorithm (64) to reconstruct the ancestral states, either "strong match," "weak match," or "background" and infer the number of gains and losses for strong and weak matches along each branch.

Although the current view supports the clade containing K. lactis, Ashbya gossypii, Kluveromyces waltii, and Saccharomyces kluyveri as a sister to the clade containing S. cerevisiae (65, 66) the placement of the species (Fig. 3 A) is not yet conclusively established (66). We therefore repeated the analysis using a multifurcation at this node and found similar results regarding the asymmetry, but observed variation in the estimates of CDK consensus gain and loss events on each branch (data not shown).

To calculate the expected number of gains in the Saccharomyces clade under the hypothesis of symmetrically distributed changes, we assume the number of background positions is large relative to the number of matches and that gains of matches are rare (no multiple hits). The expected number of gains in a subclade c is then Poisson with mean = $n g \times tc/t$, where t is the sum of the branch lengths (tree length), tc is the sum of the branches in the clade c, and ng is the number of gains inferred along the whole tree. To calculate the ancestral values of SLR, we reconstructed the ancestral positions of each

column of aligned matches by recursively assigning to each ancestor the average position of the matches in its children.

Construction of GFP Reporters and Localization Assays.

We obtained genomic DNA for C. albicans, C. glabrata, and K. lactis from D. Galgoczy (University of California, San Fransico) and A. Johnson (University of California, San Francisco) and for A. gossypii from A. Gladfelter (Dartmouth College, Hanover, NH). We amplified the region homologous to the Sc-Mcm3-CTR or the entire C terminus by PCR (Phusion; Finnzymes, Espoo, Finland), using primers (IDT Technologies, Coralville, IA) that introduced ClaI or EcoRI restriction sites into the 5′ or 3′ ends of the PCR product. Primer sequences are available on request. These PCR products were inserted between the ClaI and EcoRI sites in the plasmid pML104, a gal inducible TRP1 integrating plasmid containing the S. cerevisiae Mcm2 nuclear localization signal and three tandem copies of GFP (25). All constructs were confirmed by sequencing (MClab, South San Francisco, CA). Plasmids were transformed into YJL310 (67), grown, arrested and photographed as described in figures 4, 5, 6B, 8, and 9 of ref. 25. The cell-cycle arrests were confirmed by scoring the fraction budded for >60 cells for each strain under each condition. The GFP localization panels shown were "representative," and observations were confirmed by scoring the fraction showing nuclear staining for >60 cells for each construct under each condition.

Acknowledgments

We thank Dr. Dave Morgan, Dr. Seth Grant, Dr. Avril Coghlan, and Dr. Jean-Karim Hériché, and Dave Galgoczy for helpful discussions; Dave Galgoczy, Dr. Alexander Johnson, and Dr. Amy Gladfelter for genomic DNA; and Dr. Avril Coghlan for comments on the manuscript. This work was supported by National Institutes of Health Grants 5F31CA110268-03 (to M.E.L.) and R01 GM59704 (to J.J.L.). A.M.M. is a Sanger Postdoctoral Fellow. R.D. and the Wellcome Trust Sanger Institute are funded by the Wellcome Trust.

References:

1. Wilson AC, Maxson LR, Sarich VM (1974) Proc Natl Acad Sci USA 71:2843–2847.

2. Tautz D (2000) Cur r Opin Genet Dev 10:575–579.

3. Lev ine M, Tjian R (2003) Nature 424:147–151.

4. Wray GA (2003) Int J Dev Biol 47:675– 684.

5. Ronshaugen M, McGinnis N, McGinnis W (2002) Nature 415:914 –917.

6. Galant R, Carroll, SB (2002) Nature 415:910 –913.

7. Wray GA, Hahn MW, Abouheif E, Balhof f JP, Pizer M, Rockman MV, Romano LA (2003) Mol Biol Evol 20:1377–1419.

8. Johnson SA, Hunter T (2005) Nat Methods 2:17–25.

9. Palm D, Goerl R, Burger KJ (1985) Nature 313:500 –502.

10. Hwang PK, Fletterick RJ (1986) Nature 324:80 – 84.

11. Rittenhouse J, Harrsch PB, Kim JN, Marcus F (1986) J Biol Chem 261:3939 –3943.

12. Budovskaya YV, Stephan JS, Deminof f SJ, Her man PK (2005) Proc Natl Acad Sci USA 102:13933–13938.

13. Jensen LJ, Jensen, TS, de Lichtenberg U, Br unak S, Bork P (2006) Nature 443:594 –597.

14. Norbur y CJ, Nurse P (1989) Biochim Biophys Acta 989:85–95.

15. Nasmyth K (1993) Cur r Opin Cell Biol 5:166 –179.

16. Murray AW (1994) Chem Biol 1:191–195.

17. Kelly TJ, Brown GW (2000) Annu Rev Biochem 69:829 – 880.

18. Bell SP, Dutt a A (2002) Annu Rev Biochem 71:333–374.

19. Lee MG, Nurse P (1987) Nature 327:31–35.

20. Elledge SJ, Spottswood MR (1991) EMBO J 10:2653–2659.

21. Brown NR, Noble ME, Endicott JA, Johnson LN (1999) Nat Cell Biol 1:438 – 443.

22. Vas A, Mok W, Leather wood J (2001) Mol Cell Biol 21:5767–5777.

23. Remus D, Blanchette M, Rio DC, Botchan MR (2005) J Biol Chem 280:39740 – 39751.

24. Serber Z, Ferrell JE, Jr (2007) Cell 128:441– 444.

25. Liku ME, Nguyen VQ, Rosales AW, Irie K, Li JJ (2005) Mol Biol Cell 16:5026 –5039.

26. Jallepalli PV, Brown GW, Muzi-Falconi M, Tien D, Kelly TJ (1997) Genes Dev 11:2767–2779.

27. Pelizon C, Madine MA, Romanowski P, Laskey RA (2000) Genes Dev 14:2526 –2533.

28. Herbig U, Grif fith JW, Fanning E (2000) Mol Biol Cell 11:4117– 4130.

29. Perkins G, Dr ur y LS, Dif f ley JF (2001) EMBO J 20:4836 – 4845.

30. Weinreich M, Liang C, Chen HH, Stillman B (2001) Proc Natl Acad Sci USA 98:11211–11217.

31. Jiang W, Wells NJ, Hunter T (1999) Proc Natl Acad Sci USA 96:6193– 6198.

32. Kim J, Feng H, Kipreos ET (2007) Cur r Biol 17:966 –972.

33. Takeda DY, Par v in JD, Dutt a A (2005) J Biol Chem 280:23416 –23423.

34. Thomer M, May NR, Aggar wal BD, Kwok G, Calv i BR (2004) Development (Cambr idge, UK) 131:4807– 4818.

35. Nguyen VQ, Co C, Li JJ (2001) Nature 411:1068 –1073.

36. Komamura-Kohno Y, Karasawa-Shimizu K, Saitoh T, Sato M, Hanaoka F, Tanaka S, Ishimi Y (2006) FEBS J 273:1224 –1239.

37. Mont agnoli A, Valsasina B, Brotherton D, Troiani S, Rainoldi S, Tenca P, Molinari A, Santocanale C (2006) J Biol Chem 281:10281–10290.

38. Laman H, Peters G, Jones N (2001) Exp Cell Res 271:230 –237.

39. Moses AM, Heriche JK, Durbin R (2007) Genome Biol 8:R23.

40. Mendez J, Zou-Yang XH, Kim SY, Hidaka M, Tansey WP, Stillman B (2002) Mol Cell 9:481– 491.

41. Li CJ, Vassilev A, DePamphilis ML (2004) Mol Cell Biol 24:5875–5886.

42. Mailand N, Dif f ley JF (2005) Cell 122:915–926.

43. Miyake Y, Mizuno T, Yanagi K, Hanaoka F (2005) J Biol Chem 280:12643–12652.

44. Yan H, Merchant AM, Tye BK (1993) Genes Dev 7:2149 –2160.

45. Nguyen VQ, Co C, Irie K, Li JJ (2000) Cur r Biol 10:195–205.

46. Kearsey SE, Labib K (1998) Biochim Biophys Acta 1398:113–136.

47. Mimura S, Seki T, Tanaka S, Dif f ley JF (2004) Nature 431:1118 –1123.

48. Tak YS, Tanaka Y, Endo S, Kamimura Y, Araki H (2006) EMBO J 25:1987–1996.

49. Strickfaden SC, Winters MJ, Ben-Ari G, Lamson RE, Tyers M, Pr yciak PM

(2007) Cell 128:519 –531.

50. Nash P, Tang X, Orlicky S, Chen Q, Gertler FB, Mendenhall MD, Sicheri F, Pawson T, Tyers M (2001) Nature 414:51.

51. Lenz P, Swain PS (2006) Cur r Biol 16:2150 –2155.

52. Kim SY, Ferrell JE, Jr. (2007) Cell 128:1133–1145.

53. Ludwig MZ, Patel NH, Kreitman M (1998) Development (Cambr idge, UK) 125:949 –958.

54. Ludwig MZ, Bergman C, Patel NH, Kreitman M (2000) Nature 403:564 –567.

55. Gasch AP, Moses AM, Chiang DY, Fraser HB, Berardini M, Eisen MB (2004) PLoS Biol 2:e398.

56. Gompel N, Pr ud'homme B, Wittkopp PJ, Kassner VA, Carroll SB (2005) Nature 433:481– 487.

57. Li H, Coghlan A, Ruan J, Coin LJ, Heriche JK, Osmotherly L, Li R, Liu T, Zhang Z, Bolund L, et al. (2006) Nucleic Acids Res 34:D572–D880.

58. By rne KP, Wolfe KH (2005) Genome Res 15:1456 –1461.

59. Notredame C, Higgins DG, Heringa J (2000) J Mol Biol 302:205–217.

60. Cherr y JM, Adler C, Ball C, Cher v itz SA, Dwight SS, Hester ET, Jia Y, Juv ik G, Roe T, Schroeder M, et al. (1998) Nucleic Acids Res 26:73– 80.

61. Yang Z (1997) Comput Appl Biosci 13:555–556.

62. Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuf f J, Cur wen V, Down T, et al . (2002) Nucleic Acids Res 30(1):38 – 41.

63. Stoye J, Evers D, Meyer F (1998) Bioinfor matics 14:157–163.

64. Durbin R, Eddy S, Krogh A, Mitchison G (1998) Biological Sequence Analysis:

Probabalistic Models of Proteins and Nucleic Acids (Cambridge Univ Press, Cambridge, UK).

65. James T Y, Kauf f F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, et al . (2006) Nature 443:818 – 822.

66. Fitzpatrick DA, Logue ME, St ajich JE, Butler G (2006) BMC Evol Biol 6:99.

67. Det weiler CS, Li JJ (1998) Proc Natl Acad Sci USA 95:2384 –2389.

68. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL (2000) Nucleic Acids Res 28:263–266.

Fig. 1.

Enrichment and clustering of CDK consensus motifs in pre-RC proteins from diverse animals and fungi. Each row represents the value of the SLR statistic for the protein indicated on the left in the species indicated above the column. Gray boxes represent cases where a confident ortholog could not be identified, there was no single ortholog in that species, or the ortholog was truncated. Orthologs in i were taken from TreeFam, orthologs in ii where assigned as described in Methods, and orthologs in iii were taken from the Yeast Gene Order Browser. "P"s indicate CDK targets where consensus sites have been characterized. Diagonal bars indicate a species boundary across which reliable sequence alignments were not possible. H. sap, Homo sapiens; M. mus, Mus musculus; M. dom, Monodelphis domestica; G. gal, Gallus gallus; X. tro, Xenopus tropicalis; F. rub, Takifugu rubripes; C. int, Ciona intestinalis; D. mel, Drosophila melanogaster; C. ele, Caenorhabditis elegans; R. ory, Rhizopus orysae; U. may, Ustilago maydis; S. pom, Schizosaccharomyces pombe; N. cra, Neurospora crassa; Y. lip, Yarrowia lipolytica; C. alb, C. albicans; K. lac, K. lactis; A. gos, A. gossypii; K. wal, K. waltii; C. gla, C. glabrata; S. cas, Saccharomyces castelii; S. cer, S. cerevisiae.

Fig. 2.

Turnover of CDK consensus sites. (a) Schematic view of ORC1 orthologs. Black and gray ticks represent matches to the strong and weak CDK consensus, respectively. iii, iv, vi, and viii indicate the characterized CDK consensus sites. Thickened regions of the

sequences represent BAH and AAA pfam domains (68), respectively. Boxed region indicates the human linker region. (b) Percentage of characterized CDK consensus sites that are either not conserved (gray) or change between strong and weak consensus (white) in alignments of closely related species. (c) Alignments of seven mammals for consensus sites in the linker region of ORC1. iii, iv, vi, and vii are experimentally characterized sites (38). Mouse-A and Mouse-B indicate alternative transcripts for the mouse gene. Text above the human sequence in x indicates polymorphisms within the human population. Black and gray boxes indicate matches to the strong and weak CDK consensus, respectively; numbers are as in a. (d) Comparison of the observed value (dotted trace) of a composite statistic to the distribution obtained from simulations indicates constraint at the level of the cluster of sites. See Results for details.

Fig. 3.

Lineage-specific evolution in the C terminus of Mcm3. (a) Phylogenetic tree relating 16 Ascomycete fungi, with maximum-likelihood branch lengths in amino acid substitutions per site. Blue or cyan plus signs and minus signs above or below the branches represent inferred gain and loss events along that branch for strong or weak CDK consensus matches. Numbers are the values of SLR computed by using the extant sequences at leaf nodes, or using ancestral reconstructions at internal nodes, colored as in Fig. 1. (b) Schematic view of a multiple alignment (created by using T-Coffee) of the C-terminal region of Mcm3. Blue and cyan symbols represent matches to the strong and weak CDK consensus, respectively. Gray regions indicate gaps in the aligned sequences, and red

boxes indicate regions inserted into the reporter construct. Scale indicates the position in the multiple sequence alignment. (c) Schematic of the GFP reporter construct used in d. See ref. 25 for details. (d) GFP localization assays show nuclear localization of the C. albicans reporter construct in both alpha factor (i) and nocodazole arrested (v) cells. This is in contrast to CDK-regulated localization of the S. cerevisiae, K. lactis (ii and vi), and C. glabrata (iii and vii) constructs, which show nuclear localization in alpha factor but not nocodazole arrested cells and therefore suggest evolution of CDK regulation since the divergence of C. albicans. M., Magnaporthe; H., Histoplasma; C., Candida; D., Debaryomyces; S. klu, S. kluyveri. Other abbreviations are as in Fig. 1.

Figure 1

Figure 2

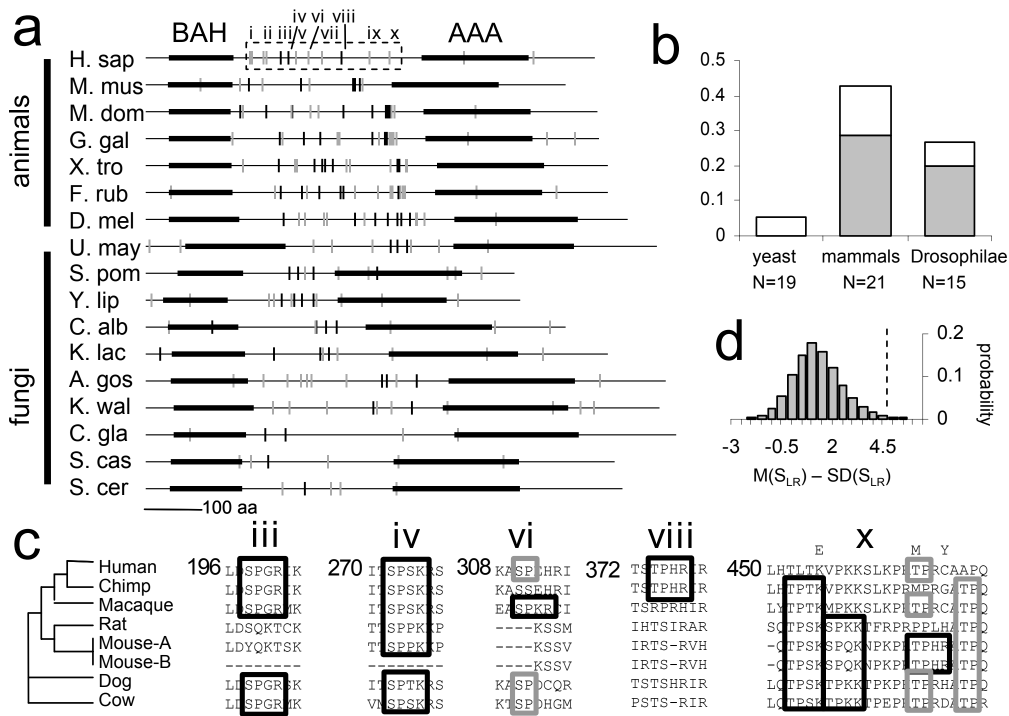
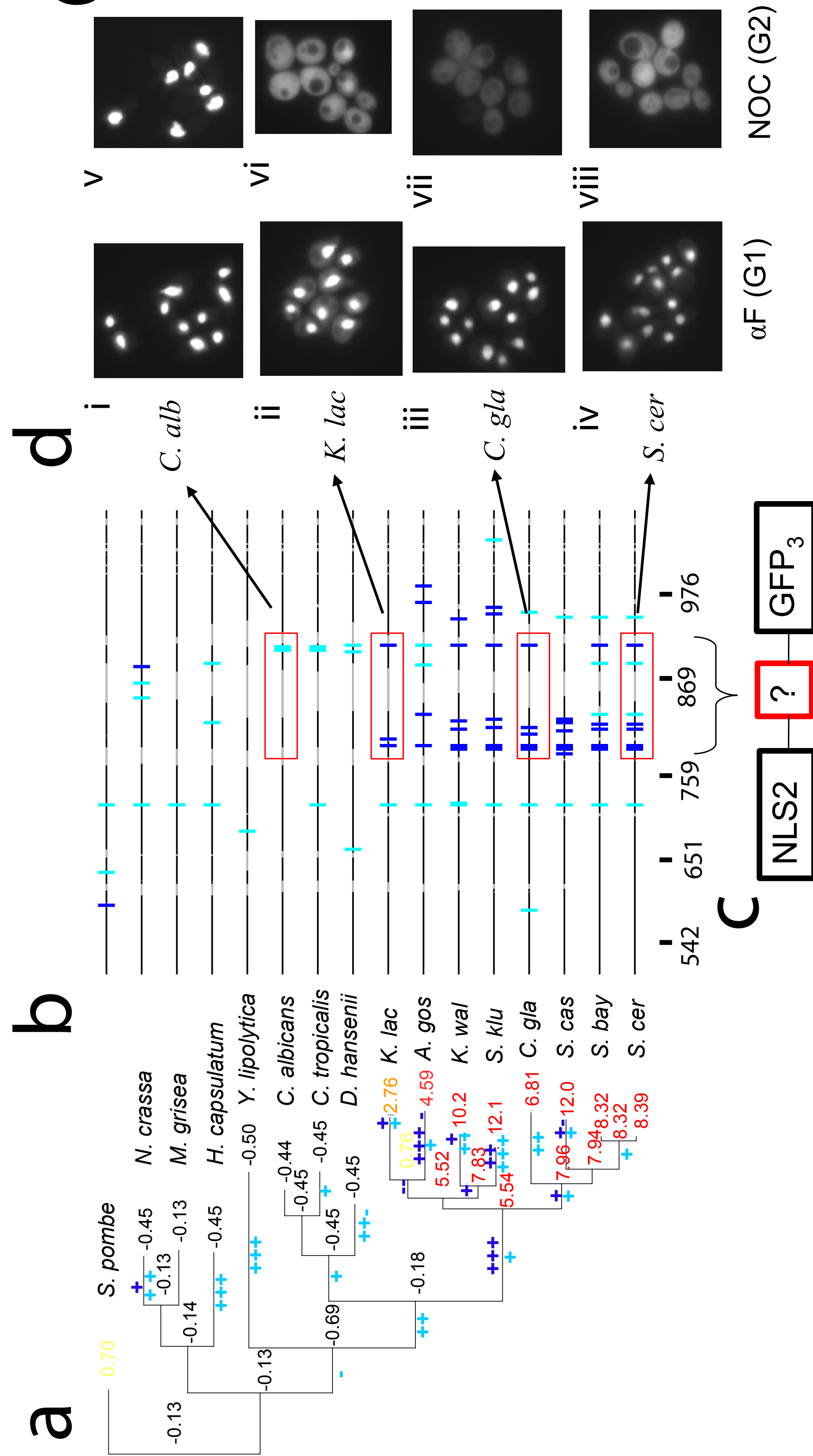


Figure 3



Supplementary Figures

SI Figure. 4. Rates of evolution in CDK consensus sequences. (a) The rate of evolution in characterized CDK sites (gray bars) or up to five flanking residues that do not contain another match to the CDK consensus (unfilled bars) for sensu stricto Saccharomyces, mammals or Drosophilae. (b) The rates of evolution in matches to S/T-P or S/T-X, where X represents any amino acid. Error bars represent twice the standard error.

SI Figure 5. Alignments of preRC-regulatory sequences. Boxes correspond to Mcm2 NLS (a), Mcm3 NLS (b), and Mcm3 NES (c) as defined in ref. 1 or recent evolution of the C-terminal CDK sites and SCF motif of CDC6 (d) as defined in ref. 2. Bold letters indicate associated strong matches to the CDK consensus. Abbreviations are as in Figs. 1 and 3, but trees represent the topology only. Sequences were manually adjusted based on a T-Coffee alignment. Red text in a indicates a nonconservative change in the A. gossypii NLS2. We note that our experiments were done in the context of a reporter that contains the S. cerevisiae Mcm2 NLS (1), which is necessary for efficient localization of the MCM complex. Changes in this sequence in other species could affect our results. Therefore, we cannot formally rule out the possibility that the C-terminal region of C. albicans protein does mediate CDK regulated export, but cannot work properly in our reporter context. Because the C. albicans, K. lactis and C. glabrata Mcm2 NLSs appear very similar to the S. cerevsiae Mcm2 NLS we consider this possibility very unlikely. However, we also tested the corresponding region from A. gossypii, and observed constitutively nuclear localization (data not shown). Because the C terminus of the A.

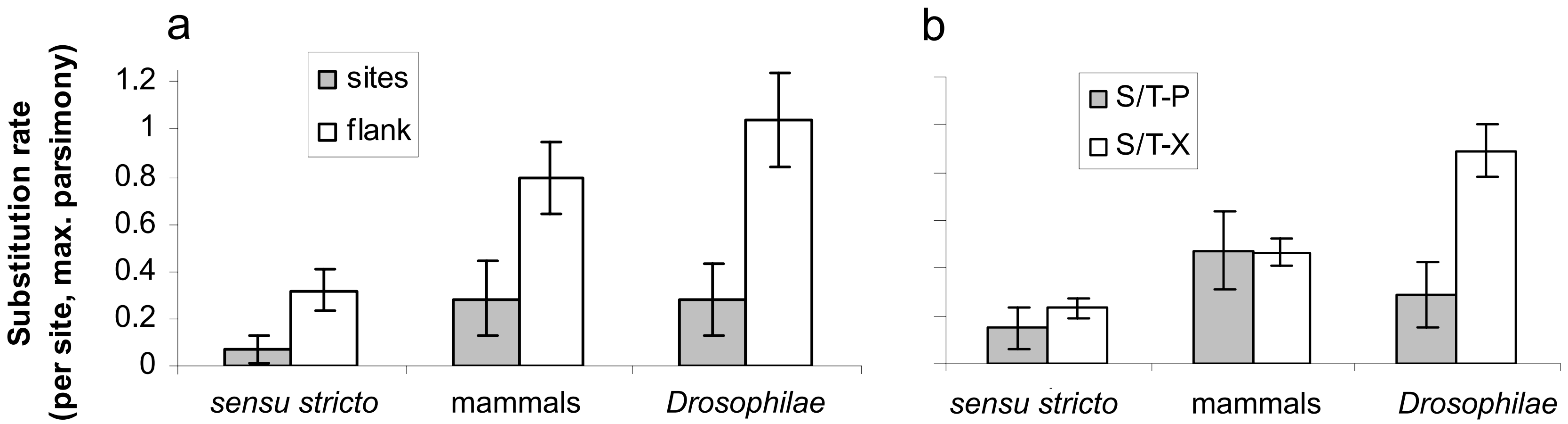
gossypii Mcm3 also contains a markedly different configuration of CDK sites, further experiments will be necessary to resolve the regulation of Mcm3 in that species. In d, the black bar indicates sensu stricto Saccharomyces, and the blue or cyan plus signs above or below the branch corresponds to the gain of a strong or weak CDK consensus sequence in this region.

1. Liku ME, Nguyen VQ, Rosales AW, Irie K, Li JJ (2005) Mol Biol Cell 16:5026-5039.

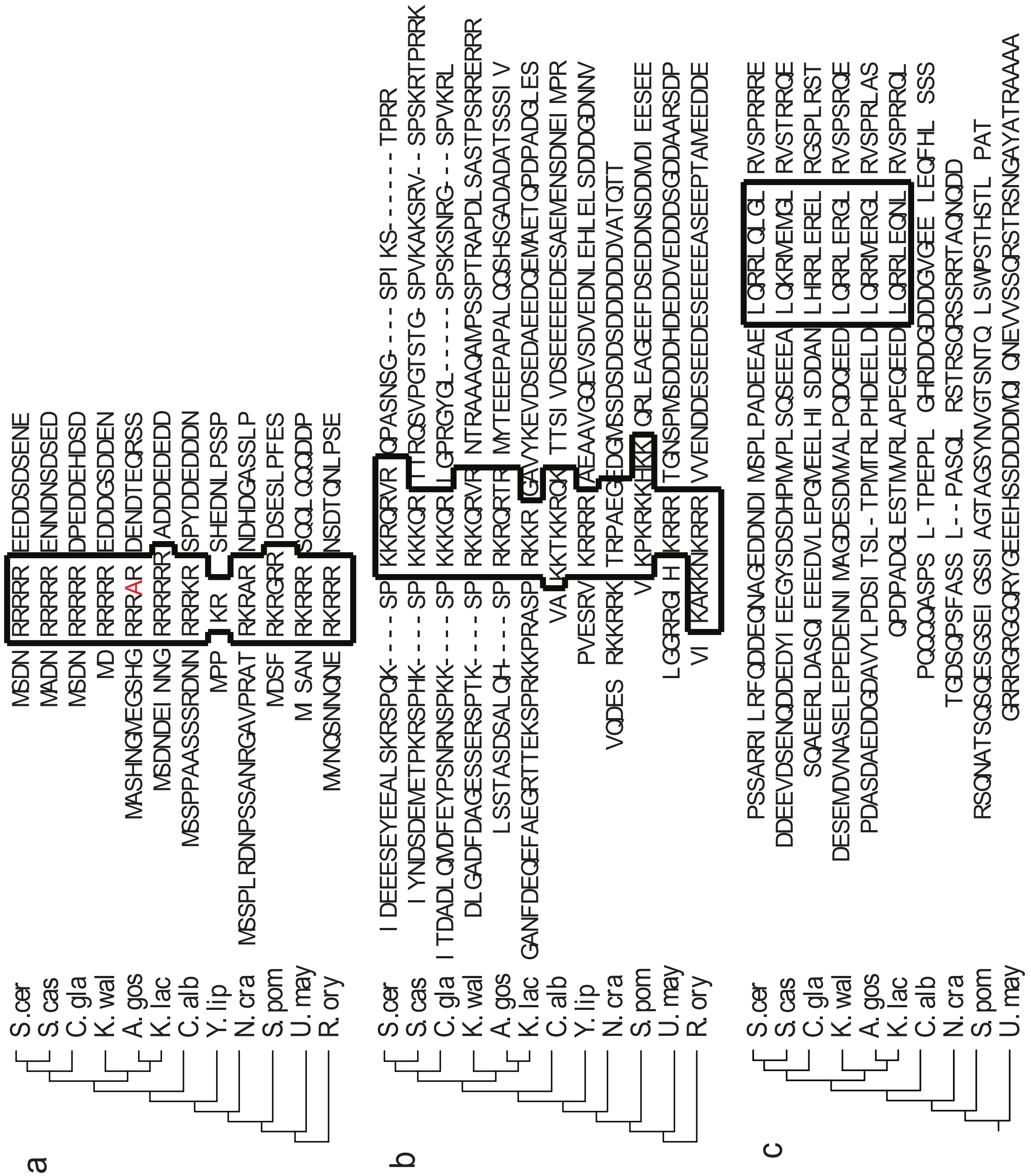
2. Perkins G, Drury LS, Diffley JF (2001) EMBO J 20:4836-4845.

SI Figure 6. Signatures of CDK regulation. (a) Each row represents the value of the SBN statistic for the protein indicated on the left in the species indicated above the column. (b) Each row represents the number of strong matches to the CDK consensus for the protein indicated on the left in the species indicated above the column. (c) Each row represents the value of the SLR statistic for the protein indicated on the left in the species indicated above the column, but using scrambled versions of the consensus sequences, P-R/K-X-S/T and P-S/T, where X represents any amino acid. Orthologs are as in Fig. 1 of the manuscript.

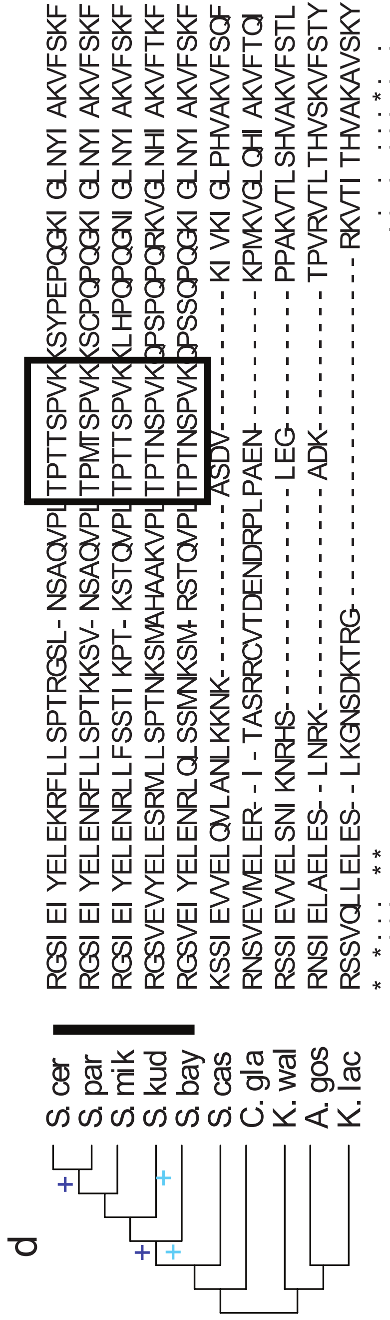
SI Figure 4



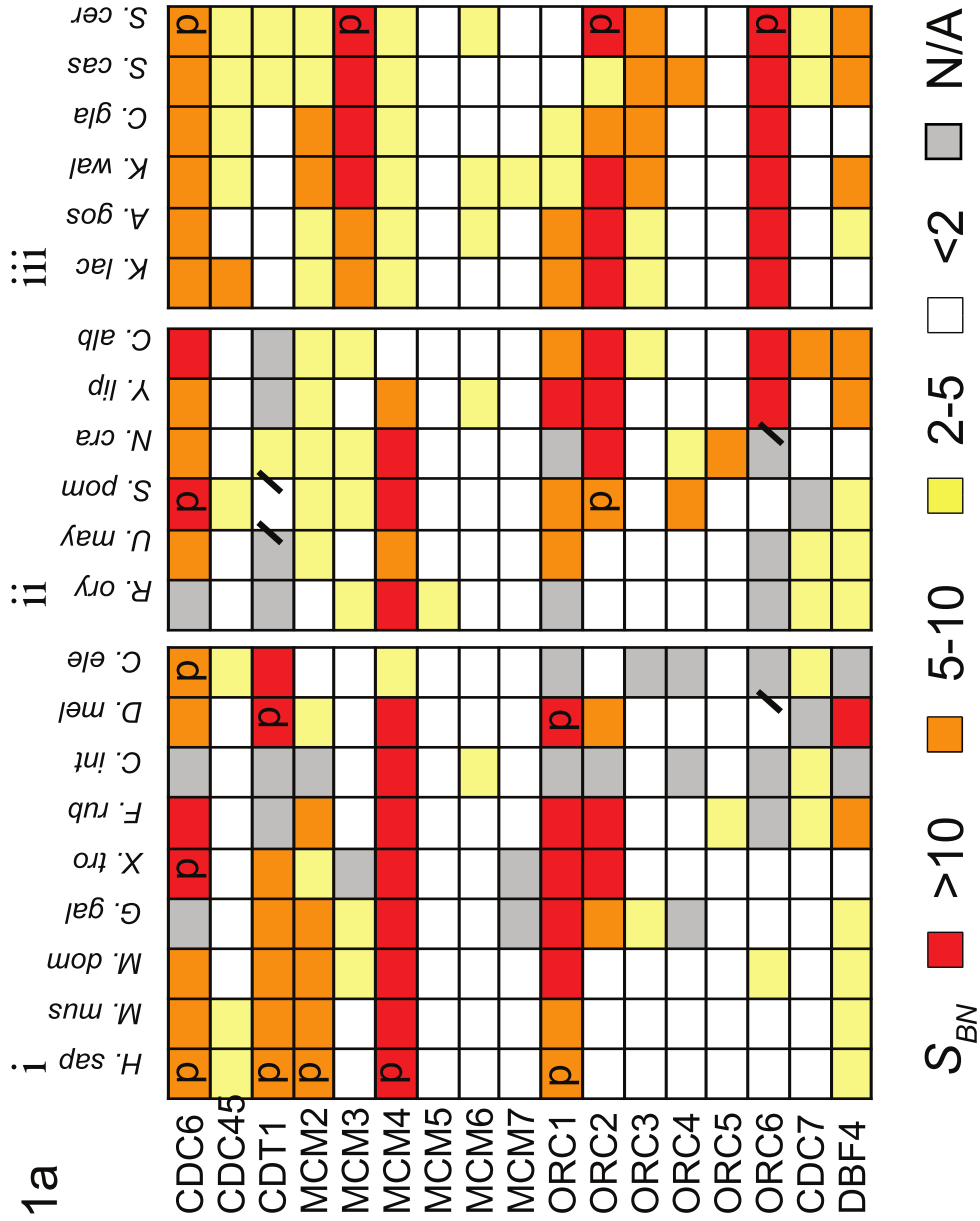
SI Figure 5a



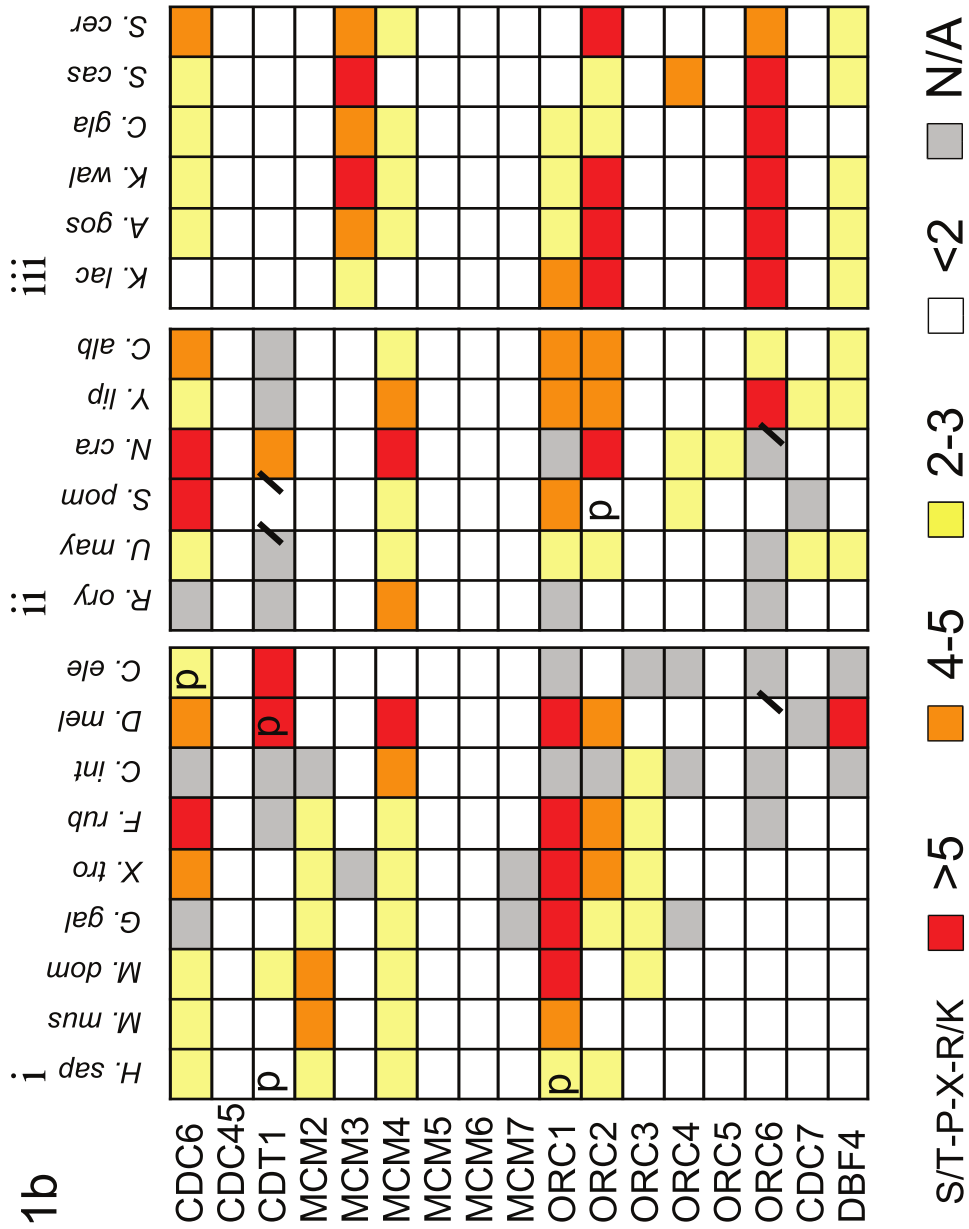
SI Figure 5b



SI Figure 6a



SI Figure 6b



SI Figure 6c

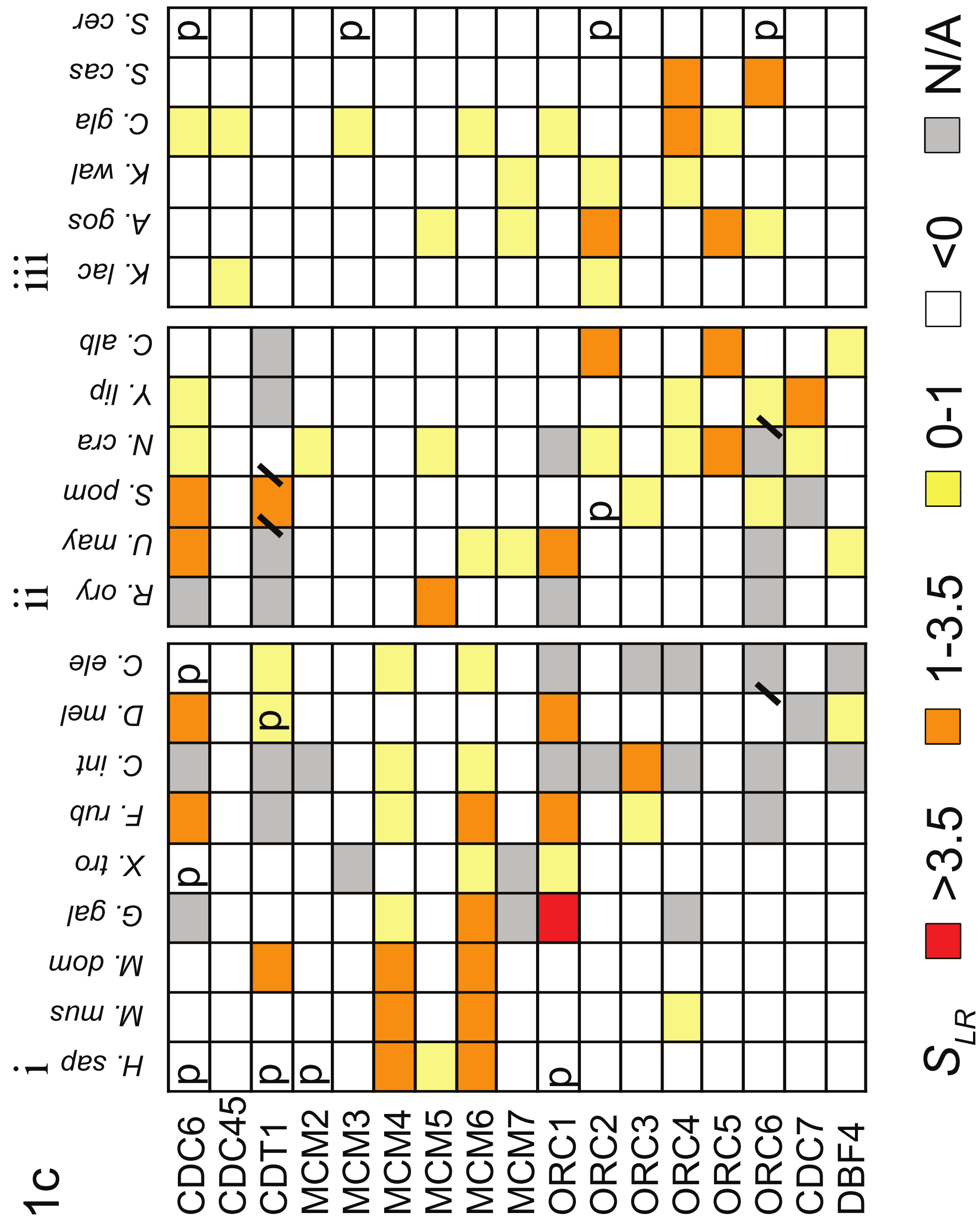


Table 1. The 55 characterized CDK sites used in Fig. 2*b* and SI Fig. 4*a*

Protein	Clade	Position (aligned)	Sequence	Evidence type	Ref.	Notes on evolution
CDC6	Mammals	45 (71)	SP	SSM	28	Not in dog
		54 (82)	SPRK	PPM	31	Conserved
		67 (95)	TP	SSM	28	Conserved
		74 (102)	SPPK	PPM	31	Conserved
		106 (134)	SPSK	PPM	31	Lost in Rat
CDC6	*Saccharomyces*	7 (6)	TPTK	SSM	30	Conserved
		23 (22)	TPPR	SSM	30	Conserved
		43 (42)	SPEK	SSM	30	Conserved
		372 (371)	SPVR	SSM	29	Conserved
CDT1	Mammals	29 (29)	TP	PPM-MS	33	Conserved
		31 (31)	SPAR	PPM-MS	33	Weak consensus in

						other mammals
CDT1	*Drosophila*	37 (36)	SP	SSM	34	Appears in *melanogaster* subgroup
		111 (114)	SPRK	SSM	34	Conserved
		158 (163)	TPTK	SSM	34	Conserved
		168 (173)	SPIK	SSM	34	Conserved
		226 (236)	SPSK	SSM	34	Conserved
		249 (263)	SPLK	SSM	34	Conserved
		256 (270)	TPTK	SSM	34	Conserved
		264 (279)	TPDK	SSM	34	Conserved
		285 (300)	SPAK	SSM	34	Conserved
		291 (306)	SPVK	SSM	34	Conserved
MCM2	Mammals	13 (12)	SP	PSA	37	Strong consensus in

						mouse
		27 (27)	SPGR	PSA	37	Conserved
		41 (42)	SPGR	PSA	37	Conserved
MCM3	*Saccharomyces*	761 (760)	SPHR	SSM	25	Conserved
		765 (764)	SPKK	SSM	25	Conserved
		781 (780)	SPIK	SSM	25	Conserved
		786 (785)	TPRR	SSM	25	Conserved
		845 (844)	SPRR	SSM	25	Conserved
MCM4	Mammals	3 (2)	SP	PSA	36	Conserved
		7 (6)	TPSR	PSA	36	Conserved
		19 (18)	TP	PSA	36	Not in cow
		32 (32)	SP	PSA	36	Not aligned in cow, but present

		54 (54)	SP	PSA	36	Conserved
		88 (88)	SP	PSA	36	Conserved
		110 (110)	TPVR	PSA	36	Conserved
ORC1	Mammals	258 (262)	SPGR	SSM	38	Lost in mouse, rat
		273 (277)	SPSK	SSM	38	Conserved
		311 (316)	SP	SSM	38	Strong consensus in macaque, lost in chimp
	Mammals	375 (380)	TPHR	SSM	38	Gained in human-chimp
ORC1	*Drosophila*	293 (314)	SP	PPM-MS	23	Conserved
		481 (516)	SPKK	PPM-MS	23	weak consensus in *D. pseudoobscura* and *D. persimilis*;

						lost in *D. virilis*
		516 (556)	TP	PPM-MS	23	Appears twice, lost once
		519 (559)	TP	PPM-MS	23	Conserved
		533 (573)	SP	PPM-MS	23	Conserved
ORC2	*Saccharomyces*	16 (15)	SPAK	SSM	35	Conserved
		24 (23)	TPKR	SSM	35	Weak consensus in *S. bayanus*
		70 (69)	TPSK	SSM	35	Conserved
		174 (174)	TPSK	SSM	35	Conserved
		188 (188)	SPLK	SSM	35	Conserved
		206 (206)	SPGK	SSM	35	Conserved
ORC6	*Saccharomyces*	104 (105)	SPMK	SSM	35	Conserved

	114 (115)	SPKK	SSM	35	Conserved
	121 (122)	SPVK	SSM	35	Conserved
	144 (145)	TPTK	SSM	35	Conserved

Sites characterized in *Saccharomyces* are characterized in *S. cerevisiae*, sites characterized in mammals are characterized in human cell lines, and sites characterized in *Drosophila* are characterized in *D. melanogaster*.

Positions in parenthesis indicate the positions in the accompanying alignments and are 0-based. Evidence types are: SSM, site specific mutagenesis; PPM, phospho-peptide mapping; PPM-MS, phospho-peptide mapping using mass spectrometry; PSA, phospho-specific antibodies. Reference numbers refer to the references in the main manuscript.

Chapter 4

CDK regulation of DNA Polymerase Alpha Primase

Introduction

As discussed in Chapter 1, we had reasons to suspect that CDK may have targets other than those discovered in the OMC strain (V Q Nguyen, Co & J J Li 2001). Briefly, first the level of re-replication observed in the OMC strain was less than doubled the G2/M DNA content. Secondly, others had discovered additional mechanisms targeting ORC6, CDC6 that promote re-replication (Green et al. 2006a; Mimura et al. 2004; Wilmes et al. 2004). The re-replication observed when assayed by microarray CGH indicates that the limitation of re-replication appears to be in part impaired elongating forks (Morreale & Joachim J Li 2008). More importantly, the microarray shows that not all origins that initiate in S-phase reinitiate and some origins exhibit inefficient reinitiation (Morreale & Joachim J Li 2008). Consequently, we hypothesized there maybe additional targets that are inhibited by the CDK to prevent reinitiation. We aimed to take a candidate approach. We considered replication proteins that are part of the initiation process containing consensus phosphorylation sites and were phosphorylated *in vitro* in a proteomic screen for CDK substrates (Ubersax et al. 2003). The consideration led us to focus our efforts on Polymerase alpha-primase complex (pol alpha).

DNA replication initiates at specific sites, origins, and elongation proceeds in a bidirectional manner. Nascent strand is synthesized by DNA polymerases. All DNA replicative polymerases require a primer on the template DNA in order to synthesize the nascent strand. These replicative polymerases are dependent on a priming enzyme to synthesize the primer. In bacteria this activity is provided by the primase enzyme dnaG,

which synthesizes an RNA primer that is less than 12nt (Baker & S P Bell 1998). In eukaryotes on the other hand the primase, Polymerase alpha-primase (pol alpha primase), synthesizes an RNA and DNA hybrid primer of up to 300nt long (Baker & S P Bell 1998). In Eukaryotes, synthesis of nascent strand on the lagging strand is performed by polymerase delta which requires repeated de novo synthesis of primers by Pol alpha primase (Nick McElhinny et al. 2008). Synthesis of nascent strand on the leading strand requires polymerase epsilon (Pursell et al. 2007). Unlike the other two major DNA replicative polymerases, Pol alpha primase does not have an intrinsic 3'-5' exonuclease activity and could incorporate incorrect nucleotides so during okazaki fragment maturation this primer is displaced (Shcherbakova, Bebenek & Kunkel 2003).

In yeast the Pol alpha primase complex is composed of four essential subunits *POL1*, *POL12*, *PRI1* and *PRI2*. The Pol1 or Cdc17 protein is approximately 180kDa and houses the DNA polymerase acitivity. Pol12 protein is approximately 86kDa and has no intrinsic enzymatic activity but is thought to play a regulatory role (Foiani et al. 1994; Foiani et al. 1995). The Pri1 protein is 48kDa and retains the catalytic primase activity (Santocanale et al. 1993). The Pri2 protein is 58kDa and is critical for bridging the interaction between Pri1 and Pol1 (S. B. Biswas et al. 2003). This critical role in replication initiation makes Pol alpha primase a choice target for regulation.

Consistent with this central role of Pol alpha primase in initiation, a number of proteins have been identified that are important in the recruitment or stability of Pol alpha-primase during DNA replication. First, in Saccharomyces cerevisiae, Mcm10 is thought to play a

chaperone role stabilizing Pol1 in the Pol alpha primase complex (Ricke & Bielinsky 2004). The HSP10 like domain of Mcm10 is thought to play the key role in this stabilizing interaction (Ricke & Bielinsky 2006). There is some evidence suggesting that in other species Cdc45 may be the protein that recruits the Pol alpha primase complex to the origin of DNA replication (Ricke & Bielinsky 2004). Another factor thought to play a recruiting role is Ctf4 (Miles & T Formosa 1992; Zhu et al. 2007). It has been reported that factors associated with chromatin modifications, yFACT components Spt16 and Pob3 along with nuclease/helicase Dna2 are thought to interact with Pol1 subunit of the Pol alpha primase complex (Tim Formosa et al. 2002; Gambus et al. 2006; Wittmeyer & T Formosa 1997; Wittmeyer, Joss & T Formosa 1999; T Formosa & Nittis 1999; Zhou & T. S. Wang 2004). The association of Pol alpha primase to origins is cell cycle regulated as assayed by ChIP studies and this association with origins is delayed in S-phase relative to late origins and occurs around the time origins initiate DNA replication (O M Aparicio, Stout & S P Bell 1999). RPA, Mcm10, Dpb11 have all been demonstrated to be critical for the association of Pol alpha primase to origins (T. Tanaka & Nasmyth 1998; Ricke & Bielinsky 2004; H Masumoto, Sugino & H Araki 2000). Lastly the origin recruitment as detected by ChIP is dependent on S-CDK activation, in part through phosphorylation of polymerase recruiting proteins, such as Sld2 and Sld3 (Zegerman & John F X Diffley 2007; S. Tanaka et al. 2007). In fact a candidate homolog of Sld2, Xenopus RecQ4, is required for chromatin binding of Pol alpha primase (Matsuno et al. 2006). Thus Pol alpha primase is indirectly activated by S-CDK phosphorylation.

In contrast to this activation by CDKs, several results raise the possibility that Pol alpha primase may also be a direct inhibitory target of S-CDKs in the block to re-replication. First, several observations indicate that S-CDKs directly phosphorylate the two largest subunits of the Pol alpha primase complex. Both Pol1 and Pol12 contain multiple conserved CDK consensus phosphorylation sites and the laboratory of David Morgan has shown that these proteins are indeed good substrates of S- and M- CDKs *in vitro* (Ubersax et al. 2003; Loog & Morgan 2005). In fact, Pol12 has been shown to be phosphorylated *in vivo* in a cell cycle dependent manner, with the hypophosphorylated form being present primarily during G1 phase when CDKs are not active (Foiani et al. 1995). More recently, mass spectrometry has detected *in vivo* phosphorylation of a subset of CDK consensus phosphorylation sites on Pol1 and Pol12, with some of these phosphorylations dependent on CDK activity (Liam Holt & David Morgan unpublished data). As will be discussed in the results, mutating all the consensus CDK phosphorylation sites on Pol1 and Pol12 has little if any effect on replication or viability. This observation suggests that CDK phosphorylation of Pol alpha primase is not required for the activation of the complex during replication initiation.

In contrast, several observations raise the possibility that CDK phosphorylation of Pol alpha primase may inhibit its function thereby helping to inhibit reinitiation and re-replication. Precedence for such an inhibitory role can be seen during the SV40 viral DNA replication in human cells. In that setting, CDK phosphorylation of the human orthologs of Pol1 and Pol12 inhibits the initiation of SV40 DNA replication. These phosphorylation events do not inhibit the intrinsic activity of Pol alpha primase but rather

its recruitment to the SV40 origin of replication through interaction with the viral initiator protein, T antigen. Interestingly, there appears to be both inhibitory (Pol12 and Pol1) and stimulatory (Pol1) phosphorylation by Cyclin A/Cdk2 (Schub et al. 2001; Voitenleitner, Fanning & Nasheuer 1997; Voitenleitner et al. 1999).

In budding yeast, Pol alpha primase displays a cell cycle dependent and CDK-inhibited association with chromatin. Consistent with CDK inhibition, this association occurs primarily in G1 phase, declines during S phase, and is absent in G2/M phase (Desdouets et al. 1998). The speculation is that the G1 association may be an important and essential early recruitment step for Pol alpha primase that is blocked by CDKs to prevent reinitiation of replication. The exact nature of the G1 association remains uncharacterized, but is presumably distinct from the origin association detected by ChIP as the latter requires CDK activity and entry into S phase (O M Aparicio, Stout & S P Bell 1999). *In* the simplest model, the CDK-inhibited G1 association of Pol alpha primase with chromatin is a prerequisite for the CDK-activated S phase association of Pol alpha primase with origins detected by ChIP. Interestingly, the Pol alpha primase G1 chromatin association is independent of Cdc6, raising the possibility that it is mediated by a direct interaction with ORC. Supporting this notion are reports of a direct interaction between ORC and Pol alpha primase in *S. pombe* (Uchiyama & T. S. Wang 2004).

The proposed CDK inhibition of Pol alpha primase can readily fit into the current paradigm for CDK regulation of replication initiation that was discussed in Chapter One. In this paradigm, replication initiation is divided into two mutually exclusive stages. In

the first stage, pre-RCs are assembled at origins in the absence of CDK activity. In the second stage, CDK activity triggers initiation at pre-RCs and simultaneously blocks any reassembly of pre-RCs. A more expansive definition of the pre-RC to include any proteins that assemble at the origin in G1 phase in preparation for initiation would allow Pol alpha primase to be considered a pre-RC component that is inhibited by CDKs.

Finally, a further layer of inhibitory regulation of Pol alpha primase is suggested by its curious cell cycle regulated localization. During G1 and S-phase the protein appears to be distributed homogenously in the nucleoplasm, however at the end of S-phase it is enriched at the nuclear periphery (Sundin et al. 2004). Such a localization could conceivable discourage re-replication by sequestering Pol alpha-primase away from the DNA origins. Thus, we have hypothesized that once Pol alpha primase completes its function in the first round of replication, CDK phosphorylation of Pol1 and/or Pol12 could both inhibit chromatin association of Pol alpha primase and promote its sequestration at the nuclear periphery.

The key to testing this hypothesis is to prevent CDK phosphorylation of Pol1 and Pol12 by mutating their CDK phosphorylation sites and seeing whether the proposed regulations are disrupted. Furthermore, if these regulations were necessary for the block to re-replication, these mutations should predispose cell to re-replication. What we found is that mutating Pol12 cdk consensus phosphorylation sites (*pol12-cdk12A*) abolishes the cell cycle dependent mobility shift of Pol12. However it didn't change the nuclear periphery enrichment in mitosis. Moreover, mutating Pol1 cdk consensus

phosphorylation sites (*pol11-cdk13A*) disrupted the nuclear periphery enrichment. However, neither the Pol12 nor Pol1mutations alone or combined caused chromatin association in mitosis or enhanced the re-replication observed beyond that seen for the OMC (*orc2-6A orc6-4A MCM7-2NLS pGAL-delntcdc6*)strain. Nonetheless, there is ample reason to suspect that it might be a target of the CDK to prevent re-replication and these reasons will be discussed in the conclusion.

Methods

Mutagenesis of Pol1 and Pol12.

All S/T-P mutated to A-P Genes synthesized with most mutations and silent restriction sites introduced as highlighted in Figure 1. Final constructs were fully sequenced

Plasmids and Strains

All plasmids are described in Table 1, all strains are described in Table 2.

Yeast media, growth and arrest

Cells were grown in YEP, synthetic complete (SC), or synthetic (S broth) medium (Guthrie and Fink, 1990) supplemented with 2% dextrose (wt/vol), 2% galactose

(wt/vol), 3% raffinose (wt/vol), or 3% raffinose (wt/vol) + 0.05% dextrose (wt/vol). The *GAL1* promoter (*pGAL1*) was induced by addition of 2% galactose and the *MET3* promoter (*pMET3*) was repressed by the addition of 2 mM methionine.

To obtain reproducible induction of re-replication, cells (YJL3249 and YJL7261) from plates were inoculated into SDC/-MUT (synthetic media with dextrose but lacking methionine, Uracil, Tryptophan) and grown for 8 hours to an OD_{600} of less than 1(it is important that cultures not get saturated otherwise re-replication will be inefficient). This starter culture was diluted into SRaff/-MUT+0.05% dextrose (it may sometime be better to make dextrose 1%--nevertheless definitely should be tweaked further) and grown overnight to reach an OD of 0.5 the following morning; after 12-15hours of growth. All experiments were performed at 30°C except where noted. For induction of re-replication in G2/M phase, cells were grown overnight (12-15 h) in SRaffC-Met,Ura,Trp + 0.05% dextrose (it may sometime be better to make dextrose 1%--nevertheless definitely should be tweaked further) were pelleted and resuspended in YEPRaff + 2 mM methionine and 15 µg/ml nocodazole. Once arrested (>90% large budded cells), galactose was added to a final concentration of 2%. The DNA content was monitered by collecting samples for FACS and budding index every hour after addition of galactose.

Flow cytometry

Cells were fixed and stained with 1 µM Sytox Green (Molecular Probes, Eugene, OR) as previously described (Haase & Reed 2002).

Chromatin Association

Chromatin Association was performed as previously described (Desdouets et al. 1998) with the following modifications. Grew 2x50mL cultures overnight and added **5λ** of 0.5mg/ml α-factor to 1 culture and **75λ** of 10mg/ml Nocodazole to 2nd culture. Cells were grown to OD_{600} of 0.4 in 50mL and Nocodazole was added at15μg/mL and incubated at least 1.5X doubling time. The cultures were harvested at 2.2krpm for 5min in a Beckman Tabletop Centrifuge (BTTC).

The cell pellet (approximately 20 OD_{600} units) was resuspended in 5mL Pipes+DTT Buffer (100mM Pipes KOH pH9.4, 10mM DTT) and incubated at 30˚C for 10 min with shaking at 100 rpms in a water bath shaker (Gyrotory Water Bath Shaker—New Brunswick Scientific Co. Inc). Next the suspension was centrifuged at 2.2Krpm in a BTTC for 5min after which the pellet was resuspended in 2mL (30 OD_{600}) of 0.6M Sorbitol Tris Buffer (0.6M Sorbitol, 1X YEPD(2% glucose), 25mM Tris HCl pH7.5). To this suspension 100μL of 4966.5U/mL Lyticase (Sigma L2524 4.3mg solid/2mL) was added and incubated at 30°C for 32 min with gentle shaking (165rpm). At this stage spheroplasting was monitored (the OD should be less than 10% of pre-lyticase treatment if not then the time was extended. Alternatively check for lack of cells or cell ghosts shells). Once greater than 95% of cells had been spheroplasted the suspensions were spun down in BTTC at 2Krpm for 5min. Then the pellets were washed with 5mL of 0.7M Sorbitol Tris Buffer (0.7M Sorbitol, 1X YEPD (2% glucose), 25mM Tris HCl pH7.5). After the wash step, the pellets were resuspended in 5mL 0.7M Sorbitol Tris Buffer and

incubated for 20 min at 30°C with gentle shaking (165rpm). Next the samples were precipitated by centrifugation at 2Krpm in BTTC; afterwards the pellet was washed 3X with 1mL cold lysis buffer 3.2Krpm 1min each in a microcentrifuge (Eppendorf 5417C) in cold room (wide bore p1000 tips were used to resuspend spheroplasted cells). Next the pellet was resuspended in 300 µL Lysis Buffer (0.4M Sorbitol, 150mM KoAc, 2mM MgoAc, 25mM Pipes pH6.8, 10% Glycerol with cocktail of protease inhibitors 10mM Benzamidine, 1µg/mL Pepstatin, 10µg/mL Leupeptin, 1mM PMSF). After measuring the volume of cell suspension (should be appoximately 340µL), 20% Triton X-100 was gently mixed in to bring the final concentration to 1%.

To make a whole cell extract (WCE), a 100µL aliquot transferred to an eppendorf tube containing 100µL Lysis Buffer and 100µL of 3X SDS Sample Buffer was added. For isolation of a chromatin enriched fraction, a 100µL aliquot was overlayed over a cushion of 30% Sucrose 1% Triton X-100 and 1mM PMSF. This suspension was microcentrifuged at 14Krpm for 10min in cold room. The resulting supernatant (approximately 190 µl) was added to 100µL of 3X SDS sample buffer. The chromatin pellet was washed with 200µL of Lysis Buffer, then resuspend in 189µL of Lysis Buffer to which 94µL of 3X SDS sample buffer was added. All samples were boiled in 100°C heat block with H_2O for 3 min before quantifying the protein by Bradford assay. 50 µg of protein were loaded for SDS PAGE.

Fluorescent Micoscopy

GFP fluorescence microscopy was performed as previously described (Liku et al. 2005).

CGH microarray

CGH Copy Number analysis was performed as described in (Green, Finn, and Li 2008 manuscript submitted) with the following modifications using EPICENTER MasterPure^{TM} Yeast DNA Purification kit. Cultures were grown in either 50mL of YEPD or SDC/-Met,Ura,Trp overnight to midlog phase then incubated in alpha factor at 50ng/mL (David King synthesized) for three hours. Next the cultures were harvested in a Beckman table top centrifuge by spinning 5min at 3Krpm, then stored at -80 degrees celcius. DNA was extracted as follows: 300microliters of Yeast Cell Lysis Solution was added to frozen pellets and vortexed to resuspend. RNAse A was added to a final concentration of 100ug/mL and incubated at 65deg C for 15'. Subsequently tubes were placed on ice for 5'. Next 150uL MPC protein precipitation Reagent was added and mixture was vortexed 10" to mix. Then the debris was pelleted by centrifugation at 14Krpm at RT for 10', with subsequent transfer of supernatants to clean eppendorf tubes and adding 500uL of isopropanol and mixed by inversion. Then the DNA was precipitated by centrifugation at 14Krpm at room temperature for 10'; Removed supernatant and washed pellets with 0.5mL 70% EtOH, vortexed then centrifuged 1' at RT at 14Krpm. Next aspirated the 70% EtOH and employed speed vacuum to dry DNA pellets. Finally re-suspend pellets in 50uL of 2mM Tris pH 7.8. 2microliters of this DNA prep was run on 0.6% Agarose to check quantity. However, regardless of the quantity 40microliters of this DNA was used to label and dye couple to Cy3 as described in (Green et al. 2006b).

The reference DNA after labeling was dye coupled to Cy5 and consisted of DNA from either YJL2067 (*orc2-6A orc6-4A MCM7-2NLS*) or YJL5834 (ura3-52::*{pGAL1, URA3}*) arrested in G1. The DNA was prepared as described in (Green et al. 2006b).

CGH re-replication analysis

CGH re-replication analysis were performed as described in (Green et al. 2006b) Except all re-replication experiments carried out here are of inductions done at metaphase arrest exclusively.

Urea protein extraction from Yeast

The following urea lysis extraction procedure was used for Figure 3A and 3B**.** 10ml of an OD=1.0 culture was harvested by centrifugation for 3 min at 3,000 rpm. Pelleted cells were resuspended in 500μl of distilled deionized water and transferred to a screw cap tube (USP #MCTS-806). Cells were precipitated by microcentrifugation (Eppendorf centrifuge 5417C) at 14Krpm for 1 min, quick frozen in liquid nitrogen and stored at -80˚C. To thaw cell pellets were placed on ice and once thawed, the cells were resuspended in 200μl of Urea buffer (20mM Tris pH 7.4, 7M Urea, 2M Thiourea, 4% CHAPS and 1% DTT). 200μl of 0.5mm glass beads were added to the tube and the cells were then lysed by two rounds of 1 min bead beating. Beads were separated from the extract by poking a hole in the bottom of the screw cap tube (using a 22 gauge needle), placing the screw cap tube into a new eppendorf tube and spinning the stacked tubes in a microcentrifuge on a setting of soft 6,000 rpm for 30 sec. The urea protein extract was incubated at 25˚C for 30 min on a nutator and then spun in a microcentrifuge at 14Krpm

for 10 min at 4˚C. The supernatant was transferred to a new tube and the concentration was determined using the Bradford assay. Then equal volume of 2X SDS sample buffer was added to samples. The samples were either run on SDS PAGE or stored at -20˚C and thawed at room temperature.

Westerns

Protein gels were run on either mini gels or large gels and were run according to Leslie E Chu and Joachim J. Li unpublished)

Westerns of extracts in which Pri1-3HA protein was being monitored were done as follows. Standard 7.5% SDS PAGE mini gels were used to separate proteins based on size. To transfer proteins from the gel onto nitrocellulose membrane (Whatman Schleicher&Schuell Protran BA 83), semidry blot transfer and wet transfers were employed. Membranes were stained with Ponceau S for 1 min, imaged to monitor protein loading and then rinsed with water. The membranes were incubated in 20ml of blocking buffer (10% Carnation dry milk, 20mM Tris pH 7.5, 0.5M NaCl, 0.1% Triton X-100) for 60 min at room temperature and then incubated for 60 min in 20ml of antibody buffer (2% Carnation dry milk, 20mM Tris pH 7.5, 0.5M NaCl, 0.1% Triton X-100). The antibody buffer contained anti-HA antibody (Convance MMS-101R mouse anti-HA 16B12) at a 1:1000 dilution. Membranes were quickly rinsed twice with 20ml of the wash buffer (20mM Tris pH 7.5, 0.5M NaCl, 0.1% Triton X-100) and then washed twice, 10 min per wash, with 20ml of the wash buffer. After washing, membranes were

incubated for 30 min in 20ml of secondary antibody buffer containing a 1:2000 dilution of sheep anti-mouse HRP secondary antibody (Amersham NA931V). Membranes were rinsed twice with 20ml of wash buffer and then washed three times, 10 min per wash. Western blots were developed using Pierce SuperSignal West Dura Extended Duration Substrate and exposed to Amersham Hyperfilm MP.

Westerns of extracts in which Pol12, Orc6, and Slk19-Myc protein was being monitored were done similarly except the gels were larger and run longer.

The following antibodies were used:

Pol12: 6D2 mouse monoclonal antibody used at 1:500 or 1:1000, a gift from Lucy Drury (Laboratory of John Diffley)

Slk19-13Myc was detected by 9E11 primary antibody and was used according to (Loog & Morgan 2005).

Orc6-3HA was detected by using the 16B12 primary antibody.

Results

CDK consensus phosphorylation sites on Pol1 and Pol12 are not essential for viability

To determine whether the CDK phosphorylation of Pol12 was necessary to prevent re-replication, we first mutated all 12 CDK consensus phosphorylation sites on Pol12. The CDK consensus phosphorylation sites for CDKs are S/T-P, where S is serine, T is threonine, and P is proline. Although the canonical consensus is S/T-P-X-R/K (Nigg 1991; Holmes & Solomon 1996), a number of notable CDK substrates, such as the replication protein Sld2, are functionally phosphorylated at the noncanonical phosphorylation site S/T-P (Hiroshi Masumoto et al. 2002). Importantly, almost all CDK phosphorylation sites mapped *in vivo* have at least a proline following the phosphorylated serine or threonine, suggesting that mutation of all S/T-P sites should prevent CDK phosphorylation of a protein. Pol12 has no canonical CDK consensus sites, but does have 12 noncanonical (S/T-P) (Figure 1). We used site directed mutagenesis to generate a *pol12-cdk12A* allele that has all serines or threonines of these 12 noncanonical sites mutated to alanine. The wild type *POL12* gene was replaced with this mutant gene by 2-step gene replacement in several strain backgrounds to yield viable strains, suggesting CDK phosphorylation of Pol12 is not essential for viability. In some cases the *pol12-cdk12A* mutations were accompanied by a C-terminal triple HA (3HA) epitope tag, yielding strains that were temperature sensitive, but we did not distinguish how much of

this temperature sensitive phenotype is due to the 3HA tag alone or its synergism with the *pol12-cdk12A* allele.

The effectiveness with which CDK phosphorylation of Pol12 is disrupted by the *pol12-cdk12A* allele can be partially tested by analyzing the mobility of the mutant pol12-cdk12A protein on SDS-PAGE. Wild-type Pol12 experiences a CDK-dependent retardation in its protein mobility due to its phosphorylation (Foiani et al. 1995). Thus, the protein migrates as a doublet (slower migrating hyperphosphorylated and faster migrating hypophosphorylated forms) when exponentially growing cells are analyzed by immunoblot (Figure 2) and as a slower migrating hyperphosphorylated form in G2/M phase. We introduced *pol12-cdk12A* allele into YJL2067 (*orc2-6A orc6-4A MCM7-2NLS*), and examined the mobility of the mutant pol12-cdk12A protein in the resulting strain, YJL6257. Immunoblot analysis using anti-Pol12 antibodies showed that the pol12-cdk12A protein migrated as the faster mobility hypophosphorylated form of Pol12 throughout the cell cycle (Figure 2). Although the mobility shift is likely not monitoring the phosphorylation of all 12 CDK consensus sites, these results confirm that CDK phosphorylation of at least some of these sites have been effectively blocked in the mutant pol12-cdk12A protein.

Similar to Pol12, Pol1 has 13 S/T-P sites, two of which satisfy the canonical consensus sequence of S/T-P-X-R/K. To prevent CDK phosphorylation of this protein, all 13 S/T-P sites were mutated to A-P, generating a *pol1-cdk13A* allele. Replacement of wild-type *POL1* with the *pol1-cdk13A* allele in several strain backgrounds yielded viable strains,

suggesting that CDK phosphorylation of Pol1 is not essential for viability. Finally, both pol12-cdk12A (in this case tagged with GFP) and pol1-cdk13A were substituted for the wild-type genes to generate a viable strain, YJL7210. As Pol1 and Pol12 are the major subunits of the Pol alpha primase complex that are phosphorylated *in vitro* by Clb2-Cdc28 (Ubersax & Morgan, unpublished), these results suggest that CDK phosphorylation of this complex is not critical for its activation during the first round of replication initiation. Thus, it seems reasonable to examine whether this phosphorylation may play a role in CDK inhibition of reinitiation, possibly by inhibiting the function of Pol alpha primase

Pol12 phosphorylation is delayed until the end of S phase

If CDK phosphorylation of Pol alpha primase were to inhibit its function, such an inhibitory effect must avoid conflicting with the indirect activation of Pol alpha primase by CDKs, which is required in the first round of replication initiation. One possible way to resolve this potential conflict is to temporally separate these two events, i.e. activate the unphosphorylated Pol alpha primase first, and then after replication is complete, inhibit Pol alpha primase. There are two basic ways one could imagine implementing such a strategy of delayed inhibition. One is to allow Pol alpha primase to be phosphorylated by CDKs when they first become active at the beginning of S phase, but somehow delay the inhibitory consequences of that phosphorylation. The other is to

simply delay CDK phosphorylation of Pol alpha primase, until the first round of replication is complete.

To determine whether the latter strategy might be employed requires monitoring the timing of Pol12 mobility shift relative to CDK activation and DNA replication during the cell cycle. Foiani et al had previously examined the timing of Pol12 mobility shift during the cell cycle, but came up with conflicting observations, with some experiments suggesting the shift occurred before S phase and others suggesting a possible delay (Foiani et al. 1995; Desdouets et al. 1998). Another group detected a possible delay in Pol12 mobility shift, but this was performed in the presence of 50 mM hydroxyurea and was used to argue that there is a checkpoint induced delay in CDK activation mediated by the CDK regulator Swe1 (Liu & Y. Wang 2006).

We monitored the mobility shift of Pol12 in cultures of YJL865 (*ORC6-3HA*) synchronously released from a G1 arrest (Figure 3). In these cells we could monitor DNA replication by flow cytometry and the activation of S-CDKs (Clb5/Clb6-Cdc28) by following the mobility shift of their substrate Orc6. Orc6 mobility shift is first detected 20 min following release and is 50% complete by 30 min. DNA replication is complete by 60 min. The hyperphosphorylated form, on the other hand, first appears at 50 min post release and does not become the predominant form until 70 min post release. Thus, the Pol12 mobility shift is delayed relative to both CDK activation and DNA replication.

To determine whether DNA replication is required for the Pol12 mobility shift we repeated the experimental time course in the presence of 200 mM HU, which inhibits DNA replication. Consistent with previous observations (Desdouets et al. 1998), blocking DNA replication also blocks the Pol12 mobility shift (Figure 3A). This cannot be due to a block in CDK activation since two CDK targets, Orc6 and Slk19 still undergo their CDK-dependent mobility shift in the HU arrest. Thus, although the Pol12 mobility shift may only be monitoring phosphorylation of a subset of its 12 consensus CDK phosphorylation sites, and although we do not have a simple way of monitoring Pol1 phosphorylation across the cell cycle, these results raise the possibility that CDK phosphorylates Pol alpha primase after DNA replication. Importantly, such a delay could ensure that the CDK activation of Pol alpha primase during the first round of replication does not conflict with the CDK inhibition of Pol alpha primase during the block to re-replication.

CDK phosphorylation of Pol1 is required for G2/M localization of Pol alpha primase to the nuclear periphery.

As discussed above, Pol alpha primase is regulated in two cell cycle dependent manners that could inhibit reinitiation of DNA replication. One of these is the potential sequestration of Pol alpha primase to the nuclear periphery in G2/M phase. Hence, we asked whether this localization to the nuclear periphery is dependent on CDK phosphorylation of Pol1 and/or Pol12.

To determine if the phosphorylation sites on Pol1 were critical for the localization of Pol alpha primase to the nuclear periphery enrichment we replaced POL1 with the *pol1-cdk13A* allele in a strain that has a *POL12-GFP* allele. In this strain, Pol12-GFP is no longer enriched at the periphery but remains in the nucleoplasm during G2/M phase (Figure 4). This suggests that CDK phosphorylation of Pol1 is required for the redistribution of the Pol alpha primase complex to the nuclear periphery.

We performed a similar experiment replacing POL12 with a *pol12-cdk12A-GFP allele*. In that strain, there was no alteration in the nuclear periphery localization of the GFP tag. Moreover, combining the *pol12-cdk12A-GFP* and *pol1-cdk13A* alleles in the same strain yielded the same localization defect as the *pol1-cdk13A* allele alone (Figure 4B). These results suggest that CDK phosphorylation of Pol12 does not contribute to the redistribution of Pol alpha primase to the nuclear periphery in G2/M phase.

The CDK block to Pol alpha primase chromatin association does not require phosphorylation of Pol1 or Pol12.

We next examined whether the CDK consensus phosphorylation sites on Pol1 and Pol12 are required to prevent the chromatin association of Pol alpha primase at metaphase. We first generated a strain that has the *pol12-cdk12A* and *PRI1-3HA* alleles in a strain background containing the *orc2-cdk6A orc6-cdk4A MCM7-2NLS* alleles, so that we could

later examine whether there polymerase alleles could exacerbate re-replication (see below). In both this strain and a congenic *POL1* wt strain, we assayed for the chromatin association of Pri1-3HA in G1 and M phase. We found that the *pol12-cdk12A* allele does not cause a detectable enrichment of Pri1-3HA on chromatin fractions at metaphase (Figure 4C). In case CDK phosphorylation of Pol12 acts redundantly with CDK phosphorylation of Pol1, we constructed a strain (YJL7220) containing the *pol1-cdk13A pol12-cdk12A PRI1-3HA* alleles in the *orc2-cdk6A orc6-cdk4A MCM7-2NLS* strain background. Chromatin association of Pri1-3HA in M phase was not enriched in this strain relative to the *pol12-cdk12A PRI1-3HA orc2-cdk6A orc6-cdk4A MCM7-2NLS* strain (Figure 4D).

We note that in our experiments we observe more residual chromatin association of Pri1-3HA in metaphase than is seen in published reports (Desdouets et al. 1998). This background level of association in metaphase arrested cells could be masking our ability to see slight or partial increases in chromatin association. Nonetheless, our data suggests that CDK phosphorylation of Pol1 and/or Pol12 is not required for the CDK inhibition of Pol alpha primase association with chromatin in G2/M phase.

Disruption of Pol1 and/or Pol12 phosphorylation does not display an increase propensity for re-replication

Although the CDK sites on both Pol12 and Pol1 appear to not be required to prevent the reassociation of the Pol alpha primase complex with chromatin in metaphase, it is possible that they may still be contribute to preventing reinitiation of DNA replication. To address this possibility we asked whether these alleles could enhance the re-replication of a strain that already re-replicates due to deregulation of ORC, MCM2-7, and CDC6 (OMC strain) (V Q Nguyen, Co & J J Li 2001). For this analysis we generated a strain with the following genotype: *pol1-cdk13A pol12-cdk12A orc2-cdk6A orc6-cdk4A MCM7-2NLS pGAL-delntcdc6 pMET-CDC20 (*OMCP1P12 strain) (Figure 5A) (YJL7261). We arrested these cells at metaphase by depleting *CDC20* and adding the microtubule disrupting agent nocodazole. Once cells were arrested in metaphase, we induced *pGAL-delntcdc6* expression to initiate DNA re-replication and monitored the increase in DNA content by flow cytometry. Our results show that both OMC and OMCP12P1 strains induced re-replication, but there were no appreciable differences in the amount of re-replication observed in these two strains. Thus, in this particular setting, we could not detect any increased propensity for re-replication associated with the *pol1-cdk13A* and *pol12-cdk12A* alleles.

Disrupting Pol1 and/or Pol12 phosphorylation in re-replicating strain background associating with genomic alterations

When performing the induction of re-replication in metaphase with several isogenic strains of OMCP1P12 we noticed differences between isolates in the level of re-replication by FACS. We decided to investigate a possible unintended and undesired genotypic difference between the isolates that could explain the difference in behavior. Using comparative genomic hybridization (CGH) we examined the genomic copy number difference between a reference DNA (YJL5834 YJL2067) and the candidate OMCP12P1 strains. We found a significant number of genomic alterations. Of note are an extra copy of *pGAL-delntcdc6* and an extra copy of chromosome XI and also quite prevalent was segmental duplications of regions on chromosome IV (Figure 6). We found that there was an association between enhanced re-replication and an extra copy of *pGAL-delntcdc6* in strains that are OMCP12 or OMCP12P1 lineage, and an extra copy of chromosome XI arose in strains in the lineage leading to OMCP1.

Discussion

Many DNA replication proteins are known to be phosphorylated by CDKs in a functionally relevant manner. In some cases, such as Orc2, Orc6, Cdc6, and Mcm3 this phosphorylation is important in preventing re-replication. In other cases, such as Sld2 and Sld3, this phosphorylation is important in triggering initiation. These observations suggest that the CDK phosphorylation of the two largest subunits of the DNA Polymerase alpha primase complex, Pol1 and Pol12, is also functionally relevant to the regulation of DNA replication. We attempted to define this functional relevance in this portion of my thesis

Our first conclusion is that CDK phosphorylation of Pol1 and Pol12 is unlikely to play a role in the triggering of replication initiation. The viability of the strains containing both *pol1-cdk13A* and *pol12-cdk12A* alleles suggests that CDK phosphorylation Pol1 and Pol12 cannot participate in an absolutely essential function such as the triggering of initiation. Although it is formally possible that phosphorylation of these subunits plays a redundant role in replication triggering with some other phosphorylation event, we do not favor such a scenario. There is no precedence for redundant targets in the triggering of initiation, and in fact, the precedence presented by Sld2 and Sld3 is for exactly the opposite. These two proteins provide non-redundant targets that are each necessary but neither sufficient for the triggering of initiation. Hence, we think it is more likely that CDK phosphorylation of Pol1 and Pol12 contributes to the block to re-initiation by inhibiting the function of Pol alpha primase.

Although CDKs phosphorylation does not directly target Pol1 and Pol12 to trigger initiation, Pol alpha primase does appear to be activated indirectly by CDK phosphorylation of Sld2 and Sld3. This indirect activation raises a potential conflict with any direct CDK inhibition of Pol alpha primase. Our investigation of the timing of Pol12 phosphorylation suggests a possible way to avoid such a conflict. We determined that there is a significant delay in CDK phosphorylation of Pol12 that appears to depend on completion of DNA replication. Although we did not have the tools to analyze the timing of Pol1 phosphorylation across the cell cycle, if Pol1 experienced a similar delay in phosphorylation, it would allow the following model to be proposed. CDK phosphorylation and thereby inhibition of Pol alpha primase is contingent on completion of DNA replication, allowing cells to avoid a conflict with the CDK activation Pol alpha primase needed to initiate DNA replication. One approach to testing the importance of the delayed phosphorylation of Pol12 (and possibly Pol1) is to examine the phenotype of the phosphomimic alleles *pol12-cdk12D* and *pol1-cdk13D*. If this delay is critical in avoiding premature inactivation of the Pol alpha primase complex, one would predict these alleles to be lethal or extremely sick, in contrast to the *pol12-cdk12A* and *pol1-cdk13A,* which have little affect on growth and replication. Finally, whether the delayed CDK phosphorylation is functionally significant or not, our observations raise the interesting question of how such a delay can be coupled to DNA replication. One wonders, for example, whether there might be phosphatases at the fork that protect the replisome components from premature CDK inhibition.

Interestingly, Pol alpha primase is not the only protein that has both a role in initiation and elongation where there is a potential for premature CDK inhibition of activity that could interfere with the first round of replication. The Mcm2-7 complex is both required for pre-RC assembly during replication initiation, and thought to act as the replicative helicase during replication elongation. As discussed in Chapter 2, CDKs prevent Mcm2-7 re-initiation during S, G2 and M phases by phosphorylating Mcm3 and promoting the net export of the Mcm2-7 complex out of the nucleus. How do CDKs make sure that this inhibition of re-initiation does not also impair replication elongation during S phase? One simple model is that the Mcm2-7 complexes somehow resist export as long as they are tethered to the chromatin at replication forks. This model is consistent with the observation that CDK activation leads to gradual loss of the Mcm2-7 complex from the nucleus in a replication dependent manner, but can be rapidly exported from the nucleus at the G1/S boundary, if it is not loaded onto chromatin (V Q Nguyen et al. 2000). Of course, the delayed export could also arise from delayed phosphorylation of Mcm complexes that are still engaged in elongation at replication forks. To investigate this possibility it would be interesting to use phospho-peptide specific antibodies to determine if the chromatin bound pool of Mcm3 or other Mcm subunits are specifically protected from phosphorylation during S phase.

We note that differences in the way Mcm2-7 and Pol alpha primase associate with chromatin may dictate different ways in which they avoid CDK inhibition during elongation. Mcm2-7 removal from chromatin is irreversible in S-phase (K Labib, Tercero & J F Diffley 2000). It can only be loaded during pre-RC formation in G1 phase,

during which it is thought to form a ring around the DNA making its association highly processive and independent of other proteins until the completion of elongation. Pol alpha primase complex, on the other hand, can reload after temporary disruptions in the replisome cause its dissociation from the fork (Ricke & Bielinsky 2004). It is tempting to speculate that Pol alpha primase is linked to the fork through its interaction with other replisome proteins and that its phosphorylation may block its interaction with other replisome proteins at the fork. In such a scenario, premature Pol alpha primase phosphorylation could be detrimental to its fork association, and thus delayed phosphorylation might be the only way to prevent its premature inactivation.

Although our results have heightened our suspicions about a role for Pol alpha primase phosphorylation in the block to re-replication, we have not been able to demonstrate such a role. The importance of CDK phosphorylation of Pol1 in the localization of Pol alpha primase to the nuclear periphery can be interpreted as mechanism to prevent re-initiation by sequestering the complex away from origins. However, we were not able to show that the CDK consensus sites on Pol1 or Pol12 are important for preventing chromatin association of Pol alpha primase or for preventing re-replication. There have considered three possible explanations for this failure:

1) The CDK consensus phosphorylation sites have no role in regulating re-replication.

2) The CDK consensus phosphorylation sites have a role in preventing re-replication but a parallel mechanism is masking its role.

3) The CDK consensus phosphorylation sites have a role in preventing re-replication, but lethality associated with this re-replication is placing selective pressure for our strains to acquire suppressors of any enhanced re-replication phenotype.

Though formally possible, we do not favor possibility #1 because we think there are still many reasons (see intro) for why it is appealing to hypothesize a role in preventing re-replication. Hence, we think it is worth pursuing possibility 2. Failure to see a significant role of CDK phosphorylation of Pol alpha primase in block to re-replication is reminiscent of repeated failure to see a role of CDK regulation of pre-RC components until multiple parallel mechanisms were disrupted (V Q Nguyen, Co & J J Li 2001).

If parallel mechanisms are masking the role of CDK phosphorylation of Pol alpha primase in the block to re-replication, there are two approaches to unmasking this role. One is the development of more sensitive re-replication assays. We note that with the advent and development of more sensitive re-replication assays it became possible to detect re-replication with deregulation of fewer pre-RC components (Green et al. 2006b). The second approach requires identification and disruption of additional parallel mechanism as was done initially in Nguyen et al.

As a first step to identifying a regulatory mechanism that operates in parallel to CDK phosphorylation of Pol alpha primase, we have speculated on a model by which this complex might be recruited to origins in G1 phase. Our speculation is based on the observation in *S. pombe* that ORC can associate with Pol1 and the observation in

cerevisiae that Pol1 chromatin association in G1 phase is independent of Cdc6. In this model ORC directly recruits Pol complex to chromatin in G1 phase, and there are at least two ways to disrupt this recruitment. One is through the mechanisms we have been studying, i.e.CDK phosphorylation of Pol alpha primase complex. The other parallel mechanism might involve the Orc6 association of Clb-Cdc28 observed by Bell lab (Wilmes et al. 2004); in theory the presence of CDK could sterically inhibit ORC-Pol alpha association. This speculation leads to several testable predictions. First, Pol chromatin association in G1 phase should be dependent on ORC. Second, an ORC-Pol alpha association should be detectable in vivo in *S. cerevisiae*. Third, disrupting Clb association with ORC in combination with mutation of *POLl* CDK consensus sites, should allow Pol chromatin association in G2/M and perhaps promote reinitiation.

Finally, possibility #3 is raised because of the following reasons:

A) During strain construction saw at least half of the strains acquired new genomic alterations (Figure 6)

B) Lethality of even very limited re-replication places strong selective pressure to suppress re-replication phenotype

C) Re-replication phenotypes in established strains show instability: can lose phenotype with prolonged incubation.

During generation of strains corresponding to OMCP1 genotype almost all strains checked by microarray CGH for genomic copy number exhibited genomic copy number variation. There appeared to be an unusually high frequency of strains that acquired

segmental duplications on Chromosome IV, corresponding to 515-620Kbp and 870-990Kbp regions. In addition Chromosome XI duplication arose during two independent derivations of the OMCP1 strains. This correlation raises suspicion about a potential deleterious effect of combining the P1 mutation with that in OM.

Furthermore, during efforts to generate the OMCP12 and OMCP12P1 strains resulted in some isolates acquiring an integration of an extra copy of *pGAL-delntcdc6*. This arose two independent times for each genotype. It appears to not be pure coincidence especially considering the observation that derivation of OMCP1 in parallel didn't result in an extra copy of *pGAL-delntcdc6*. There appears to be a selection for or against *pGAL-delntcdc6* construct. In addition at some frequency there appears to be other genomic alterations that vary from strain to strain (Figure 6).

This observation raises the question of whether there maybe selective pressure to acquire genetic alterations that reduce the toxicity by suppressing reinitiation phenotype. Furthermore, it suggests potential undetectable reinitiation occurring whose toxicity generates selective pressure. This suppression may be occurring very frequently during the course of transformation and colony growth.

Another aspect of these perturbations is that they simultaneously act as an agent that induces genomic instability. The lesions caused by reinitiation events are both inducers of genomic changes and also can be toxic when these changes take place in undesired places.

There is precedence for such high frequency suppression of a lethal mutation that also affects genomic stability. This is analogous to the extremely rapid appearance of the *sml1-1* suppressor when *MEC1* gene is mutated (Zhao, E G Muller & Rothstein 1998). The *MEC1* gene is essential and a hypomorphic mutation (*mec1-1*) reduces activity of Mec1 just above the threshold for viability but this level of activity allows for mutations in *SML1* (*sml1-1*) and selects for it. Thus during strain construction of the OMCP1 or OMCP12 or OMCP12P1 strain, it is easy to imagine suppressors being the only strains that get isolated.

Best way to detect toxic synthetic effects of pol mutations with other deregulating mutations is to combine alleles through a cross. If these mutations cause a defect in growth fitness, then it possible for the double mutants to have either synthetic dominant affect in diploid affecting mating efficiency or diploid growth.
In addition the combination of the mutations may result in synthetic lethality in tetrad analysis. Thus to see phenotype of combined alleles it is better to do cross and tetrad analysis.

CDK regulation of Pol alpha primase nuclear periphery localization raises another possibility for role of CDK sites in telomere replication or maintenance. There are several pieces of evidence that have implicated Pol alpha primase in telomere elongation and maintenance. First telomere replication and elongation occurs in late S-phase and requires

CDK activity (Ferreira 2007; Frank, Hyde & Greider 2006). This telomere elongation activity coincides with when Pol12 is hyperphosphorylated. Secondly, Cdc13 along with other components associated with telomeres are enriched at telomeres in late S-phase (Taggart, Teng & Virginia A Zakian 2002). Again this is coincident with when Pol12 is hyperphosphorylated and Pol alpha localizes to nuclear periphery. Interestingly, telomeres are present at nuclear periphery and the association of Pol alpha primase to nuclear periphery during late S-phase when cells are elongating telomeres suggests a potential link.

More importantly there are several lines of evidence suggesting a direct important role of Pol alpha primase in telomere replication and maintenance. First a mutation in *CDC17* or *POL1* (*cdc17-2*) caused abnormally long telomere (Carson & Hartwell 1985). However, inactivation of Pol1 eliminated *de novo* telomere addition (Diede & Gottschling 1999). These findings suggest an essential role for Pol alpha primase in telomere elongation and suggested a coupling of telomere replication to telomere elongation (Chan & Blackburn 2004; Adams Martin et al. 2000). Second, Pol1 is known to interact with with a number of telomeric proteins. In *S. pombe* Pol1 can co-immunoprecipitate with telomerase subunit Trt1 (Dahlén, Sunnerhagen & T. S. Wang 2003; Ray et al. 2002). In addition two other components, Cdc13 and Stn1 are known to physically interact with Pol1 and Pol12 subunits respectively (Qi & V A Zakian 2000; Grossi et al. 2004). These interactions are thought to play important roles in both telomere elongation and capping (Qi & V A Zakian 2000; Grossi et al. 2004). The defect in telomere maintenance caused by mutations in Pol alpha primase components are also associated with defects in silencing

(Dahlén, Sunnerhagen & T. S. Wang 2003; Adams Martin et al. 2000). Interestingly, Pol1 in *S. pombe* has been shown to interact with Swi6 *in vitro* and mutations in Pol1 caused defect in transcriptional silencing and recruitment of Swi6 to this region (Nakayama Ji et al. 2001). Thus it would be interesting to determine if the association of Pol alpha to telomeres and nuclear periphery is important for both silencing and telomere maintenance and if CDK regulates this association possibly through phosphorylation of Pol alpha primase.

References:

Adams Martin, A. et al., 2000. The function of DNA polymerase alpha at telomeric G tails is important for telomere homeostasis. *Molecular and Cellular Biology*, 20(3), 786-96.

Aparicio, O.M., Stout, A.M. & Bell, S.P., 1999. Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, 96(16), 9130-5.

Baker, T.A. & Bell, S.P., 1998. Polymerases and the replisome: machines within machines. *Cell*, 92(3), 295-305.

Biswas, S.B. et al., 2003. Subunit interactions in the assembly of Saccharomyces cerevisiae DNA polymerase alpha. *Nucleic Acids Research*, 31(8), 2056-65.

Carson, M.J. & Hartwell, L., 1985. CDC17: an essential gene that prevents telomere elongation in yeast. *Cell*, 42(1), 249-57.

Chan, S.R.W.L. & Blackburn, E.H., 2004. Telomeres and telomerase. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 359(1441), 109-21.

Dahlén, M., Sunnerhagen, P. & Wang, T.S., 2003. Replication proteins influence the maintenance of telomere length and telomerase protein stability. *Molecular and Cellular Biology*, 23(9), 3031-42.

Desdouets, C. et al., 1998. Evidence for a Cdc6p-independent mitotic resetting event involving DNA polymerase alpha. *The EMBO Journal*, 17(14), 4139-46.

Diede, S.J. & Gottschling, D.E., 1999. Telomerase-mediated telomere addition in vivo requires DNA primase and DNA polymerases alpha and delta. *Cell*, 99(7), 723-33.

Ferreira, M.G., 2007. Telomeres on the Cdk roller-coaster ride. *Nature Cell Biology*, 9(1), 22-3.

Foiani, M. et al., 1995. Cell cycle-dependent phosphorylation and dephosphorylation of the yeast DNA polymerase alpha-primase B subunit. *Molecular and Cellular Biology*, 15(2), 883-91.

Foiani, M. et al., 1994. The B subunit of the DNA polymerase alpha-primase complex in Saccharomyces cerevisiae executes an essential function at the initial stage of DNA replication. *Molecular and Cellular Biology*, 14(2), 923-33.

Formosa, T. & Nittis, T., 1999. Dna2 mutants reveal interactions with Dna polymerase alpha and Ctf4, a Pol alpha accessory factor, and show that full Dna2 helicase activity is not essential for growth. *Genetics*, 151(4), 1459-70.

Formosa, T. et al., 2002. Defects in SPT16 or POB3 (yFACT) in Saccharomyces cerevisiae cause dependence on the Hir/Hpc pathway: polymerase passage may degrade chromatin structure. *Genetics*, 162(4), 1557-71.

Frank, C.J., Hyde, M. & Greider, C.W., 2006. Regulation of telomere elongation by the cyclin-dependent kinase CDK1. *Molecular Cell*, 24(3), 423-32.

Gambus, A. et al., 2006. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nature Cell Biology*, 8(4), 358-66.

Green, B.M. et al., 2006a. Genome-wide mapping of DNA synthesis in Saccharomyces cerevisiae reveals that mechanisms preventing reinitiation of DNA replication are not redundant. *Molecular Biology of the Cell*, 17(5), 2401-14.

Green, B.M. et al., 2006b. Genome-wide mapping of DNA synthesis in Saccharomyces cerevisiae reveals that mechanisms preventing reinitiation of DNA replication are not redundant. *Molecular Biology of the Cell*, 17(5), 2401-14.

Grossi, S. et al., 2004. Pol12, the B subunit of DNA polymerase alpha, functions in both telomere capping and length regulation. *Genes & Development*, 18(9), 992-1006.

Haase, S.B. & Reed, S.I., 2002. Improved flow cytometric analysis of the budding yeast cell cycle. *Cell Cycle (Georgetown, Tex.)*, 1(2), 132-6.

Holmes, J.K. & Solomon, M.J., 1996. A predictive scale for evaluating cyclin-dependent kinase substrates. A comparison of p34cdc2 and p33cdk2. *The Journal of Biological Chemistry*, 271(41), 25240-6.

Labib, K., Tercero, J.A. & Diffley, J.F., 2000. Uninterrupted MCM2-7 function required for DNA replication fork progression. *Science (New York, N.Y.)*, 288(5471), 1643-7.

Liku, M.E. et al., 2005. CDK phosphorylation of a novel NLS-NES module distributed between two subunits of the Mcm2-7 complex prevents chromosomal rereplication. *Molecular Biology of the Cell*, 16(10), 5026-39.

Liu, H. & Wang, Y., 2006. The function and regulation of budding yeast Swe1 in response to interrupted DNA synthesis. *Molecular Biology of the Cell*, 17(6), 2746-56.

Loog, M. & Morgan, D.O., 2005. Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. *Nature*, 434(7029), 104-8.

Masumoto, H., Sugino, A. & Araki, H., 2000. Dpb11 controls the association between DNA polymerases alpha and epsilon and the autonomously replicating sequence region of budding yeast. *Molecular and Cellular Biology*, 20(8), 2809-17.

Masumoto, H. et al., 2002. S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. *Nature*, 415(6872), 651-5.

Matsuno, K. et al., 2006. The N-terminal noncatalytic region of Xenopus RecQ4 is required for chromatin binding of DNA polymerase alpha in the initiation of DNA replication. *Molecular and Cellular Biology*, 26(13), 4843-52.

Miles, J. & Formosa, T., 1992. Protein affinity chromatography with purified yeast DNA polymerase alpha detects proteins that bind to DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, 89(4), 1276-80.

Mimura, S. et al., 2004. Phosphorylation-dependent binding of mitotic cyclins to Cdc6 contributes to DNA replication control. *Nature*, 431(7012), 1118-23.

Morreale, R.J. & Li, J.J., 2008. The progression of replication forks set up during inappropriate
re-replication is impaired
.

Nakayama Ji et al., 2001. A role for DNA polymerase alpha in epigenetic control of transcriptional silencing in fission yeast. *The EMBO Journal*, 20(11), 2857-66.

Nguyen, V.Q. et al., 2000. Clb/Cdc28 kinases promote nuclear export of the replication initiator proteins Mcm2-7. *Current Biology: CB*, 10(4), 195-205.

Nguyen, V.Q., Co, C. & Li, J.J., 2001. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature*, 411(6841), 1068-73.

Nick McElhinny, S.A. et al., 2008. Division of labor at the eukaryotic replication fork. *Molecular Cell*, 30(2), 137-44.

Nigg, E.A., 1991. The substrates of the cdc2 kinase. *Seminars in Cell Biology*, 2(4), 261-70.

Pursell, Z.F. et al., 2007. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science (New York, N.Y.)*, 317(5834), 127-30.

Qi, H. & Zakian, V.A., 2000. The Saccharomyces telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase alpha and the telomerase-associated est1 protein. *Genes & Development*, 14(14), 1777-88.

Ray, S. et al., 2002. Interactions between telomerase and primase physically link the telomere and chromosome replication machinery. *Molecular and Cellular Biology*, 22(16), 5859-68.

Ricke, R.M. & Bielinsky, A., 2006. A conserved Hsp10-like domain in Mcm10 is required to stabilize the catalytic subunit of DNA polymerase-alpha in budding yeast. *The Journal of Biological Chemistry*, 281(27), 18414-25.

Ricke, R.M. & Bielinsky, A., 2004. Mcm10 regulates the stability and chromatin association of DNA polymerase-alpha. *Molecular Cell*, 16(2), 173-85.

Santocanale, C. et al., 1993. The isolated 48,000-dalton subunit of yeast DNA primase is sufficient for RNA primer synthesis. *The Journal of Biological Chemistry*, 268(2), 1343-8.

Schub, O. et al., 2001. Multiple phosphorylation sites of DNA polymerase alpha-primase cooperate to regulate the initiation of DNA replication in vitro. *The Journal of Biological Chemistry*, 276(41), 38076-83.

Shcherbakova, P.V., Bebenek, K. & Kunkel, T.A., 2003. Functions of eukaryotic DNA polymerases. *Science of Aging Knowledge Environment: SAGE KE*, 2003(8), RE3.

Sundin, B.A. et al., 2004. Localization of proteins that are coordinately expressed with Cln2 during the cell cycle. *Yeast (Chichester, England)*, 21(9), 793-800.

Taggart, A.K.P., Teng, S. & Zakian, V.A., 2002. Est1p as a cell cycle-regulated activator of telomere-bound telomerase. *Science (New York, N.Y.)*, 297(5583), 1023-6.

Tanaka, S. et al., 2007. CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature*, 445(7125), 328-32.

Tanaka, T. & Nasmyth, K., 1998. Association of RPA with chromosomal replication origins requires an Mcm protein, and is regulated by Rad53, and cyclin- and Dbf4-dependent kinases. *The EMBO Journal*, 17(17), 5182-91.

Ubersax, J.A. et al., 2003. Targets of the cyclin-dependent kinase Cdk1. *Nature*, 425(6960), 859-64.

Uchiyama, M. & Wang, T.S., 2004. The B-subunit of DNA polymerase alpha-primase associates with the origin recognition complex for initiation of DNA replication. *Molecular and Cellular Biology*, 24(17), 7419-34.

Voitenleitner, C., Fanning, E. & Nasheuer, H.P., 1997. Phosphorylation of DNA polymerase alpha-primase by cyclin A-dependent kinases regulates initiation of DNA replication in vitro. *Oncogene*, 14(13), 1611-5.

Voitenleitner, C. et al., 1999. Cell cycle-dependent regulation of human DNA polymerase alpha-primase activity by phosphorylation. *Molecular and Cellular Biology*, 19(1), 646-56.

Wilmes, G.M. et al., 2004. Interaction of the S-phase cyclin Clb5 with an "RXL" docking sequence in the initiator protein Orc6 provides an origin-localized replication control switch. *Genes & Development*, 18(9), 981-91.

Wittmeyer, J. & Formosa, T., 1997. The Saccharomyces cerevisiae DNA polymerase alpha catalytic subunit interacts with Cdc68/Spt16 and with Pob3, a protein similar to an HMG1-like protein. *Molecular and Cellular Biology*, 17(7), 4178-90.

Wittmeyer, J., Joss, L. & Formosa, T., 1999. Spt16 and Pob3 of Saccharomyces cerevisiae form an essential, abundant heterodimer that is nuclear, chromatin-associated, and copurifies with DNA polymerase alpha. *Biochemistry*, 38(28), 8961-71.

Zegerman, P. & Diffley, J.F.X., 2007. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature*, 445(7125), 281-5.

Zhao, X., Muller, E.G. & Rothstein, R., 1998. A suppressor of two essential checkpoint genes identifies a novel protein that negatively affects dNTP pools. *Molecular Cell*, 2(3), 329-40.

Zhou, Y. & Wang, T.S., 2004. A coordinated temporal interplay of nucleosome reorganization factor, sister chromatin cohesion factor, and DNA polymerase alpha facilitates DNA replication. *Molecular and Cellular Biology*, 24(21), 9568-79.

Zhu, W. et al., 2007. Mcm10 and And-1/CTF4 recruit DNA polymerase alpha to chromatin for initiation of DNA replication. *Genes & Development*, 21(18), 2288-99.

Figure 1A

The *pol12-cdk12A* mutant allele with the nucleotide sequence above and the amino acid sequence below. Highlighted in red are the Serine and Threonine codons that have been mutated to code for the Alanine amino acid which is highlighted in yellow. The Cyan highlighting indicates those silent nucleotides in lower case which have been mutated to either destroy restriction enzyme sites or generate new ones but maintain the amino acid residue identity.

Figure 1B-D

The *pol1-cdk13A* mutant allele with the nucleotide sequence above and the amino acid sequence below. Highlighted in red are the Serine and Threonine codons that have been mutated to code for the Alanine amino acid which is highlighted in yellow. The Cyan highlighting indicates those silent nucleotides in lower case which have been mutated to either destroy restriction enzyme sites or generate new ones but maintain the amino acid residue identity. In addition there is a single nucleotide polymorphism (SNP) at position +1394, which is highlighted in purple. S288C and W303 have a "G" at that position but our re-replicating strain background has "A" at that position. This SNP was corrected in our construct for replacing *POL1* in our re-replicating strain to ensure it was identical at all other amino acid positions with the wild-type, with the exception of CDK sites.

Figure 1A pol12-cdk12A

```
   1 ATG AGC GGA TCT ATC GAT GTC ATC ACT CAT TTT GGG CCT GAT GCA GAC AAG CCG GAA ATC ATC ACT GCA TTA GAG AAT TTA ACT AAA CTG CAT GCA TTG TCC GTC GAG GAT CTA TAC ATT  120
   1  M   S   G   S   I   D   V   I   T   H   F   G   P   D   A   D   K   P   E   I   I   T   A   L   E   N   L   T   K   L   H   A   L   S   V   E   D   L   Y   I   40

 121 AAA TGG GAG CAG TTC TCC AAT CAA AGG CGT CAA ACT CAC ACA GAT CTG ACT TCC AAG AAC ATA GAT GAG TTT AAA CAA TTT TTA CAA TTA CAA ATG GAA AAG CGT GCT AAC CAA ATT TCA  240
  41  K   W   E   Q   F   S   N   Q   R   R   Q   T   H   T   D   L   T   S   K   N   I   D   E   F   K   Q   F   L   Q   L   Q   M   E   K   R   A   N   Q   I   S   80

 241 TCG TCC TCT AAG GTA AAC ACA TCC ACA AAA AAA CCA GTG ATT AAA AAA AGT TTG AAT TCg gcg CCC TTG TTT GGC CTT AGT ATT CCA AAG gCc CCG ACT TTA AAA AAA AGA AAG TTA CAc  360
  81  S   S   S   K   V   N   T   S   T   K   K   P   V   I   K   K   S   L   N   S   A   P   L   F   G   L   S   I   P   K   A   P   T   L   K   K   R   K   L   H   120

 361 GGT CCg TTT TCA CTC AGT GAT TCC AAG CAA ACA TAT AAT GTA GGC TCT GAG GCA GAA ACA AAT GAA AAA GGA AAc tcg agT TTA AAA TTG GAG TTT gCa CCA GGt ATG GCA GAA GAT GCT  480
 121  G   P   F   S   L   S   D   S   K   Q   T   Y   N   V   G   S   E   A   E   T   N   E   K   G   N   S   S   L   K   L   E   F   A   P   G   M   A   E   D   A   160

 481 GTG GGT GAC AGT GCT CCg cTG agC CAC GCT AAA AGC TCG GAT GCC AAG gCT CCC GGC TCa TCG ACA TTC CAg gCg CCA ACA ACA AAC gCg CCT ACC ACA TCt AGa CAA AAT GTT CCT GCT  600
 161  V   G   D   S   A   P   L   S   H   A   K   S   S   D   A   K   A   P   G   S   S   T   F   Q   A   P   T   T   N   A   P   T   T   S   R   Q   N   V   P   A   200

 601 GGT GAA ATT TTG GAT TCT TTG AAC CCT GAA AAT ATT GAA ATA TCA TCT GGT AAT CCA AAT GTA GGC CTT CTT TCG ACT GAA GAA CCA TCA TAT AAC CAG GTC AAA GTA GAA CCA TTT TAT  720
 201  G   E   I   L   D   S   L   N   P   E   N   I   E   I   S   S   G   N   P   N   V   G   L   L   S   T   E   E   P   S   Y   N   Q   V   K   V   E   P   F   Y   240

 721 GAT GCT AAA AAG TAT AAA TTC cGt ACg ATG AGG CAA AAC TTA CAG GAA GCG TCT GAT GTC CTC GAT GAT CAA ATT GAA TCG TTT ACT AAG ATT ATT CAA AAT CAC TAT AAG cTA gCT CCA  840
 241  D   A   K   K   Y   K   F   R   T   M   R   Q   N   L   Q   E   A   S   D   V   L   D   D   Q   I   E   S   F   T   K   I   I   Q   N   H   Y   K   L   A   P   280

 841 AAC GAC TTC GCT GAT CCA ACT ATT CAA TCT CAG TCC GAA ATT TAT GCT GTT GGA AGG ATA GTT CCA GAc gCG CCt ACg TAt GAC AAG TTT TTG AAT CCA GAA TCG CTT TCC CTA GAG ACC  960
 281  N   D   F   A   D   P   T   I   Q   S   Q   S   E   I   Y   A   V   G   R   I   V   P   D   A   P   T   Y   D   K   F   L   N   P   E   S   L   S   L   E   T   320

 961 TCA AGA ATG GGC GGC GTT GGA AGA AGG GTA AGG TTA GAT CTA TCG CAG GTC AAC GAG CTT TCG TTT TTC TTA GGC CAA ATT GTG GCA TTC AAA GGT AAA AAT GCT AAT GGG GAT TAC TTC  1080
 321  S   R   M   G   G   V   G   R   R   V   R   L   D   L   S   Q   V   N   E   L   S   F   F   L   G   Q   I   V   A   F   K   G   K   N   A   N   G   D   Y   F   360

1081 ACC GTT AAC TCT ATT TTG CCG CTA CCT TAT CCA AAC gCa CCg GTt TCC ACt agt CAG GAA CTG CAG GAG TTT CAA GCA AAT TTG GAA GGT TCT TCT TTG AAG GTT ATC GTG ACC TGT GGC  1200
 361  T   V   N   S   I   L   P   L   P   Y   P   N   A   P   V   S   T   S   Q   E   L   Q   E   F   Q   A   N   L   E   G   S   S   L   K   V   I   V   T   C   G   400

1201 CCA TAC TTT GCA AAT GAT AAC TTC TCC CTC GAA TTG TTG CAA GAA TTT ATC GAT AGT ATA AAT AAT GAG GTA AAA CCT CAC GTC TTG ATA ATG TTT GGC CCA TTT ATA GAT ATC ACC CAT  1320
 401  P   Y   F   A   N   D   N   F   S   L   E   L   L   Q   E   F   I   D   S   I   N   N   E   V   K   P   H   V   L   I   M   F   G   P   F   I   D   I   T   H   440

1321 CCT CTG ATA GCG AGT GGT AAA TTG CCA AAT TTT CCC CAG TTC AAG ACA CAG CCC AAA ACA TTG GAC GAG CTT TTt cTt AAG TTA TTc gCg CCT ATT CTA AAA ACg ATc gCA CCA CAT ATC  1440
 441  P   L   I   A   S   G   K   L   P   N   F   P   Q   F   K   T   Q   P   K   T   L   D   E   L   F   L   K   L   F   A   P   I   L   K   T   I   A   P   H   I   480

1441 CAA ACC GTA TTG ATA CCg TCg ACA AAA GAT GCA ATT TCT AAT CAC GCC GCA TAT CCA CAA GCT TCT TTA ATT AGA AAG GCT TTA CAG CTG CCT AAA AGG AAC TTC AAA TGC ATG GCT AAT  1560
 481  Q   T   V   L   I   P   S   T   K   D   A   I   S   N   H   A   A   Y   P   Q   A   S   L   I   R   K   A   L   Q   L   P   K   R   N   F   K   C   M   A   N   520

1561 CCT TCA TCT TTT CAA ATA AAT GAG ATA TAC TTC GGT TGC TCA AAC GTT GAT ACA TTC AAA GAT CTA AAG GAA GTC ATA AAA GGC GGT ACT ACC TCA TCA AGA TAT AGA TTA GAT CGT GTT  1680
 521  P   S   S   F   Q   I   N   E   I   Y   F   G   C   S   N   V   D   T   F   K   D   L   K   E   V   I   K   G   G   T   T   S   S   R   Y   R   L   D   R   V   560

1681 TCT GAG CAT ATT TTA CAA CAG CGC AGG TAT TAC CCC ATA TTT CCA GGT AGT ATT CGT ACA AGA ATA AAA CCA AAA GAC GTA TCT ACG AAG AAG GAA ACC AAT GAT ATG GAA AGC AAA GAA  1800
 561  S   E   H   I   L   Q   Q   R   R   Y   Y   P   I   F   P   G   S   I   R   T   R   I   K   P   K   D   V   S   T   K   K   E   T   N   D   M   E   S   K   E   600

1801 GAA AAA GTT TAT GAA CAT ATA TCT GGT GCA GAT TTA GAC GTG AGT TAT CTG GGA CTA ACA GAG TTT GTT GGT GGG TTC gCt CCg GAC ATA ATG ATT ATA CCC AGT GAA TTA CAA CAC TTT  1920
 601  E   K   V   Y   E   H   I   S   G   A   D   L   D   V   S   Y   L   G   L   T   E   F   V   G   G   F   A   P   D   I   M   I   I   P   S   E   L   Q   H   F   640

1921 GCA AGA GTT GTC CAA AAC GTA GTT GTT ATA AAT CCC GGA AGA TTT ATT AGA GCA ACA GGT AAC AGG GGA TCC TAT GCA CAA ATT ACT GTC CAA TGC CCT GAT CTT GAA GAC GGG AAA TTG  2040
 641  A   R   V   V   Q   N   V   V   V   I   N   P   G   R   F   I   R   A   T   G   N   R   G   S   Y   A   Q   I   T   V   Q   C   P   D   L   E   D   G   K   L   680

2041 ACG CTT GTT GAA GGT GAA GAG CCA GTT TAT CTA CAC AAC GTG TGG AAG CGC GCT AGA GTT GAC TTG ATT GCT AGT TGA 2118
 681  T   L   V   E   G   E   E   P   V   Y   L   H   N   V   W   K   R   A   R   V   D   L   I   A   S   *  705
```

Figure 1B poll-cdk13A

```
   1 ATGTCTTCTAAAAGTGAAAAATTAGAAAAATTGAGAAAGCTGCAAGCCGCTCGAAATGGTACATCCATCGATGAT 75
   1  M  S  S  K  S  E  K  L  E  K  L  R  K  L  Q  A  A  R  N  G  T  S  I  D  D  25

  76 TACGAAGGTGACGAATCAGATGGTGACAGAATATACGACGAAATCGACGAAAAAGAATACAGGGCCAGAAAGCGC 150
  26  Y  E  G  D  E  S  D  G  D  R  I  Y  D  E  I  D  E  K  E  Y  R  A  R  K  R  50

 151 CAGGAATTGCTGCATGACGATTTTGTTGTAGATGACGATGGTGTAGGCTACGTCGATCGTGGTGTCGAAGAAGAT 225
  51  Q  E  L  L  H  D  D  F  V  V  D  D  D  G  V  G  Y  V  D  R  G  V  E  E  D  75

 226 TGGAGGGAAGTTGATAACAGTTCAAGTGATGAAGACACTGGAAACCTTGCTAGCAAGGATTCTAAAAGGAAGAAG 300
  76  W  R  E  V  D  N  S  S  S  D  E  D  T  G  N  L  A  S  K  D  S  K  R  K  K  100

 301 AATATAAAAAGGGAGAAAGACCACCAGATTACAGATATGCTGCGAACTCAACATTCCAAATCAACACTGCTAGCA 375
 101  N  I  K  R  E  K  D  H  Q  I  T  D  M  L  R  T  Q  H  S  K  S  T  L  L  A  125

 376 CATGCAAAAAAATCCCAAAAAAAAAAGTATCCCCATCGATAATTTTGATGACATTCTTGGTGAGTTTGAATCTGGT 450
 126  H  A  K  K  S  Q  K  K  S  I  P  I  D  N  F  D  D  I  L  G  E  F  E  S  G  150

 451 GAAGTAGAAAAACCCAATATTTTACTGCCATCCAAATTAAGGGAAAATTTAAATTCggCaCCGACTAGcGAATTT 525
 151  E  V  E  K  P  N  I  L  L  P  S  K  L  R  E  N  L  N  S  A  P  T  S  E  F  175

 526 AAgTCGTCAATAAAGAGAGTTAACGGAAACGATGAGTCAAGTCATGATGCCGGTATTTCTAAAAAAGTCAAGATT 600
 176  K  S  S  I  K  R  V  N  G  N  D  E  S  S  H  D  A  G  I  S  K  K  V  K  I  200

 601 GATCCAGATTCAAGTACTGATAAATATTTAGAAATCGAAagtgCACCGTTAAAATTACAAAGCAGGAAACTGCGC 675
 201  D  P  D  S  S  T  D  K  Y  L  E  I  E  S  A  P  L  K  L  Q  S  R  K  L  R  225

 676 TACGCAAATGATGTACAaGATcTATTGGATGATGTGGAAAATgCCCCgGTAGTAGCCACAAAGAGACAGAATGTA 750
 226  Y  A  N  D  V  Q  D  L  L  D  D  V  E  N  A  P  V  V  A  T  K  R  Q  N  V  250

 751 CTTCAAGATACGTTACTAGCAAATCCACCATCCGCTCAAAGTTTGGCTGATGAAGAAGATGATGAAGATAGTGAT 825
 251  L  Q  D  T  L  L  A  N  P  P  S  A  Q  S  L  A  D  E  E  D  D  E  D  S  D  275

 826 GAAGATATTATTCTGAAAAGAAGAACTATGAGAtcgGTtACCACAACTAGACGCGTGAATATAGACTCAAGAAGT 900
 276  E  D  I  I  L  K  R  R  T  M  R  S  V  T  T  T  R  R  V  N  I  D  S  R  S  300

 901 AACCCTTCAACggCACCGTTTGTCACAGCACCCGGAgCtCCAATCGGTATTAAAGGgCTAgCTCCAAGTAAATCT 975
 301  N  P  S  T  A  P  F  V  T  A  P  G  A  P  I  G  I  K  G  L  A  P  S  K  S  325

 976 TTGCAGAGCAACACCGATGTTGCTACACTTGCGGTTAATGTCAAAAAGGAAGATGTGGTAGATCCGGAAACTGAT 1050
 326  L  Q  S  N  T  D  V  A  T  L  A  V  N  V  K  K  E  D  V  V  D  P  E  T  D  350

1051 ACTTTCCAAATGTTTTGGTTGGATTACTGTGAAGTGAACAACACTTTGATTTTATTTGGTAAGGTTAAACTAAAG 1125
 351  T  F  Q  M  F  W  L  D  Y  C  E  V  N  N  T  L  I  L  F  G  K  V  K  L  K  375

1126 GATGATAATTGTGTGTCAGCCATGGTTCAAATCAACGGTCTCTGTAGAGAGCTATTTTTCCTTCCAAGAGAGGGT 1200
 376  D  D  N  C  V  S  A  M  V  Q  I  N  G  L  C  R  E  L  F  F  L  P  R  E  G  400

1201 AAggCcCCTACCGATATACACGAGGAAATAATTCCGTTACTGATGGATAAATATGGACTGGATAATATCCGCGCT 1275
 401  K  A  P  T  D  I  H  E  E  I  I  P  L  L  M  D  K  Y  G  L  D  N  I  R  A  425

1276 AAgCCTCAAAAAATGAAATATTCGTTTGAGCTTCCTGACATTCCATCTGAAAGTGATTATTTGAAAGTCTTACTT 1350
 426  K  P  Q  K  M  K  Y  S  F  E  L  P  D  I  P  S  E  S  D  Y  L  K  V  L  L  450

1351 CCATACCAggCtCCGAAGTCCAGTCGAGACACCATTCCATCGGATTTATCTAGCGATACCTTTTATCATGTTTTT 1425
 451  P  Y  Q  A  P  K  S  S  R  D  T  I  P  S  D  L  S  S  D  T  F  Y  H  V  F  475

1426 GGTGGGAATTCCAATATTTTTGAAAGTTTTGTTATTCAAAATAGGATTATGGGACCATGCTGGTTAGATATAAAG 1500
 476  G  G  N  S  N  I  F  E  S  F  V  I  Q  N  R  I  M  G  P  C  W  L  D  I  K  500
```

Figure 1C pol1-cdk13A

```
1501 GGAGCAGATTTCAATTCAATaCGTAATGCTTCTCACTGTGCGGTTGAAGTTTCGGTAGATAAGCCTCAGAATATC 1575
 501  G  A  D  F  N  S  I  R  N  A  S  H  C  A  V  E  V  S  V  D  K  P  Q  N  I  525

1576 gCGCCTACTACAACAAAGACAATGCCTAATCTGAGATGTCTATCTTTATCGATTCAGACGTTAATGAATCCCAAG 1650
 526  A  P  T  T  T  K  T  M  P  N  L  R  C  L  S  L  S  I  Q  T  L  M  N  P  K  550

1651 GAAAATAAGCAAGAGATTGTTTCGATAACcTTAagtGCATATAGAAATATCTCGTTGGAcgCcCCGATACCTGAA 1725
 551  E  N  K  Q  E  I  V  S  I  T  L  S  A  Y  R  N  I  S  L  D  A  P  I  P  E  575

1726 AATATAAAACCAGACGACTTGTGTACTTTAGTAAGGCCGCCTCAGTCAACTAGTTTCCCATTAGGTTTAGCTGCA 1800
 576  N  I  K  P  D  D  L  C  T  L  V  R  P  P  Q  S  T  S  F  P  L  G  L  A  A  600

1801 CTGGCAAAACAGAAACTACCCGGTAGAGTCAGGTTGTTCAACAACGAAAAGGCCATGCTATCTTGTTTTTGTGCT 1875
 601  L  A  K  Q  K  L  P  G  R  V  R  L  F  N  N  E  K  A  M  L  S  C  F  C  A  625

1876 ATGTTAAAAGTGGAAGACCCAGATGTTATCATAGGCCATCGTTTGCAAAATGTCTACTTAGATGTTTTAGCTCAT 1950
 626  M  L  K  V  E  D  P  D  V  I  I  G  H  R  L  Q  N  V  Y  L  D  V  L  A  H  650

1951 AGAATGCATGATCTCAACATTCCAACGTTCAGCTCTATTGGCCGTCGTTTAAGAAGGACTTGGCCTGAAAAATTC 2025
 651  R  M  H  D  L  N  I  P  T  F  S  S  I  G  R  R  L  R  R  T  W  P  E  K  F  675

2026 GGTAGAGGAAATTCGAATATGAACCATTTTTTTATTAGTGACATTTGCTCTGGTAGGCTGATATGTGATATCGCC 2100
 676  G  R  G  N  S  N  M  N  H  F  F  I  S  D  I  C  S  G  R  L  I  C  D  I  A  700

2101 AATGAAATGGGTCAATCcTTggCtCCCAAATGTCAAAGTTGGGATCTTTCAGAGATGTATCAAGTTACATGTGAA 2175
 701  N  E  M  G  Q  S  L  A  P  K  C  Q  S  W  D  L  S  E  M  Y  Q  V  T  C  E  725

2176 AAGGAGCATAAGCCGTTAGATATTGATTATCAAAATCCACAATACCAAAATGATGTAAATAGTATGACAATGGCC 2250
 726  K  E  H  K  P  L  D  I  D  Y  Q  N  P  Q  Y  Q  N  D  V  N  S  M  T  M  A  750

2251 CTACAAGAGAATATTACTAATTGTATGATTTCTGCTGAGGTGTCTTATAGAATTCAATTATTAACCTTGACTAAA 2325
 751  L  Q  E  N  I  T  N  C  M  I  S  A  E  V  S  Y  R  I  Q  L  L  T  L  T  K  775

2326 CAGTTGACAAATTTGGCTGGTAATGCATGGGCCCAAACGCTAGGCGGTACAAGAGCTGGTAGAAATGAGTATATC 2400
 776  Q  L  T  N  L  A  G  N  A  W  A  Q  T  L  G  G  T  R  A  G  R  N  E  Y  I  800

2401 TTACTACATGAATTTTCAAGAAATGGTTTTATTGTTCCTGACAAAGAAGGCAATAGAAGTAGAGCTCAGAAACAA 2475
 801  L  L  H  E  F  S  R  N  G  F  I  V  P  D  K  E  G  N  R  S  R  A  Q  K  Q  825

2476 AGACAAAATGAAGAAAATGCTGATGCACCAGTTAATTCTAAAAAGGCAAAATATCAGGGTGGTTTAGTTTTTGAA 2550
 826  R  Q  N  E  E  N  A  D  A  P  V  N  S  K  K  A  K  Y  Q  G  G  L  V  F  E  850

2551 CCTGAGAAAGGTCTTCATAAGAATTACGTTTTAGTCATGGACTTTAATTCTTTGTATCCATCTATTATCCAGGAA 2625
 851  P  E  K  G  L  H  K  N  Y  V  L  V  M  D  F  N  S  L  Y  P  S  I  I  Q  E  875

2626 TTTAATATATGTTTTACCACCGTTGATAGAAATAAGGAAGATATCGATGAGTTACCAAGCGTACCTCCTTCAGAG 2700
 876  F  N  I  C  F  T  T  V  D  R  N  K  E  D  I  D  E  L  P  S  V  P  P  S  E  900

2701 GTGGATCAGGGTGTTTTACCAAGGTTATTAGCTAATCTGGTGGATCGTCGACGTGAAGTTAAGAAGGTGATGAAA 2775
 901  V  D  Q  G  V  L  P  R  L  L  A  N  L  V  D  R  R  R  E  V  K  K  V  M  K  925

2776 ACTGAAACTGATCCCCATAAGCGAGTTCAATGTGATATTCGTCAACAAGCTCTAAAATTGACTGCCAATTCTATG 2850
 926  T  E  T  D  P  H  K  R  V  Q  C  D  I  R  Q  Q  A  L  K  L  T  A  N  S  M  950

2851 TATGGTTGTTTGGGTTATGTTAACAGTAGATTTTACGCAAAGCCGCTCGCCATGTTAGTCACTAATAAGGGTCGT 2925
 951  Y  G  C  L  G  Y  V  N  S  R  F  Y  A  K  P  L  A  M  L  V  T  N  K  G  R  975

2926 GAGATTTTAATGAATACAAGACAACTAGCTGAAAGTATGAATCTTCTTGTAGTTTATGGTGATACAGATTCGGTC 3000
 976  E  I  L  M  N  T  R  Q  L  A  E  S  M  N  L  L  V  V  Y  G  D  T  D  S  V  1000

3001 ATGATAGATACCGGTTGTGATAATTATGCGGATGCAATTAAAATTGGCTTGGGATTTAAAAGGCTAGTAAATGAG 3075
1001  M  I  D  T  G  C  D  N  Y  A  D  A  I  K  I  G  L  G  F  K  R  L  V  N  E  1025
```

Figure 1D pol1-cdk13A

```
3076 CGCTATAGATTATTGGAGATTGATATTGATAATGTTTTTAAGAAGTTACTATTACATGCCAAGAAAAAGTACGCT 3150
1026  R  Y  R  L  L  E  I  D  I  D  N  V  F  K  K  L  L  L  H  A  K  K  K  Y  A  1050

3151 GCTTTGACTGTAAATTTGGACAAAAATGGTAATGGAACTACTGTTCTAGAAGTTAAAGGGTTGGACATGAAGCGT 3225
1051  A  L  T  V  N  L  D  K  N  G  N  G  T  T  V  L  E  V  K  G  L  D  M  K  R  1075

3226 CGTGAATTTTGTCCACTTTCGAGGGACGTTTCTATACATGTTTTAAACACCATCCTGTCAGATAAGGACCCAGAA 3300
1076  R  E  F  C  P  L  S  R  D  V  S  I  H  V  L  N  T  I  L  S  D  K  D  P  E  1100

3301 GAAGCATTGCAAGAAGTGTATGATTACTTAGAAGACATCAGAATAAAAGTGGAAACCAATAACATTAGAATTGAT 3375
1101  E  A  L  Q  E  V  Y  D  Y  L  E  D  I  R  I  K  V  E  T  N  N  I  R  I  D  1125

3376 AAATATAAGATCAATATGAAGCTTTCAAAAGATCCCAAGGCCTACCCAGGTGGTAAAAACATGCCTGCAGTCCAA 3450
1126  K  Y  K  I  N  M  K  L  S  K  D  P  K  A  Y  P  G  G  K  N  M  P  A  V  Q  1150

3451 GTAGCTCTAAGAATGCGTAAGGCTGGTAGAGTTGTTAAAGCTGGCTCTGTCATCACTTTTGTGATCACAAAGCAG 3525
1151  V  A  L  R  M  R  K  A  G  R  V  V  K  A  G  S  V  I  T  F  V  I  T  K  Q  1175

3526 GATGAAATAGACAATGCAGCGGATgCcCCGGCTCTTTCTGTGGCTGAACGTGCCCATGCATTGAATGAGGTAATG 3600
1176  D  E  I  D  N  A  A  D  A  P  A  L  S  V  A  E  R  A  H  A  L  N  E  V  M  1200

3601 ATTAAAAGTAACAATTTGATACCTGATCCACAATATTATCTCGAGAAACAAATATTCGCGCCGGTAGAAAGGTTG 3675
1201  I  K  S  N  N  L  I  P  D  P  Q  Y  Y  L  E  K  Q  I  F  A  P  V  E  R  L  1225

3676 TTAGAAAGAATTGATAGCTTCAACGTGGTGCGTTTGAGTGAAGCGCTTGGTTTAGATAGTAAAAAGTATTTTAGA 3750
1226  L  E  R  I  D  S  F  N  V  V  R  L  S  E  A  L  G  L  D  S  K  K  Y  F  R  1250

3751 AGAGAGGGCGGTAATAACAATGGAGAAGATATTAATAATTTGCAACCCTTAGAAACAACAATTACAGACGTTGAA 3825
1251  R  E  G  G  N  N  N  G  E  D  I  N  N  L  Q  P  L  E  T  T  I  T  D  V  E  1275

3826 AGGTTTAAGGATACTGTAACATTAGAATTAAGTTGCCCATCATGCGATAAAAGGTTTCCATTTGGTGGTATTGTA 3900
1276  R  F  K  D  T  V  T  L  E  L  S  C  P  S  C  D  K  R  F  P  F  G  G  I  V  1300

3901 TCTTCAAATTACTATCGCGTGTCATATAATGGTTTACAGTGCAAGCATTGTGAGCAACTTTTTgCgCCTCTTCAA 3975
1301  S  S  N  Y  Y  R  V  S  Y  N  G  L  Q  C  K  H  C  E  Q  L  F  A  P  L  Q  1325

3976 TTAACTAGCCAAATAGAGCATTCTATAAGGGCACACATTTCCTTATATTACGCAGGGTGGTTACAGTGTGATGAC 4050
1326  L  T  S  Q  I  E  H  S  I  R  A  H  I  S  L  Y  Y  A  G  W  L  Q  C  D  D  1350

4051 AGCACATGCGGTATAGTTACAAGACAAGTCTCCGTTTTTGGTAAGCGTTGTTTGAATGACGGCTGCACGGGTGTC 4125
1351  S  T  C  G  I  V  T  R  Q  V  S  V  F  G  K  R  C  L  N  D  G  C  T  G  V  1375

4126 ATGAGATACAAATACAGTGACAAGCAATTGTACAATCAACTTTTATATTTCGATTCTTTATTCGATTGTGAAAAA 4200
1376  M  R  Y  K  Y  S  D  K  Q  L  Y  N  Q  L  L  Y  F  D  S  L  F  D  C  E  K  1400

4201 AATAAAAAGCAAGAATTGAAGCCAATATATCTACCCGATGATCTCGACTACCCCAAGGAACAGCTGACAGAATCA 4275
1401  N  K  K  Q  E  L  K  P  I  Y  L  P  D  D  L  D  Y  P  K  E  Q  L  T  E  S  1425

4276 TCTATTAAGGCTTTAACAGAACAAAACAGAGAACTAATGGAAACCGGCCGGAGCGTTGTTCAAAAATATTTGAAC 4350
1426  S  I  K  A  L  T  E  Q  N  R  E  L  M  E  T  G  R  S  V  V  Q  K  Y  L  N  1450

4351 GATTGTGGACGTCGCTACGTTGATATGACTAGCATATTTGATTTCATGCTAAATTAG 4407
1451  D  C  G  R  R  Y  V  D  M  T  S  I  F  D  F  M  L  N  *  1469
```

Figure 2

CDK sites are responsible for SDS PAGE gel mobility shift

Left panel: A 50mL culture of YJL3516 were grown to midlog phase in YEPD then split into two 25mL cultures with either 50nM alpha factor (G1) or 15ug/mL Nocodazole (G2/M), after cultures were more than 95% unbudded or budded I used Urea Lysis extraction method to prepare protein extracts for westerns to detect Pol12 by probing with 6D2 primary antibody (a gift of John Diffley lab).

Right panel: A 50mL culture of YJL3516 (*POL12*) and YJL6257 (*pol12-cdk12A*) were grown similarly as above except the cultures after splitting were either kept asynchronously growing (Log Phase) or treated with 15ug/mL Nocodazole to arrest in G2/M. Extracts were also prepared as described above. The samples were run on a small gel but run for approximately 3hours at 100 volts constant. After transfer Pol12 protein was detected with the primary antibody 6D2 (a gift of John Diffley lab).

Figure 2

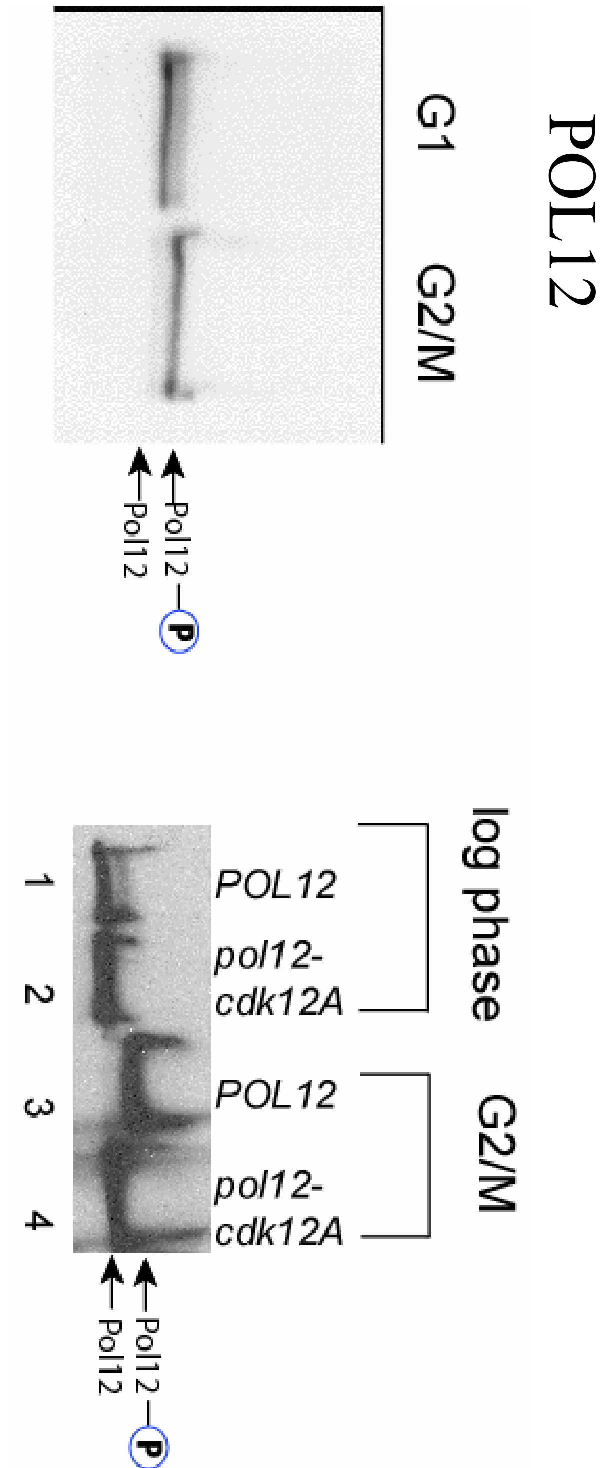
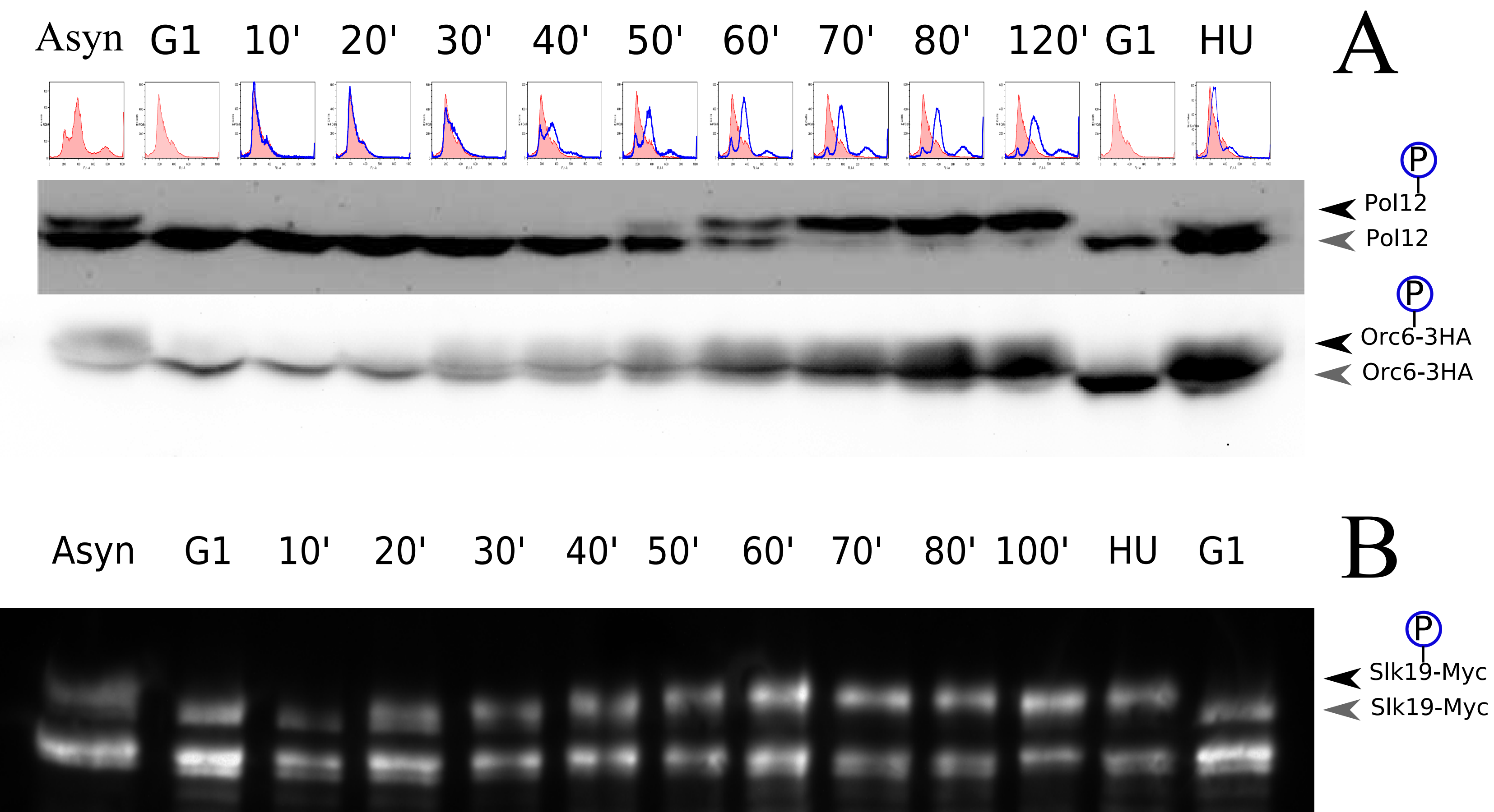


Figure 3

Pol12 exhibits a delayed mobility shift

Top panel: A 50mL culture of YJL865 (ORC6-3HA) was grown to midlog phase in YEPD then arrested in G1 by treatment with alpha factor 50ng/mL for 2hours. Once the culture was 95% unbudded the alpha factor was washed out by filtering the cells and washing once with 400mL fresh YEPD. The washed cells were then resuspended in 50mL of fresh YEPD. Samples were taken for FACS, protein extracts and budding index every 10 min after resuspension in fresh YEPD. Pol12 hyperphosphorylation begins around 50 min post release and Orc6's begins around 20 min. In parallel another culture was arrested in alpha factor and released into HU. After 2.5hours upon release into HU samples were prepared for FACS and westerns. Protein samples for western were prepared by Urea Lysis extraction method. Pol12 was detected with 6D2 primary monoclonal antibody. Subsequently the membrane was stripped and re-probed with 16B12 primary to detect Orc6-3HA.

Bottom panel: A 50mL culture of YJL6928 (*SLK19-myc13*) was also grown and processed as above. 9E11 was used to detect the 13MYC tag.



B.) The lower pair is a proteolytic band produced by separase activity (Loog et al 2005). This is a Slk19-13Myc

B.) EML9-29B rerun 3min28sec using 1:1000 9E10 (anti-myc) 11/16/06. Used according to instructions from Erika W (Morgan Lab) in TBST without milk

Figure 3

Figure 4

Pol1-cdk13A mutation disrupts G2/M nuclear periphery localization

Figure 4A: A 50mL culture of YJL5543 (*POL12-GFP(S65T)*) and also of YJL6453 (*pol12-cdk12A-mGFP*) were grown to midlog phase in YEPD then arrested in G2/M by treatment with 15μg/mL Nocodazole for no more than 2hours. Once the culture was 90% large budded a sampling was taken spun down to pellet cells and resuspended in SDC media before examining cells under GFP fluorescence microscopy. It is important to leave some residual rich media so that the cells are not shocked by new media; so it is better to just resuspend in SDC instead of washing.

Figure 4B: A 50mL culture of YJL7208 (*pol1-cdk13A POL12-GFP(S65T)::His3MX*) and YJL7210 (*pol1-cdk13A pol12-cdk12A-mGFP*) was also grown and processed as above.

Figure 4A

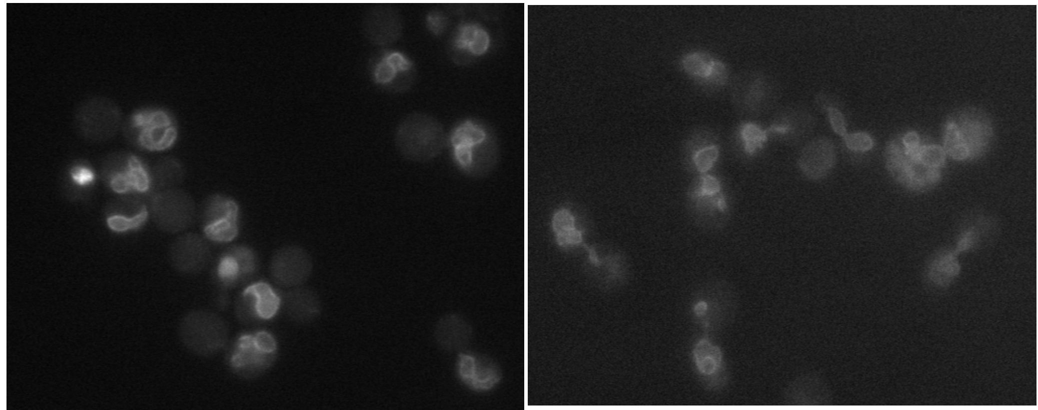


Figure 4B

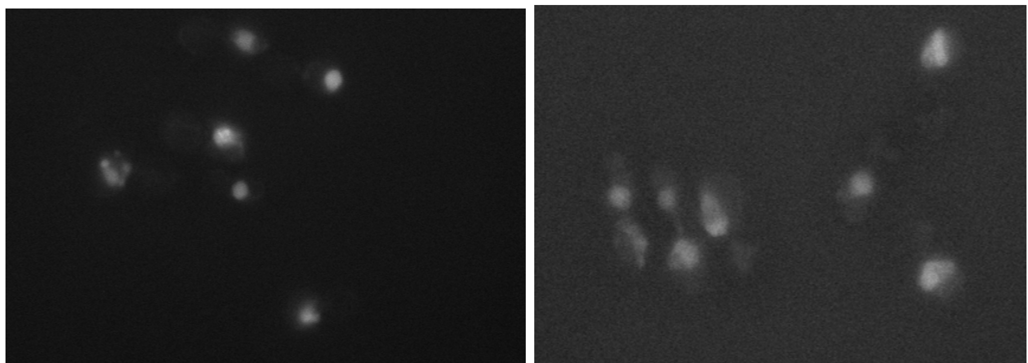
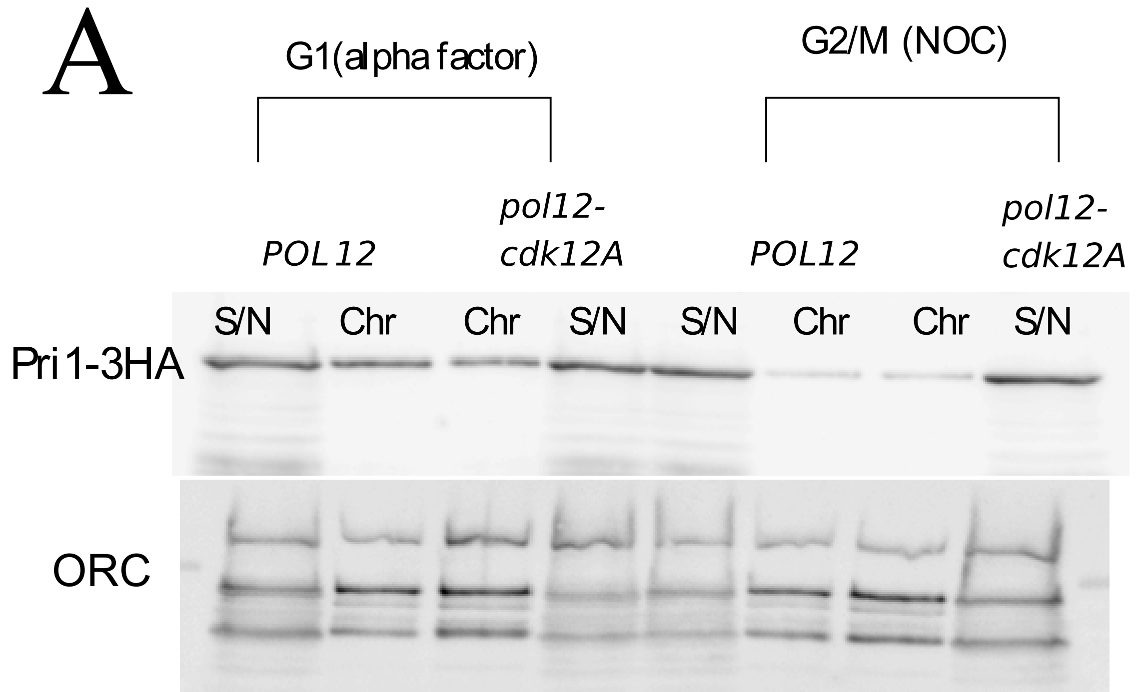


Figure 4



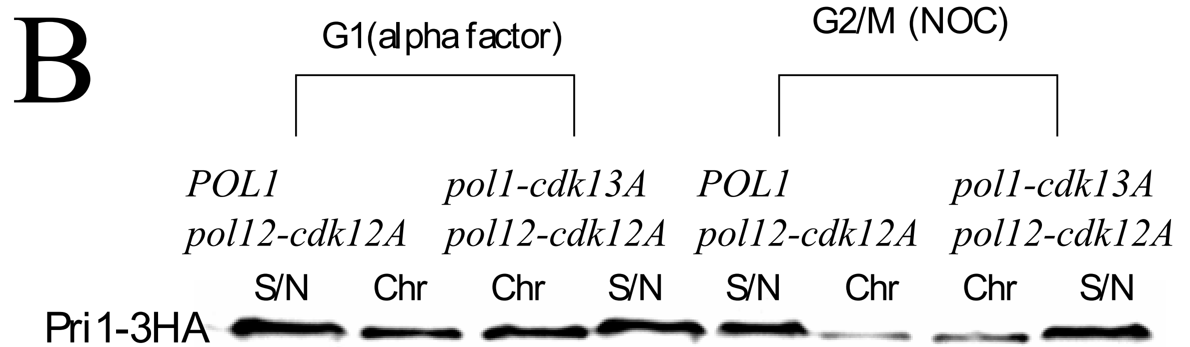


Figure 4

Pol1-cdk13A and pol12-cdk12A mutations alone or in combination do not promote G2/M chromatin association

Figure 4C: A 50mL culture of YJL7029 (*PRI1-3HA*) and also of YJL7031 (*pol12-cdk12A-mGFP PRI1-3HA*) were grown to midlog phase in YEPD then arrested in either G1 with alpha factor 50ng/mL or G2/M by treatment with 15µg/mL Nocodazole for no more than 3hours. Once cultures had arrested at the appropriate arrest point, samples were then processed using the chromatin isolation method as described in the Methods section. And only the Supernatant (S/N) and Chromatin (Chr) fractions were run on a gel. To determine level of Pol alpha on chromatin Pri1-3HA was detected with 16B12 primary antibody in each fraction.

Figure 4D: A 50mL culture of YJL7031 (*pol12-cdk12A-mGFP PRI1-3HA*) and also YJL7220 (*pol1-cdk13A pol12-cdk12A PRI1-3HA*) were also grown and processed as above.

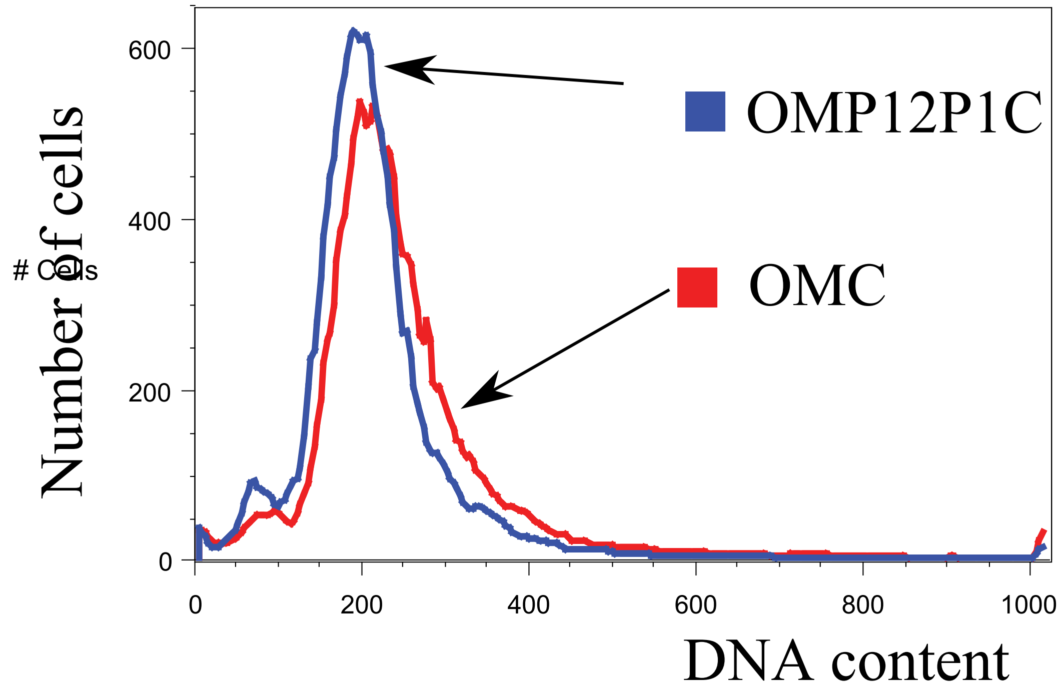
Figure 5

The addition of Pol1-cdk13A and Pol12-cdk12A mutations do not enhance the re-replication observed in OMC strain

Top panel: A 50mL culture of YJL3249 (*OMC)*) and also of YJL7261 (*OMP12P1C)* were grown for preparation for G2/M induction of re-replication. The experiment was carried out as described in the Methods section. This panel is a FACS sample taken immediately after addition of Galactose (0 hr time point). The PMTs were adjusted such that the peak height was at 200 arbitrary units for DNA content.

Bottom panel: A FACS sample of 4hours post induction of re-replication (addition of galactose). Samples were collected and processed as above. Note that the peak height for both YJL3249 and YJL7261 is below 400 arbitrary units for DNA. The presumption is that if the whole genome was duplicated completely within four hours the value of DNA content should have been at least 400 arbitrary units. More importantly, there is no significant difference between YJL3249 and YJL7261.

Figure 5



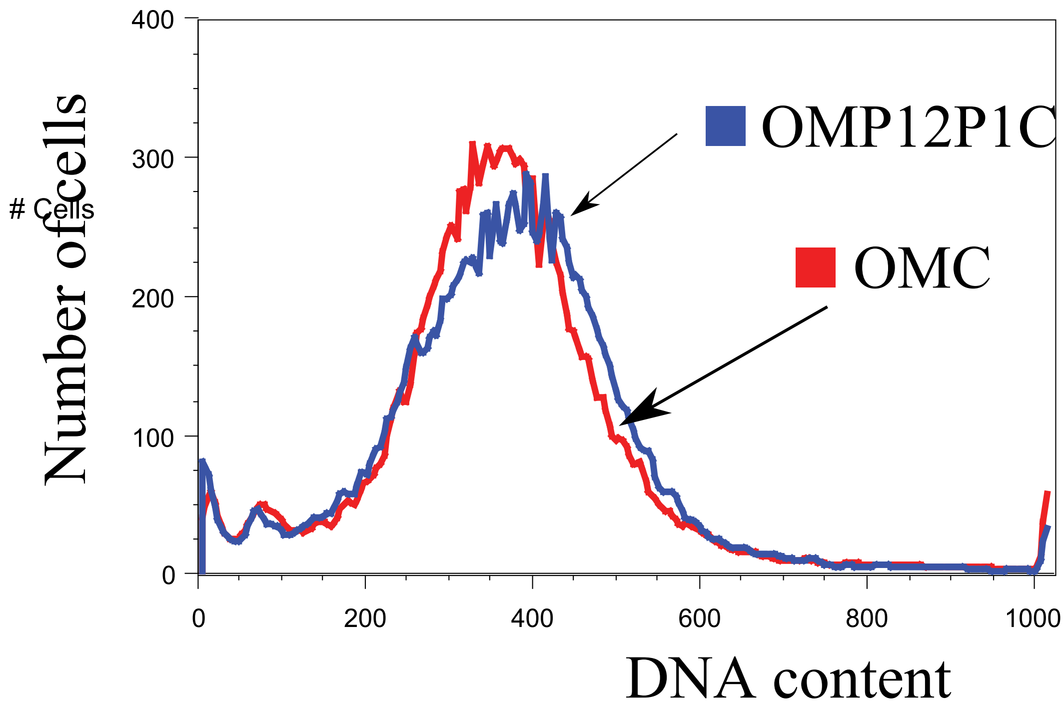


Figure 6

The addition of Pol1-cdk13A and Pol12-cdk12A to OM strain is associated with genomic instabilities

This genealogy map with annotations about genomic copy number variations is a result of several microarray CGH experiments done to determine copy number variation in the strains. The respective strains were grown and DNA extracted and prepared as indicated in the Methods section. The reference DNA came from a large 450mL culture whose DNA was harvested and prepared as described in (Green et al. 2006) from either YJL2067 or YJL5834. These two strains were shown not to contain copy number variations when compared to YJL310 (Richard Morreale personal communication).

Figure 6

OM 2067

SphI pML139 → 6164 OMP12 IV_L
YIP22 → 6211 OMP1C 3249; 6212 OMP1C >3249, 2X C; 6213 OMP12C >3249; 6214 OMP1C

→ 6215 OMP12 IV_L IV_R

A/P pML139 → 6260 OMP12 IV_R XVI_R

6257 OMP12 IV_R X_{500} XM_R
6258 OMP12 IV_R XVI_R
6259 OMP12 IV_R XVI_R

→ OMP12 PRI1-3HA 7031 (and 7032)

YIP22 → OMP12 pr MET-CDC20 7223 (and 7224-26)
pJL1489 → 7252 OMP1C <3249 III_C III_R IV_R VII_L XVI_R XVI_R; 7253 OMP12C >3249; 7254 OMP12C; 7255 OMP12C

pML149 → OMP12 pr P1snp 7220 (and 7219-22)
YIP22 → OMP12 pr P1snp C20 7227 (and 7228-30)
pJL1489 → 7260 OMP12P1C >3249; 7261 OMP12P1C 3249, XVI_R; 7262 OMP12P1C >3249, 2X C; 7263 OMP12P1C

pML151 → OMP12P1 7278 (7279-81)
YIP22 → 7326 OMP12P1C20 (and 7327-29)
pJL1489 → 7438 OMP12P1C <3249; 7439 OMP12P1C; 7440 OMP12P1C 2X C; 7441 OMP12P1C 2X C; 7442 OMP12P1C

pML151 → 7277 OMP1 IV_R; 7282 OMP1 0; 7283 OMP1 0

pJL1489 → 7307 OMP1C (7307-10) IV_R
YIP22 → 7338 OMP1C >3249, IV_R XI; 7339 OMP1C <3249, IV_R $XIII_L$; 7340 OMP1C <3249 IV_L IV_R; 7341 OMP1C IV_R $XIII_L$

YIP22 → 7322 OMP1 XI (7322-25)
pJL1489 → 7429 OMP1C >3249, IV_R XI; 7430 OMP1C >3249, IV_R XI; 7431 OMP1C >3249, IV_R XI; 7432 OMP1C IV_R XI; 7433 OMP1C IV_R XI

pJL806 → OMg 4603-5
YIP22 → OMg 5493-6
refreeze stock → 7401

pJL1489 → OMC 2099 2100
YIP22 → 3248 OMC >3249 XI; 3249 OMC; 3250 OMC; 3251 OMC

red roman numerals are chromosome #
red letters indicate region of amplification
2X C indicates integration of 2 copies of pGALdelntcdc6
pMET-CDC20 is omitted but is introduced by YIP22
=3249 indicates DNA content by FACS is similar to YJL3249
<3249 indicates DNA content by FACS is less than YJL3249
>3249 indicates DNA content by FACS is greater than YJL3249

Table 1. Integrating or integrating replacement plasmids

Plasmids	*Gene* Allele	Source
pJL921	*ORC6-HA3 (pRS306)*	
pRM120	*orc6-4A,rxl, (pRS306)*	This Study
pJL1033	*MCM7*	This Study
YIP22	*pMET-CDC20, (pRS304)*	(Nguyen, Co & Li 2001)
pJL1489	*pGAL-delntcdc6 (pRS306)*	(Nguyen, Co & Li 2001)
pML139	*pol1-cdk13A-3HA (SNP relative to YJL2067) (pRS306)*	This Study
pML146	*pol1-cdk13A-GFP (SNP relative to YJL2067) (pRS306)*	This Study
pML149	*pol1-cdk13A (SNP relative to YJL2067)(pRS306)*	This Study
pML151	*pol1-cdk13A (pRS306)*	This Study

A list of plasmids used in this study.

Table 2. Yeast Strains

Strain	Genotype	Source
YJL312	*MATa leu2-3,112 ura3-52 trp1-289 bar1::LEU2 pep4::TRP1*	(Liku et al. 2005)
YJL5542	*MATa POL1-GFP(S65T)::His3MX his3del1 leu2del0 met15del0 ura3del0*	(Huh et al. 2003) S288C
YJL5543	*MATa POL12-GFP(S65T)::His3MX his3del1 leu2del0 met15del0 ura3del0*	(Huh et al. 2003) S288C
YJL1737	*MATa orc2(6Ala,NotI) orc6(4Ala,NotI) leu2 ura3-52 trp1-289 ade2 ade3 bar1::LEU2 his7? sap3?*	(V Q Nguyen, Co & J J Li 2001)
YJL6293	*orc2(6Ala,NotI) orc6(4Ala,rxl) leu2 ura3-52 trp1-289 ade2 ade3 bar1::LEU2 his7? sap3?*	pRM120 into YJL1737[a]
YJL6453	*pol12-cdk12A-mGFP*	pML146 into YJL6293[a]
YJL7208	*pol1-cdk13A POL12-GFP(S65T)::His3MX*	pML149 into YJL5543[a]
YJL7210	*pol1-cdk13A pol12-cdk12A-mGFP*	pML149 into YJL6454[a]
YJL865	*ORC6-(HA)3*	pJL921 into YJL312 [a]
YJL6928	*SLK19-myc13-His bar1::HISG ade2-1 ura3-1 his3-11 trp1-1 leu2-3,112 can1-100*	(Loog & Morgan 2005)
YJL3155	*MATa ORC2-(NotI, SgrAI) ORC6 leu2 ura3-52 trp1-289 ade2 ade3 MCM7-2NLS bar1::LEU2 his7? sap3?*	(Green et al. 2006)
YJL3516	*ORC2-(NotI, SgrAI) ORC6 leu2 ura3-52 trp1-289 ade2 ade3 MCM7 bar1::LEU2 his7? sap3?*	pJL1033 into YJL3155[a]
YJL7029	*PRI1-3HA::KanMX6 ORC2-(NotI, SgrAI) ORC6 leu2 ura3-52 trp1-289 ade2 ade3 MCM7 bar1::LEU2 his7? sap3?*	pFA6a-3HA-KanMX6 OJL2038/39 into YJL3516[b]
YJL2067	*MAT a orc2(6Ala,NotI) orc6(4Ala,NotI) leu2 ura3-52 trp1-289 ade2 ade3 MCM7-2NLS bar1::LEU2 his7? sap3?*	(V Q Nguyen, Co & J J Li 2001)
YJL6257	*pol12-cdk12A*	pML139 into YJL2067[a]
YJL6258	*pol12-cdk12A*	Sister isolate of YJL6257
YJL6259	*pol12-cdk12A*	Sister isolate of YJL6257
YJL7031	*PRI1-3HA::KanMX6 pol12-cdk12A*	pFA6a-3HA-KanMX6 OJL2038/39 into YJL6257[b]
YJL7220	*pol1-cdk13A PRI1-3HA::KanMX6 pol12-cdk12A*	pML149 into YJL7031[a]
YJL7223	*cdc20::{pMET-3HACDC20}*	YIP22 into YJL7031[c]
YJL7252	*ura3-52::{pGAL-delntcdc6}*	pJL1489 into YJL7223[b]
YJL7253	*ura3-52::{pGAL-delntcdc6}*	Sister isolate of YJL7252
YJL7227	*cdc20::{pMET-3HACDC20}*	YIP22 into YJL7220[c]
YJL7260	*ura3-52::{pGAL-delntcdc6}*	pJL1489 into YJL7227[b]

Table 2. Yeast Strains (cont'd)

Strain	Genotype	Source
YJL7261	*ura3-52::{pGAL-delntcdc6}*	Sister isolate of YJL7260
YJL7262	*ura3-52::{pGAL-delntcdc6}*	Sister isolate of YJL7260
YJL7278	*pol1-cdk13A*	pML151 into YJL6257[d]
YJL7326	*cdc20::{pMET-3HACDC20}*	YIP22 into YJL7278[c]
YJL7438	*ura3-52::{pGAL-delntcdc6}*	pJL1489 into YJL7326[b]
YJL7439	*ura3-52::{pGAL-delntcdc6}*	Sister isolate of YJL7438
YJL7440	*ura3-52::{pGAL-delntcdc6}*	Sister isolate of YJL7438
YJL7441	*ura3-52::{pGAL-delntcdc6}*	Sister isolate of YJL7438
YJL7277	*pol1-cdk13A*	pML151 into YJL2067[b]
YJL7282	*pol1-cdk13A*	Sister isolate of YJL7277
YJL7283	*pol1-cdk13A*	Sister isolate of YJL7277
YJL7307	*ura3-52::{pGAL-delntcdc6}*	pJL1489 into YJL7277[b]
YJL7338	*cdc20::{pMET-3HACDC20}*	YIP22 into YJL7307[c]
YJL7339	*cdc20::{pMET-3HACDC20}*	Sister isolate of YJL7338
YJL7340	*cdc20::{pMET-3HACDC20}*	Sister isolate of YJL7338
YJL7341	*cdc20::{pMET-3HACDC20}*	Sister isolate of YJL7338
YJL7322	*cdc20::{pMET-3HACDC20}*	YIP22 into YJL7277[c]
YJL7429	*ura3-52::{pGAL-delntcdc6}*	pJL1489 into YJL7322[b]
YJL7430	*ura3-52::{pGAL-delntcdc6}*	Sister isolate of YJL7429
YJL7431	*ura3-52::{pGAL-delntcdc6}*	Sister isolate of YJL7429
YJL7432	*ura3-52::{pGAL-delntcdc6}*	Sister isolate of YJL7429
YJL7433	*ura3-52::{pGAL-delntcdc6}*	Sister isolate of YJL7429
YJL3248	*orc2(6Ala,NotI) orc6(4Ala,NotI) ura3-52::{Galp-delntcdc6, URA3} trp1-289 leu2 ade2 ade3 MCM7-2NLS bar1::LEU2 his7? sap3? cdc20::{MET3p-HA3-CDC20, TRP1}*	(V Q Nguyen, Co & J J Li 2001)
YJL3249	*orc2(6Ala,NotI) orc6(4Ala,NotI) ura3-52::{Galp-delntcdc6, URA3} trp1-289 leu2 ade2 ade3 MCM7-2NLS bar1::LEU2 his7? sap3? cdc20::{MET3p-HA3-CDC20, TRP1}*	(Green et al. 2006)
YJL5493	*orc2(6Ala,NotI) orc6(4Ala,NotI) ura3-52::{pGal, URA3} cdc20::{Met3-HA_CDC20, TRP1} MCM7-2NLS bar1::LEU2 leu2 trp1-289 ade2 ade3 his7? sap3?*	(Green et al. 2006)
YJL7401	A RESTREAK AND REFREEZE OF YJL5493	
YJL6164	*pol12-cdk12A*	pML139 into YJL2067[a]
YJL6211	*ura3-52::{pGAL-delntcdc6} cdc20::{pMET-3HACDC20}*	pJL1489 into YJL6164[b] YIP22 into YJL6164[c]
YJL6212	*ura3-52::{pGAL-delntcdc6} cdc20::{pMET-3HACDC20}*	Sister isolate of YJL6211

YJL6213	*ura3-52::{pGAL-delntcdc6} cdc20::{pMET-3HACDC20}*	Sister isolate of YJL6211

(a)Strain derived by two-step gene replacement using indicated plasmid

(b)Strain derived by one-step integrative tagging using indicated plasmids and also pringle tagging.

(c)Strain derived by one-step integration using indicated plasmid

Some of these strains have segmental aneusomy and chromosomal disomies refer to Figure 6 and Table 3.

Green, B.M. et al., 2006. Genome-wide mapping of DNA synthesis in Saccharomyces cerevisiae reveals that mechanisms preventing reinitiation of DNA replication are not redundant. *Molecular Biology of the Cell*, 17(5), 2401-14.

Huh, W. et al., 2003. Global analysis of protein localization in budding yeast. *Nature*, 425(6959), 686-91.

Liku, M.E. et al., 2005. CDK phosphorylation of a novel NLS-NES module distributed between two subunits of the Mcm2-7 complex prevents chromosomal rereplication. *Molecular Biology of the Cell*, 16(10), 5026-39.

Loog, M. & Morgan, D.O., 2005. Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. *Nature*, 434(7029), 104-8.

Nguyen, V.Q., Co, C. & Li, J.J., 2001. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature*, 411(6841), 1068-73.

Table 3

Chr	Left JXN	Right JXN	genes/elements
III_C	150-152		delta8-10, YRCtau1
IIIc		167-169	delta11
III	196-200		HML
III		300	HMR
IV_L	515		Ty2-1
IV_L		620	
IV_R	871-885		Ty2-2,Ty1-2
IV_R		980-993	Ty2-3,Ty1-3
X_{500}	470-486		Ty1-1,Ty1-2
X_{500}		538-539	delta15,16,17
$XIII_L$	184-190		Ty1-1
$XIII_L$		284-286	ADI1, YMR010W,ARS1310, HXT2,YMRtau1
XVI_R	~800		Ty1-2
XVI_R		850	Ty1-3,Ty1-4

A list of junctions for duplicated segments of the genome as illustrated in Figure 6. Also listed are the genes or sequence elements that are at the junction.

Chapter 5

Conclusion

Conclusion

Unlike nonliving objects, the drive for immortality of living organisms can only be achieved through propagation of their genetic material, in the form of DNA. To remain immortal this genetic material or genome must be efficiently copied or replicated with little or no alteration from one generation. A key part of preserving genome integrity is ensuring that every segment is duplicated precisely once per cell cycle, and this is done by making sure the initiation of DNA replication occurs once and only once. Unlike prokaryotes, however, eukaryotes have large genomes that can only be efficiently replicated by initiating replication at hundreds to thousands of sites along the genome. All of these sites must be prevented from re-initiating DNA replication until the cell has entered the next cell cycle. Thus, to maintain genomic integrity the cell has to balance two diametrically opposing activities: triggering initiation and blocking re-initiation at multiple origins.

A model that accommodates these two opposing activities has emerged out of work performed over several decades. Origins are first licensed to initiate, but only in G1 phase, then are triggered to initiate, but only after entering S-phase. In budding yeast, cyclin dependent kinases (CDKs) acts as a switch that allows the transition from origin licensing to triggering and simultaneously prevents any re-licensing of those origins.

Thus, CDKs enforce a single irreversible initiation event at each origin. In other organisms, additional activities work in conjunction with the CDK to prevent reinitiation (Arias & Walter 2007).

Work presented in Chapter 2 dissects one of three known mechanisms employed by CDKs in *S. cerevisiae* to prevent reinitiation: the relocalization of the replication factor, Mcm2-7, outside of the nucleus. We identify a bipartitite NLS for the Mcm2-7 complex that is split between Mcm2 and Mcm3. We also identify an NES for the complex that is adjacent to the Mcm3 NLS segment. We show that CDK consensus phosphorylation sites surrounding the Mcm3 NLS are likely phosphorylated in vivo and in vitro (Liku et al. 2005; Ubersax et al. 2003; Loog & Morgan 2005). and that this phosphorylation alters the activity of these transport signals so as to promote net nuclear export of the Mcm2-7 complex (Liku et al. 2005).

Although the phosphorylation of these sites is important to exclude the Mcm2-7 from the nucleus, they are not essential for this purpose. Mutation of the CDK sites adjacent clustered around the Mcm3 NLS and NES delays, but does not prevent the nuclear export of Mcm2-7. This result suggests that CDKs can regulate Mcm2-7 localization by targeting other proteins and sites. Candidates include several other Mcm subunits that have been shown to be phosphorylated by CDKs in vitro, including Mcm2, Mcm6 (Loog & Morgan 2005) and Mcm4 (Jeff Ubersax and David Morgan unpublished data).

The Mcm3 NES that we identified is important but also not essential for the export of the Mcm2-7. Thus, we suspect that the existence of NES(s) for the complex on other subunits. We identified segments on Mcm4 and Mcm7 that could act like *CRM1* dependent NESs, but did not see any enhanced export defect when mutations in these segments were combined with mutation of the Mcm3 NES. Thus a more exhaustive search for additional NESs may be needed to fully understand the regulation of Mcm2-7 localization.

Another possibility, however, is that CDKs regulate this localization by regulating Cdt1 association with Mcm2-7. The Diffley lab has reported that Cdt1 and Mcm2-7 are interdependent for nuclear localization (Tanaka & Diffley 2002), and the presumption has been that this interdependence is based on a direct physical interaction between Cdt1 and Mcm2-7. One can thus imagine CDK phosphorylation of one of these components disrupting their interaction and resulting in relocalization of both components to the cytoplasm. Such a prediction is clearly testable. Nonetheless, we clearly showed that one way in which CDK regulates Mcm2-7 localization is by phosphorylating sites near the Mcm3 NLS.

As noted previously and in Chapter 3 the CDK regulation of specific pre-RC proteins is not conserved across species. However, the principle of regulating the pre-RC or licensing to prevent reinitiation is conserved from Bacteria to mammals. The change in CDK targets or emergence of new mechanisms like geminin regulation across species suggests an evolution of regulation as species diverge. To determine how rapid the

evolution was occurring we examined CDK regulation of pre-RC proteins across many species. We find that in the case of ORC1 there are dramatic changes in position and number of phosphorylation sites even in closely related species. Despite the rapid changes in phosphorylation site position and number of sites CDK regulation is maintained through conservation of the cluster of phosphorylation sites.

In contrast, although Mcm2-7 localization is regulated in *Saccharomyces cerevisiae* other more distantly related organisms do not regulate its localization. We investigated the evolution of this regulation by examining the CDK sites flanking the NLS signal on Mcm3. To determine this evolution we compare 16 different species between *S. pombe* and *S. cerevisiae*. We find that a clade leading to *S. cerevisiae* after divergence from *C. albicans* acquired the regulated localization and CDK sites flanking the NLS. Interestingly the NES on Mcm3 appears to have coevolved with the CDK sites. This suggests the acquisition of regulation required both the phosphorylation of the NLS to inactivate it and an NES to export the phosphorylated protein.

Naturally, one may wonder whether species that don't regulate Mcm2-7 localization can simply make do with one fewer regulatory mechanisms, or whether they have acquired other mechanisms to compensate. The latter possibility is suggested by analysis of how *S. pombe* prevent reinitiation of DNA replication. Unlike *S. cerevisiae, S. pombe* regulates the levels of Cdt1 (Arias & Walter 2007) along with Cdc6 to prevent re-licensing. Cdt1 and Cdc6 are both required to load Mcm2-7 onto ORC bound origins. Thus, even though Mcm2-7 is constitutively nuclear in *S. pombe*, there is an additional

level of regulation to prevent it from reloading onto origins. Another possible way to prevent Mcm2-7 from promoting reinitiation is to regulate its activity. In mammals, for example, there is evidence that Mcm4 phosphorylation reduces the helicase activity of an Mcm4-6-7 subcomplex (Ishimi & Komamura-Kohno 2001). Consistent with the dependence on an alternative regulatory mechanism, Supplemental Figure 6A of Moses et al 2007 shows Mcm4 has increased conservation of CDK site clustering as you move in evolutionary distance from *S. cerevisiae* to mammals and that it is negatively correlated with a decrease in the level of CDK site clustering on Mcm3. Thus, there appears to be considerable flexibility in evolving new and different combinations of mechanisms to prevent re-replication in eukaryotes. Part of this flexibility may arise from the use of overlapping mechanisms to prevent reinitiation. Such overlap could relieve the pressure to strictly conserve any one specific mechanism, allowing a highly dynamic regulatory evolution.

Finally, in chapter 4 we find some tantalizing evidence for CDK regulation of Pol alpha primase complex. The observations lead us to hypothesize that CDK may regulate Pol alpha primase to prevent reinitiation. However, we did not collect the evidence for this possible role under the conditions we attempted. Nonetheless, the phenotypes observed during efforts to generate the strains suggests an undetectable insult that leads to the chromosomal abnormalities evident in the microarray CGH. Intriguingly, there were two isolates of OMP1 genotype that didn't have detectable genomic copy number variations (YJL7282 and YJL7283). These observations warrant investigation into a possible role of Pol alpha primase regulation by CDK to prevent reinitiation. One possible solution to this

problem of inability to detect this potential reinitiation caused by CDK site mutation sites in Pol alpha primase with the current techniques is to develop better more sensitive assay. The ideal technique should be sensitive enough and allow sufficient resolution to monitor single cell reinitiation events.

Moreover, inhibition of CDK phosphorylation of Pol1 abrogates the mitotic nuclear periphery enrichment. This has implication for both a potential role in reinitiation and telomere maintenance as discussed in Chapter 4. Pol alpha has clear roles in telomere replication and maintenance. And these activities at telomeres are cell cycle regulated. Since the master regulator of the cell cycle is the CDK, it prompts the question of whether the CDK regulates Pol alpha primase activity at telomeres indirectly or directly. The presence of CDK consensus phosphorylation sites and demonstration that Pol alpha primase subunits, Pol1 and Pol12, are phosphorylated by the CDK draws a stronger correlation towards Pol alpha being a direct substrate. As outlined in Chapter 4, there a number of compelling pieces of evidence suggesting CDK mediated cell cycle regulation of Pol alpha primase complex activities at telomeres. This warrants further investigation.

In conclusion:

The central paradigm for propagation of genetic material controlled by the cell cycle states that DNA replication occurs once and only once per cell cycle. There is an exquisite control of replication such that it is done efficiently and accurately in

coordination with the segregation machinery which acts during mitosis. The process of DNA replication involves an orchestra of events that are carried by numerous proteins. The process of DNA replication is the stage of the cell cycle in which the DNA (containing the genetic information) is most vulnerable to alterations. Any errors that result in irreversible change can be detrimental to the survival of the cell. One event that has the potential to cause alterations is reinitiation of DNA replication. Several groups including work from our lab has shown that reinitiation causes DNA damage response indicating that a potentially toxic structure has arisen (Archambault et al. 2005; Green & Li 2005).Thus it is not surprising that the cell has evolved control mechanisms to ensure faithful transmission of genetic information from one generation to the next; this requires that cells replicate their genome once and only once per cell cycle. Although there are cases of controlled whole genome duplication and localized re-replication (Edgar & Orr-Weaver 2001), vast majority of cells in the human body and other organisms replicate only once each cell cycle and divide the genome into two daughter cells. Thus it is the loss of this coordination between first round of initiation and subsequent ones that may cause problems. Polyploidy is well tolerated in many organisms (Edgar & Orr-Weaver 2001). However aneuploidy of many chromosomes and segmental aneusomy is toxic to the cells (Torres et al. 2007; Tybulewicz & Fisher 2006). This is consistent with the principle and observation that abnormal copies of chromosomes can cause imbalances in dosage of gene activities. The gene copy number imbalance can cause gene transcription and protein activity dosage imbalances which could be toxic to the cells (Torres et al. 2007; Torres, Williams & Amon 2008). Furthermore this gene imbalance can potentially lead to tumorigenesis and or developmental problems (Torres, Williams & Amon 2008).

In this regard what is most threatening to a cell is not overt duplication of the whole genome but irreparable or irreversible amplifications of genomic segments caused by reinitiation. It is therefore imperative for the cell's survival, especially multicellular organisms, to ensure that every segment of its chromosomes replicate only once.

References

Archambault, V. et al., 2005. Disruption of mechanisms that prevent rereplication triggers a DNA damage response. *Molecular and Cellular Biology*, 25(15), 6707-21.

Arias, E.E. & Walter, J.C., 2007. Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. *Genes & Development*, 21(5), 497-518.

Edgar, B.A. & Orr-Weaver, T.L., 2001. Endoreplication cell cycles: more for less. *Cell*, 105(3), 297-306.

Green, B.M. & Li, J.J., 2005. Loss of rereplication control in Saccharomyces cerevisiae results in extensive DNA damage. *Molecular Biology of the Cell*, 16(1), 421-32.

Ishimi, Y. & Komamura-Kohno, Y., 2001. Phosphorylation of Mcm4 at specific sites by cyclin-dependent kinase leads to loss of Mcm4,6,7 helicase activity. *The Journal of Biological Chemistry*, 276(37), 34428-33.

Liku, M.E. et al., 2005. CDK phosphorylation of a novel NLS-NES module distributed between two subunits of the Mcm2-7 complex prevents chromosomal rereplication. *Molecular Biology of the Cell*, 16(10), 5026-39.

Loog, M. & Morgan, D.O., 2005. Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. *Nature*, 434(7029), 104-8.

Sazer, S., 2005. Nuclear envelope: nuclear pore complexity. *Current Biology: CB*, 15(1), R23-6.

Tanaka, S. & Diffley, J.F.X., 2002. Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase. *Nature Cell Biology*, 4(3), 198-207.

Torres, E.M. et al., 2007. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science (New York, N.Y.)*, 317(5840), 916-24.

Torres, E.M., Williams, B.R. & Amon, A., 2008. Aneuploidy: cells losing their balance. *Genetics*, 179(2), 737-46.

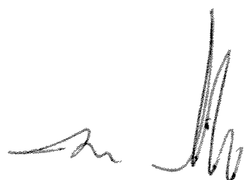
Tybulewicz, V.L.J. & Fisher, E.M.C., 2006. New techniques to understand chromosome dosage: mouse models of aneuploidy. *Human Molecular Genetics*, 15 Spec No 2, R103-9.

Ubersax, J.A. et al., 2003. Targets of the cyclin-dependent kinase Cdk1. *Nature*, 425(6960), 859-64.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

Date