2011

A Proteomic and Genomic Investigation Into Replication Fork Dynamics in Saccharomyces Cerevisiae

Matthew David Sekedat

A PROTEOMIC AND GENOMIC INVESTIGATION INTO REPLICATION FORK DYNAMICS IN *SACCHAROMYCES CEREVISIAE*

A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfilment of the Requirements for
the degree of Doctor of Philosophy

by
Matthew David Sekedat

June 2011

A PROTEOMIC AND GENOMIC INVESTIGATION INTO REPLICATION FORK DYNAMICS IN *SACCHAROMYCES CEREVISIAE*

Matthew David Sekedat, Ph.D.

The Rockefeller University 2011

In eukaryotic organisms, each chromosome must be precisely replicated every time a cell divides so that the genetic material can be passed on to the cell's progeny. The work presented here is an in-depth investigation into the dynamics of the proteins that associate with progressing replication forks in yeast. A focused proteomics approach is employed to specifically identify interactions between the replication fork-coupled GINS complex and other components of the replication machinery. The scope of this technique is extended by applying it to cells that have been synchronized within the cell cycle – revealing the cell cycle dependent interactions of the GINS complex. The results show that GINS is a stable complex throughout the cell cycle, and interacts with components of the replicative helicase and chromatin during S –phase.

Previous studies have led to a picture wherein the replication of DNA progresses at variable rates over different parts of the budding yeast genome. It is widely held that the dynamics of replication fork progression are strongly affected by local chromatin structure/architecture, and by interaction with machineries controlling transcription, repair and epigenetic maintenance. Here we adopted a complementary approach to those previously applied for assaying replication dynamics wherein we used whole genome time-resolved ChIP-chip analysis of three integral members of the replication fork – the GINS complex, Polymerase ε, and Polymerase δ. Surprisingly, our data

demonstrate that these proteins progress at highly uniform rates regardless of genomic location, revealing that replication fork dynamics in yeast is simpler and more uniform than previously envisaged. In addition, we demonstrate how the synergistic use of experiment and modeling leads to novel biological insights. In particular, a parsimonious model allowed us to accurately simulate fork movement throughout the genome and also revealed a subtle phenomenon, which we interpret as arising from low frequency fork arrest. Taken together, these experiments suggest that the progressing replication forks take precedence in the genome, and that chromatin state does not have as significant affect on the rate of fork progression, as was previously believed.

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge my family for their love and support. My parents, Ron and Carolyn, my brother Brad, and my sister Kelsey have been there for me, through all times thick and thin.

I must thank my wife Kristy for putting up with long and unusual hours in the lab, the missing weekends due to experiments that absolutely could not wait, and my midnight disappearances just to check that experiments were still running. But most importantly, I thank her for being a loving and supportive wife and friend. I also want to acknowledge my son Miles, who drives me to work hard, if for no other reason than to set a good example for him.

I owe a special thanks to Dr. Thomas Muir, who played a particular role in bringing me to this point. He took me on as a technician after I finished undergrad, and gave me the amazing opportunity to work on a very exciting research project. It was the success and thrill of this project that convinced me that I should give graduate school a try, and I am very happy with the way things worked out.

I owe a huge debt of gratitude to Dr. Mike O'Donnell and Dr. C. David Allis for counselling me for the past six years as members of my faculty advisory committee. Their advice and support made every single committee meeting enjoyable and educational events for me, which is precisely what they should be.

I want to thank all members of the Chait lab, past and present, for useful discussions and for making the lab an exceptional place to work. Particularly, Dr. Júlio Caesar Padovan, Dr. Beatrix Uberheide, Dr. Sunnie Myung, Dr. Jeffrey DeGrasse, Dr. Erica

Jacobs, Mr. Zachary Quinkert, Dr. Andrew Krutchinsky, Ms. Kelly Molloy, Mr. Herbert Cohen, Dr. Yinyin Li, Mrs. Gladys McMilleon, Mr. Roman Subbotin, Dr. Eileen Woo, and Dr. Wenzhu Zhang.

In particular, I would like to thank Dr. Alan Tackett for launching the project that ended up as my thesis work, supplying a number of the strains used in this work, and for teaching me everything to know about yeast.

I owe a singularly colossal debt to Dr. David Fenyö for designing the simulation experiments, and for happily tweaking them, even at my most frivolous whims. Beyond being the best bioinformatician that I know, he has also been a great comrade and office-mate.

I also want to thank Dr. Mike Rout and members of his lab for their helpful suggestions throughout the years.

Many thanks go to the dean's office, including Dr. Sid Strickland, Mrs. Marta Delgado, and especially Mr. Cristian Rosario, for assistance and guidance in writing and formatting my thesis, and for help throughout my graduate career.

Thanks to Dr. Johnathan Aitchison for fruitful teleconferences and discussions about the project, and his support through it all. I also thank the members of his lab who helped with microarray experiments, especially Dr. Richard Rogers and Mr. Bruz Marzolf.

A final important thank you goes to Dr. Brian Chait. For cheerfully mentoring me for six years. He taught me countless scientific facts, but more importantly, he taught me how to do science. He gave me the freedom to explore my project in any direction that it lead, but was also ever willing to discuss problems and offer encouragement at any moment. He has taught me many, many things over the years, but I think that the most

important lesson that I learned while working with him is that it is entirely possible to be a great scientist, and a great person.

I would like to dedicate this thesis to my (nearly) 2-year-old son Miles. Having him in our lives puts everything in its proper perspective. That he is delightfully oblivious to the stresses associated with writing this thesis has made it significantly less daunting. I only hope that I will always be up to the task of balancing my obligations with my Miles time. Whatever path he chooses in life, I hope at the very least, that I can impart to him a basic appreciation for the importance of science.

"For me, it is far better to grasp the Universe as it really is than to persist in delusion, however satisfying and reassuring." — Carl Sagan

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

CHAPTER II

CHAPTER III

CHAPTER IV

CHAPTER V

LIST OF TABLES

APPENDICES

ABBREVIATIONS & ACRONYMS

ARS – Autonomously Replicating Sequence

ChIP – Chromatin immuno-precipitation

ChIP-chip - Chromatin immuno-precipitation followed by microarray analysis

CDK – Cyclin dependent kinase

CMG – Cdc45, MCM, GINS complex

Da - Dalton

DDK – Dbf4/Drf1-dependent kinase

DNA – Deoxyribonucleic acid

DNase – Deoxyribonuclease

GINS – Go, ichi, ni, san (five, one, two, three in Japanese)

IgG – Immunoglobulin G

Kb – Kilobase

KDa- Kilodalton

MCM – Mini-chromosome maintenance complex

MDa - Megadalton

ORC – Origin recognition complex

PCNA – Proliferating cell nuclear antigen (sliding clamp)

Pol - Polymerase

PrA – Protein A tag originally generated from *S. aureus*

RFC – Replication factor C (clamp loading complex)

RNA – Ribonucleic acid

RPA – Replication protein A (single stranded binding protein)

RPC – Replisome progression complex

CHAPTER I – INTRODUCTION

DNA is often referred to as the blueprint of life. This analogy carries significant conceptual value, even though it should be taken with a grain of salt – one cannot entirely predict the traits of an organism by examining its DNA sequence, and in humans a majority of the blueprint appears not to diagram anything at all. It is certainly true, however, that the genome encodes the basic plans for a cell, and every cell in an organism contains the same genetic sequence. Thus, an immense amount of information is carried within the genome of even the simplest organism. Indeed, the blueprints of a simple bacterium are many orders of magnitude more complicated than even those of the Empire State Building.

Despite its daunting complexity, the concept of genetic material was surprisingly well understood well before the physical or chemical composition of genomes was uncovered. Gregor Mendel's 19th century work on pea plants revealed that traits were systematically passed from generation to generation, and Darwin explained how these traits could change gradually over time. Thus, scientists intuited that organisms must contain a blueprint to be passed on to their progeny, and this blueprint needed to be copied prior to cell division or reproduction. For this reason, the molecular component of evolution itself is in part based on the replication of genomes, as mutations caused by errors in the replication process occasionally lead to adaptive advantages for the organism.

In the physical sense, eukaryotic genomes are a collection of linear strands of DNA called chromosomes (Figure 1-1). The human genome contains 23 pairs of chromosomes, all of which must be precisely copied each time a cell divides. In humans, cell division primarily occurs during development and in tissue that receive the most wear and tear, and these cells must regenerate themselves throughout life. Problems in chromosome replication are associated with disease and aging, making them important areas of research. The replication process in higher metazoan species with large genomes and long generation times is extremely difficult to study. Fortunately, most of the components of the replication process are very highly conserved, and unicellular eukaryotes, such as *Saccharomyces cerevisiae*, divide rapidly

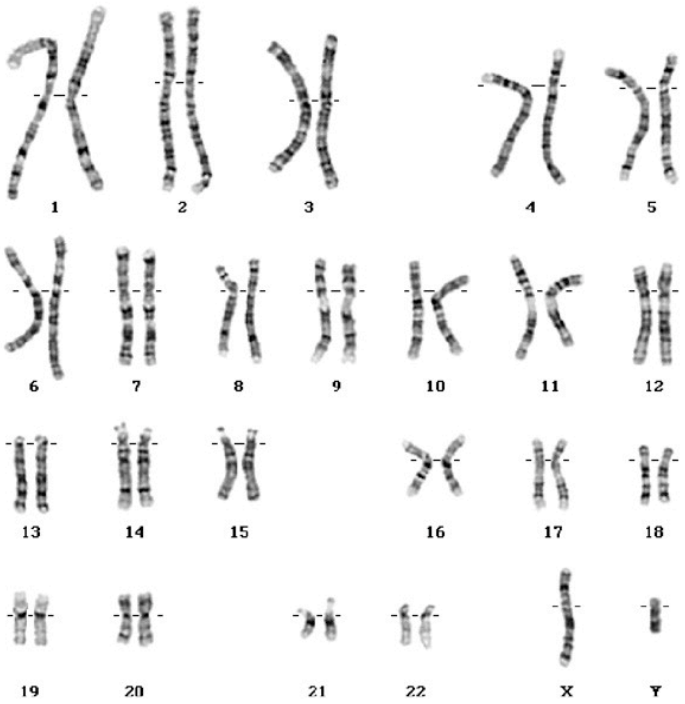


Figure 1-1. The genome of an organism is a composite of its chromosomes. A human male karyotype is shown, in which ~3.2 billion base pairs are distributed among 23 pairs of chromosomes. Every time a human cell divides, it must copy each of these chromosomes so that each of the daughter cells gets one. [from (Wippold and Perry, 2007)]

(Table 1-1). Yeast is therefore an excellent model system for investigating chromosomal replication, and indeed, a large amount of the work on eukaryotic DNA replication has been carried out in *S. cerevisiae* and *S. pombe*.

Table 1-1. Genome Sizes and Generation Times for Selected Eukaryotic Organisms

Species Name	Common Name	Estimated Total Genome Size (bp)[a]	Estimated Number of Protein Encoding Genes[a]	Generation Time[b]
Homo sapiens	human	2.9 billion	20,000 - 25,000	20 years
Mus musculus	lab mouse	2.5 billion	30,000	65 days
Drosophila melanogaster	fruit fly	170 million	14,000	12 days
Caenorhabditis elegans	nematode	95.5 million	18,000	3 days
Sacchoramyces cerevisiae	budding yeast	12 million	6,000	1.5 hours

[a] Estimates from (Van Straalen and Roelofs, 2006)
[b] Estimates from (Subramanian, 2008)

Eukaryotic chromosomes are composites of DNA and its interacting proteins - collectively referred to as chromatin. In particular, histone proteins confer a higher order structure to chromatin by tightly wrapping the DNA into nucleosomes. Histones also play an active regulatory role in gene expression by compacting the regions near gene promoters. This creates regions of the genome that are more densely packed than others.

The structure of DNA affects the way proteins can bind to it. Therefore, to understand the process of replicating DNA, we must first understand its chemical nature. Moreover, the different degrees of chromatin compaction have potentially major implications on the process of replicating the chromosomes. A picture has emerged in which some regions of the genome are more difficult to replicate due to barriers created

by dense packing of the chromatin. The task of xeroxing the blueprints of life is therefore daunting, but one that cells pull off with aplomb.

1.1 - DNA

Chromosomes are made of DNA, as demonstrated in 1944 by Avery, Macleod, and McCarty (Avery et al., 1944). DNA is composed of nucleotides, which consist of a base group, a sugar, and a phosphate group (Figure 1-2). The sugar found in all DNA is a pentose ring named 2-deoxyribose. There are four nucleotides found in DNA, distinguished only by their base groups – adenine (A), guanine (G), cytosine (C), and thymine (T) (Figure 1-2, top gray shaded area). These are grouped into two classes, purines (A and G) and pyrimidines (C and T).

Nucleotides are linked to each other via alternating sugar and phosphate residues. A phosphodiester bond links adjacent sugars via their third and fifth carbon atoms. This generates an asymmetric chain of nucleotides with a sugar-phosphate backbone that spells out the genetic code in the order of the base groups. Therefore, the 5′ end of the molecule displays a phosphate group, while the 3′ end displays a hydroxyl group.

Two strands of DNA pair together through hydrogen bonding of the base groups, creating a double strand of DNA (Figure 1-3). This double-strand is anti-parallel – the 5′ end of one molecule binds to the 3′ end of the other. The strands spiral around themselves, forming one of the most striking and elegant structures in all of nature - a double helix. The double helix is 23.7 Å in diameter, and can be extraordinarily long. In fact, human chromosome 1 is the longest strand of DNA occurring in biology at ~220 million base pairs. The helix completes a 360° right-handed turn every 10.4 – 10.5 bp (Figure 1-4). Many biological processes can impart torsional strain on the DNA. This is particularly true in the case of DNA replication, as unwinding the DNA for use as a template can impart negative supercoils upstream of the unwound DNA.

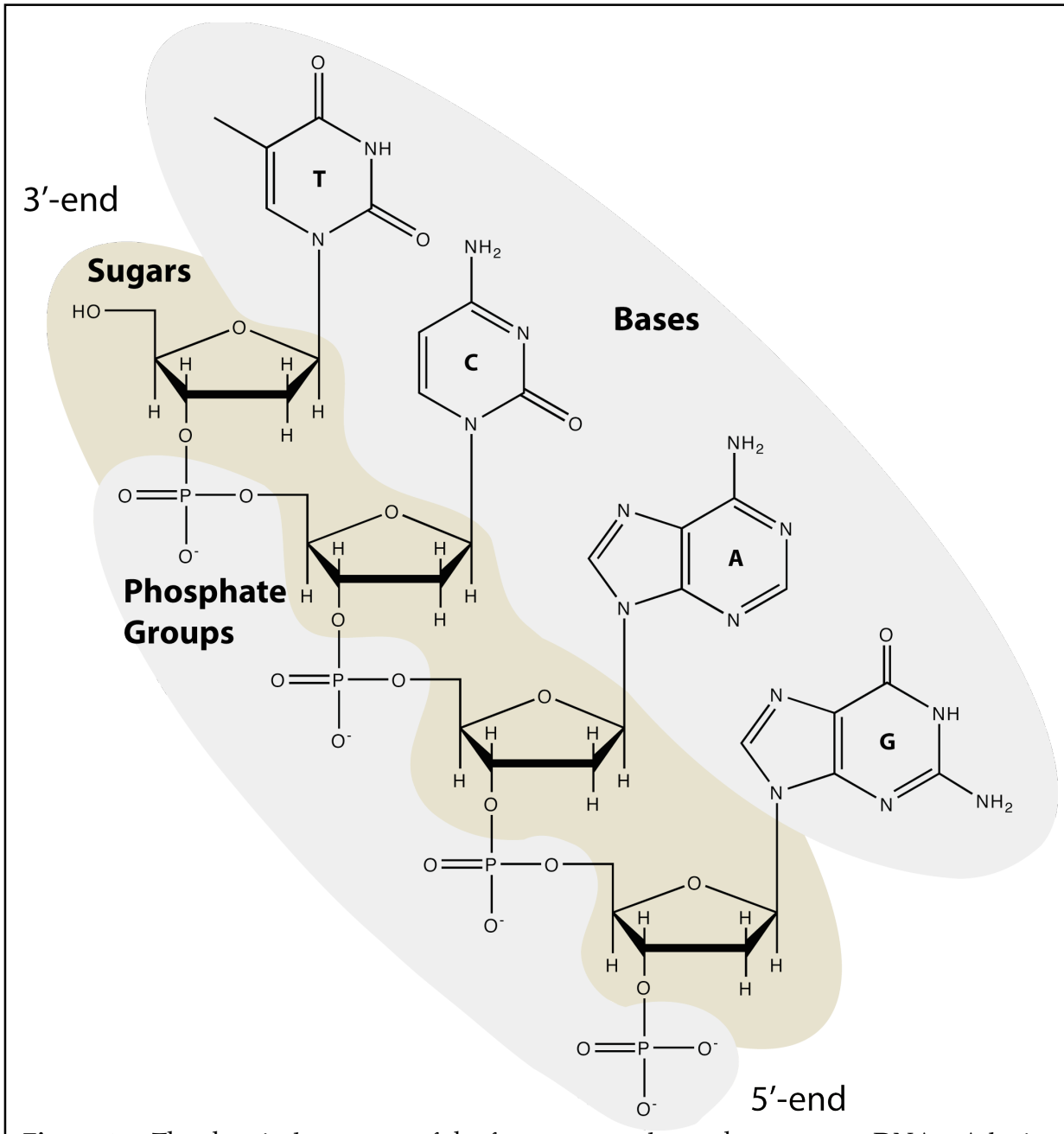


Figure 1-2. The chemical structure of the four common bases that compose DNA – Adenine (A), Guanine (G), Cytosine (C), and Thymine (T). The sugar group is a 2-deoxyribose, and when coupled to a base the resulting group is referred to as a nucleoside. When the phosphate group is coupled to the 5′ – hydroxyl group of the ribose, the resulting molecule is referred to as a nucleotide. Nucleotides are strung together when the phosphate bonds with the 3′ – hydroxyl of another nucleotide. This gives DNA its polarity, as the sugar phosphate backbone is directional (i.e., 5′ to 3′).

The discovery of the double-stranded nature of DNA was remarkable for many reasons, but particularly because it immediately hinted at the mechanism used by cells to copy the chromosomes. Indeed, Watson and Crick noted at the end of their 1953 paper "it has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material" (Watson and Crick, 1953). Essentially, the two strands of DNA may be separated, and each used as a template for synthesizing a complementary second strand. This process will yield two chromosomes (or genomes once all of the chromosomes are replicated), one of which would be passed on to the daughter cell. This so-called "semi-conservative" model of DNA replication means that in every cell, one strand of DNA is from the parent, and the other is newly synthesized.

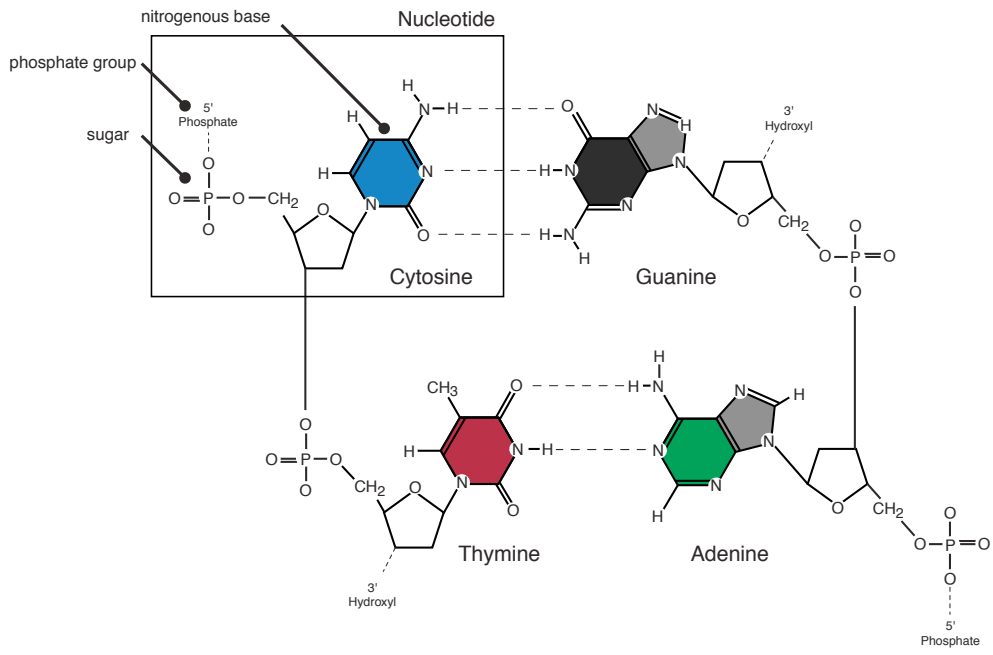


Figure 1-3. Hydrogen bonds maintain interactions between the base pairs in double-stranded DNA. Base pairs stack on top of each other and hold the sugar-phosphate backbone together. [from the NIH genome site http://www.genome.gov/glossary/index.cfm?id=143].

Haploid yeast cells contain ~12 million base pairs distributed among 16 chromosomes that must be accurately replicated every time a cell divides. The size of each chromosome ranges from ~230 kb to ~1.5 Mb, which means that a single uncoiled molecule of yeast DNA can be as long as ~0.5 mm. The nucleus of the yeast cell is 1-2 μm in diameter, so chromosomes must be heavily compacted to fit into the tiny compartment of the nucleus. DNA may carry the instructions for life, but clearly there is more to a chromosome than DNA alone.

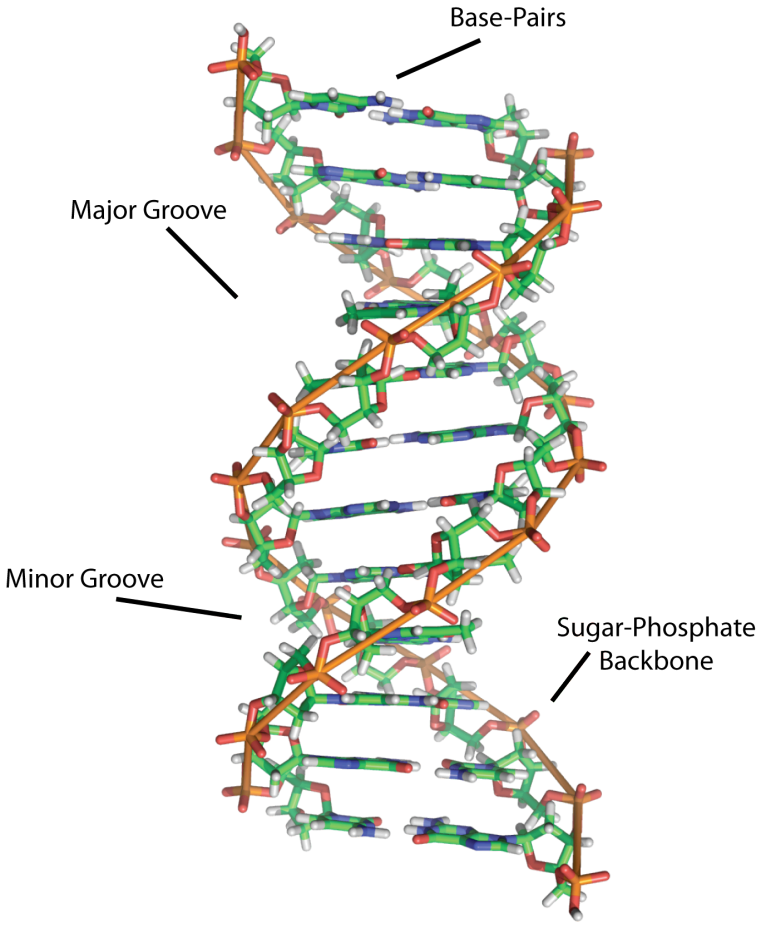


Figure 1-4. Structure of double-helical DNA. The base pairs are stacked, and the sugar-phosphate backbone is twisted around forming the double helix structure. The helix completes one turn approximately every 10.5 bp. Two grooves form along the backbone – the larger major groove, and the smaller minor groove.

1.2 - CHROMATIN

The genome can only be partially described by the chemical nature of DNA. A number of proteins integrally associated with genomic DNA affect its three-dimensional structure and function. Most notable is the wrapping of DNA around the four core histones (H2A, H2B, H3, and H4) to create nucleosomes (Figure 1-5). Nucleosomes are further coiled into tight fibers that are called chromatin, and amazingly, this process is sufficient for packaging two meters worth of human DNA within a nucleus that is only 6 μm in diameter (on average).

In *S. cerevisiae*, the nucleosome consists of 147 bp of DNA tightly wrapped twice around the histone octamer (2 copies of each core histone) (Felsenfeld and Groudine, 2003; Richmond and Davey, 2003; Segal and Widom, 2009) (Figure 1-5). Nucleosomes are typically spaced ~20 – 50 bp apart, and ~75 - 90% of DNA is wrapped up in these structures (Van Kolde, 1989). Six nucleosomes coil together, and these are then stacked on top of each other into a fiber of packed nucleosomes called chromatin. The chromatin is looped and packaged with other proteins into chromosomes (Figure 1-6). In general terms, two types of chromatin exist in the cell – densely packed heterochromatin and loosely packed euchromatin. The dense packing of heterochromatin excludes transcriptional machinery from these regions, and it is therefore a major regulatory factor for gene expression. Importantly, post-translational modifications on the tails of histones affect the state of compaction of the chromatin (Jenuwein and Allis, 2001; Vermaak et al., 2003). While this fact has major implications in the chromatin structure and gene expression, it has also been proposed that certain regions of the genome are available for the replication machinery to pass freely through,

Figure 1-5. Three-dimensional structure of a nucleosome. 147 bp of double helix DNA (green and gold helices) are wrapped nearly twice around the four core histones (red, orange, green and blue ribbons). The view on the left is looking down the superhelix axis, and the right is looking perpendicular to it. shown in ribbon structure. [From (Luger et al., 1997)]

while very tightly compacted regions are less accessible to the replication machinery. A handful of highly influential studies that measure the rates of DNA synthesis have supported this hypothesis by reporting a wide range of rates at different regions of the genome (Raghuraman et al., 2001; Yabuki et al., 2002). This view has settled in the literature (see for example (Antequera, 2004; Spiesser et al.; Tourriere and Pasero, 2007)), although the direct effect that heterochromatin has on progressing forks remains unclear.

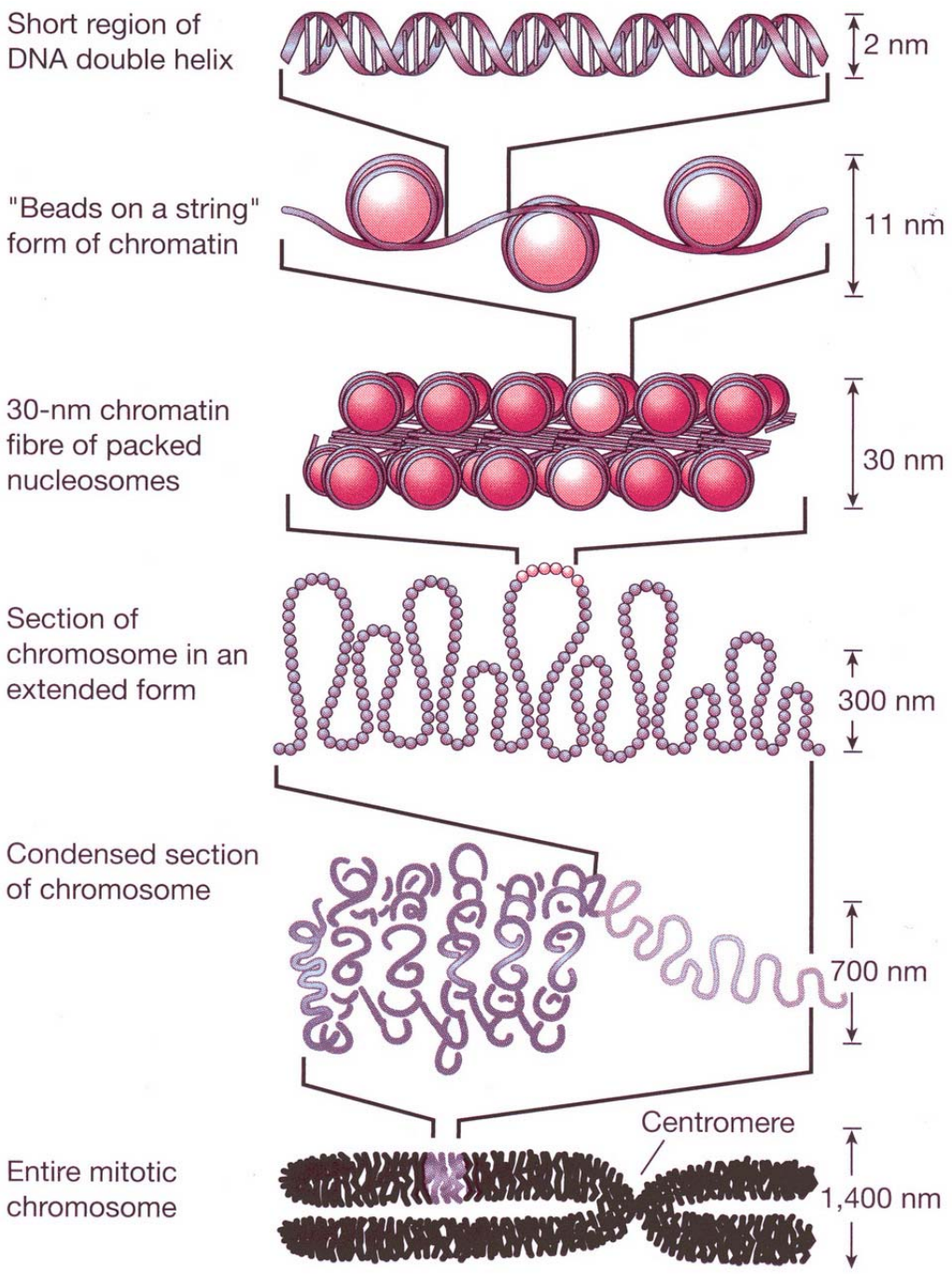


Figure 1-6. The different levels of chromatin compaction. Double-helix DNA is wrapped around histone octamers to form nucleosomes. Six nucleosomes are coiled together, and these are stacked on top of each other to form a 30 nm in diameter fiber called chromatin. Chromatin is further wrapped and condensed to form chromosomes. [From (Felsenfeld and Groudine, 2003)]

DNA replication is a fundamental aspect of life, and as such, likely evolved well before histones or other histone-like proteins began occupying the genome. Certainly, as life evolved and genomes became more and more complex, the replication machinery must have also co-evolved to cope with the sharp bending of DNA around the histone octamers. A major goal of this thesis will be to explore the extent to which heterochromatin and other obstacles affect the progression of the replication machinery. Previous genome-wide studies have suggested that some regions of the genome are more difficult to replicate than others, and have attributed this to heterochromatic impediment of the replication fork (Raghuraman et al., 2001; Spiesser et al., 2010; Tourriere and Pasero, 2007). Raghuraman et. al. monitored the incorporation of nucleotides during S phase by density transfer experiments and concluded that replication proceeds more slowly in certain regions of the genome, and that several of the origins showed asymmetrical fork migration. The results of this study will be discussed in detail in Chapter 5.

1.3 – ORIGINS OF REPLICATION & REPLICATION BUBBLES

Every time a cell divides, the genome must be precisely copied within the window of S phase. Even the simplest protists have genome sizes approaching 10 million base pairs, and the yeast genome contains ~12 million base pairs. To replicate an entire genome of that size would take several days if the process only initiated from a single site. In fact, several-hundred start points, called "origins of replication", are present in the yeast genome.

Identifying the determinants for origins in the genomes of various eukaryotic organisms has been a complicated endeavor. One defining characteristic of all eukaryotic origins is the presence of a set of six proteins called the Origin Recognition Complex (ORC) (Diffley, 2004). While the ORC proteins are very highly conserved, the DNA sequence that they bind to is not. In *Xenopus* egg extracts and *Drosophila* embryos the origins appear randomly located in the genome with respect to DNA sequence (Blow et al., 2001; Costa; Hyrien et al., 2003). Mammalian origins of replication seem to be localized to 'initiation zones' that can be several kb in length (Mesner et al., 2006). *S. cerevisiae* origins, on the other hand, contain a consensus sequence that the ORCs bind to called Autonomously Replication Sequence (ARS). At the core of the origin is a 17 bp, A-T rich consensus sequence (5'- T/A T T T A Y R T T T T T/A -3' (Bell, 1995; Newlon and Theis, 1993)), called an ARS Consensus Sequence (ACS), which is required for ORC binding to the origin.

The ORC proteins mark the origins, and recruit other replication factors in the steps leading up to S phase. The culmination of these events is the melting of the DNA at the origins, and ultimately in bi-directional movement of replication forks away from the

origin. This process is referred to as "origin firing", and the firing time can vary from origin to origin. Interestingly, the firing times of origins appear to cluster spatially in the genome, implying a spatial regulatory element exists (McCarroll and Fangman, 1988; Poloumienko et al., 2001; Raghuraman et al., 2001; Yamashita et al., 1997).

As many as 732 sites of initiation within the 16 chromosomes of the *S. cerevisiae* genome have been reported as potential origins of replication (reviewed in (MacAlpine and Bell, 2005), (Nieduszynski et al., 2007)). Three separate studies monitoring the production of newly replicated DNA identified respectively 332, 260 and 444 origins in the yeast genome (Feng et al., 2006; Raghuraman et al., 2001; Yabuki et al., 2002). In addition, experiments using chromatin immunoprecipitation combined with microarray analysis (ChIP-chip) to search for the genomic localization of MCM and ORC proteins identified 422 (Wyrick et al., 2001) and 529 sites (Xu et al., 2006). Thus, although it is likely that virtually all origins in *S. cerevisiae* have been identified, exactly which are active during S-phase remains open to debate.

Origin activation produces bi-directionally moving replication forks and expanding replication bubbles (Figure 1-7). As indicated, hundreds of these bubbles stud the genome during S phase, all merging to create two of every chromosome. The replication machinery slides along the template in the 3′ to 5′ direction, and can only synthesize DNA in the 5′ to 3′ direction. Therefore, the two melted strands of DNA at the replication fork must be synthesized in opposite directions. In the leftward moving fork in Figure 1-7, the DNA being synthesized on the top is labelled the leading strand, and the DNA being synthesized on the bottom is the lagging strand. Barring any problems, the leading strand can be synthesized continuously by a single polymerase. Conversely, the lagging strand is synthesized in the opposite direction in short bursts called Okazaki

fragments, which are annealed together. The process is referred to as semi-conservative because each of the resulting DNA molecules contains one newly synthesized strand, and a strand from the parent.

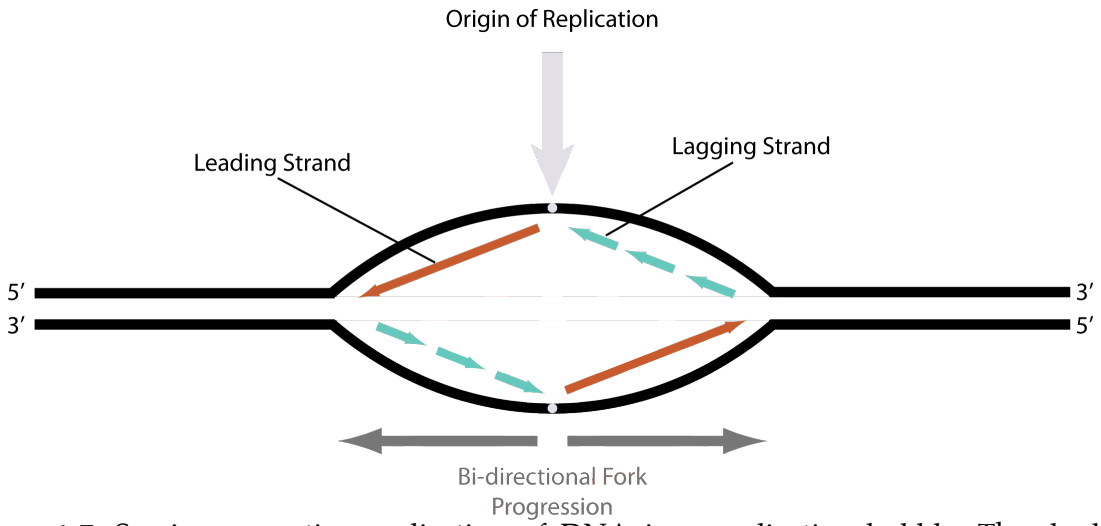


Figure 1-7. Semi-conservative replication of DNA in a replication bubble. The double-stranded DNA from the parent is pulled apart and the resulting two single strands are used as a template to replicate two new double-stranded molecules of DNA. DNA synthesis only happens in the 5′ to 3′ direction, thus the red arrows represent the leading strand synthesis, while the short, green arrows represent the Okazaki fragments on the lagging strand. One of the new molecules will remain with the parent cell, while the other is passed on to the daughter.

As a final note regarding replication forks, a long-standing hypothesis was recently confirmed when the Tanaka and Blow laboratories showed that many replication forks cluster together as "replication factories" (Kitamura et al., 2006). By their count, at least a dozen factories are present in the yeast nucleus, indicating that each contains at least fifty forks. These experiments suggest that forks that coalesce into factories are likely exposed to the same conditions and behave similarly. Interestingly, clusters of early and late firing origins are observed across the genome, and may be a by-product of this phenomenon (McCarroll and Fangman, 1988; Poloumienko et al., 2001; Raghuraman et al., 2001; Yamashita et al., 1997). Sister forks, generated from the same origin, were

shown to associate with each other during the replication process. It is somewhat surprising then, that asymmetric progression of sister forks away from the origins has been described, and used as evidence for certain regions being more difficult to replicate than others (Raghuraman et al., 2001). This study revealed that forks probably do not slide along a stable DNA substrate, as schematized in many textbook representations of the process, and in this thesis, for the purposes of simplification. Instead, many strands of unreplicated DNA enters the factories, and pairs of replicated sister strands are extruded. A large number of proteins are required for replicating each strand, and therefore these factories are very likely enormous clusters of DNA and replication fork proteins, each of which contribute an important role in the replication process.

1.4 - The Composition & Behavior of Eukaryotic Replication Forks

The eukaryotic replication fork has yet to be reconstituted *in vitro*. The prevailing explanations for this fact are 1) not all of the essential components have been identified, 2) specific post-translational modifications are required for them to function, and 3) a specific sequence of association of the players is critical, or 4) some combination of these possibilities. Conceptually, only a few enzymes are required for replication to occur - a helicase enzyme for unwinding the DNA (the MCM complex in yeast (Forsburg, 2004)), topoisomerases to remove negative supercoils from the annealed DNA, and polymerase enzymes that incorporate nucleotides into the new DNA strand (Figure 1-8). Of course, the entire picture is much more complicated and a number of accessory proteins are also required to initiate and maintain active replication.

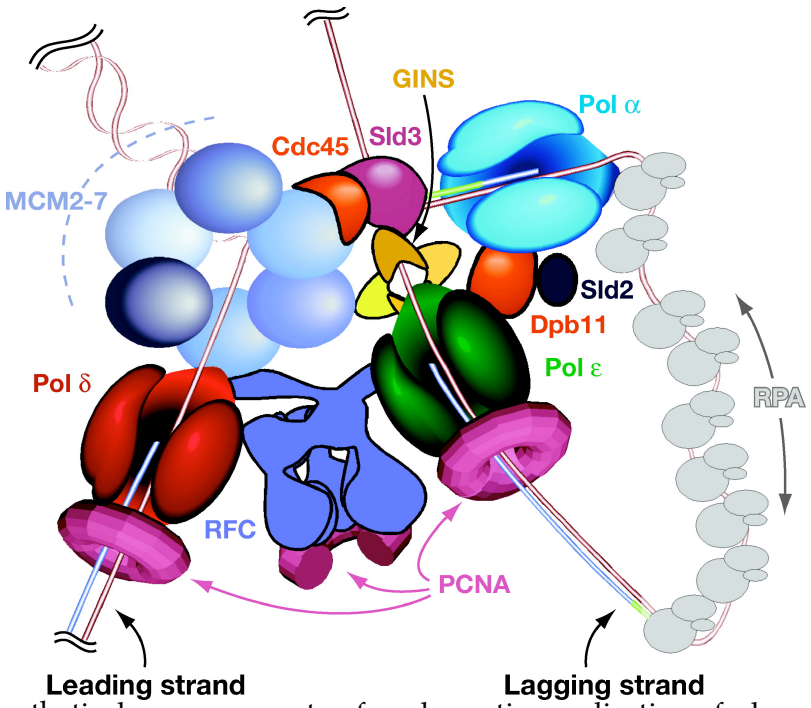


Figure 1-8. Hypothetical arrangement of eukaryotic replication fork proteins. RFC is coordinating the loading of the sliding clamp (PCNA) with the polymerases Pol ε and Pol δ. MCM2-7 is the putative helicase, while GINS, and Cdc45 are accessory proteins [From (Johnson and O'Donnell, 2005)]

Other essential proteins associated with the replication fork include single-stranded binding proteins (RPA) to prevent the re-annealing of the melted DNA, sliding clamp proteins that hold the polymerase enzymes to the DNA (PCNA), a clamp loader to load the sliding clamp with the polymerases (RFC), and a number of ancillary proteins that appear to be required for helicase activity (Cdc45 and the GINS complex).

1.4.1 - The Yeast Polymerases

Replicating the genome requires the accurate and coordinated copying of both the leading and lagging strand templates by DNA polymerases. In prokaryotic organisms, a single enzyme is responsible for replicating both strands (e.g., DNA Polymerase III in *E. coli*). At least three enzymes are required for the synthesis of DNA in eukaryotic systems like *S. cerevisiae* (Figure 1-8). Specifically, Pol α, Pol δ, and Pol ε coordinate to synthesize two new strands of DNA every round of cell division.

Pol α has both RNA primase and DNA polymerase activity within the four subunits of the complex (Garg and Burgers, 2005; Johnson and O'Donnell, 2005), but it lacks a high degree of processivity. It therefore has the unique ability to lay down RNA primers followed by short strands of DNA. On the leading strand, it performs this function only near origins, but on the lagging strand this happens continuously in the form of Okazaki fragments, as replication is moving in the opposite direction of the progressing fork (Figure 1-7).

The Pol ε holoenzyme consists of four polypeptide subunits – Pol2, Dpb2, Dpb3, and Dpb4 (Figure 1-9 middle). The essential POL2 gene encodes the catalytic subunit Pol2. ChIP studies indicate that Pol ε is loaded onto the origins prior to S phase and moves away upon origin firing (Aparicio et al., 1997; Feng et al., 2003) and reviewed in (Johnson and O'Donnell, 2005). Surprisingly, the entire catalytic domain has been found to be non-essential for growing cells (Dua et al., 1999; Feng and D'Urso, 2001; Kesti et al., 1999). Dpb2, Dpb3, and Dpb4 all appear to provide some stabilizing function for the Pol ε polymerase activity, as well as other chromatin remodelling functions (Tackett et al., 2005). The Pol ε holoenzyme seems to have a fairly high degree of processivity even

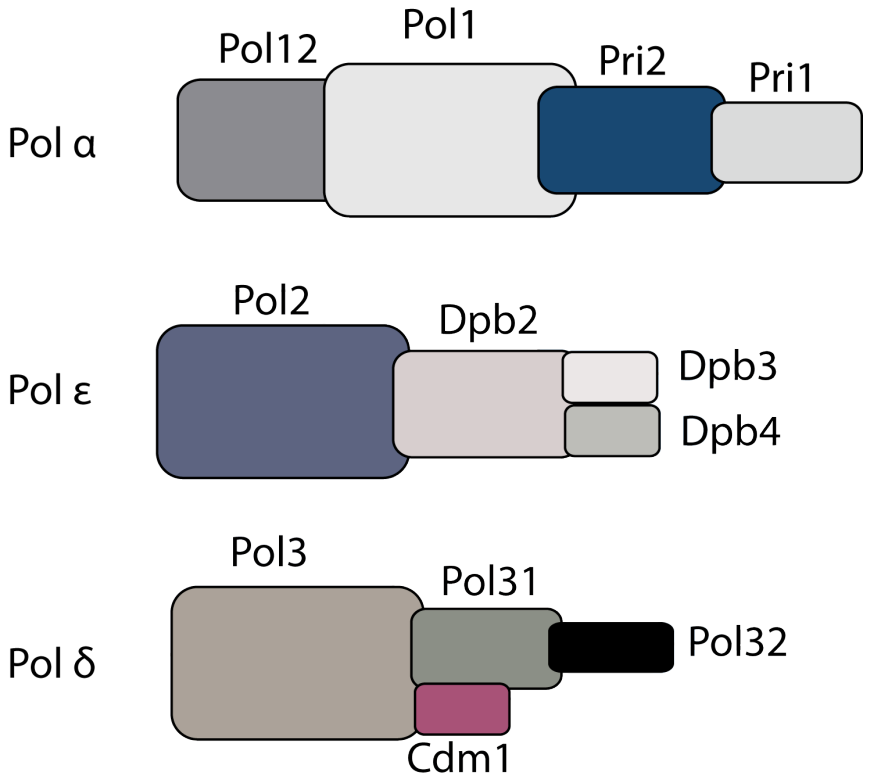


Figure 1-9. The composition of the three replicative polymerases present in yeast. **Top -** Pol α contains four subunits; the catalytic Pri1, Pri2, Pol1, and Pol12. **Middle -** Pol ε contains four subunits; the catalytic Pol2, Dpb2, Dpb3, and Dpb4. **Bottom -** Pol δ contains three subunits; Pol3, Pol31, and Pol32 (Cdm1 is only present in humans and *S. pombe*. [Adapted from (Garg and Burgers, 2005)].

without PCNA, although the sliding-clamp confers some increase in fidelity, and appears to play a role in DNA repair (Dua et al., 2002).

In *S. cerevisiae*, Pol δ consists of three proteins (Figure 1-9 bottom) – Pol3, Pol32, and Pol31. Humans and *S. pombe* both have a fourth Pol δ subunit called Cdm1 that is not present in yeast. This subunit does not appear to be essential for DNA replication (Reynolds et al., 1998). The catalytic subunit for Pol δ is encoded by the essential POL3 gene. Pol δ interaction with PCNA on DNA conveys a high degree of fidelity that Pol δ, unlike Pol ε, does not possess on its own.

The exact roles of the two processive polymerases, Pol δ and Pol ε, in replicating the genome are still being debated (reviewed in (Pavlov and Shcherbakova)). The Kunkel lab has shown that an error-prone mutant of Pol2 has a strong tendency to make T to A substitutions on the leading strand (Pursell et al., 2007), while a similar Pol3 mutant tends to misincorporate on the lagging strand (Nick McElhinny et al., 2008). This strongly suggests that Pol ε is exclusively responsible for synthesizing the leading strand, and Pol δ the lagging strand (Figure 1-10 top).

Unfortunately, the design of these experiments was such that replication was monitored only near the origins. This, and the fact that the catalytic portion of Pol2 has been shown to be non-essential has lead to an alternate hypothesis in which Pol ε is loaded onto the genome as replication initiates, but Pol δ takes over for Pol ε at sites where the forks stall (Figure 1-10 bottom) [(Stillman, 2008) and personal communication with Dr. Stillman].

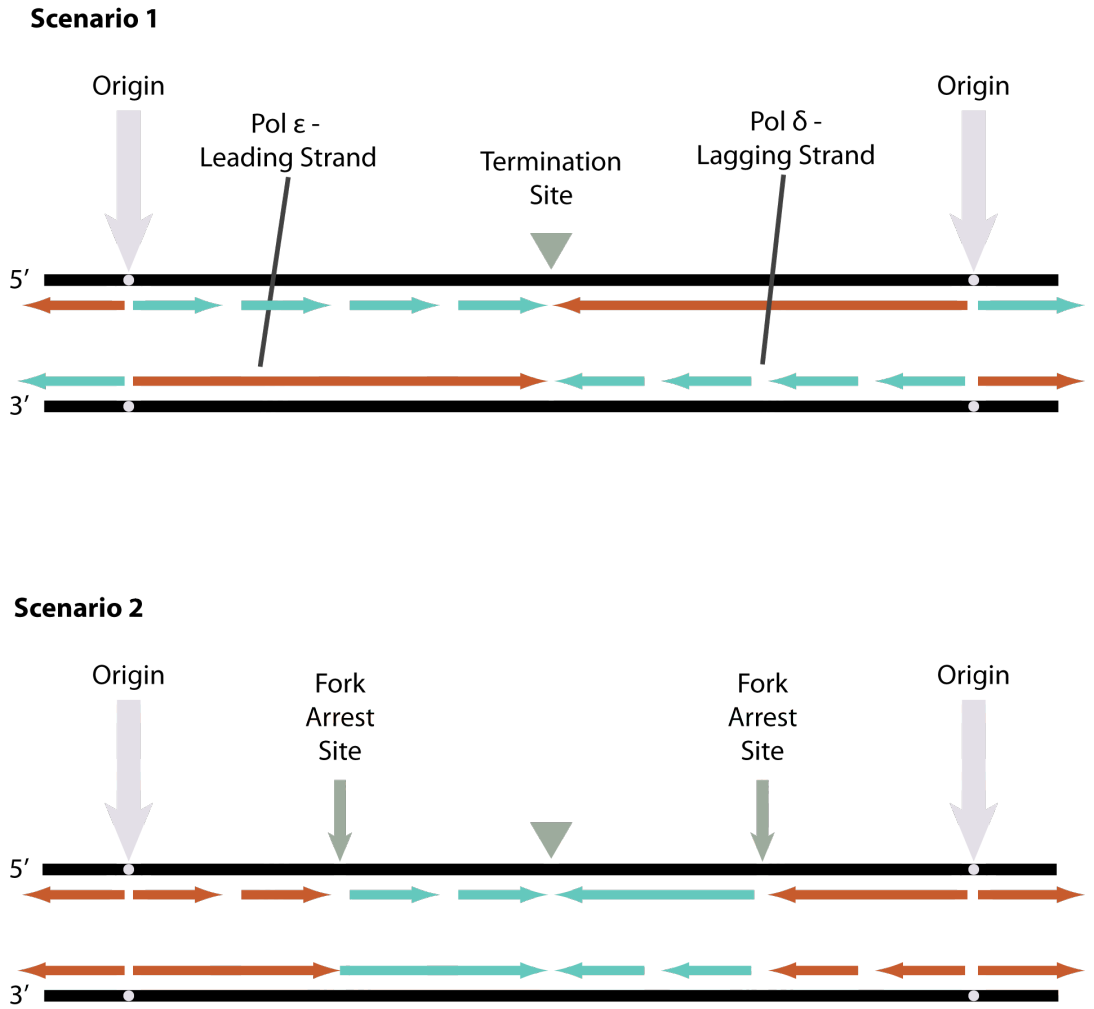


Figure 1-10. The two most plausible scenarios for polymerase usage based on the current data. **Scenario 1 –** Pol ε and Pol δ are each dedicated to a discrete strand. Based on work of the Kunkel lab, Pol ε is most likely responsible for polymerizing the leading strand, while Pol δ is responsible for the lagging strand. **Scenario 2 –** Pol ε is loaded to the origins and progresses with the fork until a difficult to replicate region is reached. At this point, Pol δ takes over and finishes the job. The observable difference between these two scenarios is in scenario 1, Pol can be seen across the genome, whereas it is localized to regions surrounding the origins in scenario 2.

Perhaps the process is so important that nature has built-in redundancy, and Pol ε can be compensated for by Pol δ. To clarify the roles of these two enzymes, one could expand on the referenced mutational studies (Nick McElhinny et al., 2008; Pursell et al., 2007) by moving the reporter genes farther away from the origins. Alternatively, simply monitoring the progression of each of these enzymes across the genome would reveal whether Pol ε is localized to the regions surrounding the origins, or if it progresses with the helicase enzymes to sites of replication termination.

1.4.2 - The Yeast Helicase

Genomic replication requires a single-stranded DNA template. An enzyme that melts double-stranded genomic DNA is essential for producing this template. In general, enzymes that carry this out are referred to as helicases. In yeast, there are 134 putative helicase open reading frames (Shiratori et al., 1999), but the most likely candidates for replicative helicases are six paralogous proteins called Mcm2-7 - collectively referred to as the MCM complex (Forsburg, 2004). These proteins are conserved in all eukaryotes, and are essential for the initiation of DNA replication.

The three dimensional structure of the eukaryotic MCM complex has yet to be solved, but electron microscopic studies of the highly orthologous MCM protein from the archaeon *Methanothermobacter thermautotrophicus* (*Mth*MCM) suggest that it forms a hexameric ring with a 3 - 4 nm pore, wide enough to allow a single or double-strand of DNA to pass through. *Mth*MCM has also been show to bind to DNA as a double hexamer, and to slide along the DNA after loading (Evrin et al., 2009).

MCM complex activity is regulated by the cell cycle, and is activated by S phase cyclin dependant kinases (including Cdk2 and Cdc7) (Sheu and Stillman, 2010). That two hexameric complexes are loaded onto the origins is in agreement with the concept of a single helicase moving in each direction away from the origin to unwind the DNA (Evrin et al., 2009). However, relative to the number of potential replication forks, there exists a vast overabundance of MCMs within cells. In *S. cerevisiae*, the number of MCMs is estimated to be ~30,000 copies per cell – a ratio of ~100 MCMs per origin (Forsburg, 2004). In human cells a majority of MCM complexes are distributed along the genome, and are not localized to sites of replication. In the literature, this excess is referred to as the "MCM paradox" – the concentration of MCMs can be drastically decreased without affecting DNA replication. However, lowered amounts of MCMs do have a negative effect on the cell, implying that MCMs are doing more than just unwinding DNA (reviewed in (Forsburg, 2004)). Possible roles for MCMs outside of replication include transcription (DaFonseca et al., 2001; Yankulov et al., 1999; Zhang et al., 1998), chromatin remodelling (Ishimi et al., 1996; Ishimi et al., 1998; Ishimi et al., 2001), and the checkpoint response (Ishimi et al., 2003; Liang et al., 1999).

Setting aside for the moment these alternate roles for MCMs, its function as the replicative helicase has been shored up in recent years. This has mainly come about from biochemical studies that finally uncovered helicase activity *in vitro* when co-purified with several accessory proteins (Ilves et al., 2010; Moyer et al., 2006).

1.4.3 - Helicase Accessory Proteins – The CMG Complex

While the MCM complex displays many characteristics of a DNA helicase enzyme (including ATPase activity, a structure that allows double stranded DNA to be passed through a 3 – 4 nm pore, and sequence homology to prokaryotic replicative helicase enzymes), *in vitro* helicase activity eluded researchers for many years. This is most likely because other accessory proteins or post-translational modifications required for the activation of this function need to be identified. Very recently, helicase activity was isolated from insect cells in a fraction that contained all seven subunits of the MCM complex, Cdc45, and the four subunits of the GINS complex (Moyer et al., 2006). Together these proteins have been labelled the CMG complex (Cdc45, MCM, GINS).

The GINS complex is a component of the pre-RC, and consists of a highly conserved set of proteins - Psf1, Psf2, Psf3 and Sld5 (Kanemaki et al., 2003; Kubota et al., 2003; Takayama et al., 2003). The four subunits form a globular complex with a small central pore (Figure 1-11). All of the subunits share a very weak homology to each other, probably representing paralogues from a single gene that was duplicated twice. In the archaeal organism *Solfolobus solfataricus*, two proteins called Gins51 and Gins23 were identified (Makarova et al., 2005; Marinsek et al., 2006). Gins51 shows a domain organization that is similar to both Psf1 and Sld5, while Gins23 has an organization that resembles Psf2 and Psf3. The archaeal GINS has been shown to directly interact with the primase and helicase, implying a role in the initiation and maintenance of replication forks.

A similar role in eukaryotic systems is very likely, as many links to the helicase and replication fork have been observed. All four of the GINS subunits are essential for cell survival, and inactivation of the individual subunits by temperature sensitive mutation show that the genes are essential for the initiation of DNA replication (Kanemaki et al., 2003; Takayama et al., 2003). In *Xenopus* egg extracts, GINS has been shown to localize at sites of unwound DNA (Pacek et al., 2006), suggestive of a role at the replication fork. In yeast, the complex associates with paused replication forks (Calzada et al., 2005), and directly interacts with several fork proteins (Gambus et al., 2006) in a complex called the Replisome Progression Complex (RPC). ChIP experiments show that the GINS complex moves away from specific ARSs at the time of initiation (Kanemaki and Labib, 2006; Takayama et al., 2003). As mentioned, biochemical fractionation of the CMG strongly implicate GINS as supporting the helicase activity of the MCM complex (Moyer et al., 2006). Taken together, these data suggest that the GINS complex is an integral component

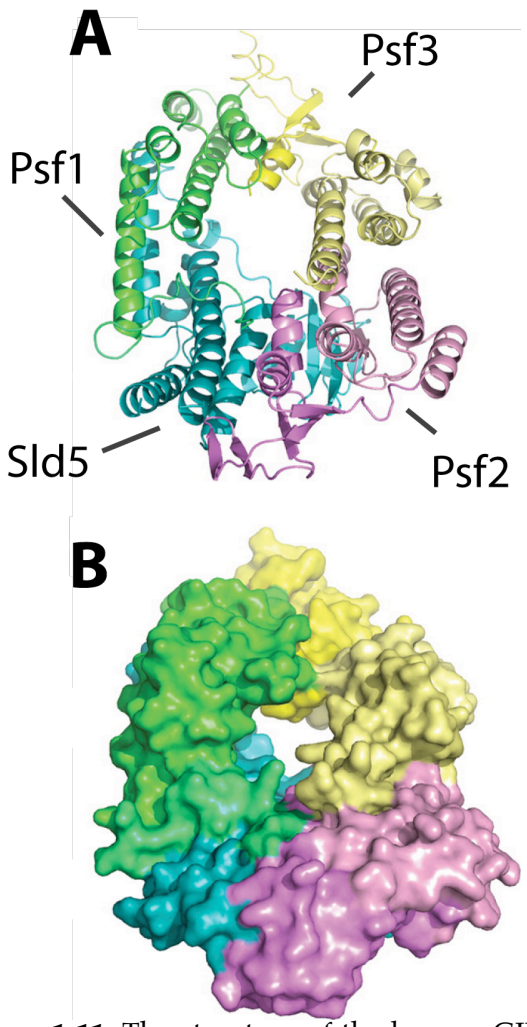


Figure 1-11. The structure of the human GINS tetramer. **A.** A ribbon model view of the central pore. **B.** A space filling model of the GINS tetramer, looking into the central pore. [From (Chang et al., 2007)]

of the replication fork, and that its interaction with the genome correlates directly to the movement of the fork (reviewed in (Labib and Gambus, 2007) and (MacNeill, 2010)).

1.4.4 – DNA Replication Fork Progression

The progressing replication forks must cope with a number of obstacles in the genome – including the hydrogen bonds that hold the DNA together, the topology of the helical DNA, and the proteins embedded in the chromatin. The timing of origin firing and the rates of fork progression have been investigated by monitoring nascent DNA synthesis (Raghuraman et al., 2001; Yabuki et al., 2002). In one approach, newly synthesized DNA was isotopically labelled and separated from unreplicated DNA before being subjected to microarray analysis (Raghuraman et al., 2001). The results are represented as a replication-timing curve, indicating the time that every microarray element is replicated. An example of what curves like this look like is shown in Figure 1-12. In this study, origin firing was observed to occur as early 14 min into the cell cycle and as late as 44 min. A wide range of nucleotide incorporation rates (0.5 – 11 kb/min) was observed, with a mean of 2.9 kb/min.

A second study monitored the copy number of each piece of DNA with microarrays (Figure 1-12, light grey line), reasoning that replicated DNA would yield twice as intense of a signal on the array. This study reported a mean rate of DNA replication of 2.8 +/- 1.0 kb/min (Yabuki et al., 2002). In addition to these velocities, replication has been inferred to progress asymmetrically from certain origins (Raghuraman et al., 2001). These data have been interpreted to mean that the dynamics of replication fork progression are strongly affected by local chromatin structure or architecture, and perhaps by interaction with the machineries controlling transcription, repair and

epigenetic maintenance (Antequera, 2004; Deshpande and Newlon, 1996; Ivessa et al., 2003; Raghuraman et al., 2001; Rothstein et al., 2000; Spiesser et al., 2010; Tourriere and Pasero, 2007).

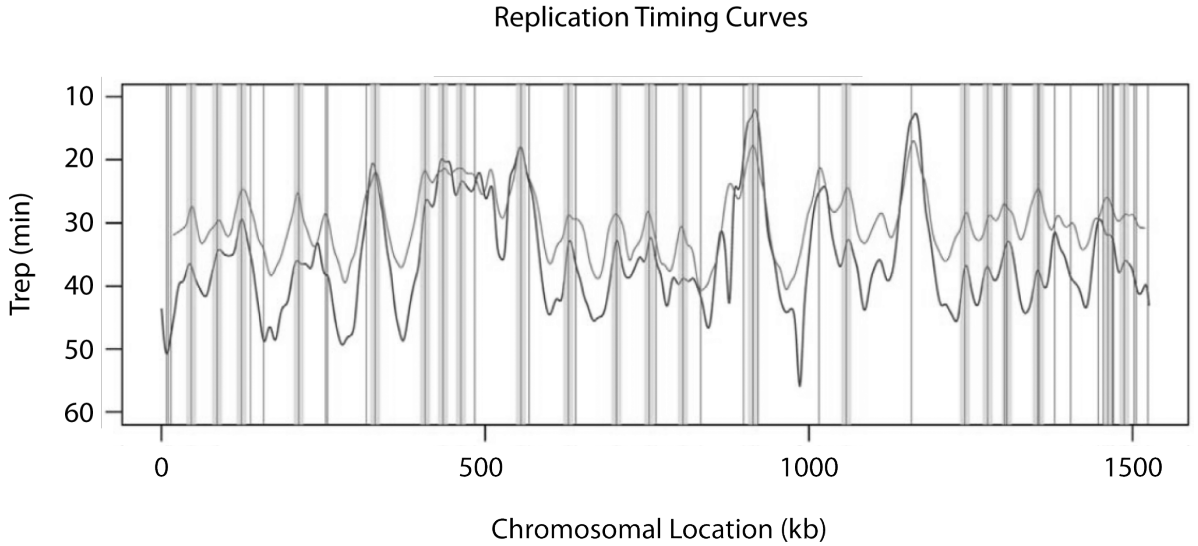


Figure 1-12. Example of Replication Timing. The replication timing curve for chromosome 4 for heavy: light analysis (dark gray line (Raghuraman et al., 2001)), and for copy number analysis (light gray line (Yabuki et al., 2002)). Regions replicated first (origins) are higher up the y-axis, while late replicating regions are at the bottom. The vertical lines indicate the locations of the origins identified by these studies. (Nieduszynski et al., 2007). [From (MacAlpine and Bell, 2005)]

1.5 - DNA Replication & the Cell Cycle

In mitotic division, cells duplicate their DNA in S-phase to ensure that the proper genetic material is passed on to their progeny. This process of DNA replication is initiated from several hundred specific sites, termed origins of replication, dispersed across the genome. It is essential for replication to initiate only once, and at the right time (during S phase), and it must finish prior to the initiation of anaphase to ensure that the genomes are properly segregated (Blow and Dutta, 2005; Machida et al., 2005). To ensure proper timing, the beginning stages of DNA replication are tightly coupled to the cell cycle through the activity of the cyclin-dependent kinases CDK and DDK in a process referred to as licensing (Masumoto et al., 2002; Nguyen et al., 2001; Sclafani and Holzen, 2007; Sheu and Stillman, 2010).

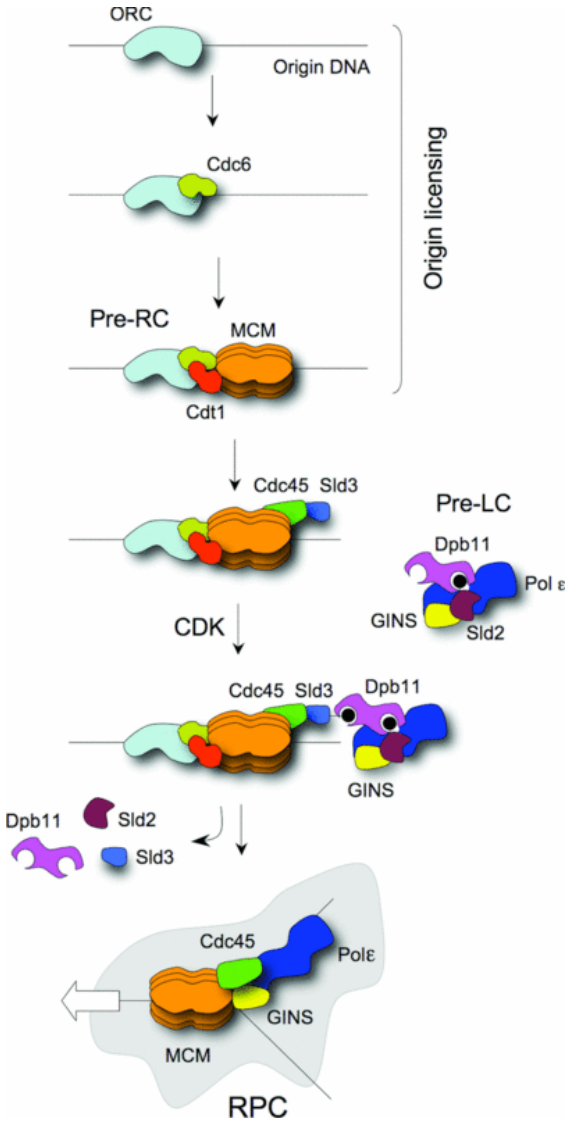


Figure 1-13. Licensing of Replication Origins in *S. cerevisiae*. A model for the stepwise assembly of the replication fork. [From (MacNeill, 2010)]

Licensing is the stepwise assembly of proteins at the origins of replication (Figure 1-13). It is initiated by the recruitment of Cdc6 to the ORC complex

at the origins in G1 (Aparicio et al., 1997). In turn, Cdc6 recruits Cdt1 and the Mcm2-7 complex, forming the licensed origin and a complex referred to as the pre-replication complex (pre-RC) ((Hua and Newport, 1998; Nishitani et al., 2000; Rowles et al., 1999) and reviewed in (Blow and Dutta, 2005),(Nishitani and Lygerou, 2004), and (Mendez and Stillman, 2003)). The pre-RC matures into the pre-initiation complex (pre-IC) upon the recruitment of several other replication fork-associated proteins, including Cdc45, Sld3, and the GINS complex (Blow and Dutta, 2005). These steps are regulated by both the CDK and DDK cell-cycle kinases (Bell and Dutta, 2002; Diffley, 2004; Masai and Arai, 2002; Sclafani, 2000; Sclafani and Holzen, 2007; Stillman, 1996).

All aspects of the cell cycle, including the licensing reaction (Figure 1-13), are controlled by the oscillating activity of CDK (reviewed in (Murray, 1989, 2004) Figure 1-14). Through most of the cell cycle, CDK activity prevents the initial steps of the licensing reaction from occurring by inhibiting the binding of Cdt1 to the ORC complex and

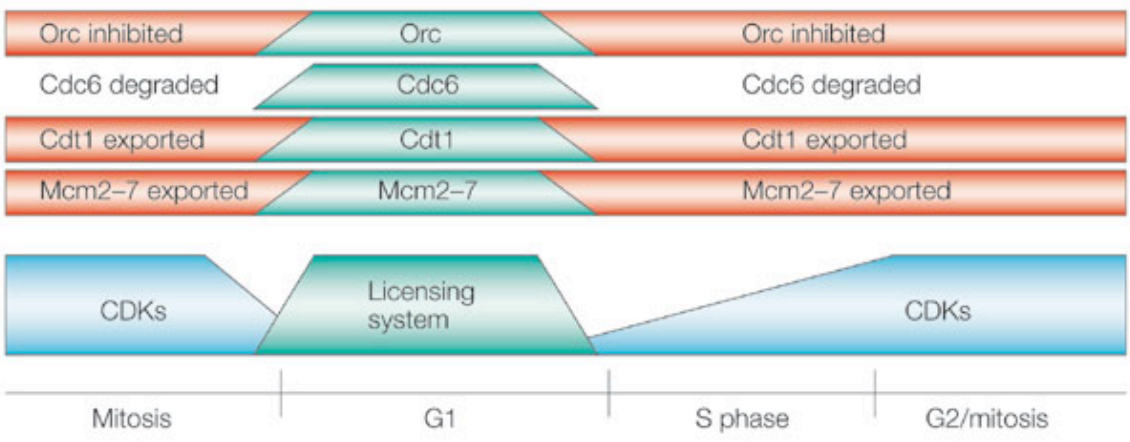


Figure 1-14. Licensing of Replication Origins in *S. cerevisiae*. CDK activity is high throughout most of the cell cycle, inhibiting the licensing of origins through a variety of mechanisms. It decreases during G1, promoting the licensing of origins, and inhibiting active replication. As S phase commences, CDK activity returns, preventing the re-loading of the origins, but promoting active replication. [From (Blow and Dutta, 2005)]

exporting Cdt1 and the MCM proteins from the nucleus (Bell and Dutta, 2002; Blow and Dutta, 2005). When cells enter a new round of mitosis (early G1), CDK activity decreases, which allows Cdc6 to bind to the ORC complex, and imports Cdt1 and the MCM proteins back into the nucleus. The licensed origins are activated through the phosphorylation of Mcm4 by DDK (Sheu and Stillman, 2010), which allows the binding of Sld3, Cdc45, and the GINS complex. These steps occur while CDK activity is still low at the end of G1. CDK activity is essential for the firing of the origin, which therefore occurs at the beginning of S phase when CDK activity rises once again (Figure 1-14) (Tanaka et al., 2007; Zegerman and Diffley, 2007).

1.6 - THE IMPORTANCE OF GENOMIC REPLICATION

1.6.1 – The Basic Components of Life and Eukaryotic Development

One of the great mysteries in biology is how life on Earth began some 3.8 billion years ago. Life as we know it today requires a blueprint in the form of a genome, and the genome in turn encodes for proteins that carry out biological functions. Proteins are necessary to replicate a genome, creating a paradox as to how the first genome could replicate itself. The best guess is that primitive blueprints were actually made from genetic heteropolymers that can both catalyze metabolism and confer heredity – an example of which is RNA, leading to the RNA world hypothesis. The precursors to prokaryotic cells are referred to as protobionts, and were able to metabolize chemicals from the environment and replicate themselves. They were solely composed of the two essential components for the "simplest" life - 1) genetic polymers and 2) membranes that separate the genetic material from the environment. Thus, copying the blueprints is one of the most fundamental aspects of life.

Protobionts gave way to prokaryotes, most of which have smaller, circular genomes containing only a single site for initiation of replication. When multicellular eukaryotes evolved DNA replication took on new responsibilities – replicating much larger genomes that were split into several chromosomes, each with multiple sites of initiation. Instead of replicating the genome for progeny, replication in mitotically-dividing somatic cells yields copies of the cells, and replication in meiotically-dividing sex cells produces gametes.

In humans, fertilization of a haploid egg cell by a haploid sperm cell sets off a cascade of events that culminates in the emergence of an adult human consisting of 10^{13} cells. How a single zygocyte, formed from the fusion of the two gametes, passes an exact copy of the genetic information to each cell over the course of 5 trillion cell divisions (DePamphilis, 2006) is a testament to the efficiency and elegance of the DNA replication process. A major difference between replication in unicellular versus multicellular organisms is that in the latter it is linked not only to cell growth, but also to differentiation. This may partially explain the fact that sequence specific origins are found in yeast, whereas multicellular organisms have evolved a more liberal origin usage.

1.6.2 - When Eukaryotic Replication Goes Bad

Cancer is a collection of diseases with a variety of associated etiologies. Needless to say, the molecular causes of these diseases are diverse and have proven difficult to remedy. Still, a single attribute underlies all cancers – namely, uncontrolled cellular growth leading to malignancy. Faithful replication of cellular DNA is part of the process that ensures normal cell growth is maintained, and defects in the DNA replication machinery can therefore lead to aberrant cell proliferation. The major causes of replication defects that result in uncontrolled cellular proliferation are the misincorporation of nucleotides, and damage or modifications to DNA.

In humans, the spontaneous mutation rate is estimated to be between 10^{-9} and 10^{-11} mutations/base pair/replication (Drake et al., 1998; Loeb, 2001). The events that ensure this extraordinary accuracy include the high degree of nucleotide selectivity of the replicative DNA polymerases, and the ability of these enzymes to proofread and correct

mismatched base pairs during the replication process, minimizing the possibility of detrimental lesions or point mutations in the genome. Despite the fine-tuned nature of this process, problems can arise that decrease fidelity, and affect the stability of genomes. If a DNA polymerase incorporates an incorrect nucleotide, and continues to extend the mismatched strand, a lesion forms and a point mutation may arise in the next round of synthesis. The major replicative polymerases - Pol ε and Pol δ - have 3' exonuclease activity that is responsible for correcting these errors, but studies of mice that are deficient in this proofreading activity show increased rates of tumorigenesis (Goldsby et al., 2002).

Overall, the process of unwinding and replicating DNA renders it particularly susceptible to damage in the form of lesions and double-stranded breaks. In turn, these features stall replication fork progression, preventing the complete replication of the genome. Cells have multiple checkpoint and repair pathways to correct these problems when they arise, and defects in any one of the repair pathways can result in decreased genome stability. For example, the ATM gene encodes for a protein that senses DNA damage associated with double-stranded breaks (Sancar et al., 2004). Mutations in this gene lead to problems in the DNA damage response, predisposing the patients to leukemias and lymphomas in a disease named ataxia-telangiectasia.

Clearly, the process of DNA replication is an integral part of genomic stability. Many other devastating diseases are the result of complications in the replication cycle, including Fanconi Anemia, Bloom's Syndrome, and Werner's Syndrome. All of these are rare, autosomal recessive conditions, and interestingly, can lead to increased rates of cancer or the appearance of aging in carriers. Understanding the process of DNA replication therefore goes beyond basic curiosity to appreciate how cells work - it is

essential for understanding the causes of cancer and other diseases associated with genomic instability.

CHAPTER II – PROTEIN-PROTEIN INTERACTIONS OF REPLICATION FORK COMPONENTS

Within cells, proteins carry out a vast assortment of cellular functions. In most, if not all cases, two or more proteins must interact with each other for their role in the cell to be fulfilled. These multiprotein complexes can be thought of as molecular machines, and can be extraordinarily large as in the case of the nuclear pore complex (456 individual proteins in yeast, totalling ~50 MDa) (Alber et al., 2007). Yet enzymes, in many cases, are composed of a single catalytic subunit bound to a handful of regulatory or auxiliary proteins.

The individual components of the replication fork are multiprotein complexes that range in size from four to seven polypeptide chains (as discussed in Chapter 1; Figure 1-8). The smaller component complexes interact with each other during S phase to form larger functioning complexes. For example, the homotrimer PCNA interacts with the three-subunit Pol δ, equipping it with the requisite processivity to fulfill its role in copying the genome (Burgers, 1991; Langston and O'Donnell, 2008). To do so, PCNA acts as a sliding clamp (Georgescu et al., 2008) and reviewed in (Jeruzalmi et al., 2002), (McHenry, 2003), and (Johnson and O'Donnell, 2005)) - wrapping itself around single-stranded DNA and forming a tight bond with the polymerase enzyme.

Because PCNA encircles single-stranded DNA, and the polymerases require a single-stranded template, a replicative helicase must unwind and melt the duplex genomic DNA in front of the PCNA-polymerase complex. The six-subunit MCM complex appears to possess this enzymatic activity only when it interacts with the four-subunit

GINS complex and Cdc45 (collectively, this is referred to as the CMG complex) (Ilves et al., 2010; Moyer et al., 2006). Thus, a minimum of eleven proteins must be in physical contact at the replication fork to produce unwound DNA, yet the details of the interactions that underlie this important complex are poorly understood. Moreover, the CMG and the replicative polymerases must coordinate with each other to some degree in order to minimize the amount of exposed, single-stranded DNA in the genome during replication.

The characterization of protein-protein interactions of these types is feasible due to advances in protein isolation and biological mass spectrometry (Beavis and Chait, 1996; Krutchinsky et al., 2001). A general workflow for such an experiment is outlined in Figure 2-1. Briefly, tiny magnetic beads (made from polystyrene embedded with paramagnetic particles) that are conjugated with the appropriate antibody can be used to quickly and directly isolate a genomically-tagged protein (the bait) from the cellular milieu. Under appropriate conditions, proteins that interact with the bait protein inside the cell will co-isolate as a complex. Peptide mass fingerprinting and mass spectrometric sequence analysis can then be used to identify these proteins from protein databases (Zhang and Chait, 2000). This powerful tool can be used to examine the function of proteins in the cell by characterizing their interaction partners. For example, the interactions of Pol ε have been previously characterized in this way (Iida and Araki, 2004; Tackett et al., 2005). In this chapter, the protein-protein interactions of GINS and Pol δ - two proteins that play key roles in the unwinding and synthesis of DNA, respectively - are described.

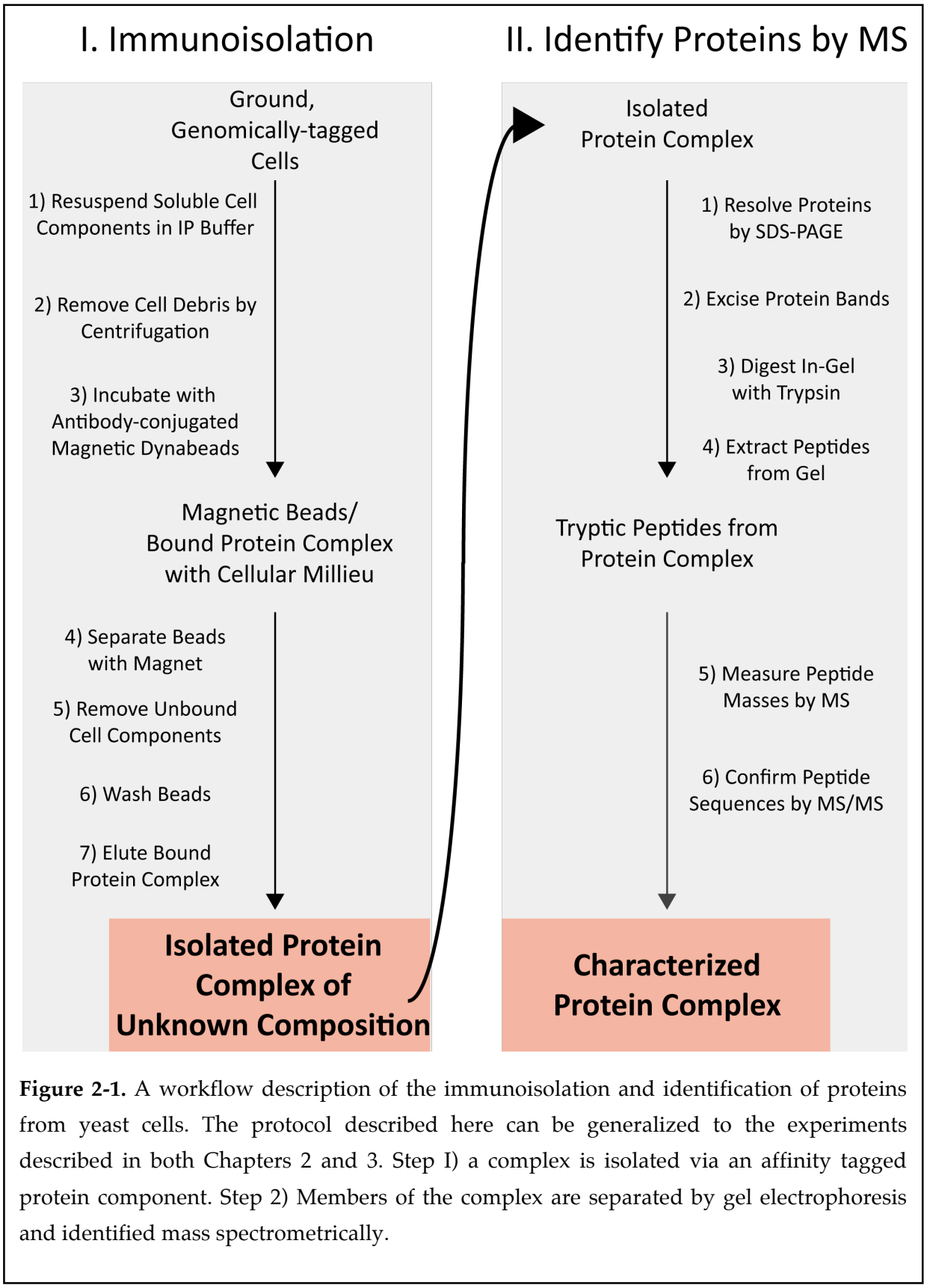


Figure 2-1. A workflow description of the immunoisolation and identification of proteins from yeast cells. The protocol described here can be generalized to the experiments described in both Chapters 2 and 3. Step I) a complex is isolated via an affinity tagged protein component. Step 2) Members of the complex are separated by gel electrophoresis and identified mass spectrometrically.

2.1 – Proteomic Characterization of the GINS Complex & its Binding Partners

Materials & Methods

Strain Construction, Storage, and Growth Conditions

S. cerevisiae strains are from the W303-1A background (see Appendix A for a list of strains used in this work, and their genetic background). The strains in use in this section (ATY112, ATY113, and ATY114) contain a genomic PrA tag at the 3' end of the ORF of each of the genes encoding the four subunits of the GINS complex (SLD5-PRA, PSF1-PRA, and PSF2-PRA, respectively). Stocks of the cells are stored indefinitely at -80°C in YEPD (10 g/L yeast extract, 20 g/L Bacto-peptone, 20 mg/L uracil, 80 mg/L tryptophan, 40 mg/L adenine hemisulfate, 20 g/L D-glucose) and 25% glycerol. When required, strains are streaked to SD –his plates (20 g/L agar, 6.7 g/L yeast nitrogen base without amino acids, 1.6 g/L yeast synthetic drop-out media minus histidine, 20 g/L D-glucose) and grown at 30°C for 2 days, or until a thick streak is observed. Plates are stored at 4°C and used to inoculate cultures for experiments for no longer than one month before re-streaking from frozen stocks. Typically, a small amount of cells is cultured in 2 mL YEPD at 30°C overnight, reaching a density of ~$2x10^8$ cells/mL. This culture is then diluted 1:100 in 5 mL YEPD and incubated at 30°C for 5 hr, until the density reaches mid-log, ~$2x10^7$ cells/mL. Finally, these cells are used to inoculate large batches of YEPD (4 – 12 L), by calculating the number of cells from the inoculum required to achieve a mid-log density culture the following morning (assuming a 90 min doubling time at 30°C).

Harvesting Cells and Cell Lysis

Once the cells reach mid-log density (~$2x10^7$ cells/mL), the culture is centrifuged for 10 min at 4°C and 10,000 RPM. The supernatant is discarded and the pellet washed with sterile, de-ionized water. The cells are centrifuged again for 5 min at 4°C and 3000 RPM. The supernatant is discarded and the pellet re-suspended in Yeast Lysis Buffer (20 mM hepes pH 7.5, 1.2% polyvinylpyrrolidone, 0.2 mg/mL PMSF, 4 μg/mL pepstatin A) at a ratio of 100 μL/g of cells (weight of the cells is an estimate based on the volume of the pellet in the tube). The cells are then dripped into liquid nitrogen, forming small frozen balls that are stored at -80°C.

The frozen cells are cryolyzed (ground at liquid nitrogen temperatures) by using a MM400 Mixer Mill (Retsch). To achieve complete lysis and minimize losses, ~4 g of cells are placed in steel jars that are pre-frozen in liquid nitrogen for 5 – 10 min. A single 5 mm stainless steel ball bearing is added and the canister oscillates at 30 Hz for 3 min. This is repeated a total of five times, with incubation of the entire canister in liquid nitrogen in between rounds of grinding. The momentum imparted to the grinding balls by the high frequency oscillations force the grinding balls to crush the samples with high energy against the rounded ends of the canisters. A small amount of powder is checked for complete lysis, as evidenced by completely broken cells with few if any intact cells in the field of a 40x objective on a phase contrast microscope (Zeiss). In all experiments the degree of lysis is estimated to be ≥95%.

GINS Complex Immunoisolation

The GINS complex is immunoisolated via the PrA affinity handle that is genomically encoded at the C-terminus of each individual subunit (Sld5, Psf1, Psf2 in strains ATY112, ATY113 and ATY114, respectively). Typically, 2 – 3 g of frozen powder is resuspended in 5 mL GINS IP Buffer per 1 g of cell grindate. The buffer includes 20 mM hepes pH 7.5, 2mM $MgCl_2$, 0.1% tween-20, 1/100 (v/v) solution P (20 mg/mL phenylmethylsulfonyl fluoride, 0.4 mg/mL pepstatin A in ethanol) and NaCl. Salt concentrations of 150 mM, 200 mM, and 300 mM were used to test for appropriate IP conditions. Additionally, 20 μg/mL of DNase I is added to half of the suspended cells, which are then incubated for 10 min at room temperature. In all other IPs, the DNase step is omitted, and the buffer contains 150 mM NaCl.

The cell powder is never allowed to thaw before being resuspended in ice-cold IP buffer. Upon resuspension, it is immediately homogenized for 10 sec with a polytron device (in earlier experiments), or by vortexing at a high setting for ~30 sec (in later experiments, which proves to be just as effective). Once thawed and fully homogenized, the suspension is centrifuged for 10 min at 4°C and 3,000 RPM. The supernatant is then passed through a 1.6 μm glass microfiber syringe filter (Whatman, Piscataway, NJ) to remove further any debris. 7.5 mg of IgG conjugated dynabeads (Invitrogen, (Cristea et al., 2005; Tackett et al., 2005); conjugation protocol can be found online at http://prowl.rockefeller.edu/protocols/conjugation_magnetic_beads.html) is added to the flowthrough and the samples are incubated at 4°C for 1 hr with constant rotation.

The beads are collected on a magnetic rack (Invitrogen) for 5 min at 4°C. The supernatant is discarded and the beads are washed five times with 1 mL of cold GINS

IP buffer. A final wash with 1 mL of wash buffer (0.1 M ammonium acetate pH 7.5, 0.1 mM $MgCl_2$, 0.02% tween-20) helps to remove salt from the sample. The complex is eluted from the beads by incubating twice with 500 µL of elution buffer (0.5 N NH_4OH and 50 mM EDTA) at room temperature for 20 min. The supernatants are pooled and dried in a speedvac (ThermoSavant) for ≥2 hr for analysis by SDS-PAGE.

SDS-PAGE Analysis

Dried protein samples were resuspended in 4.5 µL of HPLC-grade H_2O, and 5 µL of 2X TCEP-sample buffer (625 µL tris-HCL 2M, 500 µL of 0.1% bromophenol blue, 400 µL 1M tris base, 500 µL 0.2M TCEP-HCl in H_2O, 2 mL of glycerol, 4 mL 10% SDS). After mixing, samples were mixed and heated at 95°C for 5 min, and allowed to cool to room temperature. The samples were alkylated by adding 0.5 µL of 0.5 M iodoacetamide (25 mM final concentration), incubating in the dark at room temperature for 30 min. Samples were then loaded onto a NuPAGE® 4-12% Bis-Tris gel (with MOPS running buffer), and resolved at 125V for 5 min, then 200 V for 45 min. The gel was fixed in 10% acetic acid, 16% methanol for 5 min with rocking, and then washed extensively with HPLC grade H_2O. Proteins were visualized with GelCode® Blue colloidal Coomassie stain (Pierce), developed with 1% acetic acid, and scanned.

In-Gel Digestion of Protein Bands

The entire lane was cut into 1 mm bands using a Mickle gel Slicer (Brinkman Instruments). Approximately 60 bands were excised from a 10 cm gel with a small razor, and pairs of consecutive bands were transferred to a microcentrifuge tube (2 mm segments of the gel). The gel slices were destained by adding 500 µL of 50 mM

ammonium bicarbonate/50% methanol, and shaking overnight at 4°C in a vertical vortex mixer (Tomy Seiko Co.). Twice adding and removing 100 μL of acetonitrile dried the destained gel slices. Proteins are digested with 150 ng of bovine, modified, sequencing grade trypsin (Roche) for 2 hr at 37°C.

A slurry of POROS® 20 R2 beads (PerSeptive Biosystems) were used to extract the tryptic peptides.

15 μL of bead slurry was added to each gel slice, and agitated at 4°C overnight (Cristea et al., 2005; Tackett et al., 2005).

The POROS beads were removed from the sample and loaded onto ZipTips® (C18, size P10, Millipore) that were washed with elution solution (70% acetonitrile/0.1% trifluoroacetic acid). The ZipTips were washed twice with 20 μL of 0.1% trifluoroacetic acid, and eluted directly onto a MALDI target with matrix at 50%-saturation in 70% acetonitrile/0.1% trifluoroacetic acid. The standard MALDI matrix for these experiments was α-cyano-4-hydroxycinnamic acid (4-HCCA).

MALDI MS and MS/MS Protein Identification

Peptide masses were determined mass spectrometrically by using a PrOTOF (PerkinElmer Life and Analytical Sciences) orthogonal time-of-flight mass spectrometer. Peptides sequences were further validated by tandem MS/MS by using a MALDI-LTQ (Thermo Fisher). MS data is calibrated post-acquisition by using the 1673 Da trypsin auto digestion product peptide as an internal standard, improving the mass accuracy to ≤10 ppm. Identifications are made by pasting the list of peptide masses into the Profound search engine (http://prowl.rockefeller.edu/), and searching against the

yeast protein database. Similarly, the XProteo algorithm was utilized to identify proteins with the tandem MS/MS data (www.xproteo.com).

RESULTS

Genetic studies indicate that GINS is an essential component of the replication machinery. In agreement with this interpretation, GINS is found to localize to sites of unwound DNA in oocyte extracts from *X. laevis* (Pacek et al., 2006), and to interact with Cdc45 and several other replication fork proteins (Gambus et al., 2006). Our characterization of the protein-protein interactions of the GINS complex uses an orthogonal isolation strategy, and was carried out prior to the publication of this latter study. In an initial analysis, conditions for purifying GINS complex from cells expressing a PrA tagged version of Sld5 were worked out (Figure 2-2). The purity was gauged solely from the bands observed on a Coomassie stained gel, and the identities of the proteins are preliminary "guesses" for optimization purposes prior to performing the full mass spectrometric analysis. Future experiments confirmed these guesses. NaCl concentrations and the effect of DNase on the interactions were tested, and found to have little affect on the core complex.

In general, the IP appears very clean. Few, if any suspected contaminant proteins are observable by gel and the amount of IgG leaching off the beads appear low. By estimating the molecular weights from band migration, each of the four GINS subunits is observed to co-isolate with Sld5-PrA under all conditions. This indicates that the complex is extremely stable, and is not mediated by or dependent on the presence of DNA.

A fifth protein is observed to migrate ~110 KDa. This protein appears to interact stably with the GINS complex at low salt conditions, but quickly dissociates as salt concentration increases. The interaction remains consistent regardless of the presence of DNase in the IP, and is therefore unlikely to be mediated by the presence of DNA.

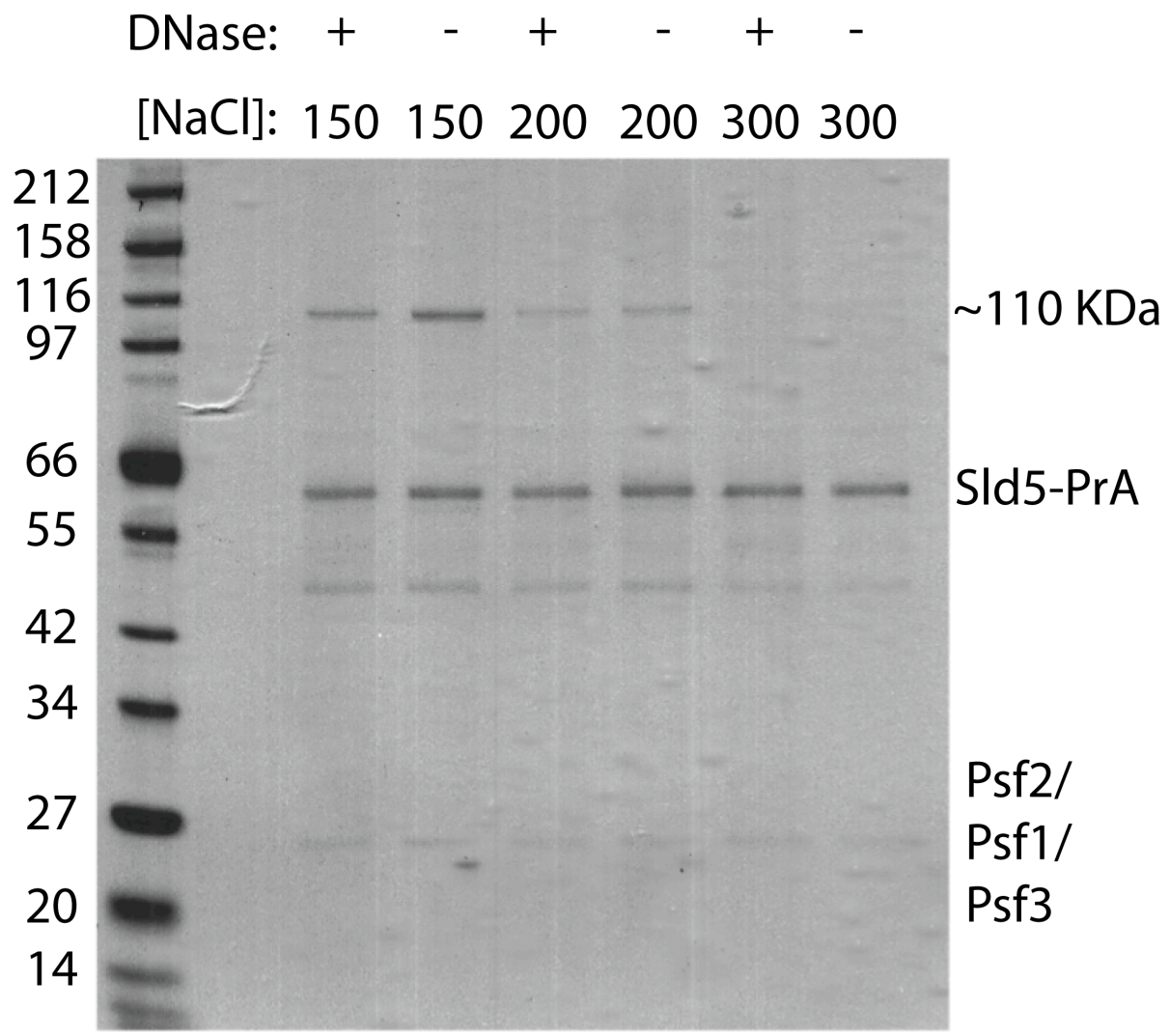


Figure 2-2. Isolation of the GINS complex by magnetic bead affinity-isolation. The Sld5 subunit of the GINS complex was tagged with PrA and 2 g of cells were immunoisolated with rabbit-IgG conjugated dynabeads. The concentration of NaCl and the presence of DNase I is indicated above each lane and the size of the molecular weight proteins are indicated on the left. The identities of the bands are based on the observed molecular weights. Psf1, Psf2, and Psf3 co-migrate as a single band in this gel.

Final conditions for routine immunoisolation of the GINS complex include 150 mM NaCl, and exclude the digestion of DNA by DNase I. In order to identify the interacting proteins by mass spectrometry the amount of samples were increased to 3 g, and 3 of the 4 subunits were isolated via their PrA tags (Sld5, Psf1, and Psf2; Figure 2-3). All four GINS proteins co-purified with the three analyzed subunits, as did the subunits of the MCM complex and Cdc45. These three proteins constitute the CMG, a complex believed to be necessary for helicase activity (Ilves et al., 2010; Moyer et al., 2006).

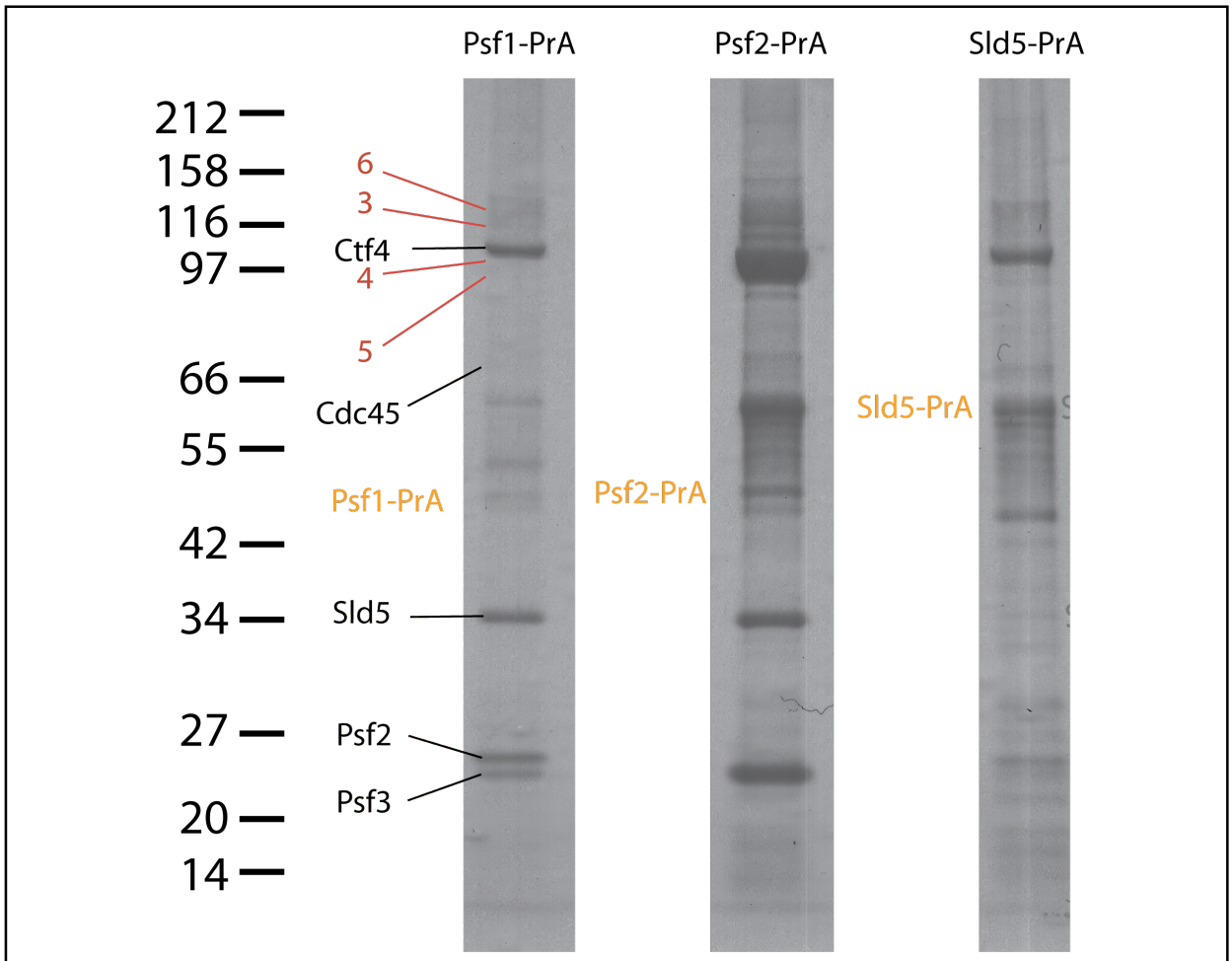


Figure 2-3. Proteomic identification of proteins interacting with the GINS complex. Psf1-PrA, Psf2-PrA, and Sld5-PrA were immunoisolated by using optimized conditions. The four subunits of the GINS complex consistently co-isolated with each of the bait proteins (indicated by yellow font). Other proteins identified include Ctf4 (105 KDa), Cdc45 (74 KDa), and members of the MCM complex (ranging in MW from 87 – 114 KDa, numbers are indicated in red font).

In addition to the isolation of the CMG complex, the cohesion-associated protein Ctf4 co-isolated with the GINS. The role of Ctf4 remains a mystery, but it is thought to be involved in bridging these two mitotic processes (replication and sister chromatid cohesion) (Formosa and Nittis, 1999; Hanna et al., 2001; Kouprina et al., 1992; Miles and Formosa, 1992). None of the three tagged proteins appeared to have any deleterious effects on the cells, but we chose to utilize Psf2 for all future isolations as this IP worked somewhat better than the other three.

Our approach demonstrates a successful attempt to isolate the CMG complex, but misses many of the components associated with the RPC (as characterized by (Gambus et al., 2006)). This, along with *in vitro* helicase assays, indicates that the CMG is a core complex, while the other RPC proteins may play an ancillary role to the unwinding of the DNA at the fork.

Another intriguing clue in these data is the lack of any hint that the three polymerase enzymes interact with GINS. Indeed, none of these proteins have ever been shown in direct complex with the CMG. This may suggest that the helicase, which by definition must travel ahead of the polymerases, is an independent complex. Native studies investigating proteins that directly associate with Pol ε failed to identify any of the CMG members (Iida and Araki, 2004; Tackett et al., 2005), although an interaction has been found when the cells were first treated with a chemical cross-linker (Muramatsu et al., 2010). This may imply that the link between these proteins is dependent on DNA, but this remains to be investigated. If a direct interaction exists, the fairly mild conditions used here to isolate the CMG complex appear to eliminate it. Regardless, the CMG must coordinate with the polymerases to some extent, and it is therefore of

interest to characterize the interactions of Pol δ, and to investigate whether it shows any interaction with these proteins.

2.2 – Proteomic Characterization of the Polymerase δ Holoenzyme

Materials & Methods

Strain Construction and Storage

The strain in use in this section (ATY44) contains a PrA tag at the 3' end of the Pol3 ORF, which is the catalytic subunit of the Pol δ holoenzyme. Cells were stored and cultured identically as in the GINS isolations.

Harvesting and Cell Lysis

The harvesting and lysis protocols were identical to those used in the GINS isolations.

Pol2 Immunoisolation

The immunoisolation of Pol3-PrA was identical to that of Psf2-PrA, except for the IP buffer (20 mM Hepes pH7.4, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.5% Na-Deoxycholate, 1/100 (v/v) solution P, 1/200 (v/v) protease inhibitor cocktail (Sigma)). These conditions were previously used to isolate Pol3 for ChIP analysis (Hiraga et al., 2005).

SDS-PAGE and Mass Spectrometric Analysis

The protocols for SDS-PAGE and mass spectrometric analysis of Pol δ were identical to those used to analyze the GINS complex.

RESULTS

Pol δ is a multiprotein complex consisting of three polypeptides – Pol3, Pol31 and Pol32 (Figure 1-9). The role of this enzyme is to synthesize DNA from a template strand, in the wake of the helicase. To date, focused proteomic analysis of the complex has not been carried out, and so interactions with other replication fork proteins have not been assigned. The results presented in Figure 2-4 A suggest that Pol3 forms a stable complex with Pol31 and Pol32 (these three proteins form the holoenzyme Pol δ), but do not directly interact with any other proteins to a significant degree (MS results can be found in Appendix B, Table B-6). No sign of any of the GINS or MCM proteins was observed to interact with Pol3 in these experiments.

There are several explanations for these results that are fully compatible with the present data. They are: 1) A large excess pool of Pol δ exists within the cells, and at any given time only a small fraction of the protein is bound to the replication fork, interacting with other fork proteins, and 2) The IP conditions in use were too stringent, stripping away less tightly bound proteins. 3) The CMG and Pol δ form completely distinct complexes within the cell.

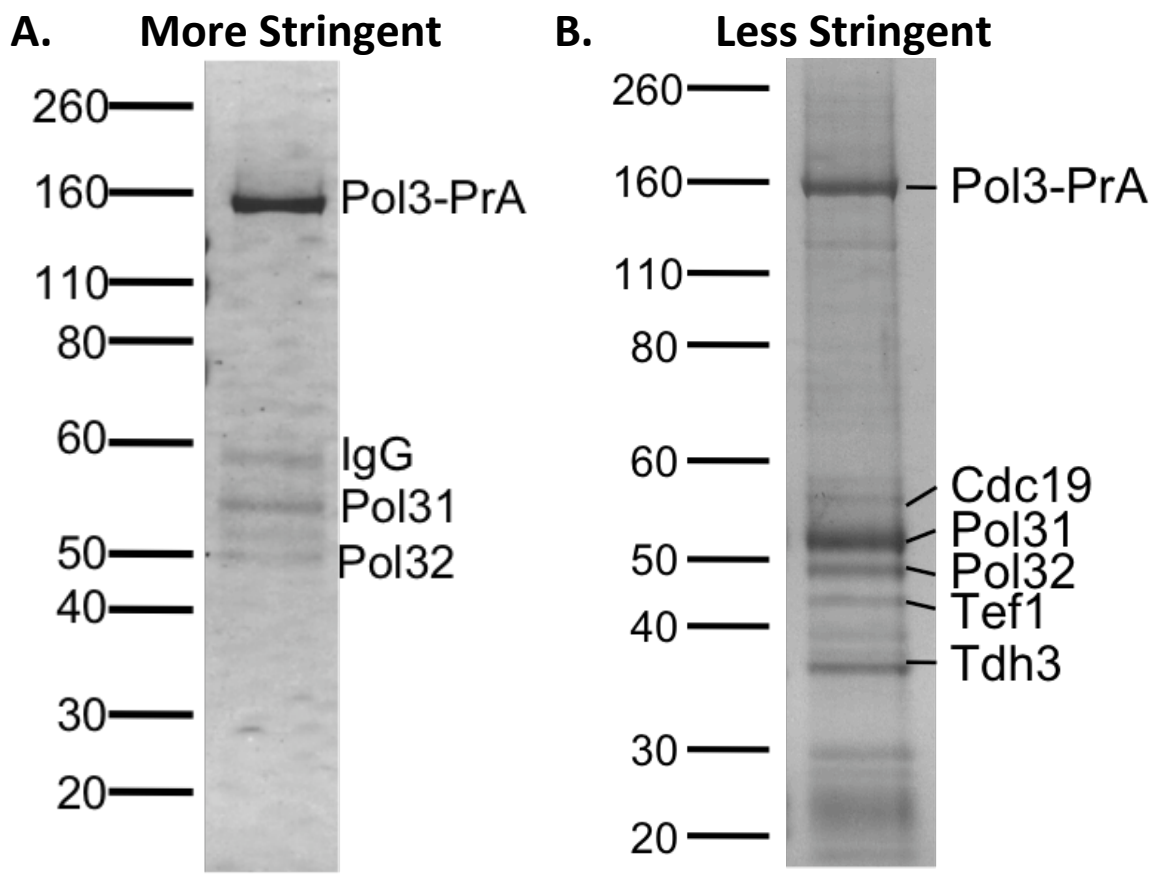


Figure 2-4. Polymerase δ was immunoisolated via a protein A tag on the c-terminus of the catalytic subunit Pol3. **A.** Under highly stringent conditions only Pol32 and Pol31 are detected. **B.** Under conditions of lower stringency, several putative contaminant proteins appear in the gel (Cdc19, Tef1, and Tdh3) along with Pol31 and Pol32, but no CMG proteins appear to be present.

By performing the immunoisolation under less stringent conditions, it is possible to test for the second interpretation (Figure 2-4 B). The results show that even in far lower detergent conditions, no CMG proteins can be found interacting with Pol3. Because the conditions are relatively mild, and no interaction has been shown in either direction (that is, via the GINS or Pol δ), it seems unlikely that the second explanation would be the case. On first glance, the third interpretation may also seem surprising since the polymerases must coordinate with the helicase to some extent, although it is possible that the distinct complexes remain near each other despite no physical interaction.

CHAPTER III – DYNAMIC PROTEIN-PROTEIN INTERACTIONS

The results described in the previous chapter demonstrate how information about protein-protein interactions for the CMG and Pol δ complexes can reveal certain aspects of about their biological relevance. The CMG is observed to form a stable complex, while other members of the RPC appear to be structurally ancillary to this core. Further, interaction studies of the CMG and the polymerases Pol ε and Pol δ have so far failed to reveal a direct connection between any of these proteins. These results suggest that if the interactions exist, they are very weak, or they are only present for a fraction of the protein pools in the cell population. One possible explanation for this observation is that the proteins only interact in particular stages of the cell cycle.

Focused proteomics experiments are typically performed from asynchronous cultures of cells. This approach is extremely effective at identifying stable interactions that persist throughout the cell cycle. However, in some cases the compositions of multiprotein complexes are altered by trans-acting stimuli, such as the phosphorylating activity of cell cycle dependant kinases (CDK). CDK activity regulates a vast number of cell cycle dependent processes within the cell, including DNA replication. Proteomic studies of specific replication complexes may therefore benefit from an approach where the interactions of these proteins were dissected into discrete stages of the cell cycle.

This chapter describes experiments that couple classical techniques for synchronizing cells with state-of-the-art immunoisolation of protein complexes and mass spectrometric readout. The focus is on the GINS complex in these experiments because a number of proteins were identified in the affinity isolation from asynchronous cells

(Chapter 2) that appear to interact less tightly than the tetrameric core GINS complex. This suggests that these loosely bound proteins are either peripheral to the core GINS complex, or that the interaction is not persistent in all cells, and may be regulated by the cell cycle.

3.1 - Proteomic Characterization of the Dynamic Protein-Protein Interactions of the GINS Complex

Materials & Methods

Strain Construction

Strain MSY1 was constructed from strain ATY114 by replacing the BAR1 gene, which encodes for the α-factor protease with the KANMX4 selection cassette (see appendix A). This produced a Psf2-PrA strain that was rendered hypersensitive to α-factor arrest in G1.

Cell Growth and Harvesting

10 L of cells were grown in a 12 L BioFlo 410 fermenter (New Brunswick) in rich YEPD medium for synchronization at 30°C. When the cells reached a density of ~7.5×10^6 cells/mL), α-factor was added to a concentration of 50 nM and the culture was allowed to incubate for another 3 hr. When fully blocked, as assessed by the complete appearance of the shmoo morphology, the cells were very quickly harvested through a drain valve at the bottom of the fermenter and centrifuged for 10 min at 10,000 RPM at 4°C. The cell pellets were washed twice with ice-cold YEPD, and resuspended in 10 L of

fresh YEPD (no α-factor) in the fermenter at 25°C. The initial cell density for all experiments was ~1x10^7 cells/mL. Upon release from G1 block, cell density, DNA content, and the budding index were all used to monitor uniform progression of the population throughout the cell cycle.

500 mL of cells were harvested every 15 or 30 minutes (two separate experiments) directly into a 1 L centrifuge tube that was filled halfway with ice to slow down cellular processes. Each time point was pelleted for 3 min at 5000 RPM in a JLA-8.1000 rotor in an Avanti J-26 XP centrifuge (Beckman Coulter). The pellet was quickly washed with sterile dH_2O, centrifuged again for 3 min at 3500 RPM, and quickly resuspended in 100 µL of yeast grind buffer. The suspension was dripped into liquid nitrogen forming very small, frozen balls. The process from harvesting to freezing took no more than 10 min, and the cells were kept on ice at all times to minimize progression or loss of synchrony. Cells were stored at -80°C until all time points were ready for grinding.

Flow Cytometry and Budding Index

10 mL of cells from each time point were fixed in 70% ethanol for 1 hr at room temperature, and prepared for FACS analysis following previously described methods (Haase and Lew, 1997). Cells were stained with 50 µg/mL propidium iodide, and DNA content was analyzed using a Facscalibur I (BD Biosciences). Data analysis was performed with FloJo 8.3.3 (http://www.flowjo.com).

Budding indices were calculated for all time points by counting the percentage of cells in each time point that contained small or large buds. The reported budding index is the total of both small and large bud counts.

Cell Lysis and Immunoisolation of Protein Complexes

Each time-point of frozen cells was cryolyzed as described for the asynchronous GINS and Pol δ immunoisolations. 3 g of grindate from each time point was resuspended in GINS IP buffer and immunoisolated as in the asynchronous IP experiments (Section 2.1).

SDS-PAGE and Mass Spectrometric Analysis

The SDS-PAGE and mass spectrometric analyses of the proteins present in each time point Psf2-PrA IP were essentially the same to those of the asynchronous experiments. Here, each gel lane was sliced into 2 mm sections for digestion and MS analysis. Thus, even slices with no observable bands were searched for the presence of proteins. This is important because the sensitivity of the MS instruments far exceeds that of Coomassie stain.

RESULTS

3.1.1 - Lower Time-Resolved Focused Proteomics of GINS

As a proof of principle for a time-course focused proteomics analysis, a preliminary experiment in which time points were harvested every 30 min for 2.5 hr is presented in Figures 3-1 and 3-2. Cell density and the budding index were monitored for each time point, and indicate that the cells are properly released from G1, and cycle through with relative synchrony (Figure 3-1). DNA content was not monitored in this experiment, and so an approximation of the cell cycle is indicated above the growth curve and IP

results. Briefly, cells enter S phase ~15-20 min after being released from G1, and finish mitosis ~80-90 min after release.

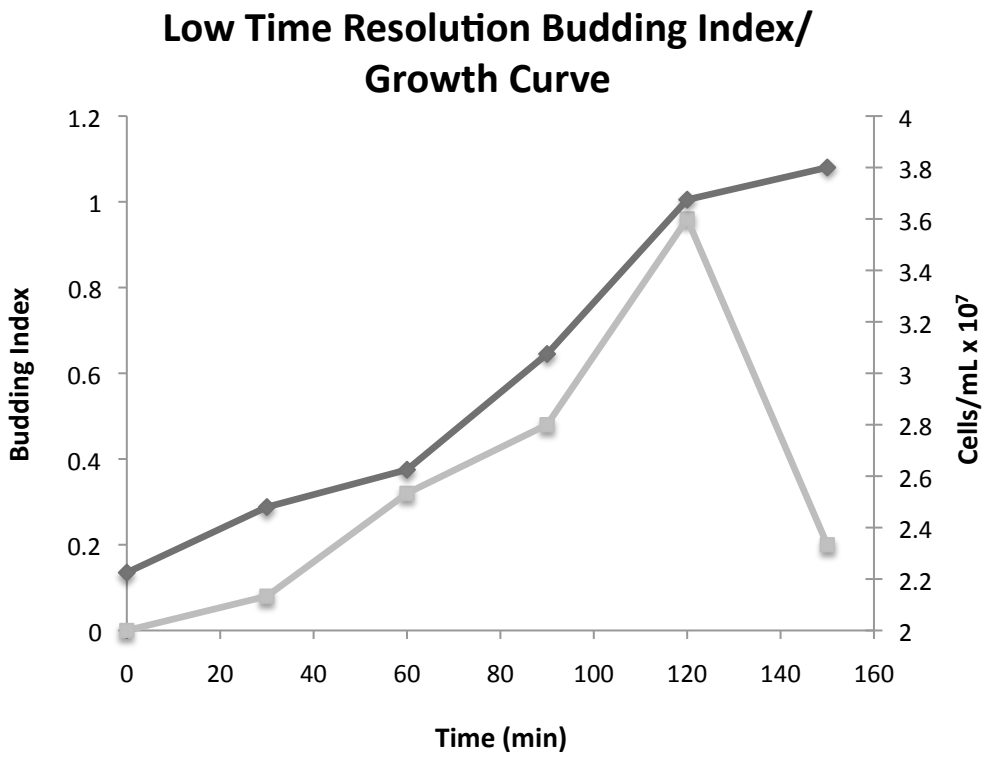


Figure 3-1. Growth curve and budding index for 30 minute time-course experiments. The dark line corresponds to the cell density (right axis), and the light line corresponds to the budding index (left axis). The cells are in S phase in the 30 and 60 min time point, and finish dividing by the 120 min time point.

The successful immunoisolation of the GINS complex for all six time points is shown in Figure 3-2 (detailed information on the MS and MS/MS identification can be found in Appendix B, table B-4). The results show that the tagged version of Psf2-PrA pulls down each of the remaining three subunits (Sld5, Psf1, and Psf3) in all time points, confirming that GINS is a stable, heterotetrameric complex throughout the cell cycle. The bands corresponding to the four GINS subunits appear to be of roughly equal intensity on the gel.

The remaining members of the CMG complex - Cdc45 and several of the MCM proteins - were identified in the 60 and 150 min time points. The 60 min sample roughly corresponds to cells that are in S phase, indicating that the CMG complex is formed exclusively in this stage of the cell cycle. Importantly, the time point that is 150 min after release from G1 corresponds to divided cells and likely re-entered the next round of DNA replication (Note: After one division, cell synchrony is partially lost due to the asymmetric division of yeast, so this later time point only stands to confirm that the cells are cycling properly). The four core histones appear to interact with the GINS complex in these same S phase time points, forecasting a role of the complex in chromatin during replication. This is highly compatible with the prediction that the CMG works as the helicase at the replication fork.

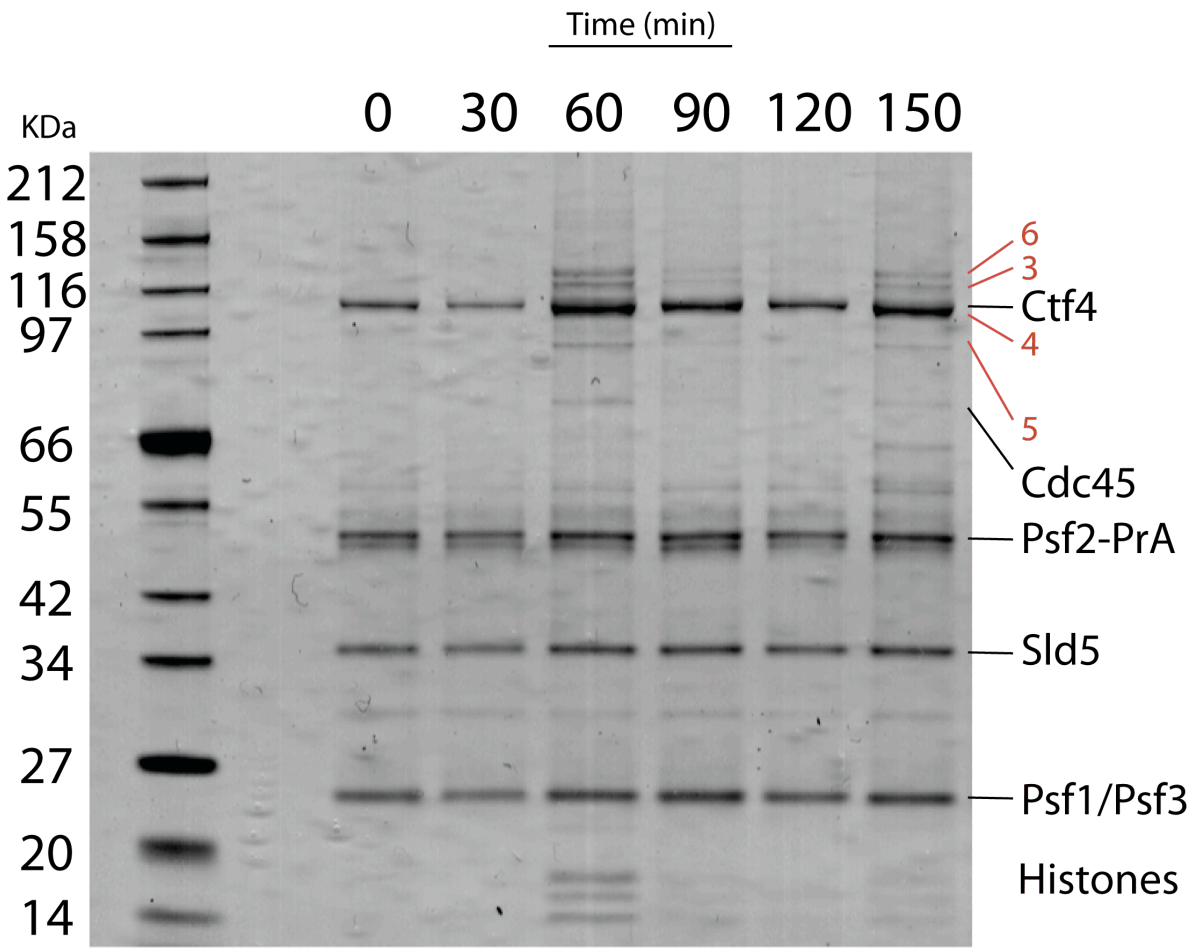


Figure 3-2. 30-Minute time-course proteomic characterization of the interactions of the GINS complex. Psf2-PrA was immunoisolated from synchronized cells corresponding to the time indicated along the top of the gel. The bands for the MCM complex are indicated by their corresponding number in red font. Proteins were identified by MS after in gel digestion with trypsin. The MS scores can be found in Appendix B, Table B-4.

Surprisingly, the only other RPC protein to co-purify with the CMG is the cohesion protein Ctf4. The fact that it does so during all time points suggests that this interaction persists throughout the cell cycle. However, no other components of the RPC are observed in any time point, and neither Pol ε or Pol δ can be identified in any time point. It remains possible that the isolation conditions are too stringent for the purification of this entire complex.

3.1.2 - Higher Time-Resolved Focused Proteomics of GINS

Given the ~90 min duration of the cell cycle, harvesting time points every 15 minutes provides considerably more information than the 30 min time points described above, and allows for more specific assignment of interactions to specific stages in the cell cycle. Because the CMG complex formation is likely to be an indicator for GINS functionality, its formation was carefully monitored.

The cells cycled appropriately (Figure 3-3), as indicated by the budding index and flow cytometry. The overall interaction picture for the 15 min time-course was identical to that in the 30 min experiment – the stable GINS complex and Ctf4 interacted with the MCM complex, Cdc45, and the four core histones specifically during S-phase (Figure 3-4). The higher time resolution revealed that the CMG complex formed sometime after the 15 min time point, but before the 30 min time point. The interaction peaked in the 45 min time point, and disappeared after the 75 min time point. These observations indicate that the CMG complex forms specifically in S phase, consistent with the view that the CMG complex is necessary for unwinding the DNA at the replication fork. It is interesting to note that the remaining members of the RPC – Mrc1, Csm3, Mcm10, Spt16, and Top1 - were not identified in any time point in either the 30 min or 15 min experiments. This is likely due to the fact that the fractionation of the RPC (Gambus et al., 2006) was done under very specific conditions to preserve these interactions, and these components very likely interact less stably than the core CMG components.

The hope for this experiment was that isolating GINS complexes from cells at discrete points in the cell cycle would help enrich for less stable or frequent interactions. Indeed

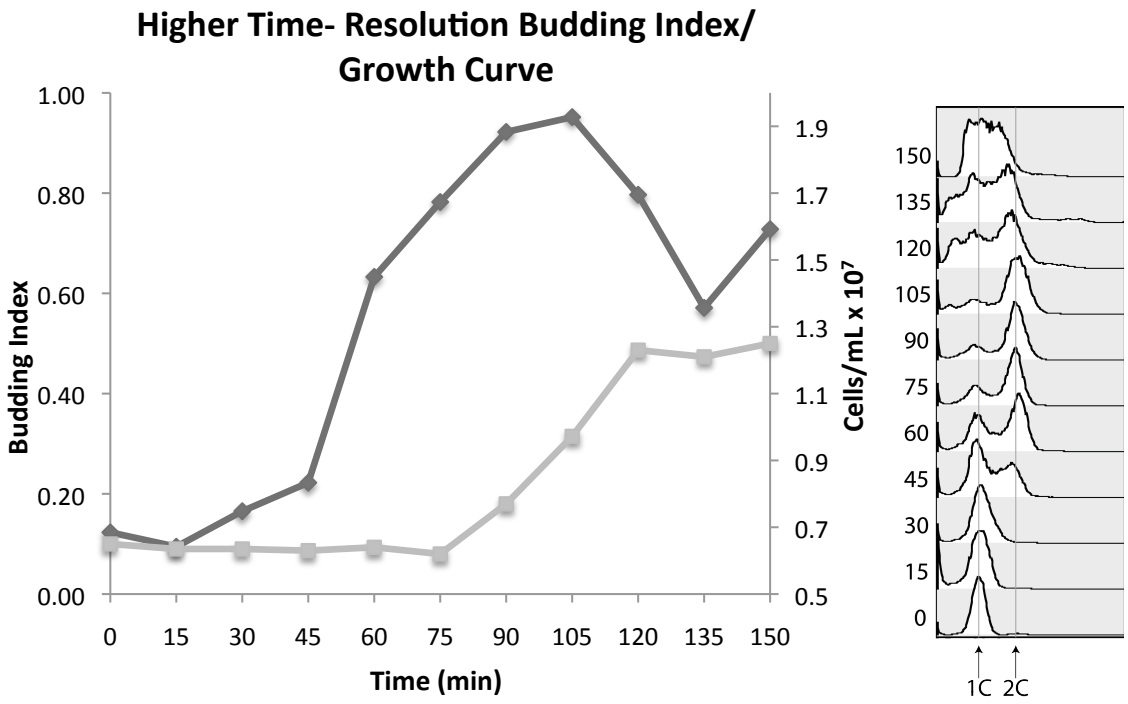


Figure 3-3. Monitoring the progression of synchronized cells for time-resolved focus proteomics analysis of the GINS complex. **A.** The budding index (black line) and cell count (grey line) for time points harvested every 15 minutes for focused a proteomics experiment. **B.** The DNA content for the cells in each time point. The y-axis indicates the time point, and the relative amount of DNA is indicated with arrows.

the relative intensities of the MCM proteins to the GINS proteins is much increased in the 30 min time point compared to asynchronous cells. Still, we were not able to detect the presence of Pol ε or Pol δ. The lack of any hint of an interaction between the GINS complex and Pol ε or Pol δ in these studies or the Pol2 and Pol3 isolations is striking. Importantly, it has been shown that GINS and Pol ε can be co-isolated from cells that have been treated with formaldehyde in order to capture transient or indirect interactions (Muramatsu et al., 2010). On first glance our results likely imply that the GINS-Pol ε interaction they have described is indirect, occurring through DNA or

another protein. They speculate that this interaction occurs via the essential C-terminal portion of Pol2, and that it is not dependent on DNA - although no data in support of this was shown in the paper. A direct interaction between the GINS complex and Pol ε must occur at a very weak level, if it occurs at all.

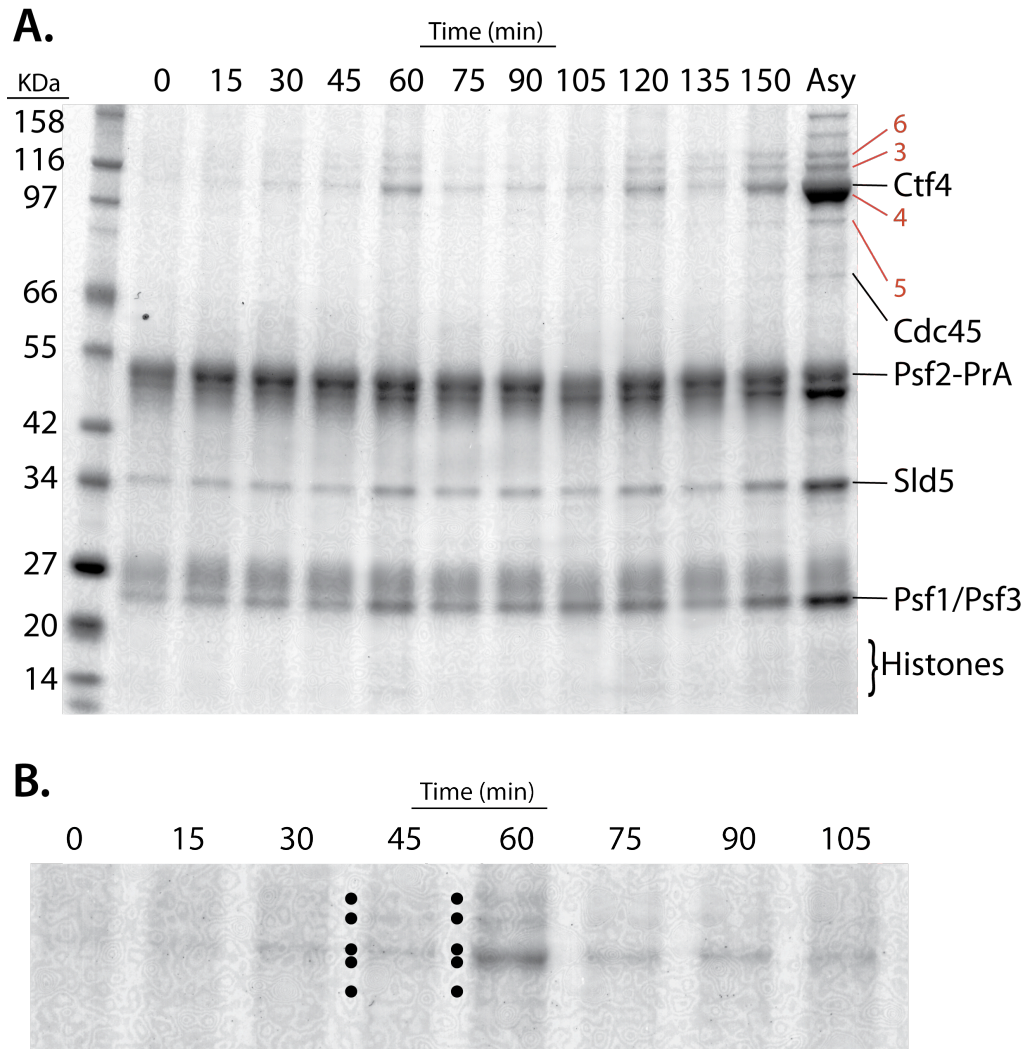


Figure 3-4. 15-Minute Time-course Proteomic Characterization of the interactions of the GINS complex. **A.** Psf2-PrA was immunoisolated from synchronized cells corresponding to the time (in minutes after release) indicated along the top of the gel. An IP from asynchronous cells is indicated (Asy). The bands MCM complex are indicated by their corresponding number in red font. Proteins were identified by MS after in gel digestion with trypsin. The mass spec scores can be found in Appendix B, Table B-5. **B.** Expanded look at the region around 105 KDa in the gel indicates which time points show interactions between Psf2-PrA and the MCM proteins (black dots).

It is fascinating that GINS co-isolates with the sister chromatid cohesion factor Ctf4. Ctf4 has been shown to interact genetically with several other replication proteins, but its role at the fork remains a mystery. Experiments in which the GINS complex was immunoisolated with various concentrations of salt (Figure 2-2) revealed that this interaction was less stable than those between the GINS subunits. Taken together, these results paint a picture wherein the GINS is constantly in a complex with Ctf4, but the interaction may be continuously exchanging.

Setting aside for the moment the details of its interaction with the polymerases, it is clear that GINS interacts with chromatin at the replication fork, exclusively during S phase. The next phase of this project aimed to understand that relationship, and determine if GINS could reveal anything interesting about the progressing replication fork.

CHAPTER IV – GENOME-WIDE LOCALIZATION OF REPLICATION FORK COMPONENTS THROUGHOUT THE CELL CYCLE

The results from the dynamic proteomic studies emphasized two major attributes of replication forks. First, the GINS complex is an integral member of the replication fork and interacts with chromatin during S phase. In light of this fact, a novel set of experiments is described in this chapter (Section 4.1) in which the genome-wide localization of the GINS complex is monitored by ChIP-chip throughout the cell cycle. The results of this study suggest that monitoring GINS progression is a valuable strategy for exploring replication fork dynamics.

The second aspect about the replication fork revealed by the proteomics study is that the relationship between the GINS and the replicative polymerases, Pol ε and Pol δ, is intriguingly complicated. No evidence for a direct interaction has been observed in any immunoisolation experiment – except when cross-linking reagents are utilized to capture the interactions. Thus, it is possible that the CMG complex progresses independently of the polymerases, and that the observed interaction is via genomic DNA covalently linking the two complexes. Alternatively, the interactions between the CMG and the polymerases are so dynamic – that is, they exchange rapidly and continuously throughout the replication cycle – they are extremely difficult to observe by standard techniques. This unresolved issue prompted an investigation into polymerase movement across the genome, with the aim to compare the dynamics of all three proteins - GINS, Pol ε and Pol δ (Section 4.2). A final step, in which two of the

proteins in question are directly compared by co-ChIP-chip experiment, is also described (Section 4.3)

4.1 - Genome-Wide & Time-Resolved ChIP-chip Localization of the GINS Complex

Materials & Methods

Yeast Strain Construction and Growth Conditions

Strains MSY1 is described in Section 3.1 (see also strain table A-1 in Appendix A). The strain is genomically modified so that the C-terminus of the Psf2 protein carries a PrA-tag.

Cell Synchronization

MSY1 cells are cultured and synchronized as in the GINS proteomics experiments in Section 3.1.

Time-course Sampling and Cell Lysis

Time points are collected in two separate experiments – a) once every 15 min after release from the G1 block from 0 min – 105 min and b) every 5 min beginning 20 min after block (through S-phase) from 20 - 50 min. For each time-point, a 1 L sample is cross-linked by incubation in 1% formaldehyde at room temperature for 20 min. Samples for analyzing budding index and DNA content by FACS are harvested at the same time.

The cross-linking reaction is quenched by adding 125 mM glycine and rocking for 5 min. The formaldehyde treated cells are centrifuged at 5,000 RPM for 4 min at 4°C, washed with ice-cold tris pH 7.8, and centrifuged again at 3000 RPM for 3 min at 4°C. The pellet is re-suspended in 500 µL of cold Lysis Buffer (50 mM Hepes pH 7.5, 1.2% polyvinylpyrrolidone) and the cells are dripped into liquid nitrogen forming small, frozen balls. The frozen cells were stored at -80°C until use.

Flow Cytometry and Budding Index

The protocols for flow cytometry and calculating the budding index are identical to those described in Section 3.1.

Cell Lysis and Chromatin Immunoprecipitation

Cryogenic grinding of the ChIP time points is identical to that for the proteomic samples (Section 3.1). ChIP is performed as previously reported (Tackett et al., 2005). Briefly, 0.5 g of each frozen, ground time point is suspended in 1 mL of ChIP lysis buffer (50mM Hepes-KOH pH7.5, 140mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate, 1/100 (v/v) solution P), and sonicated at 4˚C, to an average size of ~400 bp. 25 µL of the supernatant is reserved as the whole cell extract (WCE). The remaining sample (IP) is incubated for 2 hr at 4˚C with 4.5 mg of IgG - conjugated, 3 µm dynabeads (Invitrogen, conjugation protocol discussed in Section 2.1). The beads are collected and washed 2x with ChIP lysis buffer, 2x with high-salt lysis buffer (50mM Hepes-KOH pH7.5, 500mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate, 1/100 (v/v) solution P), 1x with TE (10 mM Tris pH 8.0, 1 mM EDTA). The samples are eluted by incubation with TE/1% SDS (50 mM Tris pH8.0, 10 mM EDTA, 1% SDS) at 65°C for 15 min. The eluate is transferred to a fresh tube and the

beads washed with TE/0.67% SDS (50 mM Tris, pH8.0, 10 mM EDTA, 0.67% SDS) and the eluates pooled.

The eluted samples are incubated overnight at 65°C to reverse the cross-links. Samples are then treated with proteinase K and the DNA is extracted with phenol/chloroform, followed by ethanol precipitation. RNase A is used to digest the RNA in the sample, and DNA is purified with a PCR cleanup kit (Qiagen).

Real-Time PCR

Real-Time PCR is typically performed as sample validation prior to array hybridization. 1 μL of undiluted sample is analyzed in 20 μL reaction volume with 900 nM forward (ARS306taqfor – 5′-TCGTCTAAGTCCTTGTAATGTAAGGTAAGA-3′) and reverse primers (ARS306taqrev – 5′-GCTTGGGTTTGTGACTTACTAACG-3′), 250 nM probe (ARS306taqprobe – 5′-FAM-TGCAAGCATCTTGTTTGTAACGCGATTG¬TAMRA-3′), and 1x Taqman Universal PCR Master Mix (Applied Biosystems). Sample analysis is carried out in a 7900HT Sequence Detection System (Applied Biosystems). The PCR program is for the following 45 cycles: denaturation at 94˚C for 15 s, annealing at 43˚C for 30 s, and extension at 72˚C for 30 s. Results are normalized to ACT1 levels (primer sequences are ACT1taqfor – 5′-CTCCGTCTGGATTGGTGGTT-3′, ACTtaq1rev – 5′-TGG ACCACTTTCGTCGTATTCTT-3′, ACT1taqprobe – 5′-FAM-TTGACTACCTTCCAACAA-TAMRA-3′).

Microarrays

A linker-mediated amplification and labelling protocol is used to prepare the samples for microarray hybridization (Agilent protocol). Briefly, 40 μL of sample DNA (IP) and

1.5 μL of whole cell extract DNA (WCE) are blunt-ended with T4 DNA polymerase (New England Biolabs). Blunt ended DNA is purified by phenol/chloroform extraction, and ligated with annealed linkers (5'-GCGGTGACCCGGGAGATCTGAATTC-3') by using T4 DNA ligase. The DNA is then subjected to two rounds of Taq polymerase amplification with primers complementary to the annealed linker. Amplified DNA is precipitated with ethanol and labelled with Cy5-dUTP (IP samples) or Cy3-dUTP (WCE samples) using a CGH labelling kit (Invitrogen).

Labelled DNA is hybridized to yeast 4x44k whole genome microarrays (Agilent) with average spatial resolution of ~290 nt, and analyzed as previously described (Ren et al., 2000; Tackett et al., 2005), and following the manufacturers recommendations. The scanner quantifies the intensity for the two fluorophores at each array element, which corresponds to the amount of Cy5-labeled DNA (IP samples) and Cy3-labeled DNA (WCE samples). The analytics software creates a list consisting of the log2 ratios of these intensities for every spot on the array. The default setting on the data analysis software sets the median ratio as the zero, giving 50% of the intensities positive values, and 50% negative values. For better visualization, the intensities were scaled so that only 3% of the data was negative. This yields a robust zero that is independent of occasional large negative outliers in the data. In order to normalize the individual time points to each other, the intensities are scaled so that the noise within each time point is equivalent (discussed in detail in Section 5.1). The program SignalMap (Nimblegen) was used to visualize the data. The list of previously documented origins was obtained from the OriDB website (Nieduszynski et al., 2007).

Calculating Fork Velocities

The exact location of a particular GINS complex at any given fork is blurred within the population due to slight losses in synchrony and random molecular fluctuations within the cells. Still, the average location for a given replication fork can be approximated from the peaks present in each time point. In many cases the edges of these peaks were well resolved from other peaks, and we were able to estimate replication fork velocities across the genome by measuring the distance between the peak edges in successive time points and dividing that number by 5 minutes (Figure 4-6). When edges began to merge with each other we ignored these data points. The average rates are reported in kilobases per minute (kb/min).

RESULTS

There is no doubt that chromatin compaction states affect the ability of proteins to bind across the genome. For example, the ability of transcription factors to bind to specific regions of the genome is regulated by the level of compaction at that location – tightly compacted heterochromatin excludes these proteins from the genome, while the less dense euchromatin allows them to bind.

This effect has lead to speculation that heterochromatic regions of the genome may be more difficult to replicate because the dense chromatin would be more difficult for the helicase to unwind (Raghuraman et al., 2001). In agreement with this, several studies reported widely varying rates of replication fork progression during S phase, a sign that the rate of replication was highly dependent on genomic location. To determine the rates, these studies monitored nucleotide incorporation, but a more direct measurement

of helicase progression would monitor the locations of members of the CMG complex throughout S phase.

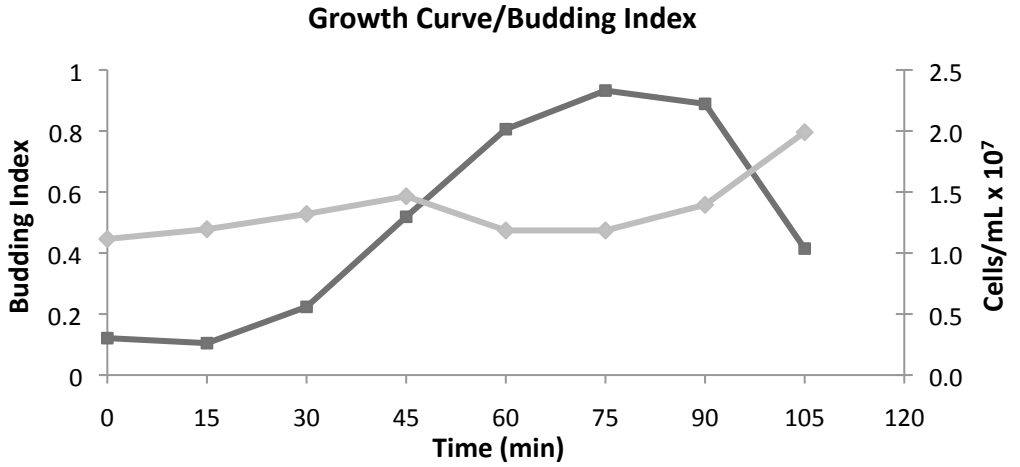
Here, we monitor the spatial and temporal association of a protein that assembles at origins with the pre-RC and has been inferred to travel with the advancing replication forks as a member of the CMG complex. For this purpose, we focus on Psf2, a component of the GINS complex (Takayama et al., 2003), and collect ChIP-chip (Ren et al., 2000; Tackett et al., 2005) data as a function of time throughout the cell cycle.

Properly synchronized cells are vital to the success of this approach. Therefore, the samples are carefully monitored for synchrony by quantitating the appearance of buds (Figure 4-1) and DNA content (Figure 4-2) for each time point. To ensure that the cell cycle does not progress during the incubation with formaldehyde, budding indices from before and after formaldehyde incubation are compared (Figure 4-1 C). These data indicate that the cells are properly synchronized through the first round of cell division and, at least by external appearance, formaldehyde arrests the cell cycle to an acceptable degree.

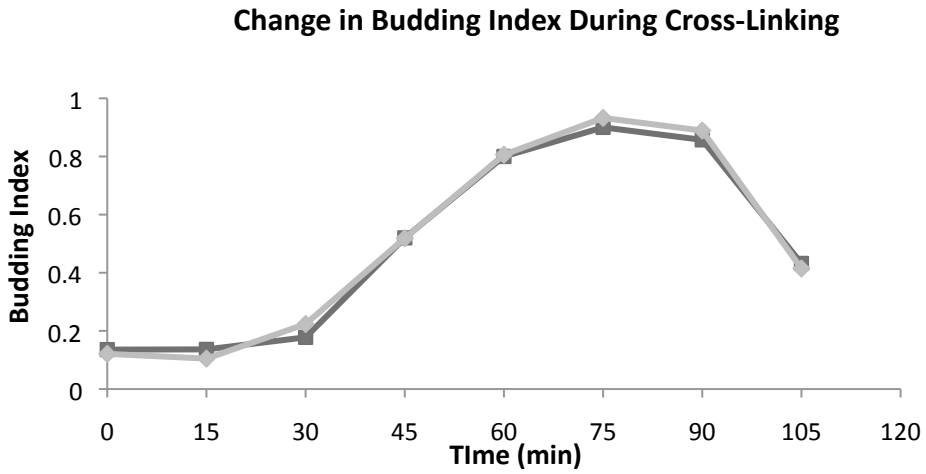
Figure 4-1 (Next Page). Cell growth for typical time-course experiments. A. Growth curve (light line) and budding indices (dark line) for each time point in a 15 min time course experiment. **B.** Comparison of budding indices before (light line) and after (dark line) formaldehyde treatment during a 15 min time course experiment. **C.** Growth curve (light line) and budding indices (dark line) during a 5 min time-course experiment carried out through S phase (20 – 50 min after release from α-factor block).

Figure 4-1.

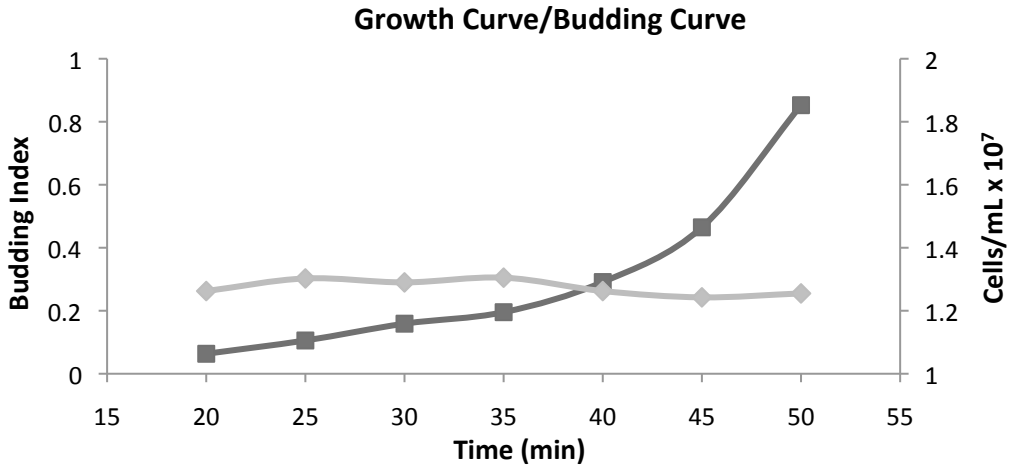
A.



B.



C.



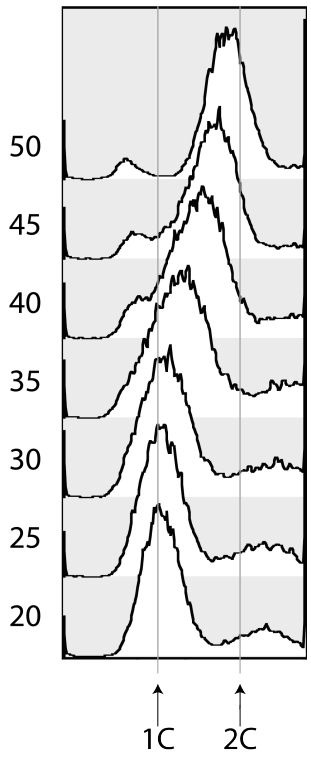


Figure 4-2. DNA content for the cells during a 5 min time-course experiment.

ChIP-chip methods are utilized to monitor progression of the GINS complex across the genome. By shearing the DNA to ~400 bp fragments (Figure 4-3) prior to immunoisolating the Psf2 subunit, relatively small stretches of genomic DNA are co-isolated with the GINS complex. After isolation and amplification, the DNA from each time point is then hybridized to yeast whole-genome microarrays. In doing so, the relative binding of GINS to precise regions of the genome can be assessed for each time point. Because GINS is a member of the CMG and progresses with helicase, a picture of fork progression emerges (Figure 4-4 and 4-5).

At early times (≤ 20 min), the GINS complex does not exhibit any observable interaction with the genome as evidenced by the lack of peaks in the ChIP-chip signal (Figure 4-4). As the cell cycle progresses, the GINS complex begins to interact with specific chromosomal sites (indicated by the peaks and the dashed lines in Figure 4-4), nearly all of which correspond to previously posited origins (Appendix F (Nieduszynski et al., 2007). The height of the peaks provides a measure of average occupancy, and their spreading indicates that some fraction of the origins within the cell population has fired. The data for chromosome 16 (Figure 4-4) indicate that 12 origins are occupied to some extent in the population, almost all of which have fired by 25 min. By 30 min the number of occupied origins increases to ~28, after which time no additional origins appear to be populated. Bi-directional GINS progression from each of these origins (averaged over the population of cells) can be inferred from the spreading

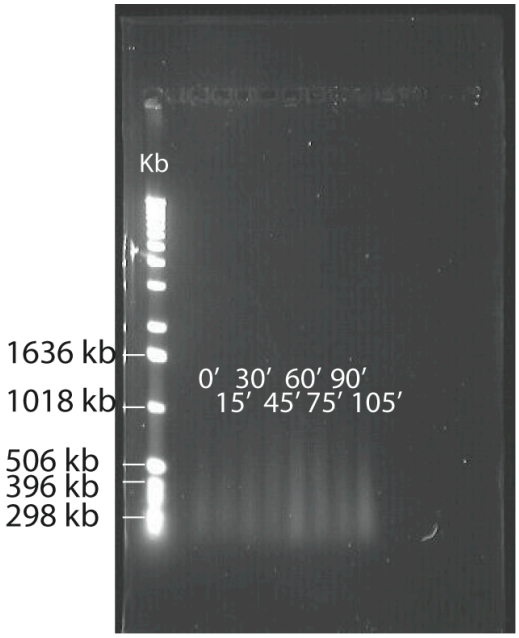


Figure 4-3. Confirming the Shearing of DNA. 1 μL of input DNA after shearing (cross-links reversed), analyzed by 0.9% agarose gel stained with SYBR-Gold. Each time point is labelled above the well. Similar fragment lengths were observed for all ChIP experiments.

of the edges of the corresponding peaks. By 35 min, many of these edges have merged with those from adjacently spreading GINS and fewer locations on the chromosome remain wholly unoccupied. Spreading continues with time such that by 50 min GINS has progressed to most regions of the chromosome, and reduced occupancy is observed in the regions surrounding the origins. By 60 min the overall occupancy has fallen sharply across the chromosome indicating that the replication process for most of the population is nearing completion and the GINS complex has been released. The GINS re-associates with the origins at the beginning of the next cell cycle (data not shown). Thus, the Psf2 reporter provides an animated view of the GINS progression across an entire chromosome.

Similar data to that shown in Figure 4-4 and 4-5 were obtained for all 16 yeast chromosomes (Appendix C). At the time resolution of the experiment (5 min), we discern three broad categories that describe the association of GINS with origins. The first includes those origins that fire in the interval 20-25 min. The second includes those that fire in the interval 25-30 min. The final category includes origins to which the GINS complex binds, but from which bidirectional spreading is not observed. While others have reported similar early and late firing origins (Poloumienko et al., 2001; Raghuraman et al., 2001; Yabuki et al., 2002; Yamashita et al., 1997), our data reveal the

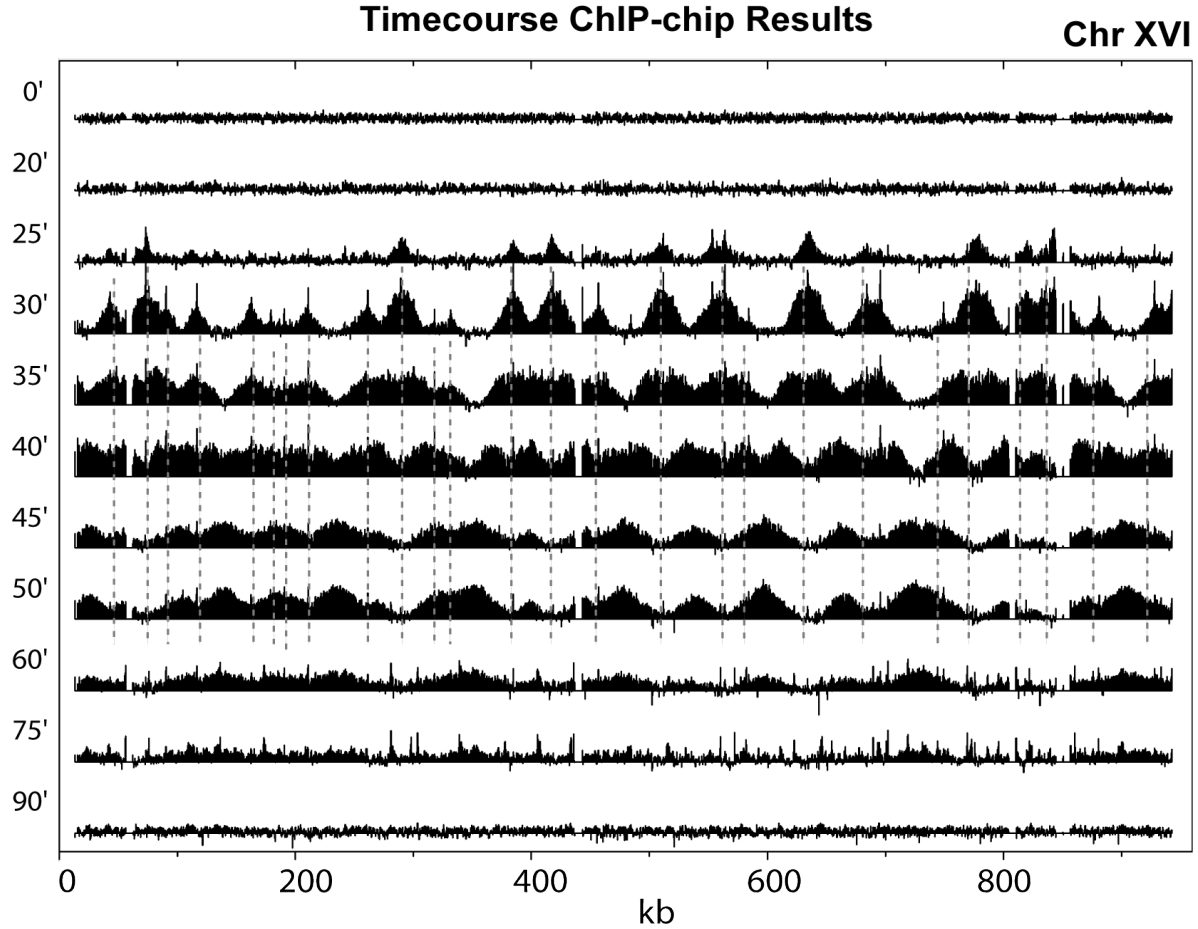


Figure 4-4. Time-course ChIP-chip analysis of GINS complex binding to chromosome 16 through the cell cycle. Strain MSY1 (Psf2-PrA) was synchronized in YPD media with α-factor. Starting from G1, ChIP-chip was performed at eleven time points through the cell cycle (see Materials and Methods). The GINS complex was immunoisolated at each time point. Associated DNA was amplified and hybridized to yeast whole genome microarrays. Relative occupancies of the GINS complex for sites across chromosome 16 are shown for each time point. Locations of the initial binding sites are noted with red dashes.

vast majority of origins fire in a much narrower time window – i.e., 15 min versus 30 min. Applying this classification system to chromosome 15 (Figure 4-5), we identified a total of 29 origins, of which 17 are category 1, nine are category 2, and three are category 3. In the whole genome, we identified 168, 135, and 24 origins respectively in categories 1, 2 and 3 (Appendix F, which also provides a detailed comparison with previous studies). Overall, we observe 303 origins that give rise to active GINS progression (i.e., categories 1 and 2). Although a small number of late firing origins may be obscured by proximal early firing origins, our data were determined with sufficiently high resolution and signal-to-noise to provide an accurate map of the majority of origins that are active in the cell cycle.

Most of the 303 active origins fire within a short time window (~10 min) (Figures 4-4, 4-5, and Appendix C). The progression rates for 278 spreading peaks emanating from 199 of these 303 origins are determined by direct measurement of the positions of the peak edges as a function of time (Figure 4-6). We obtain an average progression rate of 1.6±0.3 kb/min, a value that is considerably lower and more narrowly distributed than previous replication rate estimates (2.9 kb/min with a range between 0.5 – 11 kb/min (Raghuraman et al., 2001) or 2.8 +/- 1.0 kb/min (Yabuki et al., 2002). The bidirectional spreading that emanates from the origins progressed to the left and right at closely comparable rates. From these data, we infer that this movement is largely symmetric and occurs at a highly characteristic rate, implying that, in contrast to previous inferences relating chromatin position to replication rate, there are few obvious chromosomal features that locally alter uniform GINS progression.

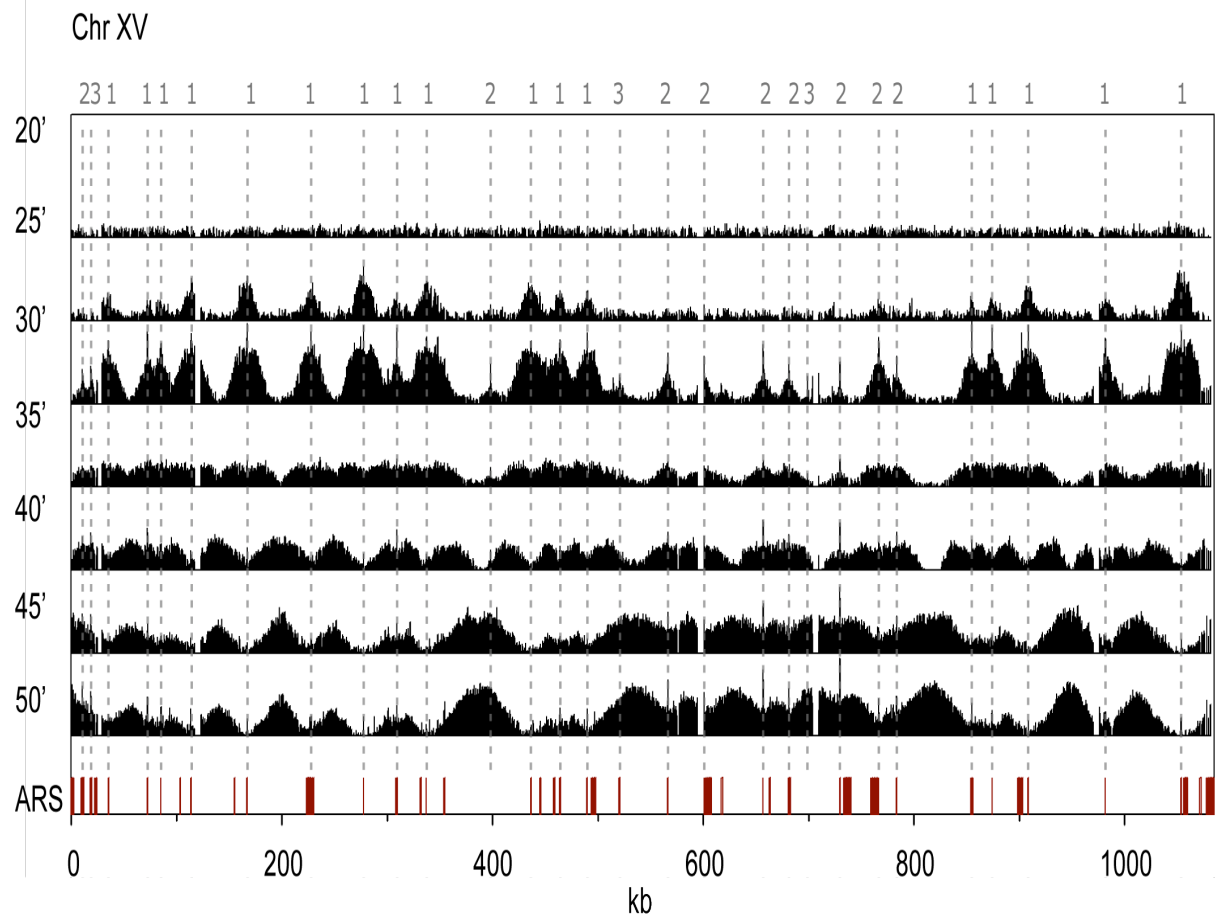


Figure 4-5. Time course ChIP-chip analysis of GINS complex binding to chromosome 15 through S-phase. At the beginning of S-phase (20 min time point) the GINS complex has not yet associated with any part of the chromosome. As S phase progresses, replication initiates at the indicated origins (25 - 30 min time points) whereupon the GINS complex moves away bi-directionally with the replication forks. Replication of a particular chromosomal region is finished when adjacent forks collide, giving rise to broad peaks in the inter-origin spaces. The locations of previously defined ARSs (excluding dubious origins as defined by (Nieduszynski et al., 2007) are shown as red columns in the bottom panel.

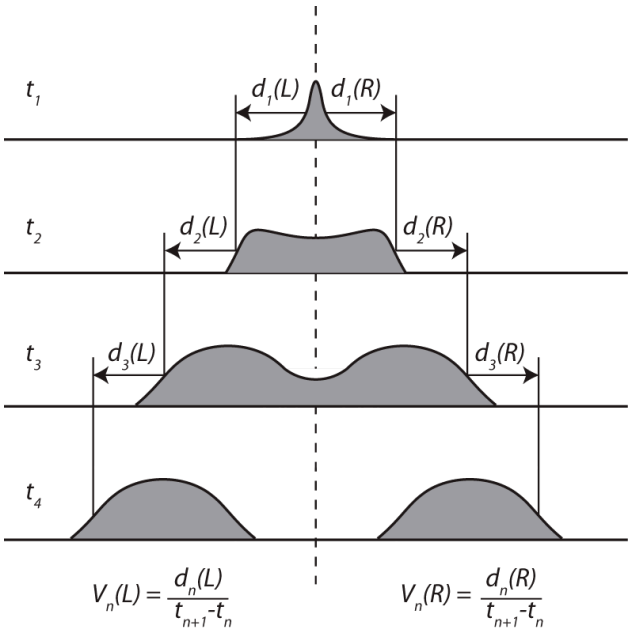


Figure 4-6. Schematic of fork progression rate calculations

The method is sufficiently sensitive to detect small heterogeneities in replication dynamics and in the process sheds new light on the phenomenon described as 'replication pausing' – short duration stalling of forks at numerous specific sites in the genome (Azvolinsky et al., 2006; Azvolinsky et al., 2009; Deshpande and Newlon, 1996; Ivessa et al., 2003). For example, pausing intervals at tRNA genes have previously been estimated at ~10 sec (Deshpande and Newlon, 1996; Ivessa et al., 2003) – i.e., ~4 times longer than the interval required for unimpeded replication fork transit of a tRNA gene. Although such events would seem too subtle to detect with 5 min resolution (Azvolinsky et al., 2009), we observed sharp features at 267 out of 275 these tRNA genes (discussed in more depth in Chapter 5), indicating that our methodology is sensitive enough to detect small perturbations, and that GINS movement can indeed be hindered to some degree (Figure 4-7). This phenomenon was also observed at 81 of the 83 snoRNA and snRNA genes, and 95 of the 100 other most highly transcribed genes in the genome (Figure 4-7), and appears to be independent from the direction of transcription.

These data represent the highest time resolution ChIP-chip experiments to date. The data contradict much of the literature regarding replication fork behavior and progression velocities. The discrepancies will be discussed in detail in Chapter 5.

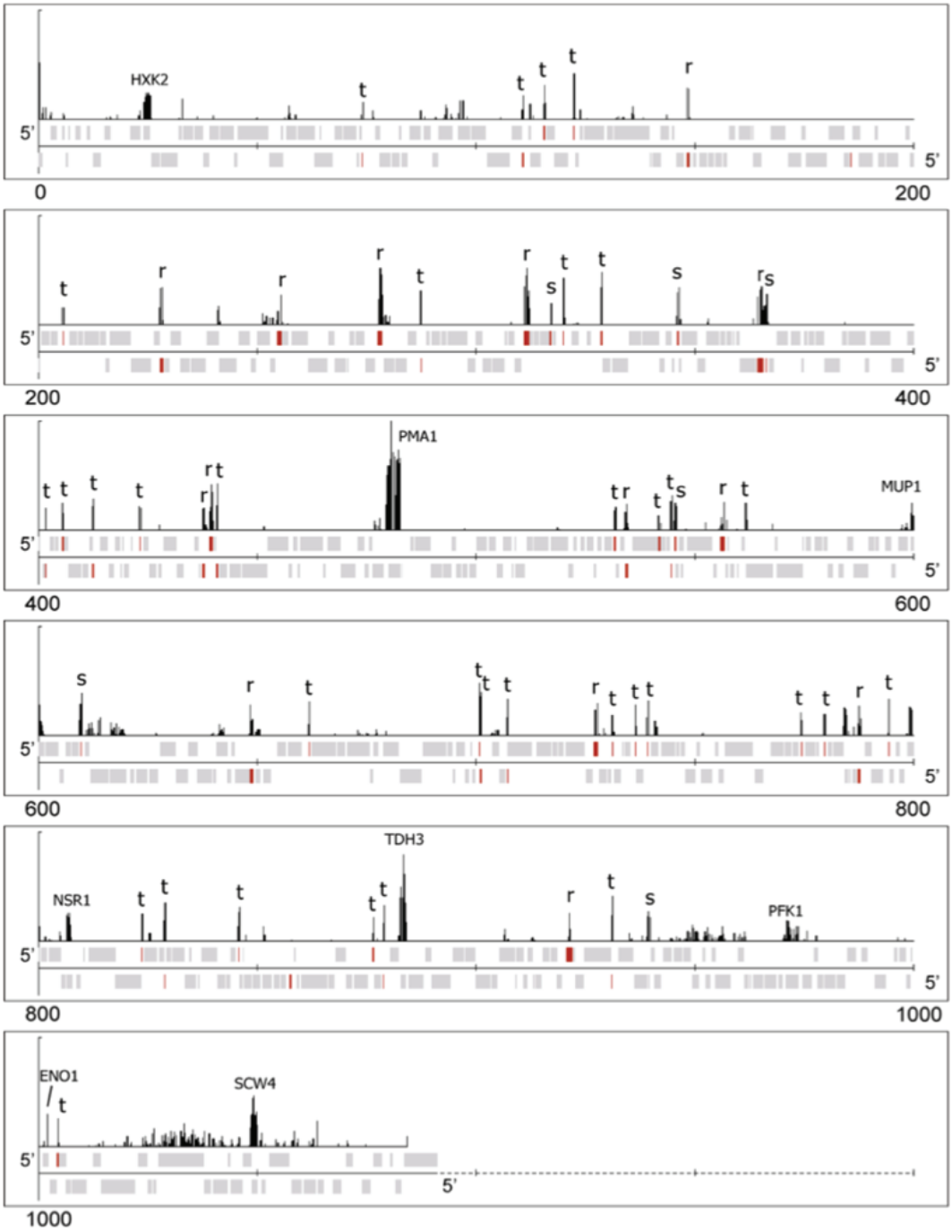


Figure 4-7. Persistent features associated with GINS progression through highly transcribed genes. ChIP-chip data for the GINS with the entire chromosome 7 are shown for the 75 min time point. The baseline has been raised to emphasize peaks. Persistent features can be observed at all of the tRNA genes (t), snoRNA genes and snRNA genes (s), rRNA genes (r), as well as 6 out of the other 9 most highly expressed genes on this chromosome (indicated by the gene name).

4.2 - GENOME-WIDE & TIME-RESOLVED CHIP-CHIP LOCALIZATION OF POLYMERASE ε AND POLYMERASE δ

MATERIALS AND METHODS

Yeast Strain Construction and Growth Conditions

Strains MSY9 (POL2-PRA, bar1Δ) and MSY10 (POL3-PRA, bar1Δ) were constructed from strains ATY27 by genomically tagging the POL2 and POL3 genes, respectively (see appendix A). The BAR1 gene has been deleted, rendering these strains hypersensitive to α-factor arrest in G1.

Cell Synchronization

Strains MSY9 and MSY10 are cultured and synchronized as in the GINS proteomics experiments in Section 3.1, and the time-course ChIP-chip experiment in Section 4.1.

Time-course Sampling, Cell Cycle Monitoring, Chromatin Immunoprecipitation

Time points were collected and cross-linked as in the GINS time-resolved ChIP-chip experiment (Section 4.1). Budding index and FACS analysis was performed as in Sections 3.1 and 4.1. ChIP was performed as previously reported (Ren et al., 2000; Tackett et al., 2005), and as described in the GINS time-resolved ChIP-chip experiments (Section 4.1).

Real-Time PCR and Microarray Analysis

Real-Time PCR and microarray hybridization and analysis was performed as in the GINS time-resolved ChIP-chip experiments above (Section 4.1). The microarray results were normalized as in GINS time-resolved ChIP-chip experiments, and fork velocities were calculated as in Section 4.1

RESULTS

The GINS complex proved to be an invaluable marker for monitoring the progression of replication forks as they traversed the genome. Intriguingly, the overall rates of fork progression and the distribution of velocities differed significantly from previous reports (Raghuraman et al., 2001). One possible explanation for this is the differences in experimental design – prior studies monitored nucleotide incorporation, while this ChIP-chip approach focuses on a protein that was ancillary to the actual process of DNA synthesis. To account for the possibility that the CMG complex is de-coupled from the polymerase at some regions, a follow-up study in which the progression of the replicative polymerases are monitored in the same way that the GINS complex is described here. Differences in the progression would indicate that replication forks behave differently from the polymerase enzymes. Because the roles of Pol ε and Pol δ remain somewhat ambiguous, both were monitored.

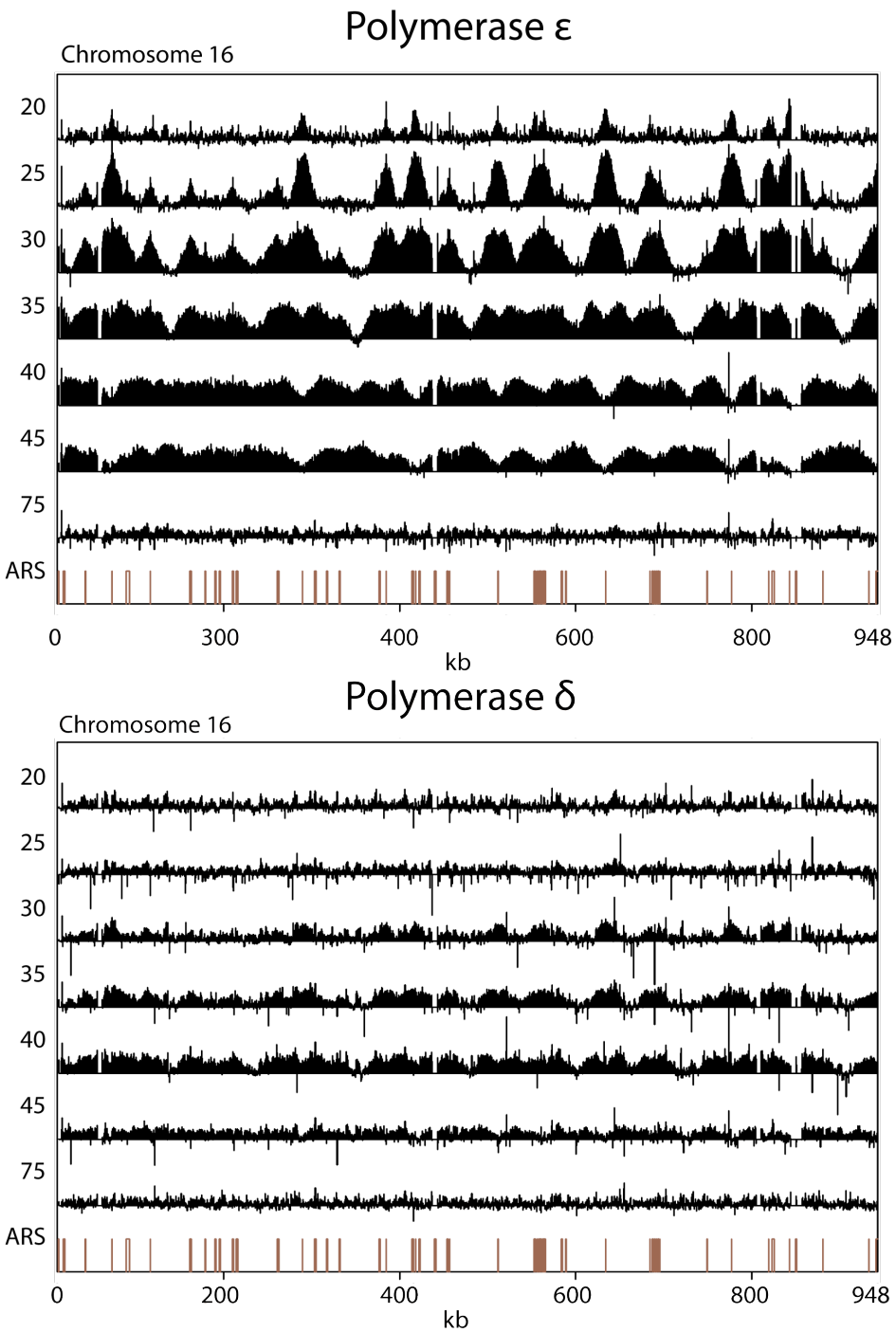


Figure 4-8. The time-resolved ChIP-chip analysis of Pol2-PrA (Pol ε, top panel) and Pol3-PrA (Pol δ, bottom panel) progression during S phase, shown for chromosome 16. Neither polymerase is localized to any particular region of the genome. The Pol3-PrA data is noisier than the Pol2 results, but the overall picture of Pol δ progression can be see and looks similar to that of Pol ε.

A second motivation for performing this experiment is the fact that the precise roles that Pol ε and Pol δ have in replicating the genome are debated. It has been assumed that one would be utilized on the leading strand, while the other on the lagging strand, but the evidence for this has remained elusive. The only competing theory that can account for all of the currently accumulated data is one in which Pol ε replicates near the origin and Pol δ takes over after some distance (Figure 1-10). By monitoring the movement of the polymerases in the same manner described for monitoring GINS complex progression (Section 4.1), we are able to test the veracity of this alternate hypothesis.

As in the Psf2 time-course ChIP-chip experiment, both Pol2 (Pol ε) and Pol3 (Pol δ) bind to sites of origins and move away bi-directionally once an origin fires (Figure 4-8 and 4-9). The rates of progression are measured to be 1.6 ± 0.3kb/min. No evidence for asymmetric fork progression was observed in either experiment, and the overall profiles of Pol ε and Pol δ progression look identical to that of Psf2. In terms of quality of the experiments, the Pol3 ChIP-chip data (bottom panels in Figures 4-8 and 4-9) do not appear to be as strong as is observed for Psf2 (Figures 4-4 and 4-5) or Pol2 (top panels in Figures 4-8 and 4-9). The overall level of noise is much higher in this experiment, but the patterns of progression are apparent.

In general, the profiles observed in the Pol2 experiment appear to be a few minutes earlier in the cell cycle than those observed for either Psf2 or Pol3 (compare the appearance of peaks in the two panels in Figure 4-7, and the appearance of peaks in Figure 4-4). If this observation is to be believed, then Pol ε arrives at the origins significantly earlier than the GINS complex or Pol δ, and progresses ahead of the

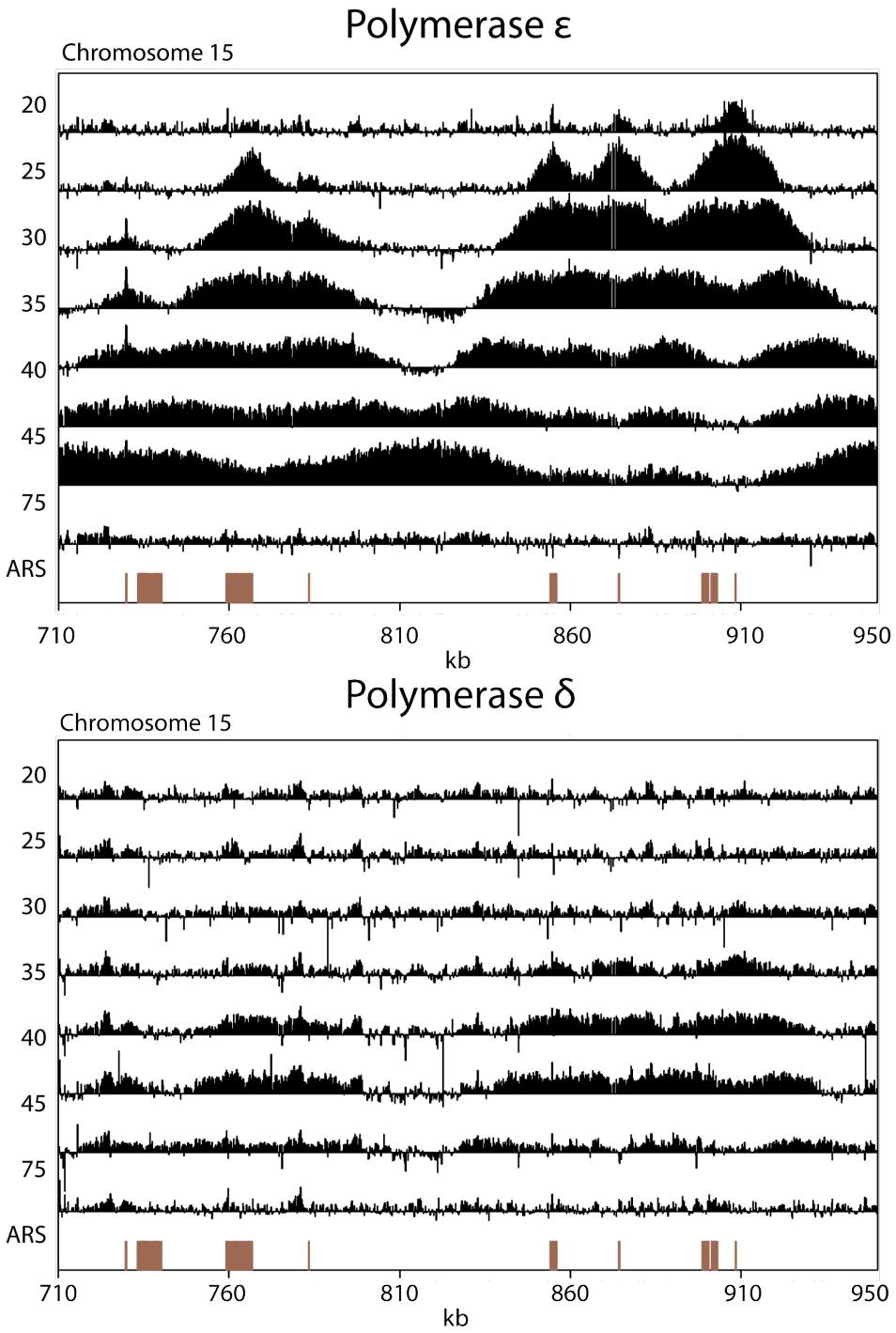


Figure 4-9. The time-resolved ChIP-chip analysis of Pol2-PrA (Pol ε, top panel) and Pol3-PrA (Pol δ, bottom panel) progression during S phase, shown for the 710-950 kb region of chromosome 15.

helicase. A much more likely explanation is that the cell cycles between the experiments have become slightly displaced, resulting in profiles with slightly different start times across the genome.

4.3 – GENOME-WIDE & TIME-RESOLVED CHIP-CHIP CO-LOCALIZATION OF THE GINS COMPLEX & POLYMERASE ε

MATERIALS & METHODS

Strain Construction

For information on the strains used in this sections see Appendix A, Table A-1. Strain MSY10 and MSY11 were constructed from strain ATY27 by genomically tagging the C-terminus of the Pol2 and Psf2 proteins with the PrA tag. Strain MSY16 was constructed from the MSY10 strain by genomically tagging C-terminus of the Pol3 open reading frame with a 13xMyc tag. Similarly, strains MSY13 and MSY17 were constructed from strain MSY11 by genomically tagging C-terminus of Pol2 and Pol3 with the 13xMyc tag.

Cell Growth and Synchronization

For the asynchronous antibody test IPs, growth conditions were identical to those used in the asynchronous IPs of Psf2-PrA and Pol3-PrA in Chapter 2. For synchronized antibody tests by ChIP-chip, growth conditions were identical to those used in the Psf2-PrA time-resolved ChIP-chip experiment in Section 4.1, except only 25 and 30 min time points were harvested.

Antibody Test Immunoisolations

Strains MSY11 (PSF2-PRA, bar1Δ) and MSY12 (PSF2-MYC, bar1Δ) were utilized to evaluate a variety of α-Myc antibodies for their cross-reactivity with PrA (see table 4.1 for a list of the antibodies tested). Protein extracts from 3g of asynchronous cells of both

strains were split seven ways and incubated with either rabbit IgG or one of six α-Myc antibodies. The immunoisolation protocol was identical to that used in the asynchronous Psf2-PrA IPs (Chapter 2). Half of each sample was analyzed by western blot using a HRP-conjugated goat polyclonal to Myc-tag (Abcam), while the remaining half was analyzed by SDS-PAGE stained with the Pierce Silver Stain Kit (Thermo).

Table 4-1. Antibodies Tested for Cross-Reactivity

Antibody	Antigen Recognized	Clone ID	Source	Manufacturer
Mouse-IgG	C-Myc	9E10	Mouse	Sigma
Mouse-IgG	C-Myc	Myc.A7	Mouse	Abcam
Rat-IgG	C-Myc	JAC6	Rat	Abcam
Rabbit-IgG	C-Myc	NA	Rabbit	Sigma
Chicken-IgY	C-Myc	NA	Chicken	Genway
Chicken-IgY	C-Myc	NA	Chicken	ICL, Inc.

Chromatin Immunoprecipitation

In general, immunoisolation of the cross-linked protein-DNA was carried out in a similar manner to the GINS ChIP-chip experiment (Section 4.1). After shearing the DNA, the samples were split into two tubes for incubation with each of the antibodies. Washing, eluting, preparation for hybridization, and data analysis was identical to the protocol described in Section 4.1.

RESULTS

Monitoring the GINS complex revealed several important aspects of replication fork behavior (Section 4.1). Most intriguing is the observation that replication forks appear to move at highly uniform velocities that are independent of genomic location. Monitoring the progression of Pol ε and Pol δ (Section 4.2) suggests that these enzymes behave very similar to each other while traversing the genome, dispelling the possibility that the uniform progression of GINS is specific to the helicase, and that this behavior cannot be generalized to the replication machinery as a whole. However, some minor differences in the timing of the experiments, as discussed at the end of Section 4.2, were detected. These could be caused by the discrete behaviors of each of the proteins in question, or by experimental differences, such as slight delays in the release of the cultures from the G1 block.

To address the possibility of experimental differences, and determine conclusively whether the polymerase movement is coupled to the helicase, we have chosen to monitor the genome-wide localization of Pol ε and the GINS. To do so, "mixed-tag" strains were constructed in which one of the proteins of interest was genomically-tagged with PrA, and the other with the C-Myc epitope.

A general concern with this strategy is that most commercial antibodies for C-Myc are isoforms of IgG, most of which bind to the PrA tag via its constant region. Therefore, it was necessary for us to identify a suitable antibody that does not cross-react with PrA-tagged protein. Once this was accomplished, time-resolved ChIP-chip studies could be performed that allowed us to directly compare the movement of the GINS and Pol ε.

4.3.1 - Antibody Tests

Of the six α-Myc antibodies tested, three originated from mouse, one from rabbit, one from rat, and two from chicken embryos (Table 4-1). Based on the known interactions between PrA and IgGs of different species (Figure 4-10), the rabbit and mouse α-Myc should interact with Psf2-PrA from strain MSY11 to a high degree, while rat and chicken α-Myc should not. This was confirmed by incubating extract from this strain with each of the antibodies in question (Figure 4-11, left panel). The results show that rat and chicken α-Myc will not cross-react with the PrA tag.

A second test that incubates MSY12 extract with the various antibodies determines whether the antibody is good for immunoisolation (Figure 4-11, right panel). This

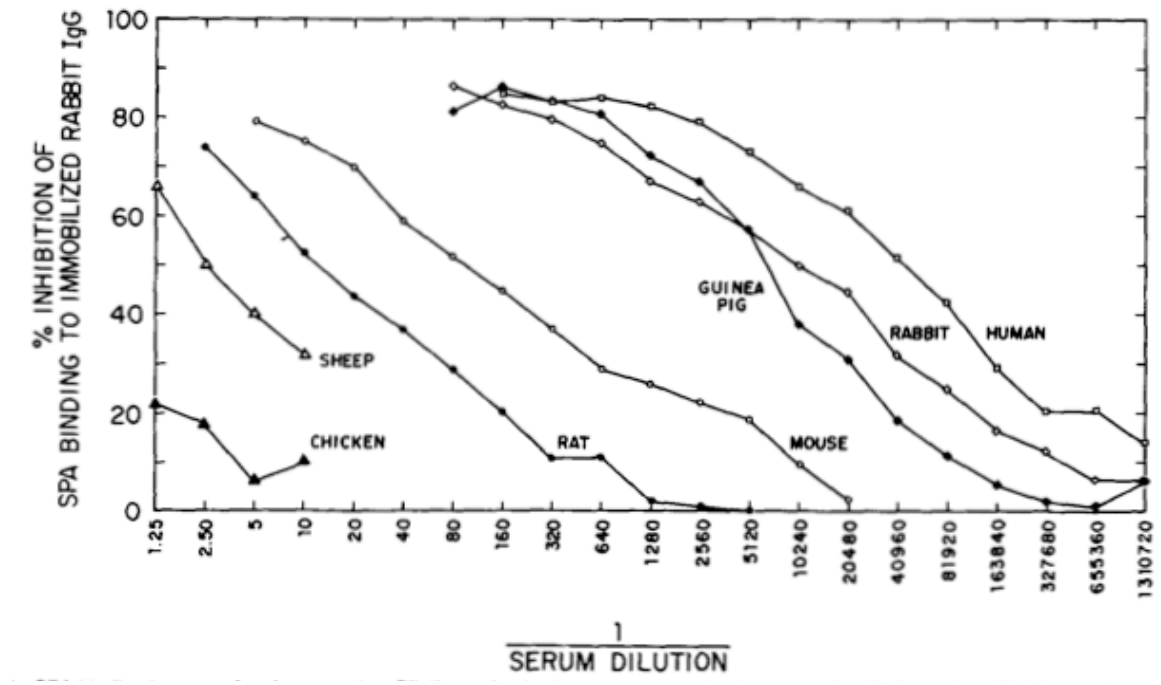


Figure 4-10. Cross-reactivity of PrA with IgG from different species (adapted from (Richman et al., 1982)). PrA from *S. aureus* binds extremely well to serum from humans and rabbits, moderately well to serum from mice and rat, and binds poorly to serum from chicken.

shows that antibody JAC6 (from rat), 9E10 (from mouse), and rabbit polyclonal α-Myc pulled out Psf2-PrA to a degree that was comparable to the standard rabbit IgG that is used for PrA isolations (Figure 4-11, left panel, PrA IP lane). A.7 and 33 (both from mouse) and both chicken α-Myc antibodies show a decreased level of interaction with Psf2-PrA, consistent with predictions based on the known PrA interactions.

In the right panel of Figure 4-11, the amount of Psf2-Myc present in the isolation with 33 was negligible. Isolations performed with A.7, JAC6, and 9E10 showed a reasonable level of Psf2-Myc, and the isolations with both of the chicken IgYs yielded very large amounts of Psf2-Myc. The signal in the chicken lanes is so strong that it used up the reagent before the film was developed.

SDS-PAGE analysis of the complex verified that the chicken antibody was successful at immunoisolating the GINS complex from MSY12 strain, but did not pull-down GINS from the MSY11 strain (Figure 4-12). The rat (JAC6) antibody pulled down some Psf2-Myc, but also cross-reacted with Psf2-PrA.

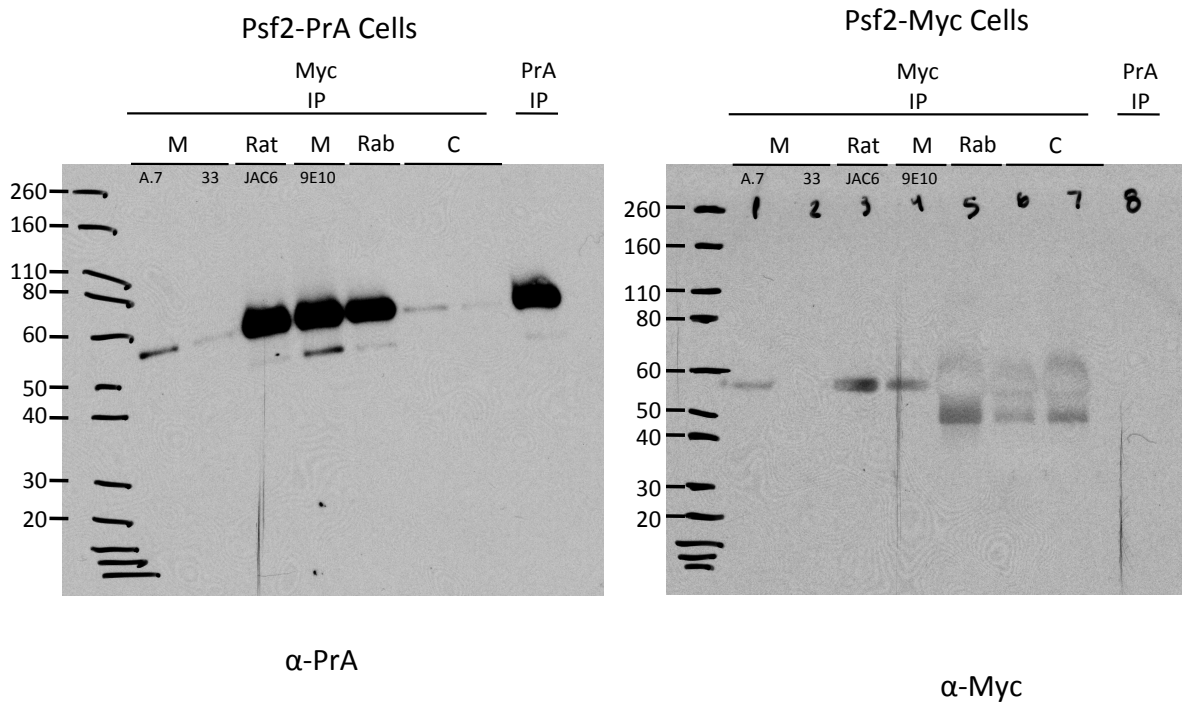


Figure 4-11. Testing the specificity of various α-Myc antibodies. Extracts from MSY11 (left panel) and MSY12 (right panel) cells were incubated with beads conjugated with one of each of the six α-Myc antibodies described in Table 4-1 (and rabbit-IgG as a control, last lane in each blot). The beads were washed and eluted, and the samples were analyzed by western blot. The MSY11 samples were probed with rabbit-IgG to test for the presence of Psf2-PrA, while the MSY12 samples were probed with a general α-Myc to test for the presence of Psf2-Myc. The letters at the top of the blot indicate the animal that the antibody originates from (M – mouse; Rat – rat; Rab – rabbit; C – chicken). The clone ID is listed below where applicable. The dark bands in the Rat, M, and Rab lanes in the MSY11 blot indicate that these antibodies cross-react with PrA. The overexposed bands in the Rab, and two C lanes, indicate that these antibodies pull down large amounts of the Myc-tagged protein.

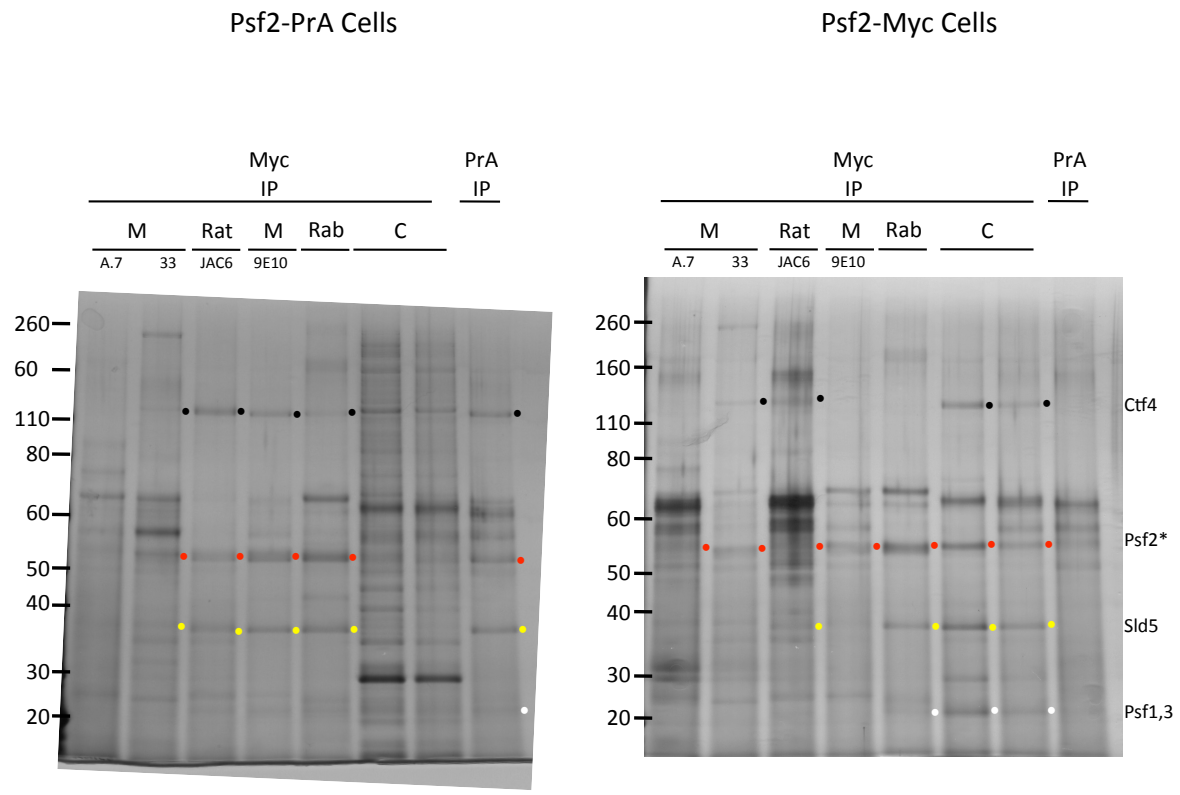


Figure 4-12. Testing the specificity of various α-Myc antibodies. The same samples described in Figure 4-10 were analyzed by SDS-PAGE, visualized with silver stain. The black dots indicate the presence of Ctf4; the red dots indicate the presence of tagged Psf2; the yellow dots indicate the presence of Sld5; the white dots indicate the presence of Psf1 and Psf3. Extracts from MSY11 (left panel) and MSY12 (right panel) cells were incubated with beads conjugated with each of the six α-Myc antibodies described in Table 4-1 (and rabbit-IgG as a control, last lane in each blot). The staining pattern in the left panel indicates that the mouse (33), rat, mouse (9E10), and rabbit antibodies cross-react with Psf2-PrA. The staining pattern in the right panel indicates that The rat, rabbit and the two chicken antibodies adequately pull down the CMG complex via Psf2-Myc.

Based on these two tests, either of the chicken α-Myc antibodies appeared to be good candidates for use in a mixed-tag experiment, since they showed excellent specificity and reasonable enrichment for Psf2-Myc as compared to the other options. For all co-ChIP-chip experiments, the chicken α-Myc antibody from ICL, Inc was utilized.

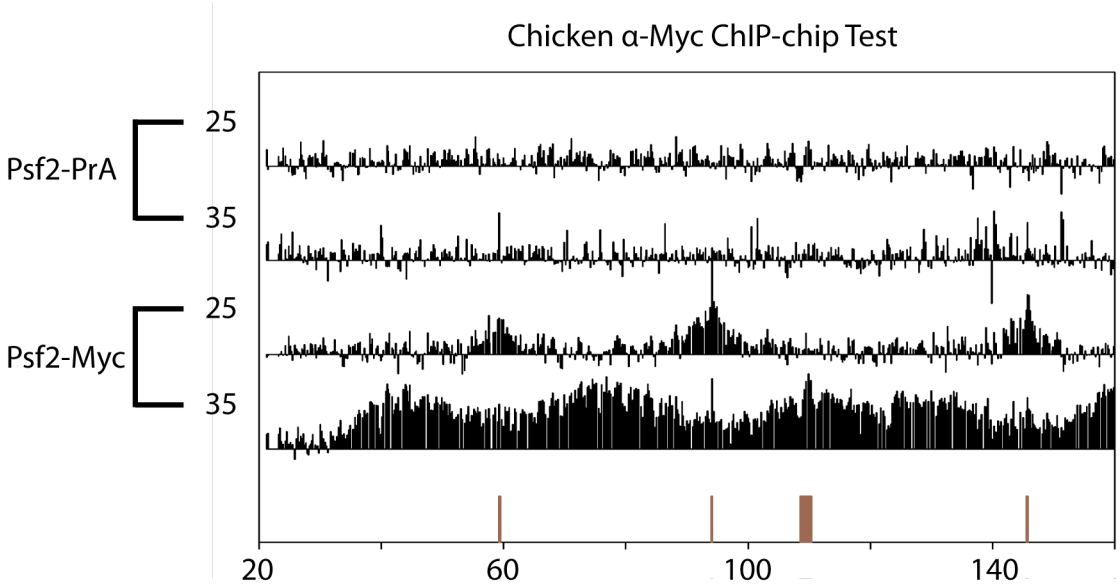


Figure 4-13. Testing the specificity of the chicken α–Myc antibody in ChIP-chip experiments. MSY11 (Psf2-PrA) and MSY12 (Psf2-Myc) cells were synchronized and samples were cross-linked 25 and 35 min after release from G1 arrest. The samples were immunoisolated with chicken α-Myc AB from ICL. No signal is observed in the PrA tagged strain, while robust replication fork progression is seen in the Myc tagged strain.

As a final test to verify that the ICL chicken α-Myc was truly specific for the C-Myc tag under the desired experimental conditions, strains MSY11 (Psf2-PrA) and MSY12 (Psf2-Myc) were subjected to time-resolved ChIP-chip with this antibody (Figure 4-13). S phase time points (25 and 30 min after release from G1) of each strain were subjected to formaldehyde cross-linking and immunoisolated with the ICL chicken α-Myc antibody. The results show that the antibody produced bi-directional movement from the origins

in strain MSY12, but no signal from strain MSY11. This confirms that this antibody is suitable for use in the mixed-tag ChIP-chip experiments.

4.3.2 - Psf2-PrA and Pol2-Myc Co-ChIP-chip

Strain MSY13 (Psf2-PrA, Pol2-Myc) was utilized to assess the differences in the progression of the GINS complex and Pol ε. Figure 4-14 shows the results for all of chromosome 16. Despite the high level of noise in the Psf2-PrA IP (Figure 4-14, top panel), the overall patterns demonstrate that the progression of the two complexes is identical.

This can be more easily seen in a close-up region of chromosome 15 (Figure 4-15). Peaks can be seen in the 30 min time point at a category 1 origin (~910 kb) in both experiments. Bidirectional spreading is observed through the duration of S phase, and termination of fork progression occurs at sites intermediate to the origins in the 50 min time point. The progression of both proteins is identical.

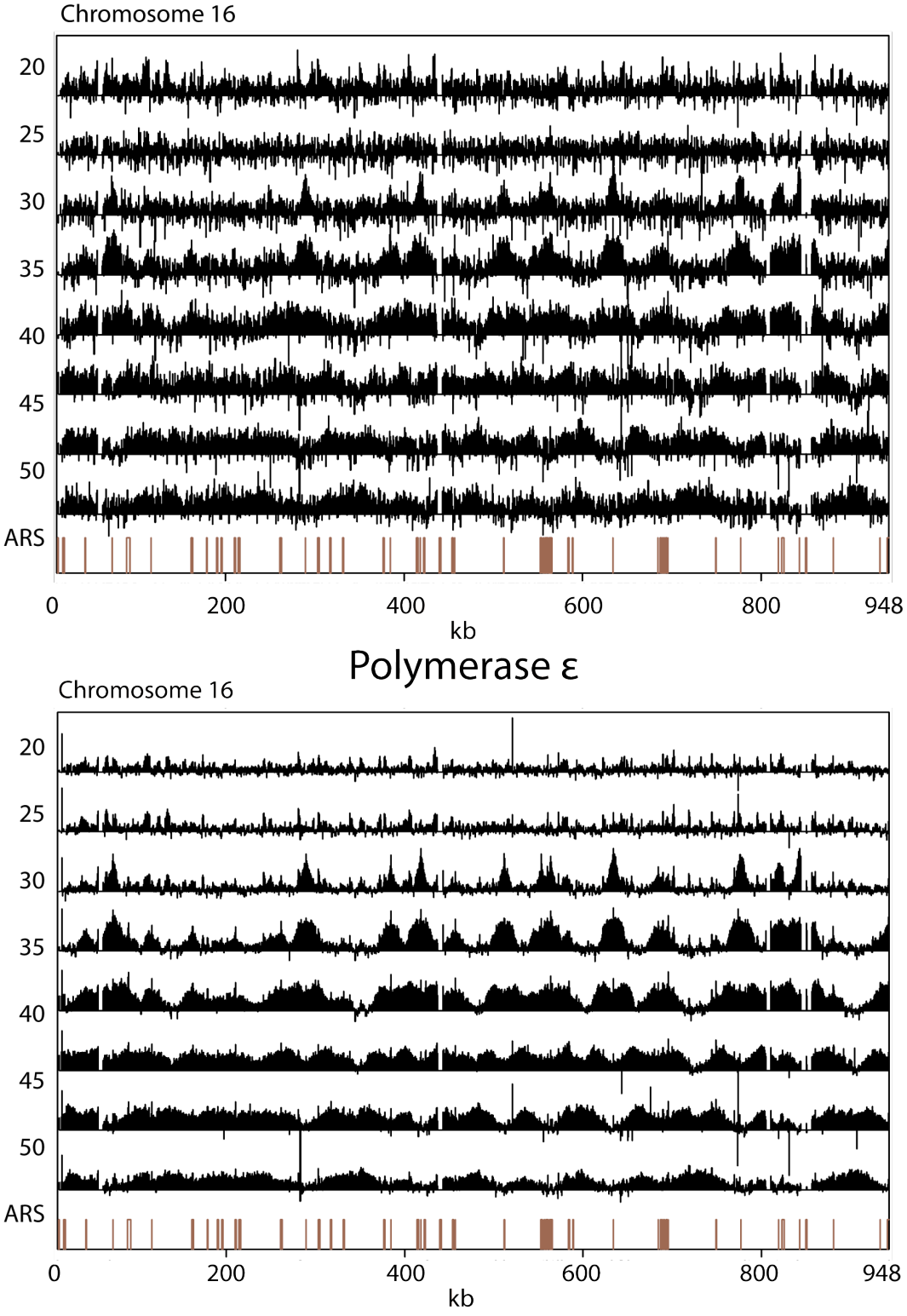


Figure 4-14. The time-resolved co-ChIP-chip analysis of Psf2-PrA (top panel) and Pol2-Myc (bottom panel) progression during S phase, shown for chromosome 16.

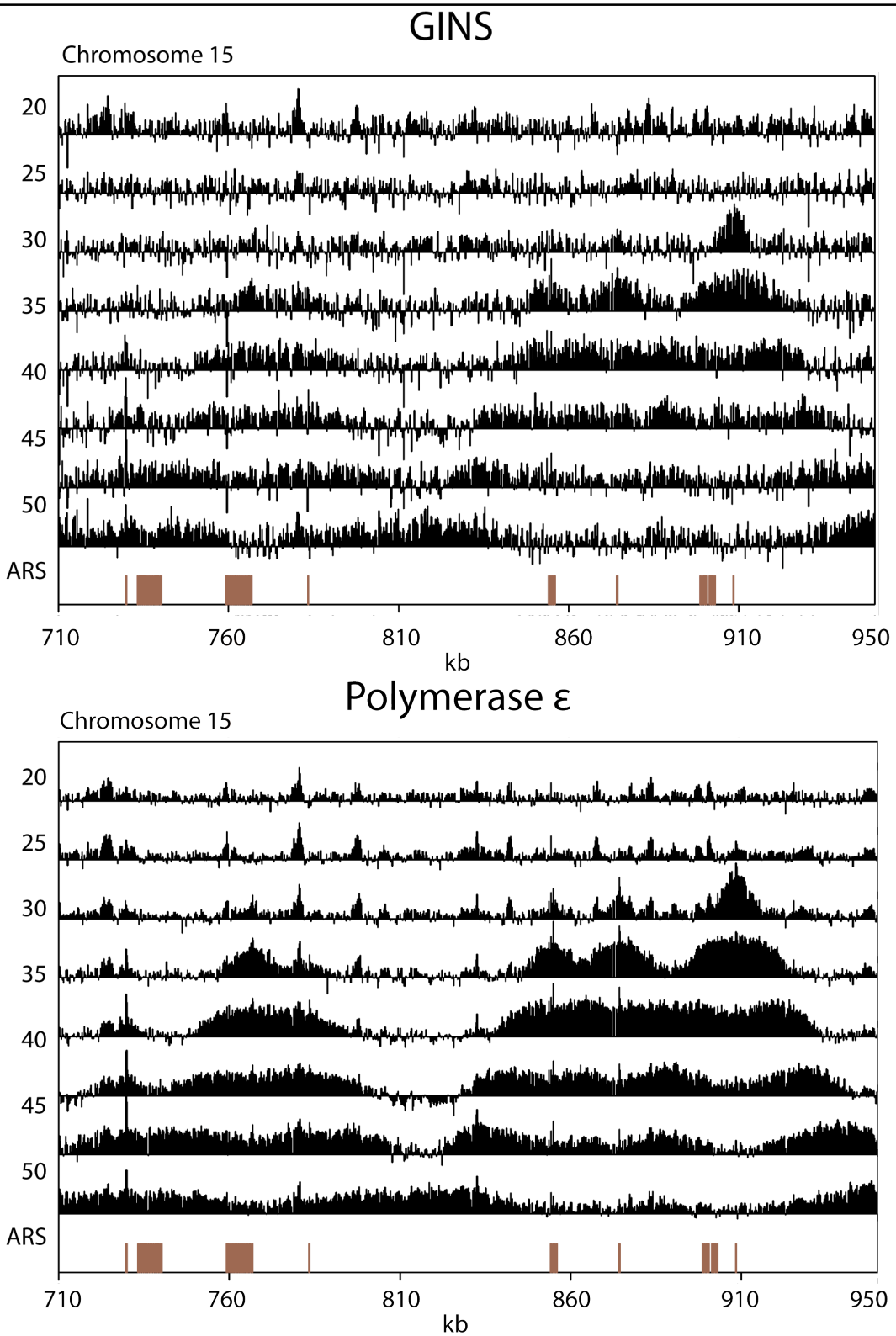


Figure 4-15. The time-resolved co-ChIP-chip analysis of Psf2-PrA (top panel) and Pol2-Myc (bottom panel) progression during S phase, shown for chromosome 15 region 710-950 kb.

CHAPTER V – DATA PROCESSING & ANALYSIS

The experiments described in Chapters 3 and 4 all couple a ChIP protocol (to isolate the DNA associated with three different replication fork proteins) with microarray technology (for visualizing the relative occupancies of these proteins). The substantial amount and complex nature of the data requires a significant degree of analysis in order to extract useful, quantitative information about the localization and progression of GINS, Pol ε and Pol δ. Importantly, several processing steps are required to extract accurate details from the data, and the confident interpretation of the time-resolved ChIP-chip data requires careful normalization of signal intensities across the genome (discussed below in Section 5.1). These steps help to reveal the locations of the forks within individual time points, and allow us to deduce the behavior of the forks as they traverse the genome by comparing different time points.

From the resulting time-resolved ChIP-chip data, we infer that this movement is largely symmetric and occurs at a highly characteristic rate, implying that, in contrast to previous inferences relating chromatin position to replication rate, there are few obvious chromosomal features that locally alter uniform GINS progression (Chapter 4). To test the generality of this rather simple view of GINS progression throughout the genome, we use our measurements to generate an iterative model of this process (discussed below in Section 5.2).

Finally, the differences between our findings and previous calculations of fork progression rates were too striking to dismiss. We have therefore carefully re-analyzed the most influential nucleotide incorporation study (Raghuraman et al., 2001) in an

attempt to comprehend at which steps our results differ. Our findings help to explain why certain aspects of the replication program have been misinterpreted for nearly a decade, and why the direct monitoring of replication fork proteins is a more suitable technique for making these calculations.

5.1 – Dynamic Range Extrapolation & Signal Normalization

For clarity's sake, the following description regarding data processing will refer to the time-resolved GINS ChIP-chip experiments describe in Section 4.1, but the remaining experiments were analyzed in an identical manner. Microarray results are returned as log2 ratios of the IP and WCE fluorescence. In one of the data sets, the fluorescence signals for all of the IP and WCE samples appeared saturated to an unacceptable degree (~25% of the data). The slides were therefore re-scanned at a lower gain setting, and the mean intensity ratios for the high and low scans were determined from the raw data (Figure 5-1). The intensity of the saturated data in the high gain setting was re-calculated by multiplying those signals in the low gain data by the mean intensity ratio, improving the dynamic range of the data, and providing an accurate picture of genome-wide GINS occupancy for each time point (Dudley et al., 2002).

Next, the extended dynamic range data was normalized. The data were binned using a 5000 base pairs bin size. The distributions for the average intensities and the RMSD for these bins can be seen in Figure 5-2 A. The first normalization of the data was carried out by finding the intensity where 3% of the bins had a negative average intensity. This gives a robust zero level that is independent of occasional large negative outliers in the data. The centroid of the RMSD distribution was then found for each time point and the intensities were scaled in such a way that these centroids were set to 1 (Figure 5-2 B). The second normalization was performed by finding an intensity threshold that separated the signal from the noise, and by normalizing the centroid of the RMSD distribution of the noise (Figure 5-2 C). An additional correction was used for time

points where the data was dominated by signal. This correction (1.5 fold for the 40 min time point) was based on the assumption that the number of forks in a wave is constant after firing of that origin has ceased and before the front of the wave has reached an adjacent wave of forks traveling in the opposite direction (see, for example, the 35 and 40 min data in Figure 4-4 and 4-5).

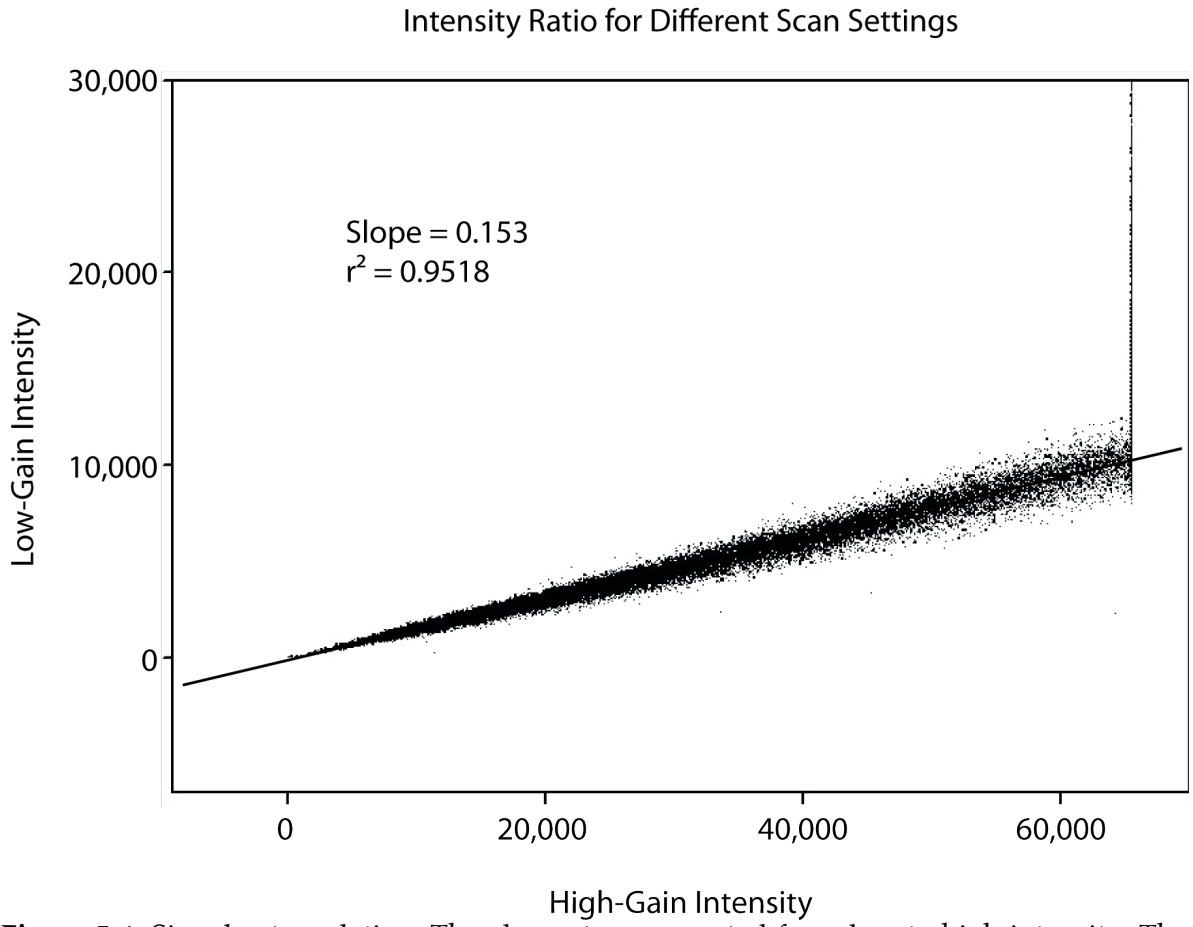


Figure 5-1. Signal extrapolation. The elements were sorted from low to high intensity. The intensity value of each array element from the low-gain scan (y-axis) was plotted against the intensity of the same element from the high-gain scan (x-axis). A large number of elements were saturated in the high-gain setting (evidenced by the straight vertical line on the right-hand side of the chart). A regression line was used to find the slope, which was then used to scale the low-gain intensities of the array elements that were saturated in the high-gain data. For this experiment the slope was ~1/6.5, so the low gain intensities were scaled by a factor of 6.5. This yielded data with improved dynamic range than would be obtained by using the low-gain data alone.

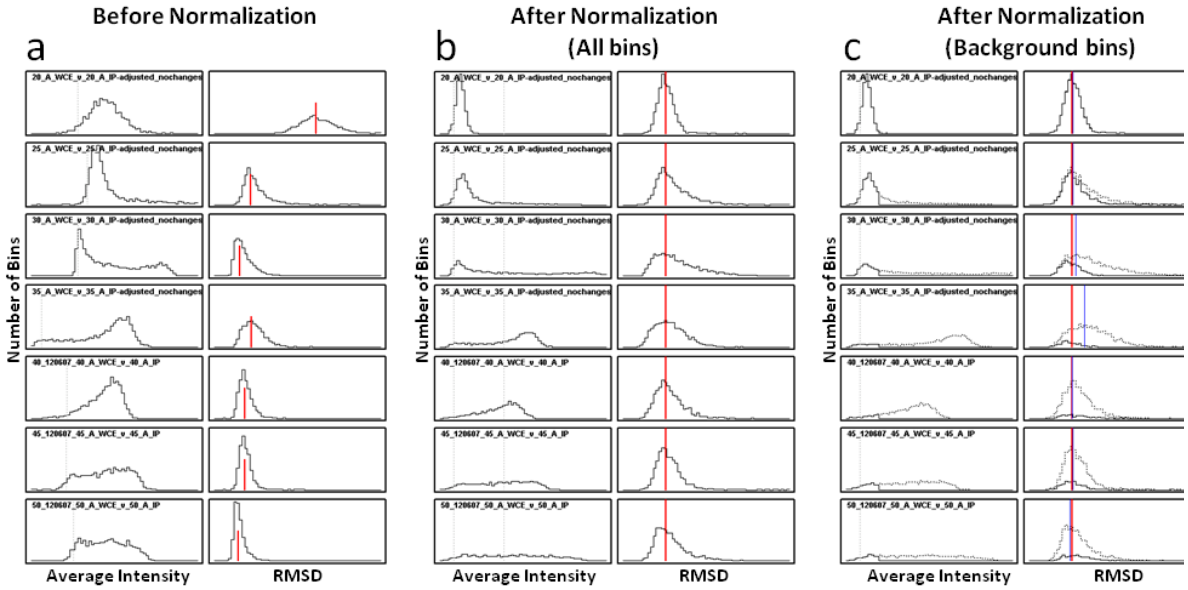


Figure 5-2. Data Normalization**.** The distributions for the average intensities and the RMSD for 5000 base pair bins. The average intensity of 3% of the bins is below the black dashed lines in the left panels. The red lines in the right panels show the centroid of the RMSD distribution. (a) before normalization; (b) normalization using all intensities: the zero is set so the average intensity of 3% of the bins is negative, and the intensities are scaled so the centroid of the RMSD distribution is one; (c) normalization using background intensities (lower than the blue line in b).

The extended dynamic range and normalization process yielded robust data sets that were easily analyzed. The peak areas appear relatively consistent between time points, implying that the GINS binding is consistent as it traverses the genome through S phase.

5.2 - SIMULATIONS

We utilize computer simulations to test our interpretation that replication fork progression is uniform across the genome. In accordance with the simplicity of our interpretation, our model uses a parsimonious set of assumptions: 1) the start of replication at origin *i* is normally distributed with an average start time t_i and a standard deviation σ; 2) replication progresses at a constant velocity, v = 1.6 kb/min, for all replication forks and over the whole genome; 3) each origin has an associated efficiency, e_i; 4) pausing might occur at a pause site *j* in a fraction (f) of the cells with a probability (p_j) and a duration (d_j).

The mean start times, the standard deviation of the start times, and origin efficiencies were determined from our data by minimizing the sum of the square differences between simulation results and experimental data. For each pair of adjacent origins, this minimization procedure used five slices of the data. These slices were chosen to encompass each of the origins and three equidistant regions between them, and the width of the slices was chosen to be 10% of the distance between the neighboring origins. The median was calculated for each of these slices, and used to calculate the sum of square differences. Each origin was assumed to have its own mean start time, ti, but the standard deviation of the start time was assumed to be constant in each simulation region. We also assumed that the GINS complex does not linger at the origin for any significant period of time, and after binding moves away 1.6 kb/min.

Finally, the GINS complexes fall off when adjacent forks collide. Each simulation was performed 10,000 times (to simulate 10,000 cells), using a random number generator to

determine start times within the Gaussian distribution for a particular origin. Examples of the effects of changing the parameters used in the optimizations are presented in Figures 5-3 (varying the standard deviation), 5-4 (varying the firing times), and 5-5 (varying the velocities). By varying the assumed fork velocity, our model also allows us to estimate the accuracy of velocities inferred from the time-resolved ChIP-chip analysis. Velocity changes as small as 0.4 kb/min can readily be discerned.

Examples comparing simulations based on our model with experimental data are provided in Figure 5-6. The top panel (Figure 5-6 A) compares a region containing a single category 1 origin and two adjacent category 2 origins. The model accurately recapitulates the features exhibited in the experimental data. The same is true of the more complex region containing seven category 1 origins shown in Figure 5-6 B. It also accounts for fork movement between the most widely spaced origins (~100kb apart) (Figure 5-6 C). Such regions are completely filled in at our experimentally determined average rate of 1.6 kb/min, within the duration of S-phase (~30 35 min at room temperature). Figs. 5-6 D and 5-6 E show that regions proximal to the telomeres and centromeres are as easily modeled as any other region of the genome. Category 3 origins were also readily modeled by allowing a small amount of binding without bidirectional spreading (Figure 5-6 F). We conclude that our simple model successfully recapitulates GINS movement throughout the genome, indicating that this movement is largely uniform irrespective of location.

Figure 5-3 (Next page). The effect of varying the standard deviation (σ_t) of the firing time distribution. Simulations with σ_t = 4, 8, and 12 min compared to the observed data for region 230 – 370 kb of chromosome 8 are shown here. A σ of 8 min was determined to give the best fit, on average, across the genome. Smaller standard deviations result in peaks that are too sharp, and wider standard deviations give peaks that are too broad and do not fill in the gaps properly. Origins are noted with dashed lines.

Figure 5-3

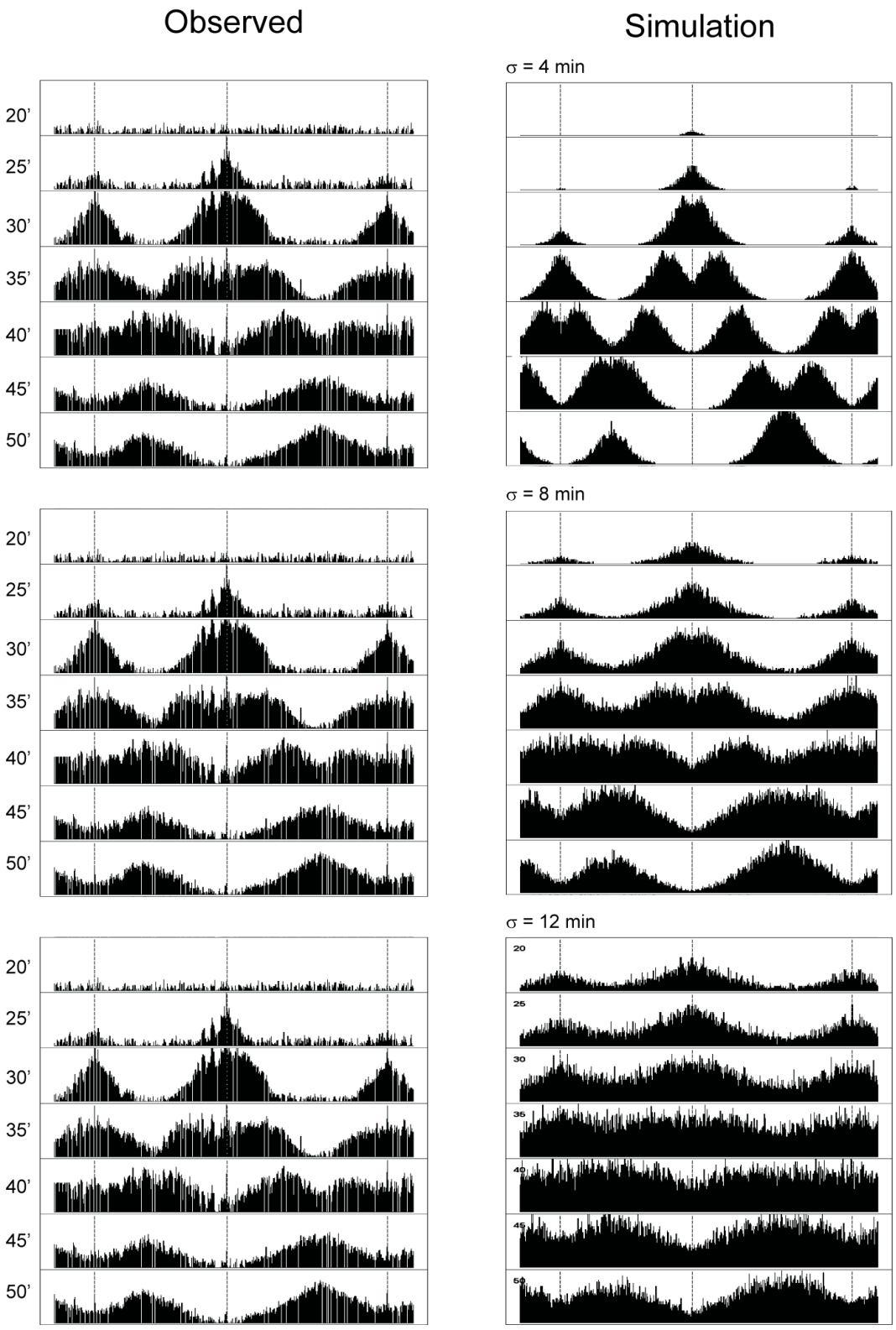


Figure 5-4 (Next page). The effect of varying the firing times for simulations. The mean of the optimized firing times (t) was shifted -3, 0, and +3 min for the region between 230-370 kb of chromosome 8. When start times were shifted -3 min, the peaks began to spread too early. When start times were set to + 3 min, adjacent peaks merged too late.

Figure 5-4

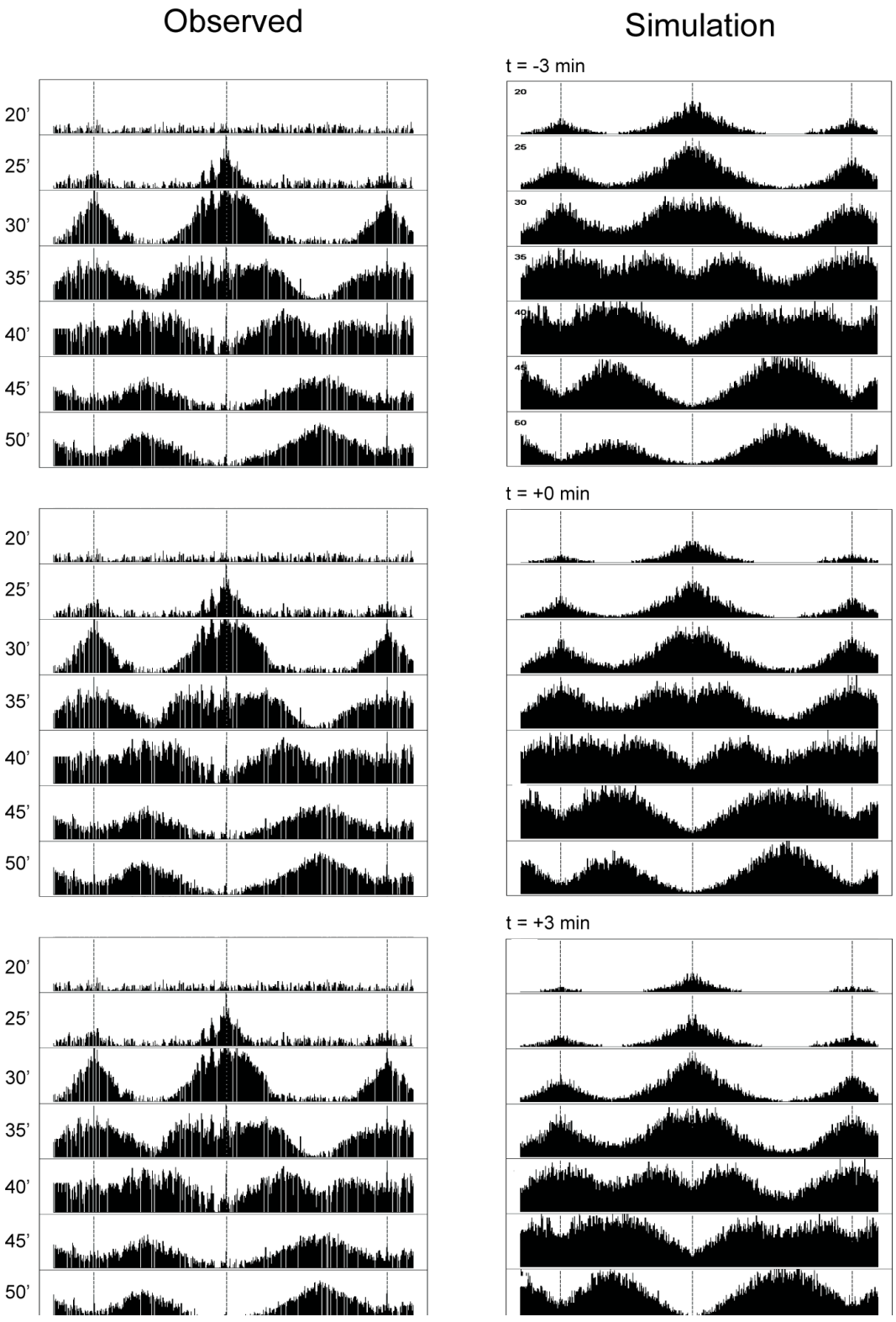

Observed
Simulation
t = -3 min
t = +0 min
t = +3 min
20'
25'
30'
35'
40'
45'
50'
20
25
30
35
40
45
50

Figure 5-5 (Next page). The effect of varying the fork velocities. Simulations with a series of different rates (1.2, 1.6 and 2.0 kb/min shown here) are compared to the observed data for region 230 – 370 kb of chromosome 8. When the rates were set to 1.2 kb/min, the peak edges did not merge fast enough, and peaks at origins lingered there for too long. When the rates were set to 2.0 kb/min, the peaks spread too fast, and the gaps filled in too early. The measured velocity of 1.6 kb/min gives the best fit between the experimental data and the simulations.

Figure 5-5

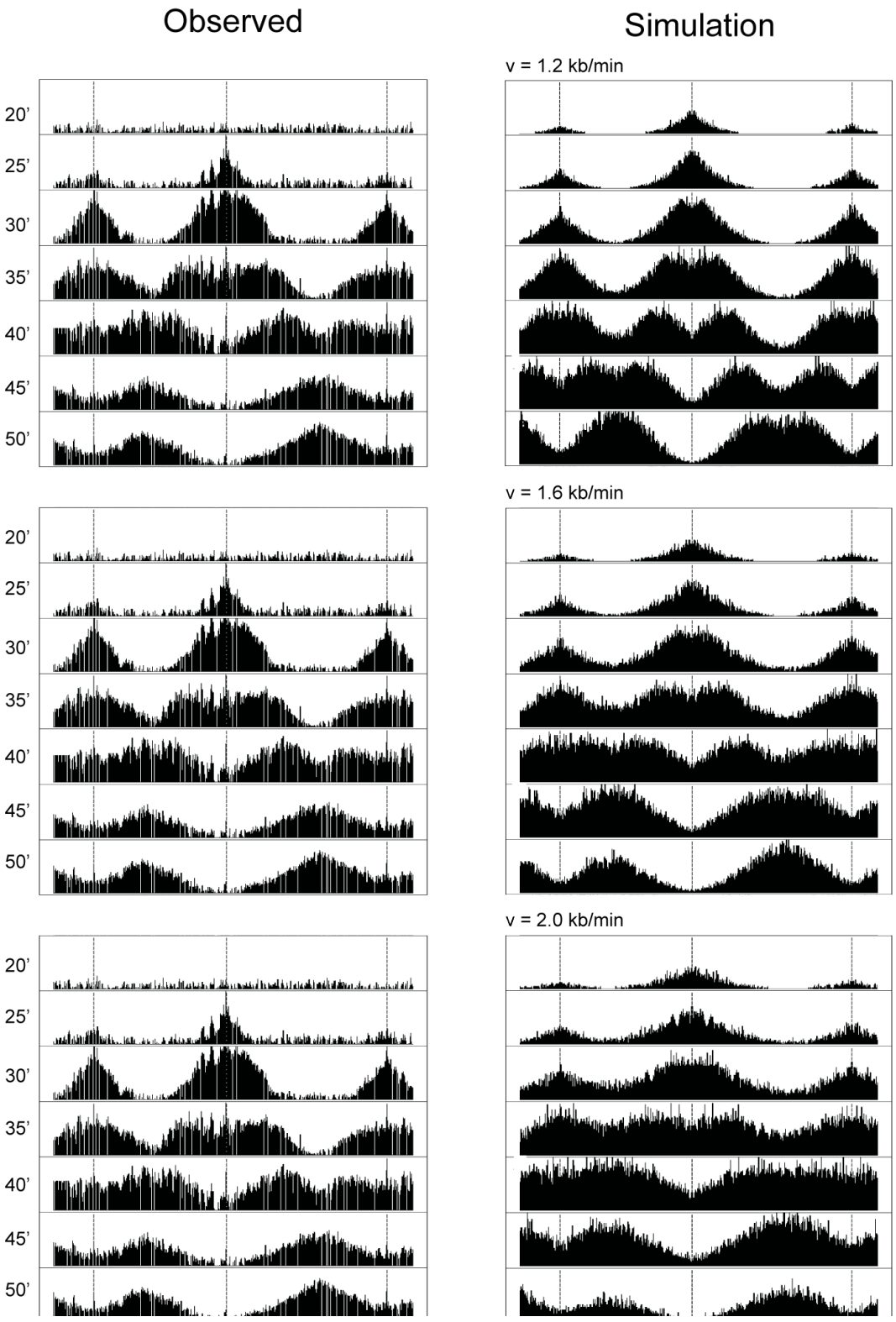
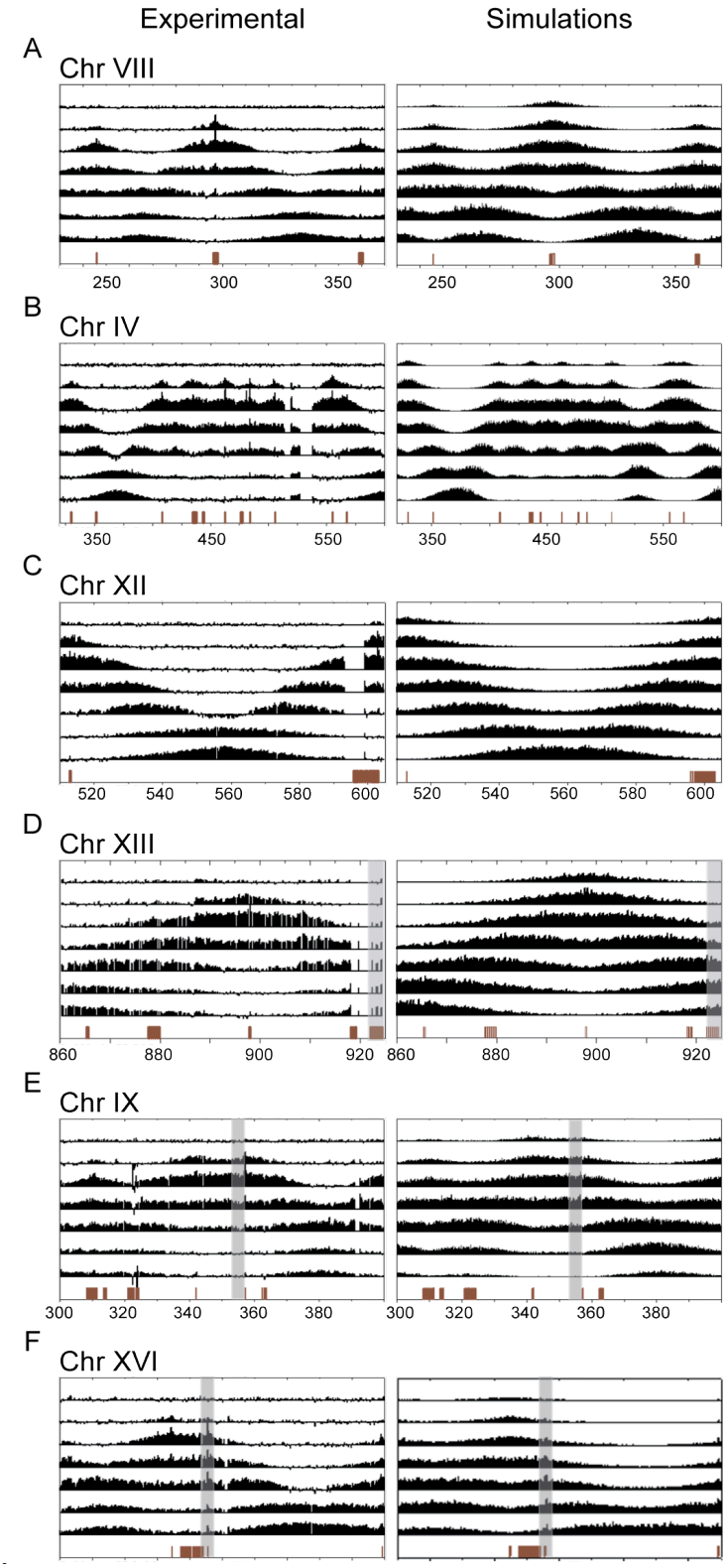

Observed
Simulation
v = 1.2 kb/min
v = 1.6 kb/min
v = 2.0 kb/min
20'
25'
30'
35'
40'
45'
50'

Figure 5-6 (Next page). S-phase time course data (experimental and simulations) for several genomic regions of interest. a, Chr 8 coordinates 240-370 kb. A category 1 origin flanked by two category 2 origins. b, Seven category 1 origins spaced unevenly between 320-600 kb of Chr 4. c, Chr 12 coordinates 510-600 kb. The gap in between the two noted origins is 97 kb, among the longest inter-origin distance observed in the yeast genome. d, The right-hand telomere region of Chr 13. The telomere is shaded. e, The centromeric region of Chr 9 (shaded grey). f, A category 3 origin (shaded grey) located at ~695 kb (ARS1625) in Chr 16.

Figure 5-6


Experimental
Simulations
A
Chr VIII
250
300
350
250
300
350
B
Chr IV
350
450
550
350
450
550
C
Chr XII
520
540
560
580
600
520
540
560
580
600
D
Chr XIII
860
880
900
920
860
880
900
920
E
Chr IX
300
320
340
360
380
300
320
340
360
380
F
Chr XVI

For simplicity, all origins illustrated in Figures 5-6 A-E are assumed to fire with unit efficiency. A refined model should include the possibility that different origins fire with different efficiencies. For example, replication is known to initiate at a reduced efficiency in many late firing origins (Poloumienko et al., 2001; Yabuki et al., 2002; Yamashita et al., 1997). In addition, there is a pattern of alternating groups of early and late firing origins across the genome (McCarroll and Fangman, 1988; Poloumienko et al., 2001; Raghuraman et al., 2001; Yabuki et al., 2002; Yamashita et al., 1997). We observe similar trends in the present work (see Appendix F). As an example, we consider the cluster of late firing origins located on the right arm of chromosome 15 between 500 - 800 kb (Figure 4-5). We find that our model fails to accurately simulate this region if the firing efficiency for these late firing origins is too high (100%) or too low (30%). However, we can accurately simulate fork dynamics in this region if we assume that the efficiencies for these sites are decreased to 50% relative to the surrounding early firing origins (Figure 5-7). Thus, not only does the model provide an overall picture of the dynamic nature of GINS, but also a means to extract quantitative details concerning factors such as firing efficiency.

Indeed, the method is sufficiently sensitive to detect small heterogeneities in replication dynamics and in the process sheds new light on the phenomenon described as 'replication pausing' – short duration stalling of forks at numerous specific sites in the genome (Azvolinsky et al., 2006; Azvolinsky et al., 2009; Deshpande and Newlon, 1996; Ivessa et al., 2003). Although such events would seem too subtle to detect with 5 min resolution (Azvolinsky et al., 2009), we observed sharp features at 267 out of 275 these

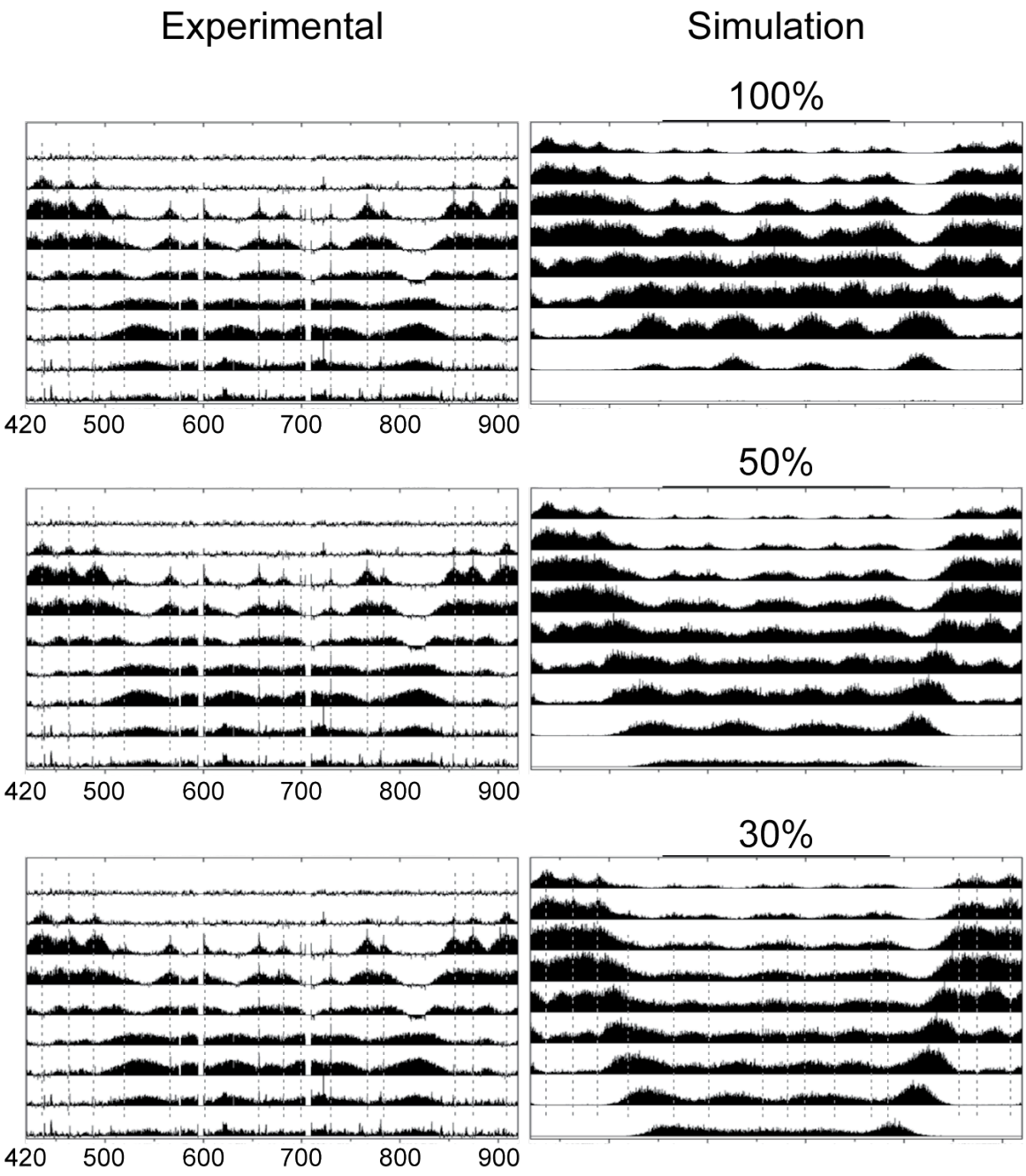


Figure 5-7. Varying efficiencies applied a group of origins located in the right arm of Chr 15. The six category 1 origins on either end are given 100% efficiencies, whereas the efficiency applied to the cluster of seven category 2 origins in the middle of this region is indicated above each panel.

tRNA genes (Figure 4-7, Figure 5-8, and Appendix G), indicating that our methodology is sensitive enough to detect small perturbations, and that GINS movement can indeed be hindered to some degree. However, rather than a brief pause of each fork as it passes through these sites, these features persist late into the cell cycle. Simulation of several different scenarios (Figures 5-9 and 5-10) suggests that these features may represent infrequent long-term arrest events, occurring with a probability of <0.5% for any given

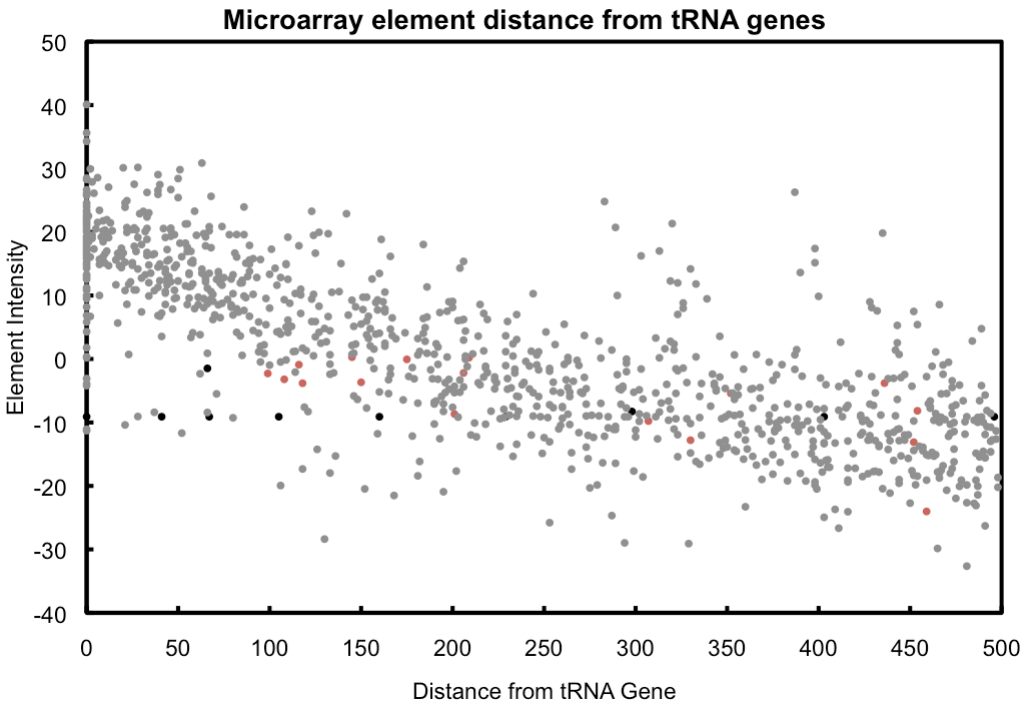


Figure 5-8. 8 of the 275 tRNA genes encoded in the yeast genome were not observable as persistent features in our analysis. 3 of the 8 genes were too close to either long terminal repeats or the LEU2 gene (not present in the strain used), leaving an obvious empty space in the array readout (black circles). To understand why the other 5 sites did not show up, we examined the relationship between the intensities of peaks at elements within 500 bp of all tRNA genes and the distance of these elements from the gene. The elements flanking the 5 remaining tRNA genes are highlighted as red circles in the above scatter plot. From this analysis we conclude that the absence of features at these tRNA genes is likely due to the fact that the nearest array elements are not close enough to these genes to give a large signal.

fork passing through a tRNA gene (Figure 5-9). This phenomenon was also observed at 81 of the 83 snoRNA and snRNA genes, and 95 of the 100 other most highly transcribed genes in the genome (Figure 5-10 and Appendix G), and appears to be independent from the direction of transcription.

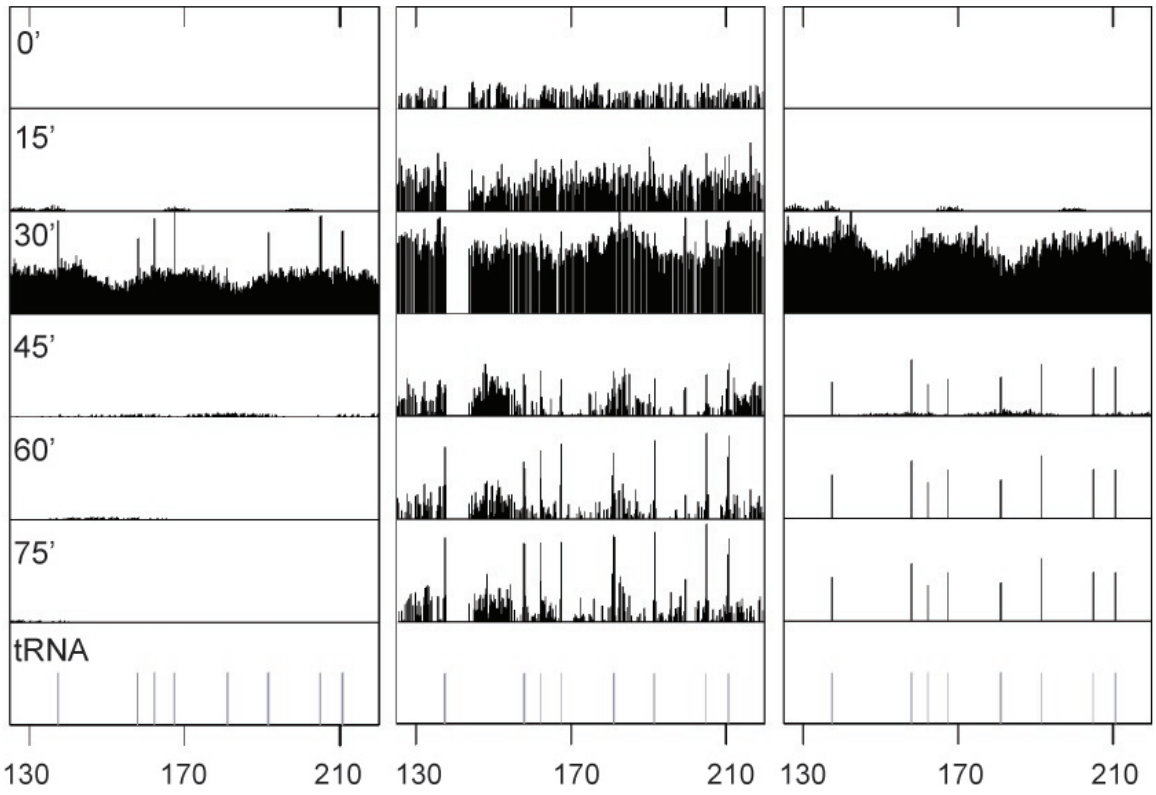


Figure 5-9. The GINS complex undergoes long-term arrest at tRNA genes. (Center) ChIP-chip data for the association of the GINS complex with a region of chromosome 6 (coordinates 125–220 kb) containing eight tRNA genes. Data are shown in 15 min increments through the cell cycle. During late S phase sharp spikes are observed, which coincide with the position of the tRNA genes. We observe similar long-lived spikes that coincide with 267 of the 275 tRNA genes in the yeast genome. Two alternative models for explaining these spikes are presented: (left) in each cell, the replication forks are assumed to pause for 6 s at each tRNA gene; (right) the replication forks are assumed to have a 0.2% chance of long-term (t ¼N) arrest at each tRNA gene. Additional model parameter conditions are shown in the Supplementary information.

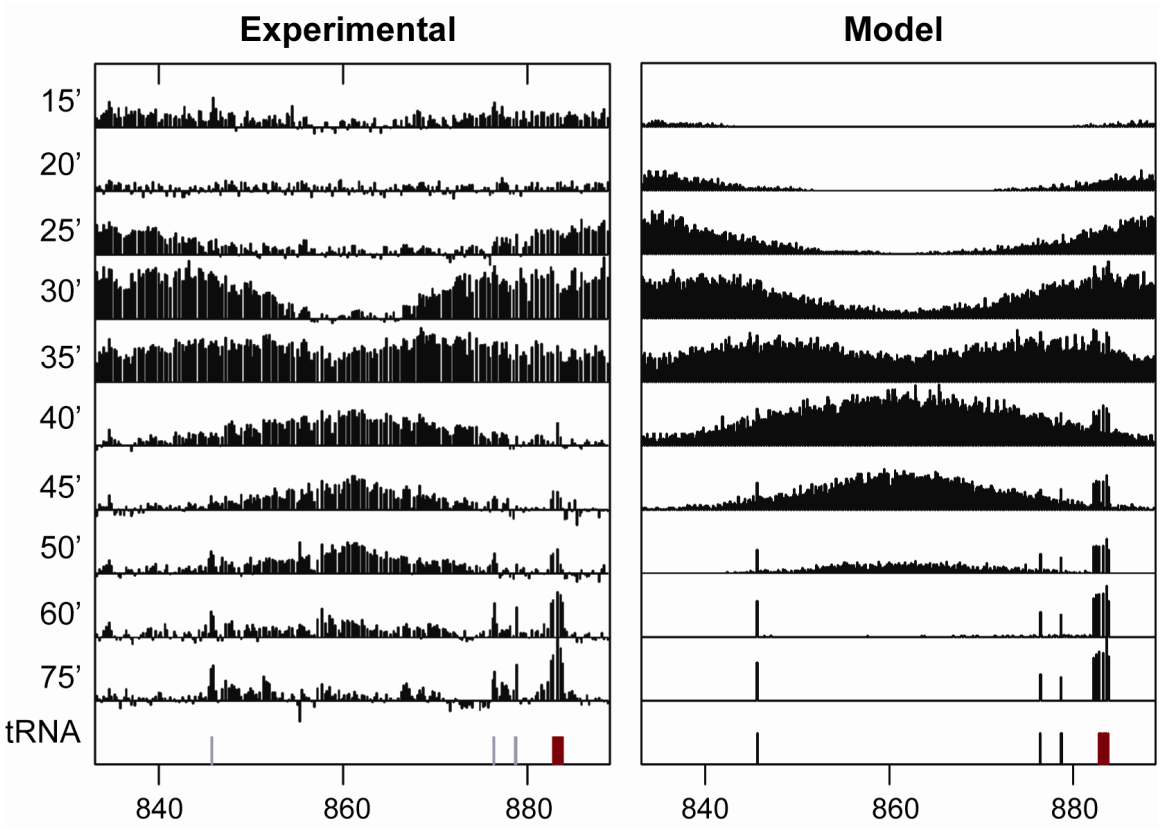


Figure 5-10. The GINS complex undergoes long-term arrest at highly transcribed genes. Time-course ChIP-chip data are shown for a region of Chr 7 (coordinates 833 – 889 Kb, left panel). The locations of the highly-transcribed TDH3 gene (red bar) as well as the tRNA genes (grey lines) in this region that overlap with the chip elements present in our arrays (see Supplementary Fig. S9). We have simulated this region (right panel) to show that the model of low probability (0.2%), long-term arrest can explain how these persistent features accumulate at highly transcribed genes.

5.3 – RE-ANALYSIS OF NUCLEOTIDE INCORPORATION STUDIES

5.3.1 - On the Integration of Replication Dynamics

Replication fork movement can be monitored in at least two different ways. In the first, the movement of the fork is directly monitored as in the present work (Figure 5-11, left panel, 'Non-Integrative'). In the second, the incorporation of nucleotides is monitored using a density transfer approach during the formation of nascent DNA (Raghuraman et al., 2001) (Figure 5-11, right panel, 'Integrative'). The first approach monitors the transit of forks, while the second approach integrates this process over time – that is, each time point accumulates signal.

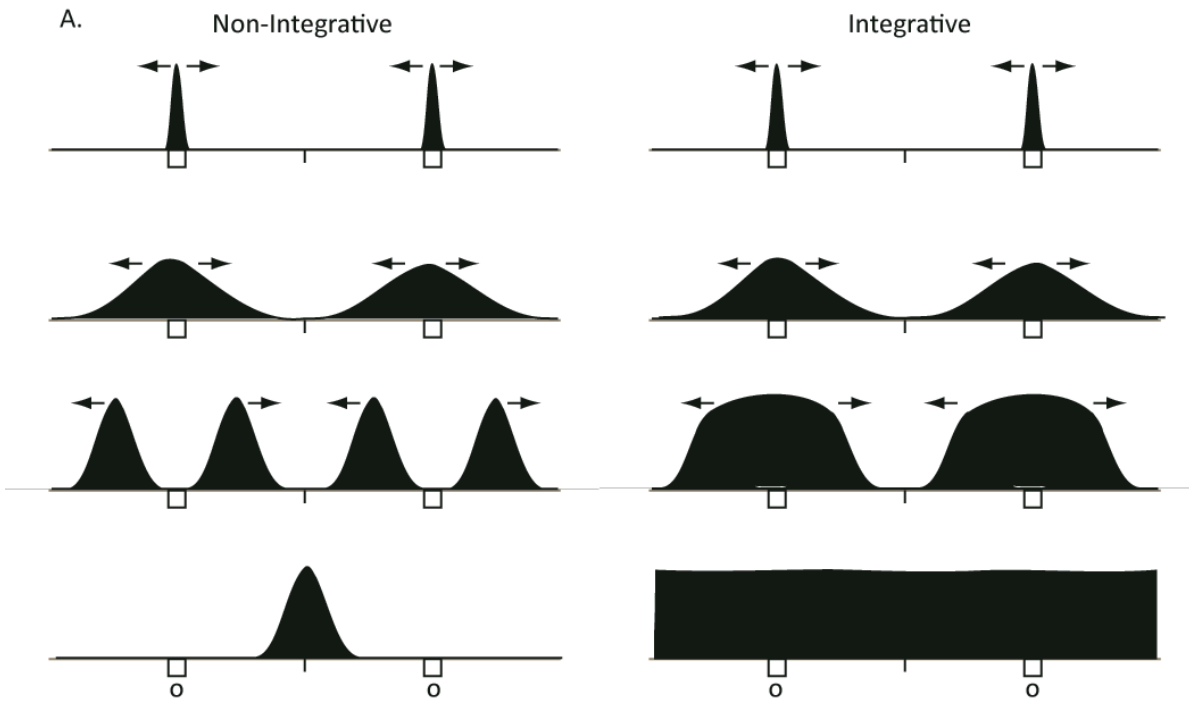


Figure 5-11. Integrative versus non-integrative monitoring of replication fork progression. Data is accumulated in the integrative approach, whereas forks can be observed leaving sites in a non-integrative approach.

The non-integrative approach more accurately differentiates between discretely firing origins within a population than the integrative approach. By analyzing our data for chromosome 16 in these two different ways (Figure 5-12), the accuracy of the non-integrative approach is demonstrated. The left panel shows the results for the non-integrative approach and the right panel the results for the integrative approach (as per (Raghuraman et al., 2001)), in both cases using the same raw time-resolved ChIP-Chip data. In this example, the origin ARS609 (indicated by the red arrow) fires later than ARS607 and with a much lower efficiency – estimated to be ~30% (Yamashita et al., 1997). The region immediately to the left of ARS609 can be replicated in two directions (either by forks originating from ARS609 or from ARS607), and at two different times: 30% of the time it is replicated early from right to left, while 70% of the time it is

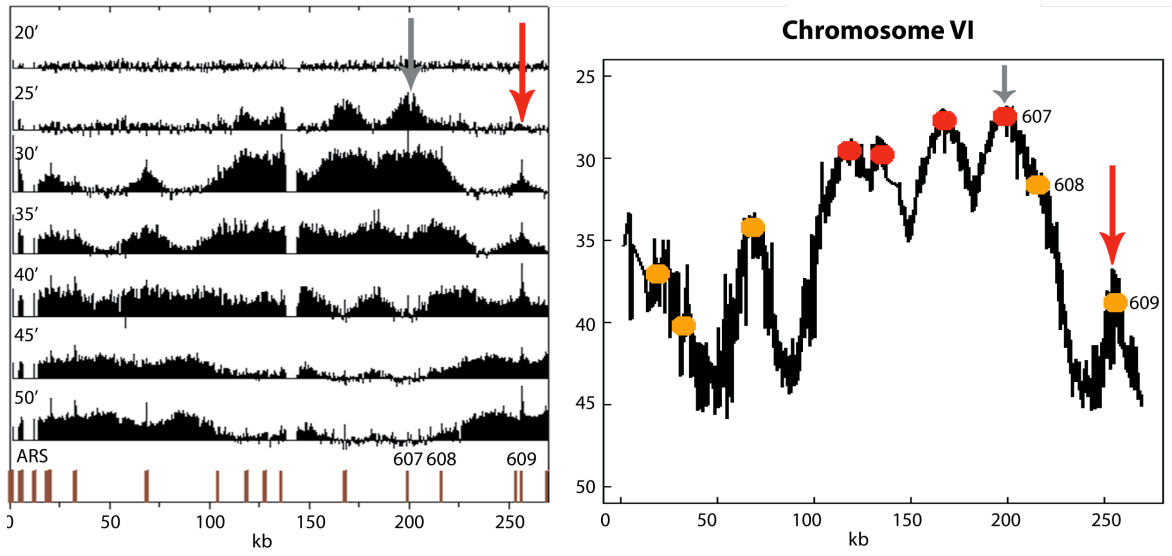


Figure 5-12. Integrative versus non-integrative monitoring of replication fork progression. In cases where an origin fires with less than unit efficiency, the observed replication time can be skewed by integration. In the case of ARS609 (red arrow in both panels), the replication time is observed to be very late in the right panel, most likely because much of the time that region is being replicated by forks emanating from an adjacent origin (gray arrow in both panels).

replicated late (since the fork emanating from ARS607 has much further to travel (~55kB)) and from left to right. Thus, integrating the replication time for this region yields an average that does not represent the actual properties of forks emanating from ARS609. This example shows that the average time of replication (T_{rep}) for a region surrounding an inefficient origin is dependent on its distance from adjacent origins (firing time and efficiency of the adjacent origins also have an effect). The change in T_{rep} affects the shape of the replication-timing peak, and because rates are calculated by measuring the slopes of curves generated from the T_{rep}s, this bias may account for some of the variation in rates across the genome when nascent DNA is monitored (Raghuraman et al., 2001).

5.3.2 - The Quality of the Density Transfer Data

The experimental details of nucleotide incorporation studies are explained in Figure 1 of that paper (Raghuraman et al., 2001) (recreated in Figures 5-13 and 5-14). Briefly, a given stretch of DNA will switch from the HH band in a CsCl gradient to the HL band once it is replicated. This process should follow a sigmoid curve, which is then used to calculate the T_{rep} for that particular stretch (the stretches are identified by microarray). The T_{rep} for every stretch is then used to generate a plot that shows the replication timing profile for every chromosome. Thus, an accurate replication profile is heavily dependent on smooth, sigmoidal replication curves.

We generated timing curves for every stretch of chromosomes 6 and 10 form the raw data provided by the authors in the downloadable supplementary data Section of the paper. Two examples are provided in Figure 5-13. There is a clear difference between the ideal sigmoid T_{rep} curve, which should begin at 50% HL, and convert to 100% HL by

the end of the curve (Figure 5-13, left panel). In fact, the HH curve nearly always increases before dropping to zero, and the HL curve never reaches above 50%. The cause of these systematic behaviors is not known, but the effects could easily include skewing the peak shapes in the T_{rep} plots to create artificially high or low slopes at various regions of the genome.

Indeed, the mean velocity reported in this study is 3.0 kb/min – nearly twice as fast as the rates observed by monitoring the GINS progression, and the rate that is typically assumed when discussing fork dynamics. However, the modes of the two rate distribution are much closer to each other (2.0 kb/min observed by Raghuraman et. al.

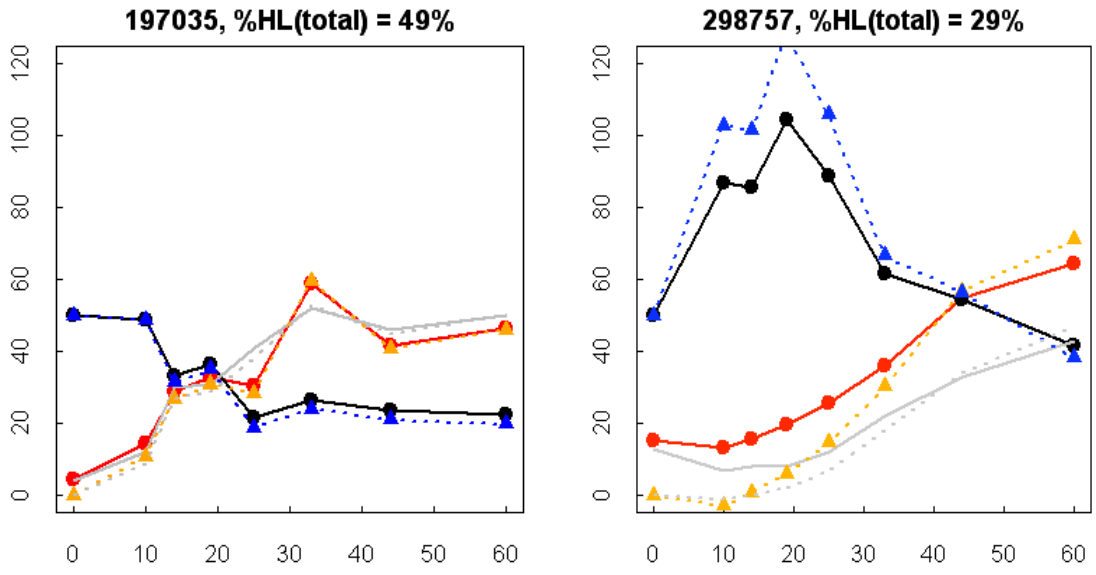


Figure 5-13. Examples of timing curves for individual array elements in Raghuraman et .al. The y-axis is percent, and the x-axis is minutes. The numbers above each panel indicate the region of the genome being examined. The %HL is the pre cent at ½ replication for each curve, an indicator of replication time. The dark lines represent the percentage of HH DNA for the region indicated above each box. The light (orange) line indicates the amount of HL DNA for this region. HH DNA should drop from 50% to 0%, while HL rises from 0% to 100%.

versus 1.6 kb/min measured by GINS progression). Together with our observation that fork progression is in fact a uniform process, this suggests that perhaps the mean is skewed by a number of outliers.

Overall, this study earlier study appears to have been very successful at identifying the locations of the origins across the genome. Unfortunately, it seems likely that the technique was not appropriate for quantitating velocities due to systematic errors in the experiment. These include skewed replication times for regions with origins that fire with less than unit efficiency, and the surprising behavior of the density transfer curves. These observations suggest that the direct monitoring of replication fork proteins is a more accurate approach for deducing the details of replication fork progression.

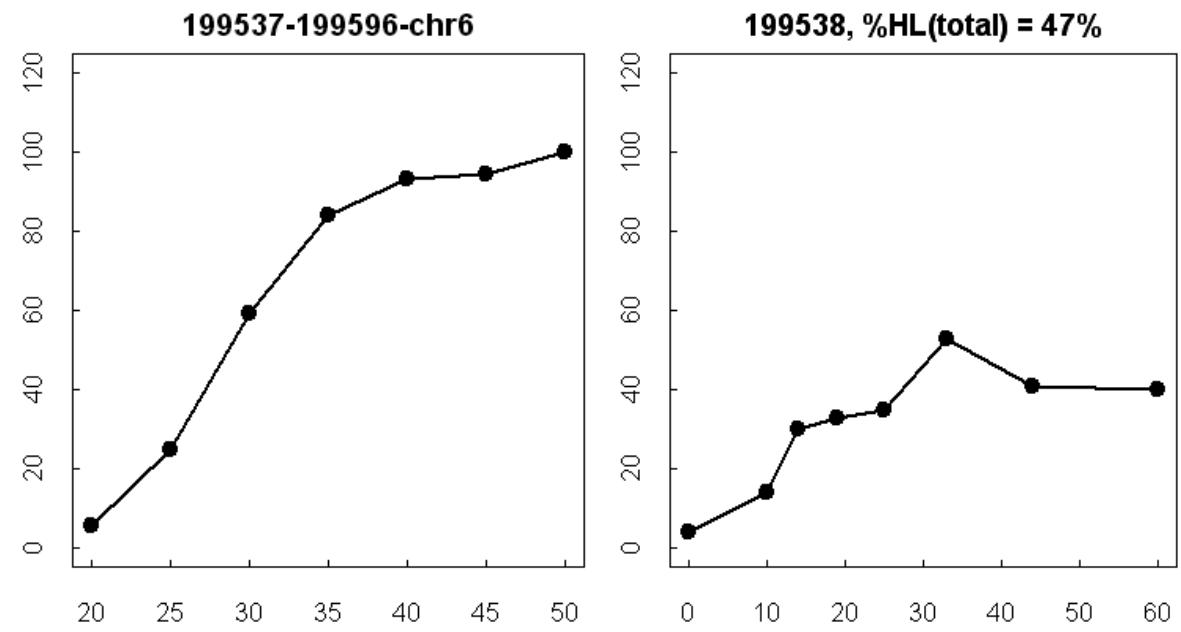


Figure 5-14. Examples of timing curves for individual array elements in Raghuraman et al. The y-axis is percent, and the x-axis is minutes. The numbers above each panel indicate the region of the genome being examined. The left panel shows a timing curve using our time-resolved ChIP-chip data, and the right panel shows a density transfer experiment (Raghuraman et al., 2001) for the same approximate region.

CHAPTER VI – DISCUSSION

6.1 – DNA POLYMERASE & HELICASE PROGRESSION DURING REPLICATION

All theories of abiogenesis – the study of how the earliest life rose from inanimate matter - include replication as an integral feature, firmly presenting replication as an indispensible aspect for all of life and indicating that the process of genome replication is fundamental to biology. Another indicator of the importance of this process is the myriad of genetic health issues associated with problems in DNA replication. These include, but are not limited to, most types of cancer as well as aging. Many of these complications arise due to the fact that the genome is particularly exposed during this process. Investigating the details of the unwinding and synthesis of DNA is therefore an essential task that will not only help us appreciate how life came to exist, and perhaps contribute to solving important human health problems.

Sites in the genome at which DNA synthesis takes place are called replication forks. Previous studies have suggested that progression of the forks is greatly affected by the different states of chromatin compaction, and from this, that the individual velocities of fork progression in yeast are highly variable. The reported velocities range from 0.5 – 10 kb/min and asymmetric progression around origins has been described (Raghuraman et al., 2001). Importantly, these experiments were carried out nearly a decade ago, and microarray technology has progressed significantly since this time. As described here in Section 5.3, we have carefully re-assessed the quality and behavior of this data. This re-analysis showed that the timing curves used to create the replication profiles in the

Raghuruman study do not behave as predicted for DNA in the process of being replicated. In fact, the shapes of the replication timing curves appear rife with systematic errors, making it difficult to extract quantitative information.

Several studies claim evidence that "hard to replicate" regions of the genome can give rise to "paused" replication forks. This pausing phenomenon has been described as short duration stalling of forks at numerous specific sites in the genome (Azvolinsky et al., 2006; Azvolinsky et al., 2009; Deshpande and Newlon, 1996; Ivessa et al., 2003). For example, pausing intervals at tRNA genes have previously been estimated at ~10 sec (Deshpande and Newlon, 1996; Ivessa et al., 2003) – i.e., ~4 times longer than the interval required for unimpeded replication fork transit of a tRNA gene. The time-resolved monitoring of replication fork proteins presented here reveals no evidence of pausing. In fact, our results suggest that instead of short pauses, the observation for pausing could equally be interpreted as infrequent replication fork arrest.

The experiments presented in this dissertation take a different approach in comparison to previous attempts at analyzing the dynamics of DNA replication. By monitoring the progression of three replication fork-associated proteins along the genome during the cell cycle with time-resolved ChIP-chip techniques, we have come to significantly contrasting conclusions regarding the details of the replication program. The proteins monitored included a member of the CMG helicase complex (the Psf2 subunit of the GINS complex), and two replicative polymerases (the Pol2 subunit of Pol ε and the Pol3 subunit of Pol δ). The aims of this study were to understand in greater detail how chromatin affects the movement of the replication machinery.

6.1.1 - GINS Progression

The data presented here, along with previously determined interactions of the GINS complex, strongly suggest that GINS moves with the helicase at the replication fork. Our time-resolved ChIP-chip data reveal that GINS binds to active replication origins and spreads bi-directionally and symmetrically as S-phase progresses. A similar approach had previously been used to monitor Pol ε localization on yeast chromosome 3, albeit at much lower time resolution (Hiraga et al., 2005). As monitored by GINS movement, the majority of origins appear to fire in the first ~15 min of S-phase. A small fraction (~10%) of the origins to which GINS binds show no evidence of spreading (category 3 origins), although it remains possible that these peaks represent passively fired origins (Shirahige et al., 1998). Once an active origin fires, the GINS complex moves at an almost constant rate of 1.6 ± 0.3 kb/min. Its movement through the inter-origin regions is consistent with that of a protein complex associated with a smoothly moving replication fork. This progression rate is considerably lower and more tightly distributed than those inferred from previous genome-wide measurements assayed via nascent DNA production (Raghuraman et al., 2001; Yabuki et al., 2002).

6.1.2 - Progression of the Polymerases

It is of interest to consider potential sources of these discrepancies between our observed fork progression rates and those reported in the literature. Some of the differences can be explained by the integration inherent to monitoring nascent DNA accumulation (discussed in Section 5.3). Such problems are avoided by our direct measurement of a specific fork-associated protein Psf2. However, it may also be conjectured that the discrepancies discussed above can be explained by the possibility

that the replication forks are not tightly coupled to the replicative polymerases with respect to their dynamics (Pacek et al., 2006; Walter and Henry, 2004).

Using similar methods, the enzymes that synthesize DNA were monitored to address this possibility. Two general results were expected – 1) if the polymerases were significantly affected by chromatin structure, the progression profiles would look drastically different in certain regions of the genome; or 2) the CMG complex and the polymerases are spatially, if not physically coupled to the genome, and the progression profiles would look identical.

The data reveal the progression profiles of both polymerases look extremely similar to that for the GINS complex. The mean and standard deviation of the velocity were identical, and no evidence of asymmetric progression was observed. Moreover, the data revealed uniquely unambiguous results that both Pol ε and Pol δ progress from the origins to the termination sites. This resolves a long-standing debate on the roles that these two enzymes play in the replication process. Indeed, it is now very likely that Pol ε is responsible for synthesizing the leading strand, while Pol δ is responsible for the lagging strand.

The only observable differences in the progression profiles of the three proteins came in the form of slight shifts in the timing – that is, the origins fired slightly earlier in the Pol ε experiments compared to the GINS or Pol δ studies. This was a confusing result because it seemed inconceivable that a polymerase could progress ahead of the helicase. Experimental variations in the timing of the cell cycle seemed a much more plausible explanation. To test this, the progression of Pol ε was directly compared to that of the GINS complex by using a dual-tagged strain (Psf2-PrA, Pol2-Myc). After carefully

testing that the tags would not cross-react with the antibodies used to isolate the proteins, the resulting data confirmed that Pol ε and the GINS complex indeed move in unison. Thus, the polymerase proteins appear to traverse the genome in unison with the CMG complex and the differences in velocities between studies do not arise from any distinct behavior being attributable these complexes. Moreover, our re-analysis of the nucleotide incorporation experiments (Section 5.3) suggests that the direct monitoring of replication fork proteins is a more reliable method for measuring velocities. These results raise important questions regarding the interpretation that specific regions of the genome are particularly difficult to replicate.

6.1.3 – Replication Fork Stalling

The phenomenon described as pausing at tRNA genes is a predicted consequence of regions that are difficult to replicate. We observe sharp features at 267 out of 275 these tRNA genes (Figure 4-7), indicating that our methodology is sensitive enough to detect small perturbations, and that GINS movement can indeed be hindered to some degree. However, rather than a brief pause of each fork as it passes through these sites, these features persist late into the cell cycle. Simulation of several different scenarios (Figure 5-9 and 5-10) suggest that these features may represent infrequent long-term arrest events, occurring with a probability of <0.5% for any given fork passing through a tRNA gene (Figure 5-9). This phenomenon was also observed at 81 of the 83 snoRNA and snRNA genes, and 95 of the 100 other most highly transcribed genes in the genome (Figure 4-7 and Appendix G).

Previously, pausing of forks had been observed at such sites, where in certain cases the pause appeared to be coordinated with head-on collisions between the replication and

transcriptional machineries (Deshpande and Newlon, 1996). Here we do not observe any significant directional dependence of fork arrest at the tRNA genes, 40% of which transcribe in the same direction as the movement of the replication fork. Thus, the presently described features at highly transcribed genes do not appear to exclusively correlate with head-on collisions. The extremely low frequency of these events in wild type cells suggests they are due to low probability stochastic occurrences during the replication process.

6.2 - UNIFORM FORK PROGRESSION & CHROMATIN

We have made extensive use of modeling to test a number of different hypotheses and assumptions. In particular, iterative modeling allowed us to infer that GINS progression is uniform and smooth throughout the genome. We have also demonstrated the potential of simulations for estimating firing efficiencies. In the future, extending such firing efficiency simulations to the whole genome should allow us to make correlations with chromosomal features such as nucleosome occupancy. Such correlations may help in determining factors that govern the probability of replication initiation throughout the genome. Finally, simulations of the persistent features that correlate with the locations of paused replication forks suggest that this phenomenon is not a particularly prominent characteristic of fork progression. Instead, a large percentage of forks in a majority of cells appear to progress with great uniformity, only stalling on very rare occasions (with <0.5% frequency).

Overall, the data and simulations lead us to the conclusion that replication forks in yeast cells traverse the genome at a relatively constant rate of 1.6 kb/min. The state of

chromatin compaction does not appear to affect fork progression to any observable degree. The different rate distributions can be attributed to systematic errors in the prior published data (Raghuraman et al., 2001), and fundamental issues in the prior experimental design (Raghuraman et al., 2001).

An overwhelming majority of the yeast genome is wrapped up in the tight packaging of nucleosomes. Yet by any measurement the replication fork does not appear to have any problem coping with these structures. Whether the fork removes and replaces histones as it traverses the genome, or uses another mechanism to get through nucleosomes is a matter of intense research. In either case, our observation that the state of chromatin compaction does not have a major effect on the rate of fork progression may not be particularly surprising for two major reasons – 1) Replication forks already have the means to confront the presence of nucleosomes without significantly affecting their progression. 2) The fact that replication occurs in factories (Kitamura et al., 2006) suggests that groups of replication forks, particularly sister forks that originate from the same origin, should perhaps have similar behavior.

6.3 – THE POLYMERASES & THE HELICASE COMPLEX

From these time-resolved ChIP-chip experiments we conclude that three of the major components of replication forks – the CMG complex, Pol ε, and Pol δ – traverse the genome in unison. We have concluded that the simplest interpretation of the data is that GINS progresses with the helicase, while Pol ε, and Pol δ synthesize the leading and lagging strands, respectively. If there is a lag in space between the helicase and the polymerases, it must be relatively small since the replication profiles look nearly

identical for these proteins. Further, the presence of an extended amount of unwound, single-stranded DNA would leave these regions of the genome particularly exposed to damage. The fact that the GINS complex does not co-isolate with either polymerase raises further questions about the relationship between these three proteins.

Recent experiments showing that Pol ε proteins can be detected in GINS immunoisolations when treated with formaldehyde may help to clarify the relationship between these proteins (Muramatsu et al., 2010). It is possible that the helicase and polymerases are in direct contact at the replication fork, but that the binding is necessarily weak in order to allow for polymerase replacement under certain conditions. It is also possible that these proteins interact indirectly through DNA, but their progression is coordinated in some other way to prevent exposed DNA.

Currently, the dual-tagged approach described in Section 4.3 is being extended to co-localize all three components together. These experiments will confirm whether progression of these three important replication fork proteins – GINS, Pol δ, and Pol ε - are indeed directly coordinated. We have shown that this technique is highly reproducible, and it appears to be a reliable approach for measuring the rates of progression across the genome. These experiments suggest that the progressing replication forks take precedence in the genome, and that chromatin state does not have as significant of an affect on it as previously believed.

APPENDICES

APPENDIX A – *Saccharomyces cerevisiae* Strains Used in this Work

Table A-1

Strain	Genotype	Source
W303-1a	Mat**a** ura3-1 leu2-3,112 his3-11,15 trp1-1ade2-1 can1-100 rad5- 535	R. Rothstein
ATY112	Mat**a** ura3-1 leu2-3,112 his3-11,15 trp1-1ade2-1 can1-100 rad5- 535 SLD5-PRA-HIS5	A. Tackett
ATY113	Mat**a** ura3-1 leu2-3,112 his3-11,15 trp1-1ade2-1 can1-100 rad5- 535 PSF1-PRA-HIS5	A. Tackett
ATY114	Mat**a** ura3-1 leu2-3,112 his3-11,15 trp1-1ade2-1 can1-100 rad5- 535 PSF2-PRA-HIS5	A. Tackett
ATY44	Mat**a** ura3-1 leu2-3,112 his3-11,15 trp1-1ade2-1 can1-100 rad5- 535 POL2-PRA-HIS5	A. Tackett
ATY46	Mat**a** ura3-1 leu2-3,112 his3-11,15 trp1-1ade2-1 can1-100 rad5- 535 POL3-PRA-HIS5	A. Tackett
ATY27	Mat**a** ura3-1 leu2-3,112 his3-11,15 trp1-1ade2-1 can1-100 rad5- 535 bar1Δ	A. Tackett
MSY1	Mat**a** ura3-1 leu2-3,112 his3-11,15 trp1-1ade2-1 can1-100 rad5- 535 PSF2-PRA-HIS5 bar1::KANMX4	Constructed from ATY114
MSY9	Mat**a** ura3-1 leu2-3,112 his3-11,15 trp1-1ade2-1 can1-100 rad5- 535 POL3-PRA-HIS5 bar1Δ	Constructed from ATY27
MSY10	Mat**a** ura3-1 leu2-3,112 his3-11,15 trp1-1ade2-1 can1-100 rad5- 535 POL2-PRA-HIS5 bar1Δ	Constructed from ATY27
MSY11	Mat**a** ura3-1 leu2-3,112 his3-11,15 trp1-1ade2-1 can1-100 rad5- 535 PSF2-PRA-HIS5 bar1Δ	Constructed from ATY27
MSY12	PSF2-MYC-KANMX4 bar1Δ	Constructed from ATY27
MSY13	Mat**a** ura3-1 leu2-3,112 his3-11,15 trp1-1ade2-1 can1-100 rad5- 535 PSF2-PRA-HIS5 POL2-MYC-KANMX4 bar1Δ	Constructed from MSY11

APPENDIX B – Mass Spec Identifications and Scores

The log expectation values are from searches against the yeast genome database by using ProFound. Some proteins were not identified in the initial peptide search, but were only found by searching the MS/MS fragmentation data with Xproteo. A Xproteo d' value of 4 means the hit probability is 0.99 with a false alarm rate of 0.5.

Table B-1

Asynchronous Psf1-PrA IP Identifications

Protein ID	Mass (kDa)	log Protein Expectation Value	# of identified peptides	% Coverage	Xproteo d' - value
Mcm6	113.0	-3.62	9	12	16.4
Mcm3	107.5	-5.68	18	26	17.7
Ctf4	105.2	-17.85	30	51	158.5
Mcm4	105.0	-7.26	8	12	6.3
Mcm5	86.4	-11.47	20	33	18.4
Cdc45	74.2	-5.42	17	33	11.8
Sld5	34.2	-8.09	10	43	46.9
Psf2	25.1	-9.39	7	56	43.4
Psf1*	24.2	-6.72	13	58	24.1
Psf3	21.9	-2.66	8	60	33.5

*tagged protein

Table B-2

Asynchronous Psf2-PrA IP Identifications

Protein ID	Mass (kDa)	log Protein Expectation Value	# of identified peptides	% Coverage	Xproteo d' - value
Mcm6	113.0	-3.62	10	13	18.4
Mcm3	107.5	-3.77	11	18	13.8
Ctf4	105.2	-17.85	30	51	76.9
Mcm4	105.0	-2.37	5	5	14.4
Mcm5	86.4	-6.64	13	23	27.2
Cdc45	74.2	-6.77	16	31	12.3
Sld5	34.2	3.01	7	34	38.2
Psf2*	25.1	-10.92	9	67	39.5
Psf1	24.2	-3.60	12	57	14.4
Psf3	21.9	-0.82	7	63	18.5

*tagged protein

Table B-3

Asynchronous Sld5-PrA IP Identifications

Protein ID	Mass (kDa)	log Protein Expectation Value	# of identified peptides	% Coverage	Xproteo d' - value
Mcm6	113.0	-6.64	14	19	28.2
Mcm3	107.5	-3.74	17	25	35.5
Ctf4	105.2	-17.85	30	51	21.6
Mcm4	105.0	-3.80	7	9	20.1
Mcm5	86.4	-4.92	17	28	26.5
Cdc45	74.2	-8.60	15	29	36.1
Sld5*	34.2	-4.02	9	39	17.9
Psf2	25.1	-3.89	4	36	24.3
Psf1	24.2	-4.72	13	57	9.0
Psf3	21.9	-2.37	9	77	29.2

*tagged protein

Table B-4

Time-Resolved Psf2-PrA IP Identifications (30 min)

Protein ID	Mass (kDa)	log Protein Expectation Value	# of identified peptides	% Coverage	Xproteo d' - value
Mcm6	113.0	-13.17	25	32	15.7
Mcm3	107.5	-5.54	20	31	34.0
Ctf4	105.2	-11.51	32	49	18.0
Mcm4	105.0	-3.46	16	27	14.8
Mcm5	86.4	-9.24	35	51	23.7
Cdc45	74.2	-10.11	21	38	39.7
Sld5	34.2	-6.01	11	48	33.3
Psf2*	25.1	-9.93	8	59	17.8
Psf1	24.2	-7.21	12	53	34.9
Psf3	21.9	-4.60	9	64	22.8
Hht1	15.4	-0.60	2.00	31	11.7
Htb1	14.2	Not Observed in MS			15.1
Hta1	13.9	Not Observed in MS			16.8
Hhf1	11.4	Not Observed in MS			5.3

Table B-5

Time-Resolved Psf2-PrA IP Identifications (15 min)

Protein ID	Mass (kDa)	log Protein Expectation Value	# of identified peptides	% Coverage	Xproteo d' - value
Mcm6	113	-12.00	26	33	32.3
Mcm3	107.5	-4.80	23	33	19.0
Ctf4	105.2	-10.11	32	49	37.6
Mcm4	105	-9.10	11	18	17.8
Mcm5	86.4	-10.85	38	53	29.0
Cdc45	74.2	-8.60	23	42	29.8
Sld5	34.16	-4.51	9	35	33.4
Psf2*	25.1	-11.54	10	66	28.1
Psf1	24.2	-4.55	15	60	12.4
Psf3	21.9	-2.19	10	80	21.0
Hht1	15.4	Not Observed in MS			17.7
Htb1	14.2	-3.74	4	34	6.7
Hta1	13.9	Not Observed in MS			8.1
Hhf1	11.4	-2.64	8	53	6.5

Table B-6

Asynchronous Pol3-PrA IP Identifications (High Stringency)

Protein ID	Mass (kDa)	log Protein Expectation Value	# of identified peptides	% Coverage	Xproteo d' - value
Pol3*	125.73	-27.70	47	52	35.2
Pol31	55.91	-13.54	17	46	24.7
Pol32	40.58	-4.80	7	18	28.7

*tagged protein

Table B-7

Asynchronous Pol3-PrA IP Identifications (Low Stringency)

Protein ID	Mass (kDa)	log Protein Expectation Value	# of identified peptides	% Coverage	Xproteo d' - value
Pol3*	125.73	-13.1	20	31	40.8
Pdc1	61.02	-6.4	8	25	23.1
Cdc19	54.93	-4.3	7	27	7.6
Pol31	55.91	-15.2	18	49	31.3
Tef2	41.55	-6.8	8	30	31.1
Rpl3	43.72	-3.1	6	15	27.4
Pol32	40.58	-5.4	9	27	25

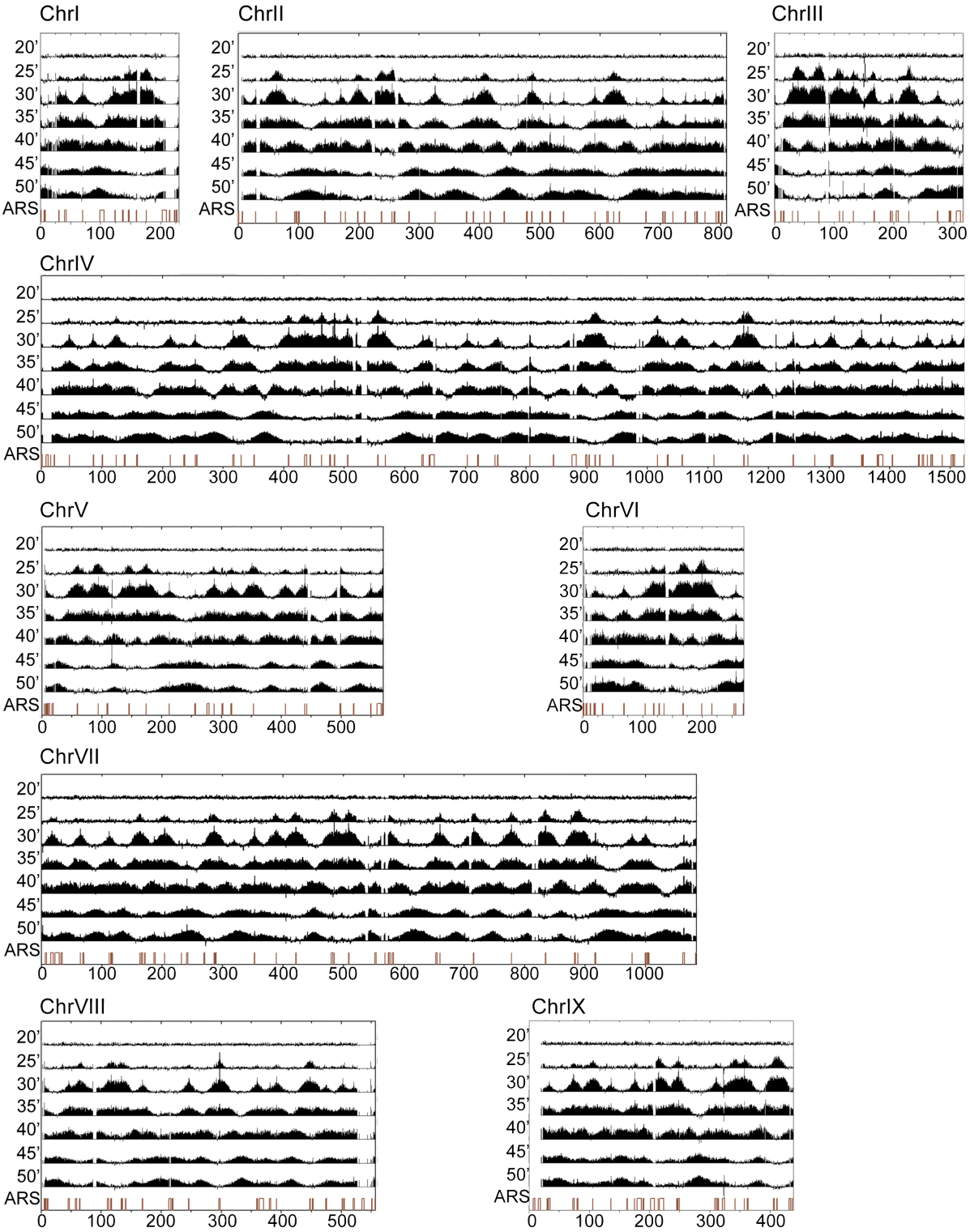
gray rows indicate likely contaminant proteins

*tagged protein

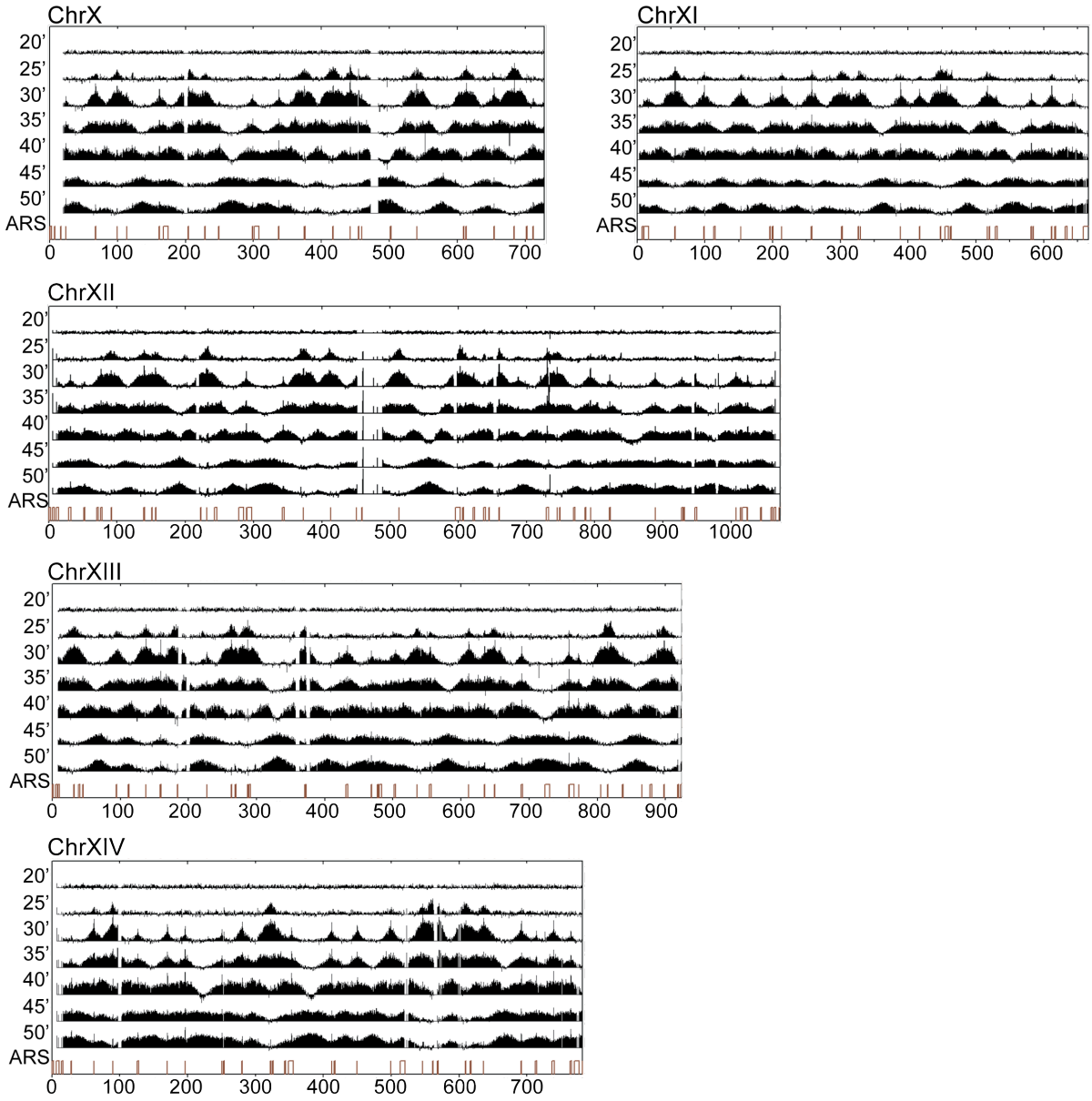
APPENDIX C — Time-Course Psf2 ChIP-chip Results Over S-phase for all 16 Chromosomes in the Yeast Genome.

The horizontal axes provide chromosomal locations in kb. Red vertical bars at the bottom of each chromosome indicate the locations of "confirmed" and "likely" ARS sites as defined in the reference (Nieduszynski et al.). The raw data is available at the following web address (prowl.rockefeller.edu/data/yeast_repl).

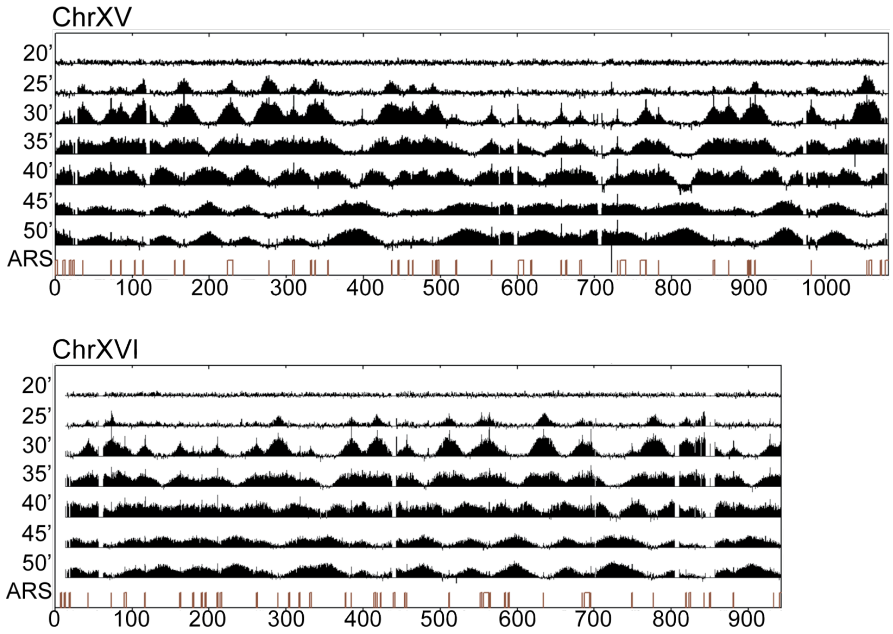
Appendix C Continued



Appendix C Continued

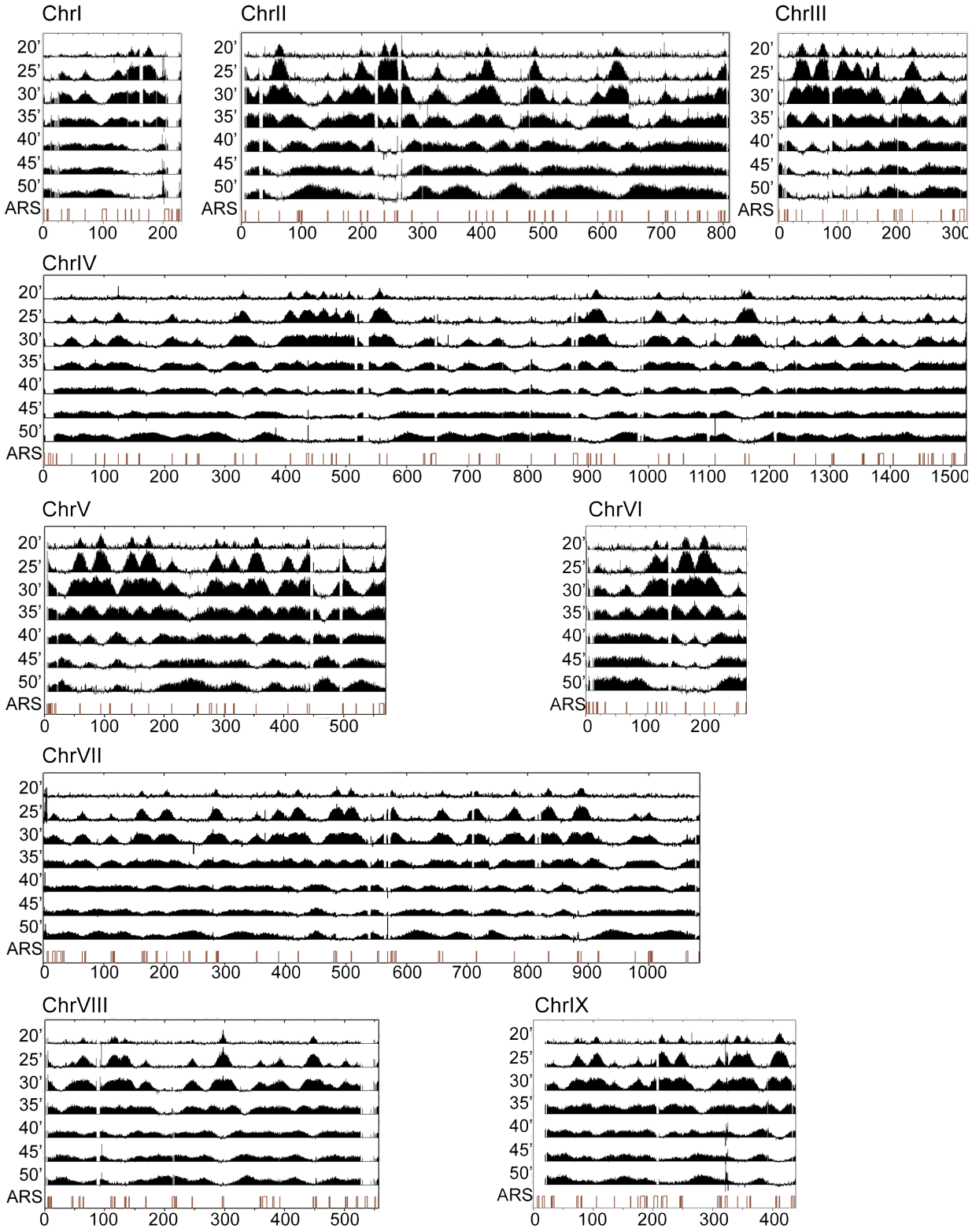


Appendix C Continued

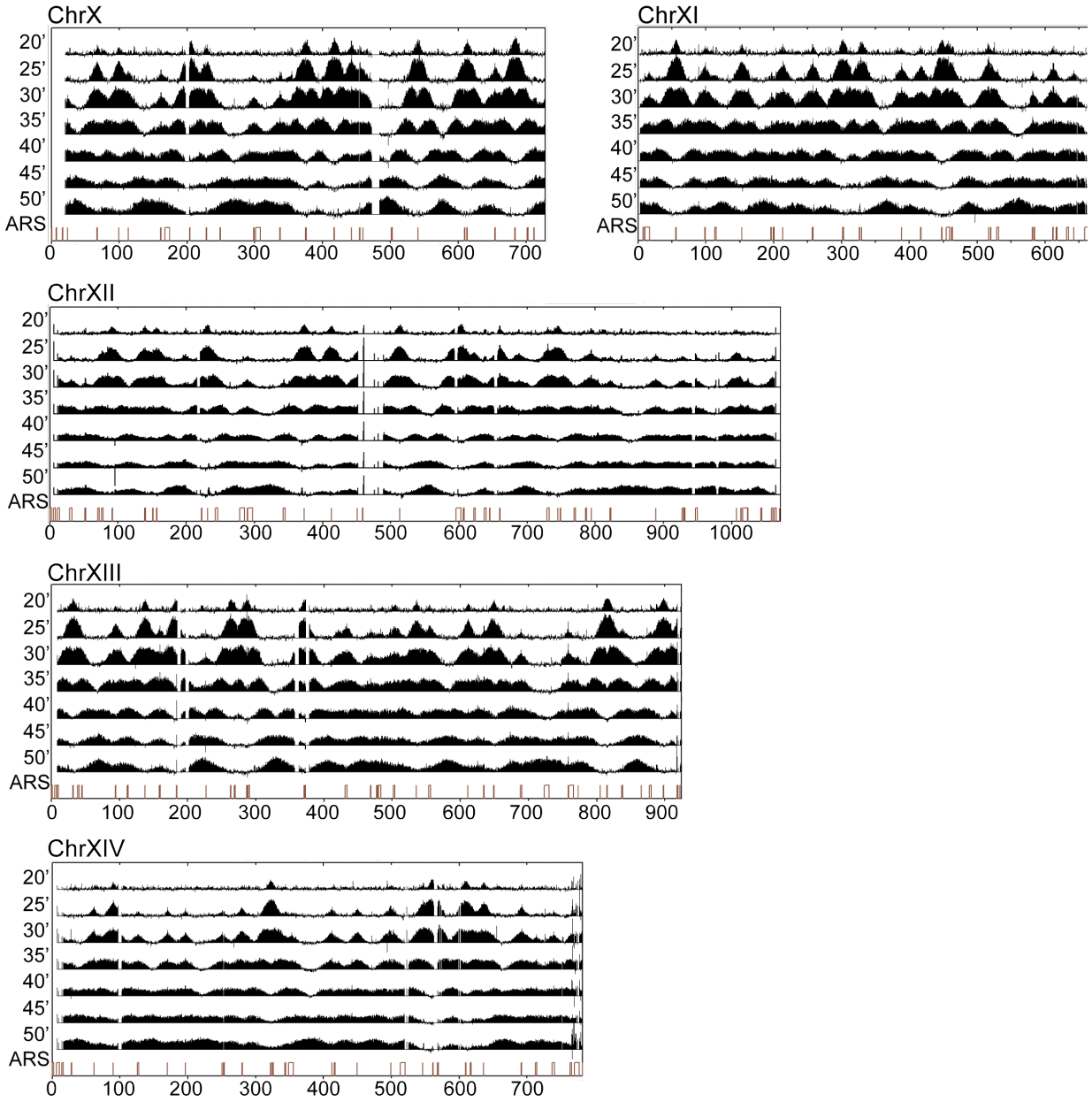


APPENDIX D — Time-Course Pol2 ChIP-chip Results Over S-phase for all 16 Chromosomes in the Yeast Genome.

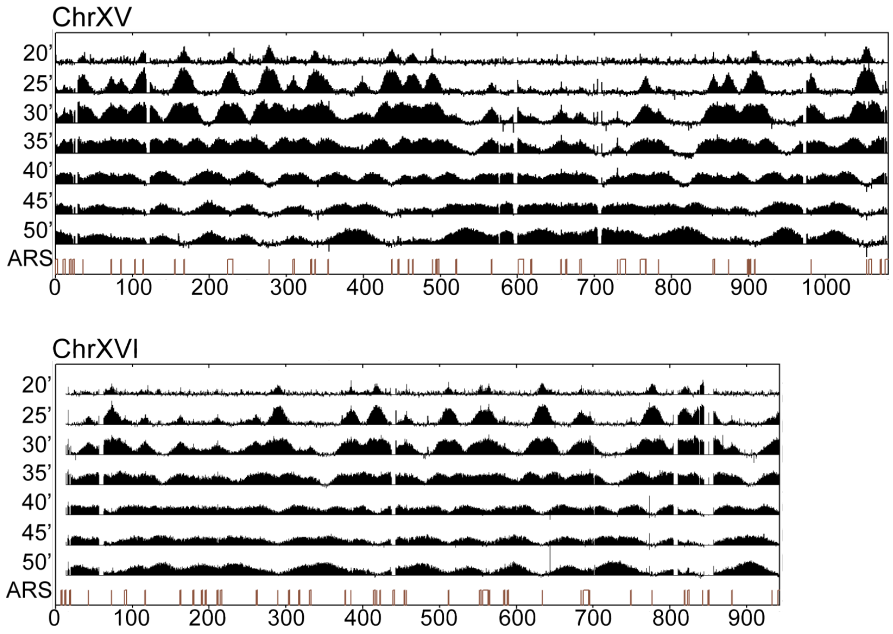
Appendix D Continued



Appendix D Continued

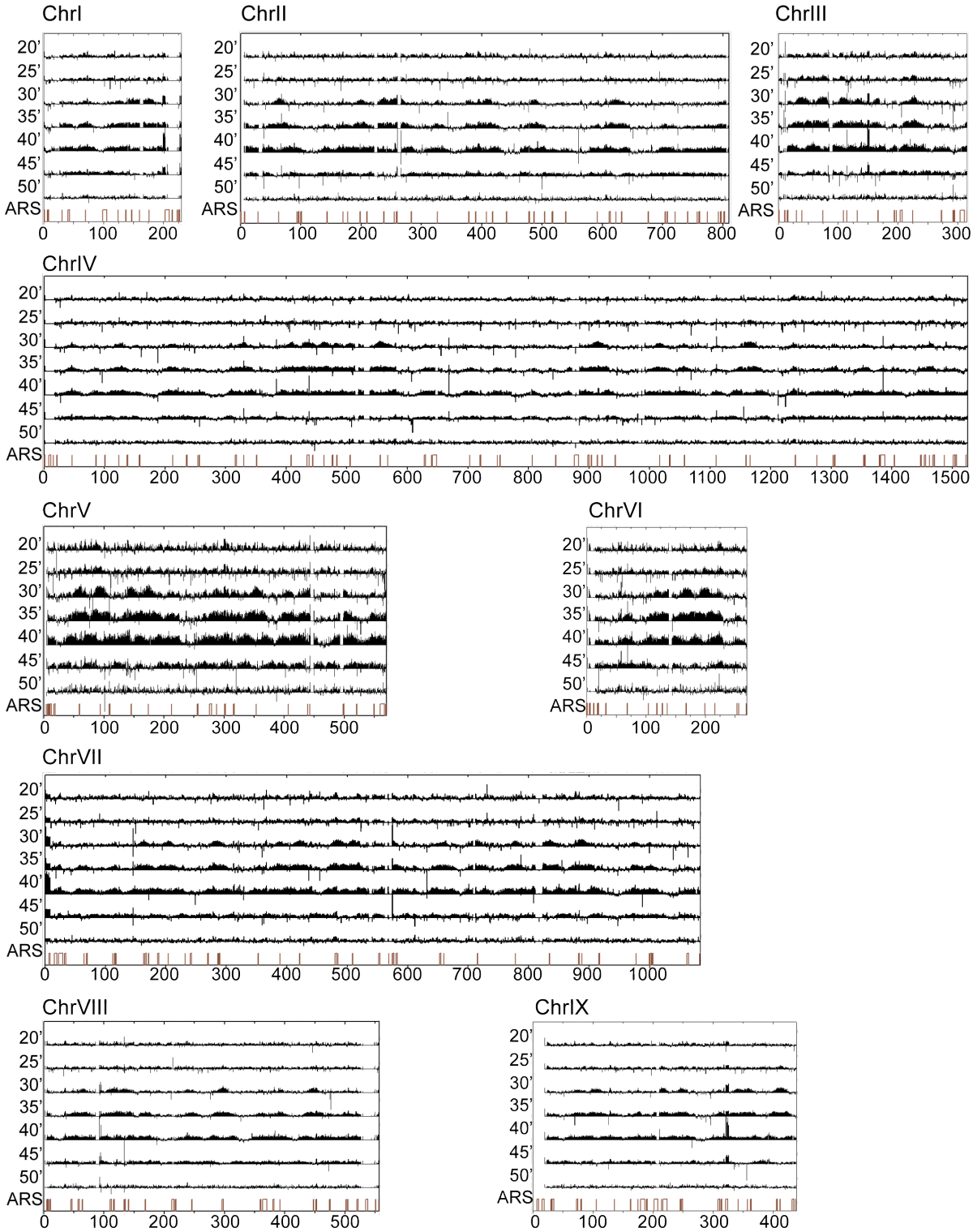


Appendix D Continued

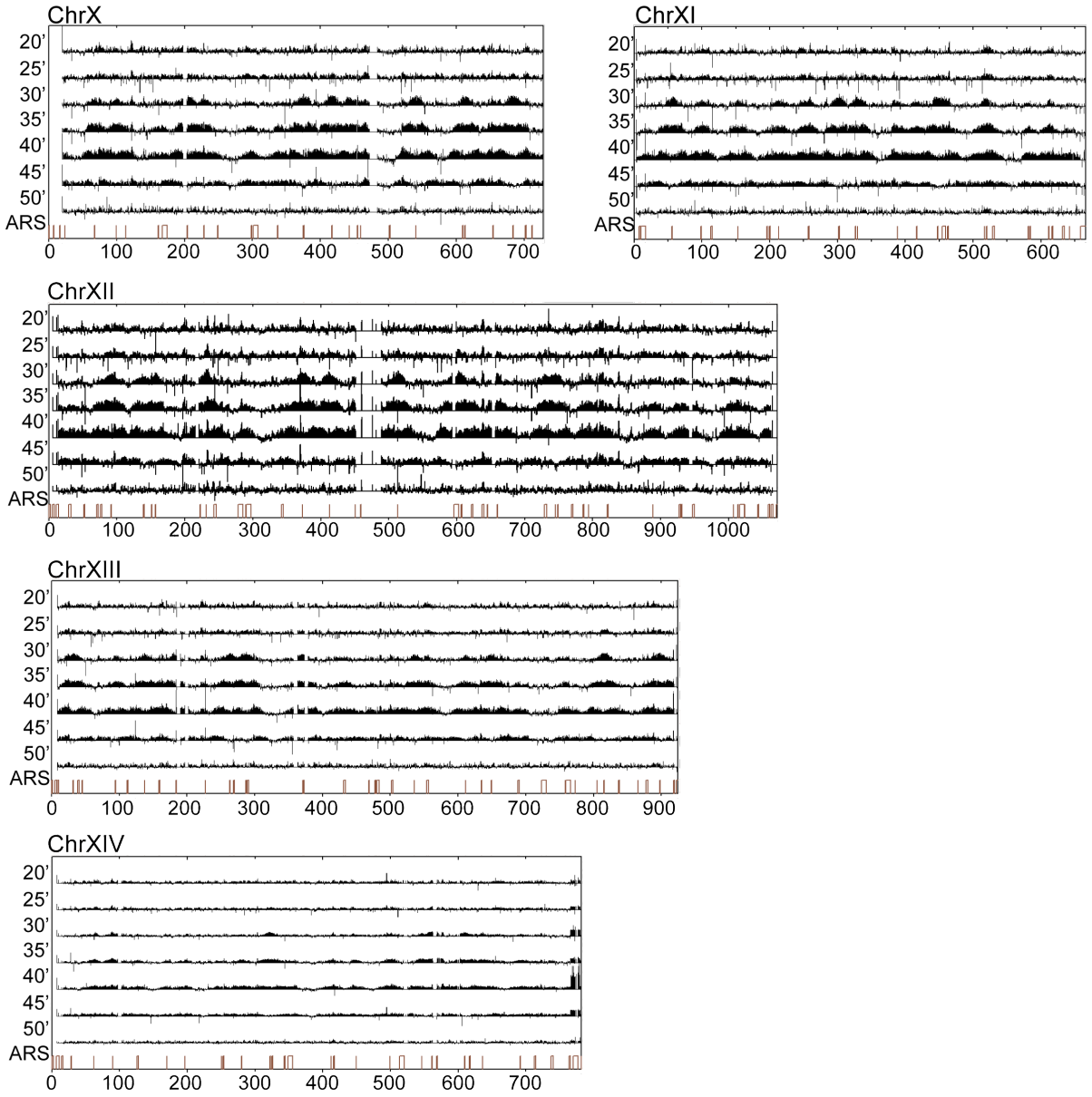


APPENDIX E — Time-Course Pol3 ChIP-chip Results Over S-phase for all 16 Chromosomes in the Yeast Genome.

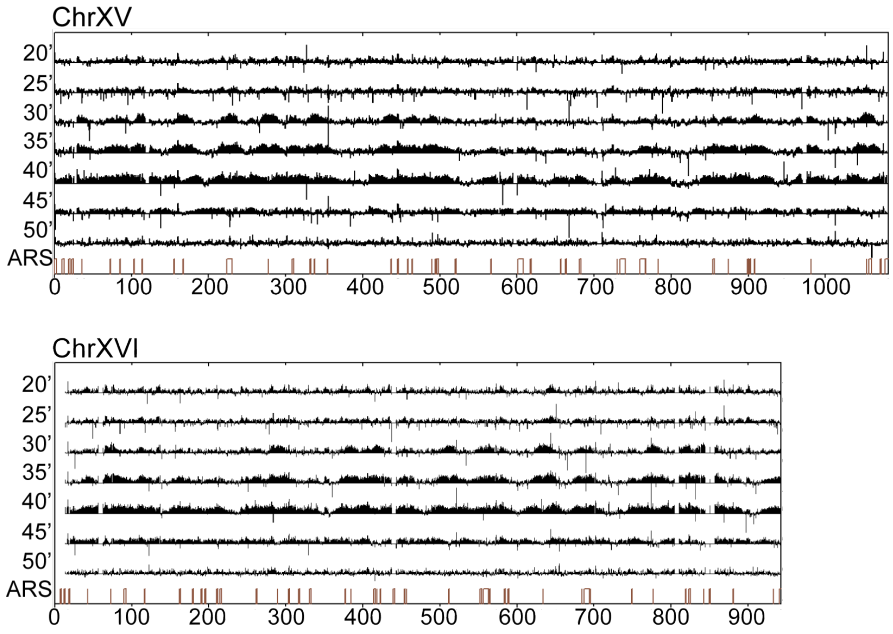
Appendix E Continued



Appendix E Continued



Appendix E Continued



APPENDIX F — Comprehensive Origin List

Origins identified in from the Psf2 time-course study listed by chromosome number, systematic name (where available), and start and end coordinates. The category number (1, 2 or 3) that we have assigned to each is listed in column 5. Prior identification of origins and their correspondence with the present study are listed in columns 6 to 10. A designation and reference for each of these 5 prior studies is given at the top of the column, and the reference number is listed in the appropriate box where a correspondence with the present work is observed. Asterisks represent potential passively firing origins.

Table F-1

Chr	Name	Start	End	Category	H/L[1]	MCM[2]	Copy#[3]	ssDNA[4]	MCM[5]
1	ARS102	650	1791	2		2			5
1	ARS103	7998	8548	3					5
1	ARS104	30946	31184	2		2	3	4	5
1	ARS105	40716	43300	2	1	2	3	4	5
1	ARS106	70258	70491	2	1	2		4	5
1	ARS107	124350	124599	1	1	2	3	4	5
1	ARS108	146703	147690	1	1	2	3	4	5
1	ARS109	159906	160127	1		2		4	5
1	ARS110	176154	176402	1	1		3	4	5
2	ARS201.5	28933	29152	2					5
2	ARS202	63186	63421	1	1	2	3	4	5
2	ARS203	93410	93811	2	1	2		4	5
2	ARS206*	142868	144016	2		2	3	4	5
2	ARS207*	170049	170298	2	1	2	3	4	5
2	ARS207.5	198193	198434	1	1		3	4	5
2	ARS208	237644	237879	1		2	3	4	5
2	ARS209	254890	255136	1		2		4	5
2	ARS211	326099	326355	2	1	2	3	4	5
2	ARS213	389245	390368	2		2		4	5
2	ARS214	407831	408064	1	1	2	3	4	5
2	ARS216	486661	486909	1	1	2	3	4	5
2	ARS217*	516805	517805	2				4	5
2	ARS218*	539137	539699	2		2		4	5
2	ARS219*	591424	591713	2		2	3	4	5
2	ARS220	622625	622894	1	1	2	3	4	5
2		660540	660587	3					
2	ARS221.5	675947	676667	2	1		3		5
2	ARS222*	704250	704521	2		2	3	4	5

2	ARS224	741512	741802	2		2	3	4	5
2	ARS225	757390	757621	2	1	2		4	5
2	ARS227	773918	774348	2	1	2		4	5
2	ARS228	792116	792340	2	1	2		4	5
2	ARS229	801930	802617	1	1	2			5
3	ARS305	39158	39706	1	1	2	3	4	5
3	ARS306	74457	74677	1	1	2		4	5
3	ARS307	108755	109291	1	1	2	3	4	5
3	ARS309	131978	132322	1	1	2	3	4	5
3	ARS310	166494	167340	1	1	2	3	4	
3	ARS314	197369	197601	3	1	2	3	4	
3	ARS315	224807	225053	1	1	2		4	5
3	ARS316*	272844	273088	2	1	2	3	4	5
4	ARS404	46181	46237	2	1	2	3	4	5
4	ARS405	85945	86177	2	1	2	3	4	
4	ARS406	123617	123902	1	1	2	3	4	5
4	*	157458	158807	3	1		3	4	
4	ARS409	212420	212669	2	1	2	3	4	5
4	ARS410*	253789	254038	2	1		3	4	5
4		315773	318097	3	1		3	4	
4	ARS413	329564	329813	1	1	2	3	4	5
4	ARS414	408070	408312	1	1	2	3	4	5
4		434223	437917	1	1	2	3	4	5
4	ARS416	462430	462700	1	1	2	3	4	5
4	ARS417	483846	484091	1	1	2	3	4	5
4	ARS417.5	505336	505578	1	1	2		4	5
4	ARS418	555224	555461	1	1	2	3	4	5
4		628203	630547	3	1	2	3	4	5
4	ARS421	639859	640108	2	1		3	4	
4		667955	668014	3					
4	ARS422	702879	703125	2	1	2	3	4	5
4		719873	721227	3	1	2		4	
4	ARS423	753159	753391	2	1	2	3	4	5
4	ARS425*	806044	806270	2	1	2	3	4	5
4	ARS428	913780	914029	1	1	2	3	4	5
4	ARS430	1016624	1016922	1	1	2	3	4	5
4	ARS430.5	1033069	1034182	3				4	
4	ARS431	1057828	1058076	1	1	2	3	4	5
4	ARS431.5	1109955	1110196	2	1	2		4	5
4	ARS432	1159250	1159499	1		2	3	4	
4	ARS432.5	1165998	1166221	1	1	2		4	
4	ARS433*	1240869	1241098	2	1	2	3	4	
4	ARS434	1276212	1276440	2	1	2	3	4	
4	ARS435	1302579	1302819	2	1	2	3	4	5
4	ARS437	1353494	1353667	1	1	2	3	4	5
4		1381291	1388291	1	1	2		4	5

4	ARS440	1404277	1404512	2	1		3	4	5
4	ARS441	1447298	1448928	2	1	2	3	4	
4	ARS442	1461849	1462161	2	1	2	3	4	5
4	ARS446	1486905	1487149	2	1	2	3	4	
4		1501448	1504702	2	1		3	4	
4		1522298	1526152	2	1	2	3	4	
5	ARS503	6464	7230	2		2		4	5
5	ARS507	59282	59516	1	1	2	3	4	5
5	ARS508	93977	94218	1	1	2	3	4	
5	ARS510	145539	145782	1	1	2	3	4	5
5	ARS511	173636	173874	1	1	2	3	4	5
5	ARS512*	212381	212630	2	1	2	3	4	5
5		255623	256957	2	1			4	5
5	ARS514	287504	287750	1		2	3		5
5		316043	317307	1	1			4	5
5	ARS516	353504	353751	1	1	2	3	4	5
5	ARS517	406747	406949	1	1	2	3	4	5
5	ARS518	438929	439178	1	1	2	3	4	5
5	ARS520	498417	499343	1	1	2	3	4	5
5	ARS522	549560	549809	1	1	2		4	5
5	ARS523	569020	570085	1	1	2		4	5
6	ARS600.4	19669	21862	2	1	2		4	5
6	ARS602	32472	32995	2				4	5
6	ARS603	68690	68869	2	1	2	3	4	5
6	ARS603.5	118631	118952	1	1	2	3	4	5
6	ARS605	135979	136080	1	1	2	3	4	5
6	ARS606	167606	168041	1	1	2	3	4	5
6	ARS607	199382	199493	1	1	2	3	4	5
6	ARS608	216344	216692	2				4	5
6	ARS609*	256263	256418	2		2	3	4	5
6	AR610	269404	270022	2		2			5
7		15327	19624	1	1			4	5
7		31980	34434	3					5
7	ARS702	64279	64528	2	1	2	3	4	5
7	ARS704	112080	112319	2	1	2	3	4	5
7	ARS707	163180	163447	1	1	2	3	4	5
7		186583	188207	3		2		4	5
7	ARS710	203917	204159	1	1	2	3	4	5
7		240393	242127	2		2		4	5
7	ARS714	285951	286246	1	1	2	3	4	5
7	ARS716*	352695	352917	2	1	2	3	4	5
7	ARS717	388658	388892	1	1	2	3	4	5
7	ARS718	421093	421342	1	1	2	3	4	5
7		479993	482537	1					5
7	ARS719	484932	485160	1	1	2	3	4	5
7	ARS720	508729	508978	1		2	3	4	5

7		552053	553847	3					5
7	ARS722	574622	574916	1	1	2		4	5
7		607547	607605	3					
7	ARS727	659809	660054	1	1	2	3	4	5
7	ARS728	715273	715556	1	1	2	3	4	5
7	ARS729	777967	778216	1	1	2	3	4	5
7	ARS731	834492	834736	1	1	2	3	4	5
7	ARS731.5*	888350	888599	1	1		3	4	5
7		916233	917857	2		2		4	5
7	ARS733	977730	977979	2	1	2	3	4	5
7	ARS734	999448	999695	2	1	2	3	4	5
7		1061998	1064852	2	1			4	5
8	ARS131n	5158	6168	2		2		4	5
8		45113	47112	2		2		4	5
8	ARS805	64255	64489	1	1	2	3	4	5
8		110423	112247	1					5
8	ARS807	133347	133591	1	1	2	3	4	
8	ARS809*	168531	168773	2	1	2	3	4	5
8	ARS813	245719	245968	1	1	2	3	4	5
8		296233	298222	1	1	2	3	4	5
8	*	358953	360647	1	1	2		4	5
8		380153	382157	3	1	2			5
8	ARS818	392148	392391	1	1	2		4	5
8	ARS820	447619	447853	1	1	2	3	4	5
8		474023	475267	2		2		4	5
8	ARS822*	501751	501992	1	1	2		4	5
8		519478	521607	2		2		4	5
9		32509	34859	2	1	2		4	5
9		72923	75194	1	1	2		4	5
9	ARS909	105821	106048	1	1	2	3	4	5
9	ARS911	136094	136355	2		2	3	4	5
9	ARS912*	175034	175355	1	1	2		4	5
9		189523	190747	2					5
9	ARS913	214675	214826	1	1	2	3	4	5
9	ARS914	247579	247800	1		2	3	4	5
9		308193	311477	1	1	2		4	5
9	ARS919	341853	342096	1		2	3	4	5
9	ARS920	357156	357393	1	1	2	3	4	5
9	ARS922	411817	412053	1	1	2	3	4	5
10	ARS1004	23661	24158	2	1	2		4	5
10	ARS1005	67476	67949	1	1	2	3	4	5
10	ARS1006	99359	99796	1		2	3		5
10	ARS1007	113226	113828	2	1	2		4	5
10	ARS1007.5	161435	161860	2	1			4	5
10	ARS1008	203729	204614	1	1	2	3	4	5
10	ARS1009	228248	228740	1	1	2		4	5

10	ARS1009.5	248600	249300	3					5
10	ARS1010	298471	298952	2	1	2		4	5
10	ARS1011	336976	337225	2	1	2	3	4	5
10	ARS1012	374575	374818	1	1	2	3	4	5
10	ARS1014	416888	417134	1	1	2	3	4	5
10	ARS1015	442248	442658	1		2	3	4	5
10	ARS1016	454276	455248	3	1			4	5
10	ARS1018	540239	540474	1	1	2	3	4	5
10	ARS1019	612542	612975	1	1	2	3	4	5
10	ARS1020	654069	654309	2	1	2		4	5
10	ARS1021	683328	683817	1	1	2	3	4	5
10	ARS1022	711590	711837	2	1	2		4	5
11		10002	17002	2	1	2			5
11	ARS1103	55670	55917	1	1	2	3	4	5
11	ARS1104.5	98329	98568	1	1		3	4	5
11		113073	115127	3		2		4	5
11	ARS1106	152934	153173	1	1	2	3	4	5
11	ARS1106.3	196038	196284	2				4	5
11	ARS1106.7	213080	213385	1	1		3	4	5
11		256478	258327	1	1	2	3	4	5
11		301463	303017	1	1		3	4	5
11	ARS1109	329322	329571	1	1	2	3	4	5
11	ARS1112	388607	388902	1	1	2	3	4	5
11	ARS1113	416822	417055	1	1	2	3	4	5
11	ARS1114	447657	447892	1	1	2	3	4	5
11	ARS1114.5	454453	459197	1	1				5
11	ARS1116	516653	516902	1	1	2	3	4	5
11		528843	531447	2	1	2	3	4	5
11	ARS1118*	581468	581706	2		2	3	4	5
11	ARS1120	611874	612107	1	1	2	3	4	5
11	ARS1123	642355	642602	2	1	2	3	4	5
12		28749	32505	2	1	2		4	5
12		51239	52929	3		2		4	5
12	ARS1206	91417	91659	1	1	2	3	4	5
12		139293	140447	1	1	2	3	4	5
12	ARS1209	156646	156883	1	1	2	3	4	5
12	ARS1211	231179	231422	1	1	2	3	4	5
12	ARS1212	289220	289469	2	1	2		4	5
12		342443	344717	2				4	5
12	ARS1213	373156	373400	1	1	2	3	4	5
12	ARS1215	412668	412897	1	1	2	3	4	5
12	ARS1216	450485	450726	2		2	3	4	5
12	ARS1217	512868	513117	1	1	2	3	4	5
12	ARS1218	602938	603155	1	1	2		4	5
12		622103	623937	2		2		4	5
12	ARS1220	659823	660072	1	1	2		4	5

12		729123	732237	1		2		4	5
12	ARS1223	744942	745179	1	1	2	3	4	5
12	ARS1226	794020	794269	2	1	2	3	4	5
12		821403	823002	2	1	2	3	4	5
12	ARS1227.5	888569	888810	2	1		3	4	5
12		927043	929292	2	1	2		4	5
12		947123	949407	2		2		4	5
12	ARS1232	1007180	1007470	2	1	2	3	4	
12	ARS1234	1023967	1024207	2	1	2		4	5
12		1058848	1060802	2	1	2			5
13	ARS1303	31687	31935	1	1	2	3	4	5
13	ARS1305	94216	94463	1	1	2	3	4	5
13	ARS1307	137299	137548	1	1	2	3	4	5
13	*	158563	159797	3	1			4	5
13	ARS1308	183793	184037	1	1	2	3	4	
13		227023	227487	2	1			4	5
13	ARS1309	263062	263296	1	1	2	3	4	5
13	ARS1310	286782	287067	1	1	2	3	4	5
13	ARS1312	370976	371221	1	1	2	3	4	5
13		431453	433827	2	1	2	3	4	5
13	ARS1316*	468177	468468	2	1	2	3	4	5
13		502233	504197	1	1	2	3	4	5
13	ARS1320	535595	535843	1	1	2	3	4	5
13		553563	556237	1		2		4	5
13	ARS1323	611273	611488	1	1	2	3	4	5
13	ARS1324	634479	634714	1		2		4	5
13	ARS1325	649307	649551	1	1	2	3	4	5
13		687903	689917	2		2	3	4	5
13		723896	730574	3	1		3		
13	ARS1327*	758222	758470	1		2		4	5
13	ARS1328	772629	772878	2		2		4	5
13	ARS1329	805116	805338	2		2			5
13	ARS1330	815341	815567	1	1	2	3	4	5
13		836823	838167	2		2			5
13	ARS1332	897804	898040	2	1	2	3	4	5
14	ARS1405	28467	28699	2		2		4	5
14	ARS1406	61597	61894	2		2	3	4	5
14	ARS1407	89528	89802	1	1	2	3	4	5
14		125623	127807	2	1	2		4	5
14	ARS1411	169566	169804	2		2	3	4	5
14	ARS1412*	196055	196291	2	1	2		4	5
14	ARS1413	250259	250506	2	1	2	3	4	5
14	ARS1414	279875	280108	2	1	2	3	4	5
14	ARS1415	321917	322210	1	1	2	3	4	5
14		348469	355469	2	1		3	4	
14	ARS1417	412263	412493	2	1	2	3	4	5

14	ARS1419	449343	449588	2		2	3	4	5
14	ARS1420	498987	499232	2		2	3	4	5
14	ARS1421	545966	546201	1		2		4	5
14	ARS1422	561106	561384	1	1	2	3	4	5
14	ARS1424	609458	609706	1	1	2	3	4	5
14	ARS1426	635660	635901	1		2	3	4	5
14	ARS1427	691482	691727	1	1	2	3	4	5
14		712633	714697	2		2		4	5
14		726215	726274	3					
14		737003	740237	2			3	4	5
14		763353	765237	2	1	2		4	5
15		9761	12672	2		2		4	5
15		18367	20227	2		2		4	5
15	ARS1506.5	35667	35903	1				4	5
15	ARS1507	72636	72872	2		2		4	5
15	ARS1508	85195	85444	1	1	2	3	4	5
15	ARS1509	113843	114084	1	1	2	3	4	5
15	ARS1510	166974	167220	1	1	2	3	4	5
15		223471	230471	1	1		3	4	
15	ARS1511	277529	277778	1	1	2	3	4	5
15		308463	310157	1	1	2	3	4	5
15	ARS1513	337279	337528	1	1	2	3	4	5
15	ARS1513.5	436732	436966	1	1		3	4	
15		463698	464877	1	1		3	4	5
15	ARS1514	489645	490129	1	1		3		5
15		496203	498207	2					5
15		520113	521217	2	1				5
15	ARS1516	566409	566643	2	1	2		4	5
15	ARS1517	600885	600960	2	1	2		4	5
15	ARS1519	656632	656876	2	1	2	3	4	5
15		680823	683197	2		2	3	4	5
15	ARS1521	729739	729969	2		2		4	5
15	ARS1523	766617	766862	1	1	2		4	5
15	ARS1524	783344	783563	2		2	3	4	5
15		854173	856027	1	1	2		4	5
15	ARS1526	874190	874434	1	1	2	3	4	5
15	ARS1528	908288	908537	1	1	2		4	5
15	ARS1529	981454	981690	1	1	2	3	4	5
15		1022548	1023652	3					
15	ARS1529.5	1053490	1053901	1	1		3	4	5
16	ARS1604	42976	43212	1	1	2	3	4	5
16	ARS1605	73038	73283	1	1	2	3	4	5
16		89758	92975	2	1	2		4	5
16	ARS1607	116505	116765	2	1	2	3	4	5
16		161693	163447	2	1	2		4	5
16		178643	179827	2	1				5

16		189963	191457	2	1			4	5
16		210143	211437	2	1	2		4	5
16		261168	262687	2	1	2	3	4	5
16	ARS1614	289483	289704	1	1	2	3	4	5
16		316743	318397	3		2	3	4	5
16		331043	332607	2	1	2		4	5
16		376723	377937	2					5
16	ARS1618	384536	384784	1	1	2	3	4	5
16	ARS1619	418132	418359	1	1	2	3	4	5
16	ARS1620.5	456557	456805	2	1		3	4	5
16	ARS1621	511619	511940	1	1	2	3	4	5
16		552403	554287	1	1			4	5
16	ARS1622	563822	564061	1	1	2		4	5
16	ARS1623	633868	634117	1	1	2	3	4	5
16	ARS1624	684383	684632	1	1	2		4	5
16	ARS1625	695432	695681	2	1	2		4	5
16	ARS1626*	749094	749341	2		2		4	5
16	ARS1626.5	776921	777152	1	1		3	4	5
16	ARS1627	819153	819393	1	1	2	3	4	5
16	ARS1628	842646	842894	1	1	2		4	5
16	ARS1630	880854	881102	2		2	3	4	5
16		927939	927998	3					
16	ARS1631	932976	933223	2	1	2		4	5
16		940923	943157	1	1	2		4	5

References for Table F-1

1) (Raghuraman et al., 2001)

2) (Wyrick et al., 2001)

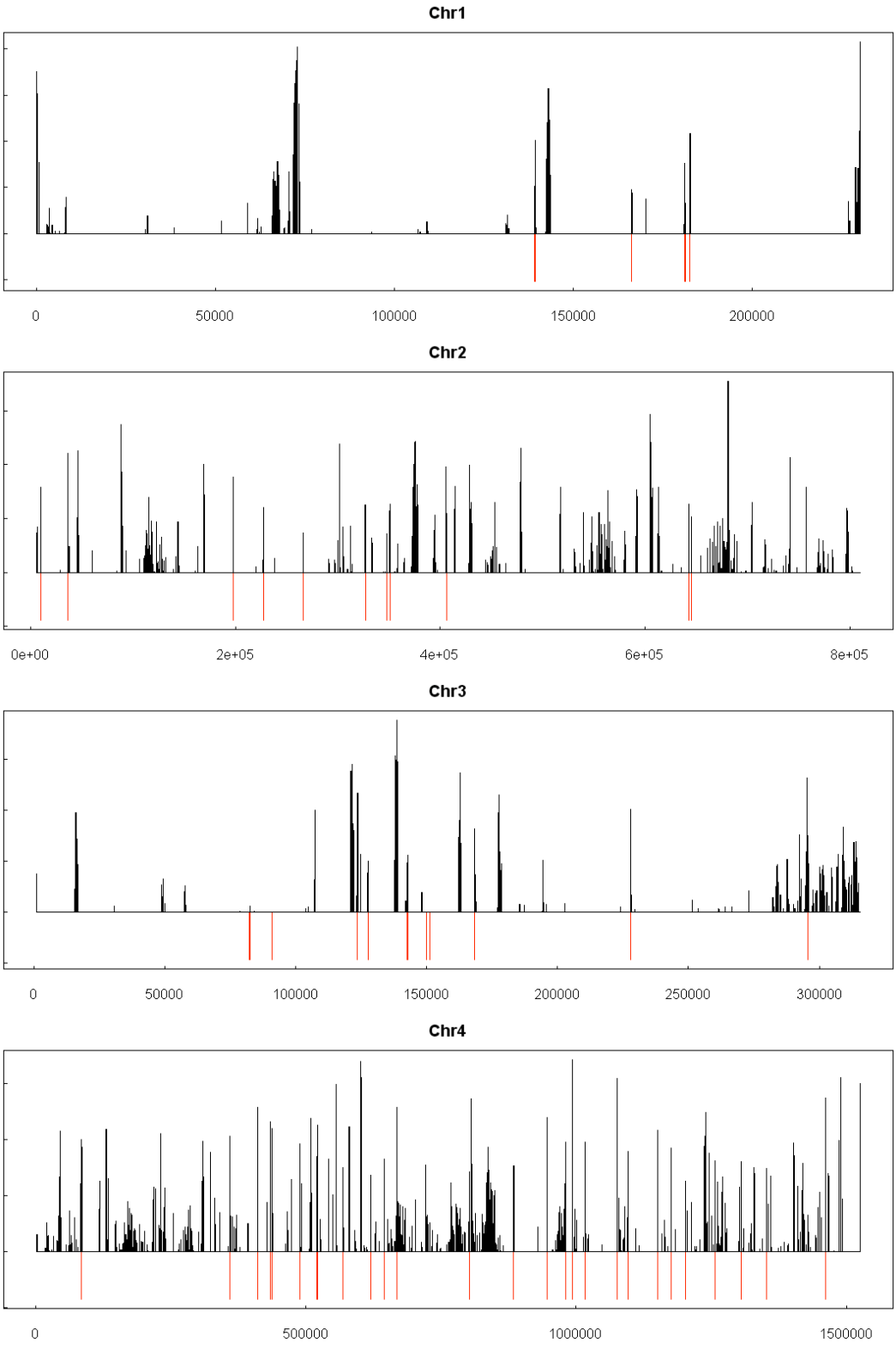
3) (Yabuki et al., 2002)

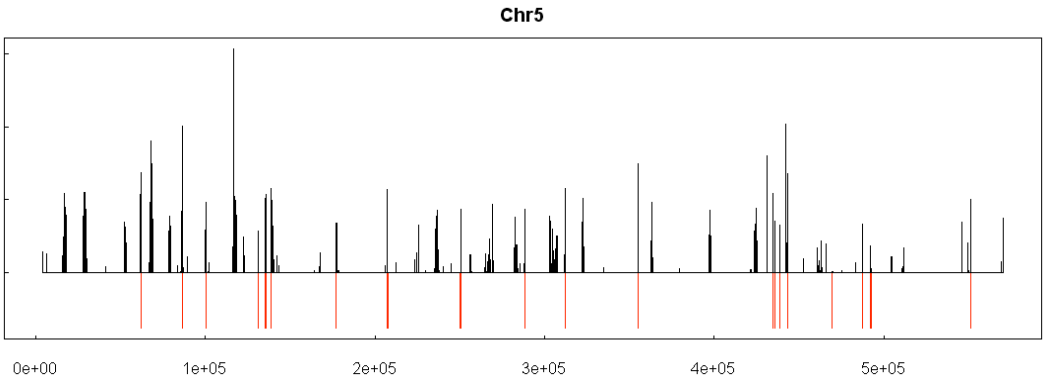
4) (Feng et al., 2006)

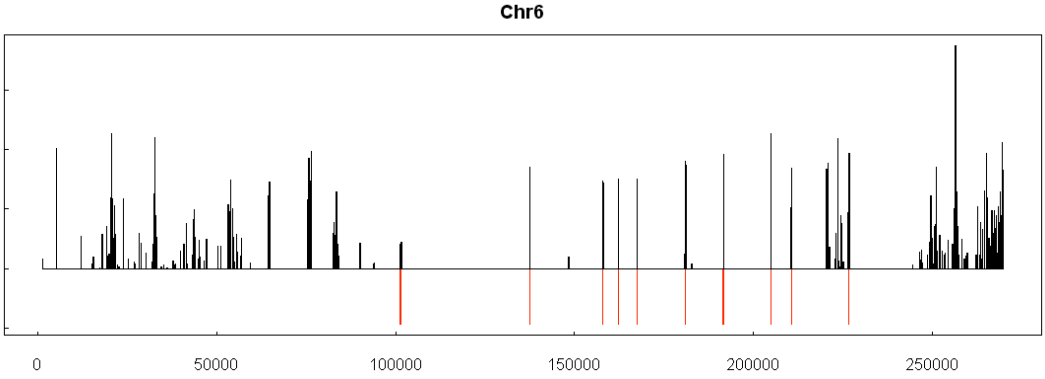
5) (Xu et al., 2006)

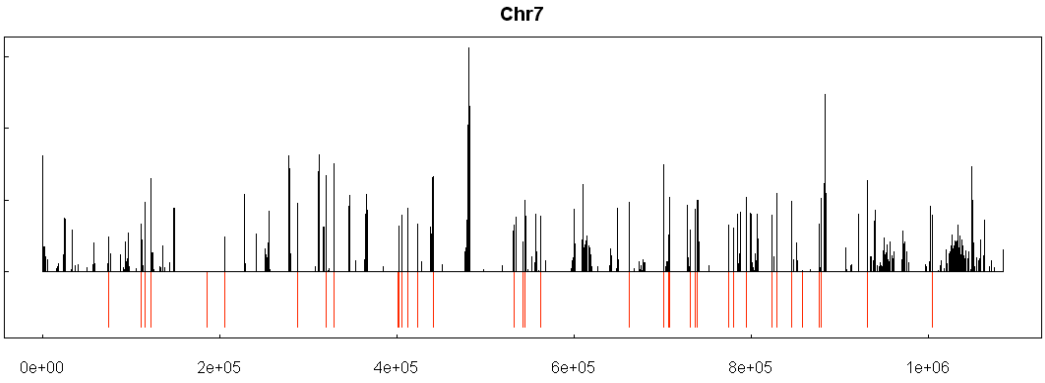
APPENDIX G — The Locations of tRNA Genes and Arrest Sites for the Entire Yeast Genome.

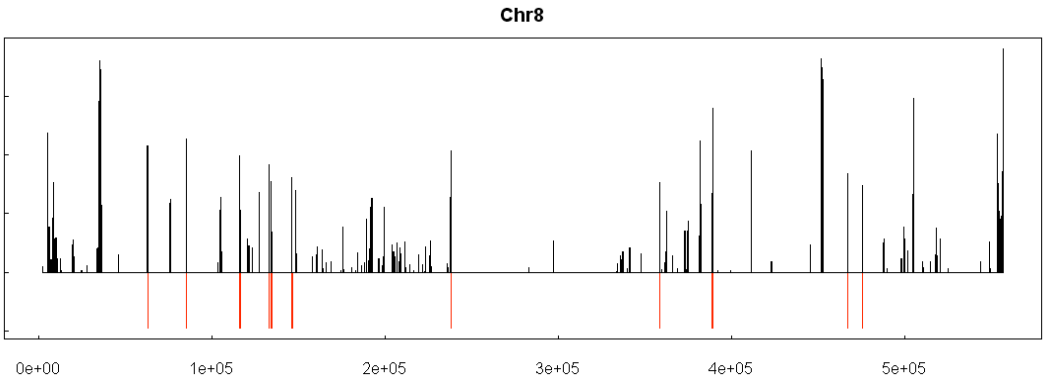
The 75 min time point is shown for each chromosome. Red lines indicate the locations of tRNA genes, while peaks in the data indicate sites of long-term fork arrest in the genome.

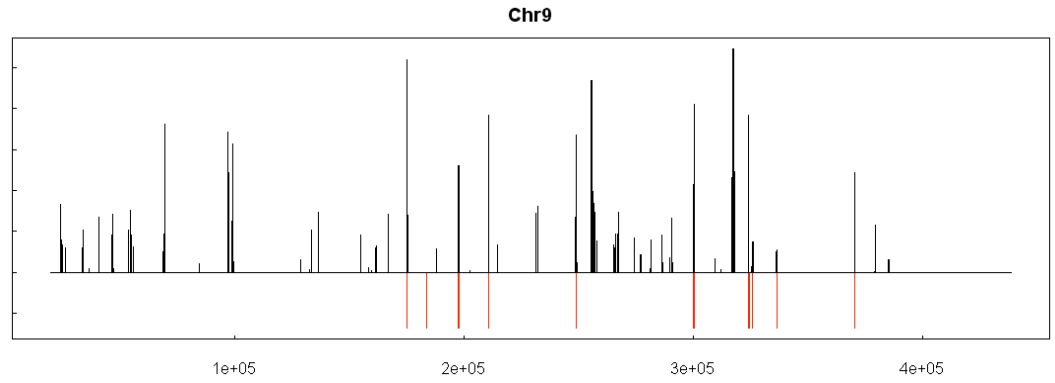

Chr1
0
50000
100000
150000
200000
Chr2
0e+00
2e+05
4e+05
6e+05
8e+05
Chr3
0
50000
100000
150000
200000
250000
300000
Chr4
0
500000
1000000
1500000

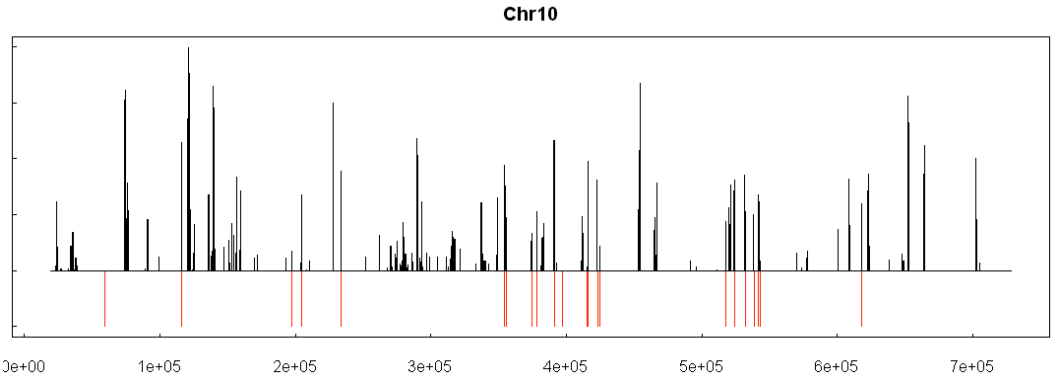

Chr5
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05

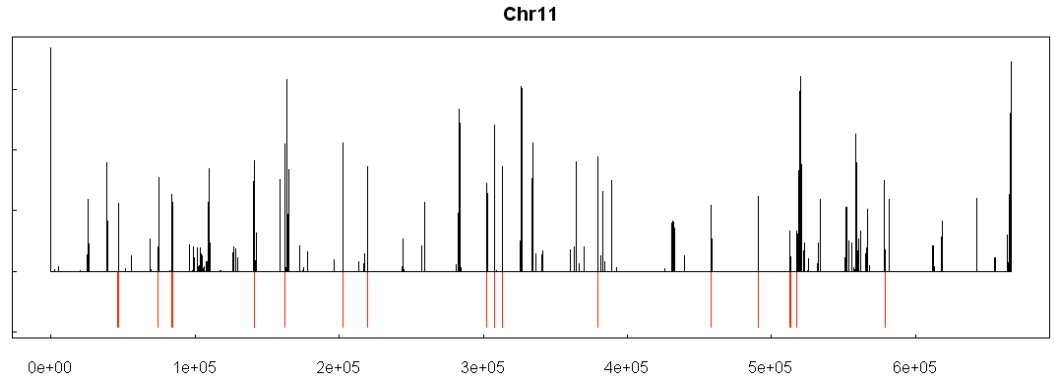

Chr6
0
50000
100000
150000
200000
250000

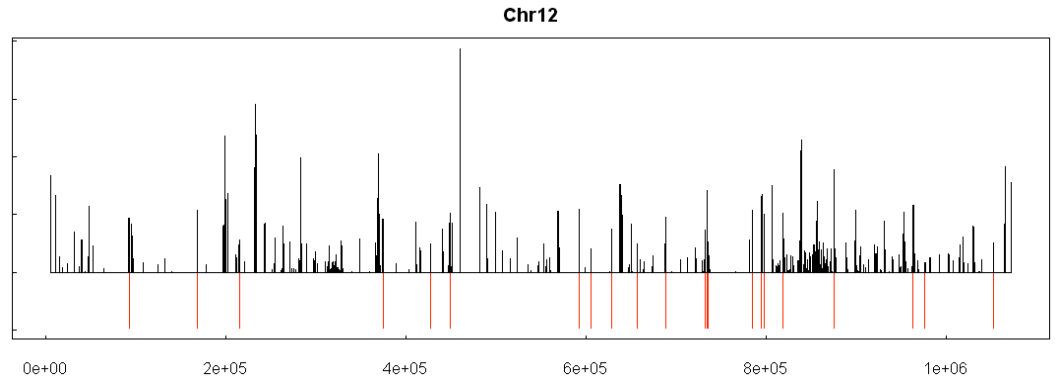

Chr7
0e+00
2e+05
4e+05
6e+05
8e+05
1e+06

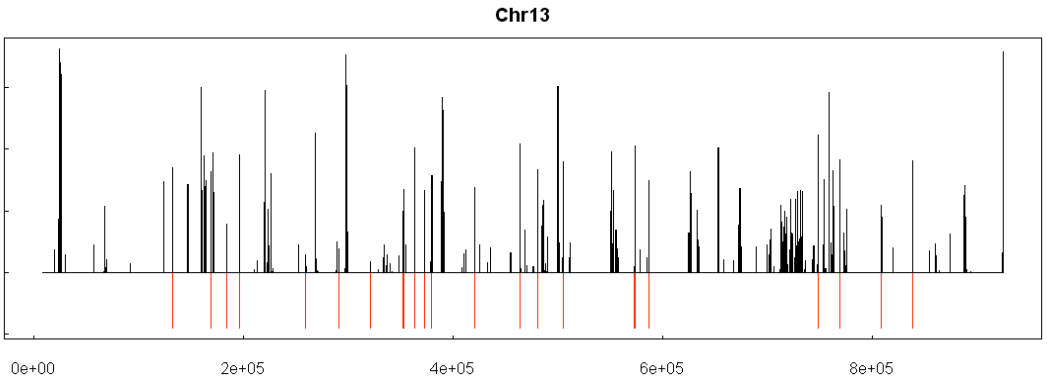

Chr8
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05

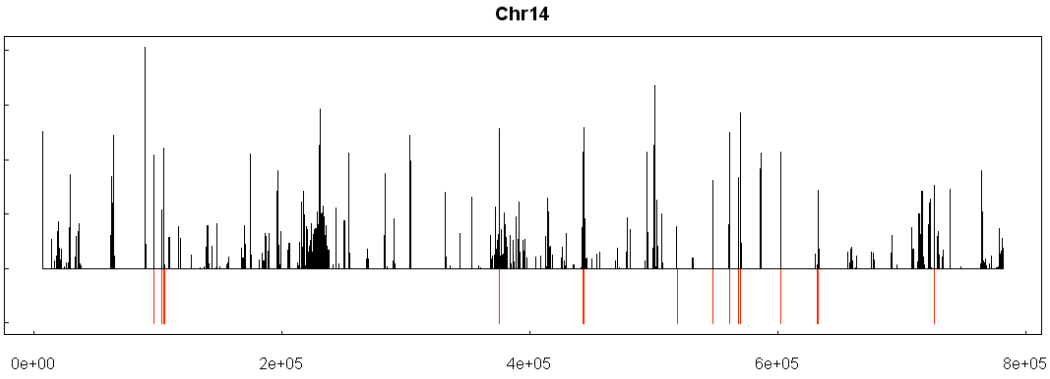

Chr9
1e+05
2e+05
3e+05
4e+05

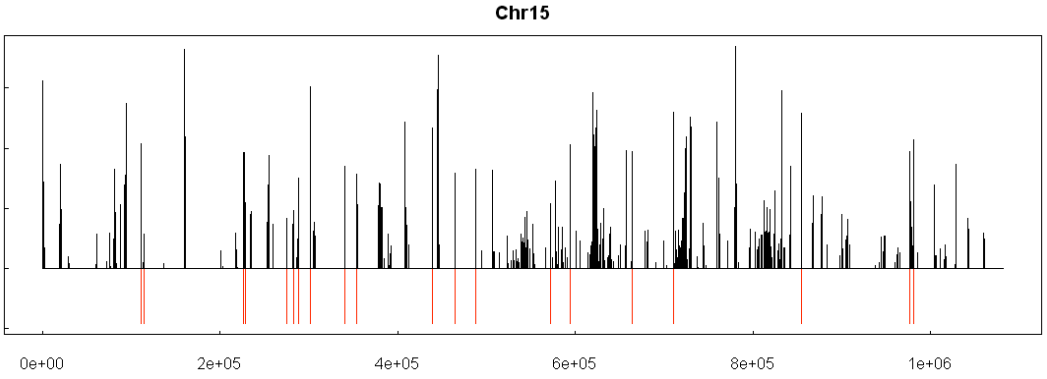

Chr10
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
6e+05
7e+05

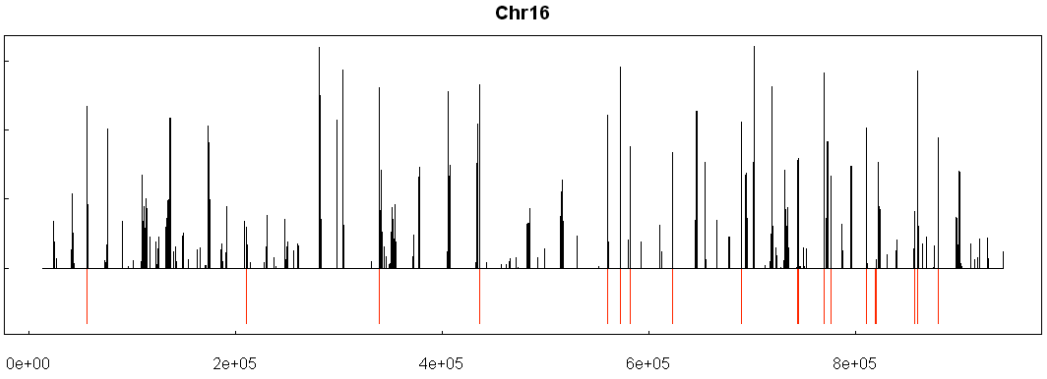

Chr11
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
6e+05


Chr12
0e+00
2e+05
4e+05
6e+05
8e+05
1e+06


Chr13
0e+00
2e+05
4e+05
6e+05
8e+05


Chr14
0e+00
2e+05
4e+05
6e+05
8e+05


Chr15
0e+00
2e+05
4e+05
6e+05
8e+05
1e+06


Chr16
0e+00
2e+05
4e+05
6e+05
8e+05

REFERENCES

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., *et al.* (2007). The molecular architecture of the nuclear pore complex. Nature *450*, 695-701.

Antequera, F. (2004). Genomic Specification and Epigenetic Regulation of Eukaryotic DNA Replication Origins. Embo J *23*, 4365-4370.

Aparicio, O.M., Weinstein, D.M., and 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*, 59-69.

Avery, O.T., Macleod, C.M., and McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J Exp Med *79*, 137-158.

Azvolinsky, A., Dunaway, S., Torres, J.Z., Bessler, J.B., and Zakian, V.A. (2006). The S. cerevisiae Rrm3p DNA helicase moves with the replication fork and affects replication of all yeast chromosomes. Genes Dev *20*, 3104-3116.

Azvolinsky, A., Giresi, P.G., Lieb, J.D., and Zakian, V.A. (2009). Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae. Mol Cell *34*, 722-734.

Beavis, R.C., and Chait, B.T. (1996). Matrix-assisted laser desorption ionization mass-spectrometry of proteins. Methods Enzymol *270*, 519-551.

Bell, S.P. (1995). Eukaryotic replicators and associated protein complexes. Curr Opin Genet Dev *5*, 162-167.

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

Blow, J.J., and Dutta, A. (2005). Preventing re-replication of chromosomal DNA. Nat. Rev. Mol. Cell. Biol. *6*, 476-486.

Blow, J.J., Gillespie, P.J., Francis, D., and Jackson, D.A. (2001). Replication origins in Xenopus egg extract Are 5-15 kilobases apart and are activated in clusters that fire at different times. J Cell Biol *152*, 15-25.

Burgers, P.M. (1991). Saccharomyces cerevisiae replication factor C. II. Formation and activity of complexes with the proliferating cell nuclear antigen and with DNA polymerases delta and epsilon. J Biol Chem *266*, 22698-22706.

Calzada, A., Hodgson, B., Kanemaki, M., Bueno, A., and Labib, K. (2005). Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork. Genes Dev. *19*, 1905-1919.

Chang, Y.P., Wang, G., Bermudez, V., Hurwitz, J., and Chen, X.S. (2007). Crystal structure of the GINS complex and functional insights into its role in DNA replication. Proc Natl Acad Sci U S A *104*, 12685-12690.

Costa, S.B.J.J. The elusive determinants of replication origins. Embo Rep *8*, 332-334.

Cristea, I.M., Williams, R., Chait, B.T., and Rout, M.P. (2005). Fluorescent proteins as proteomic probes. Mol Cell Proteomics *4*, 1933-1941.

DaFonseca, C.J., Shu, F., and Zhang, J.J. (2001). Identification of two residues in MCM5 critical for the assembly of MCM complexes and Stat1-mediated transcription activation in response to IFN-gamma. Proc Natl Acad Sci USA *98*, 3034-3039.

DePamphilis, M.L. (2006). DNA Replication and Human Disease. Cold Spring Harbor Laboratory Press.

Deshpande, A.M., and Newlon, C.S. (1996). DNA replication fork pause sites dependent on transcription. Science *272*, 1030-1033.

Diffley, J.F. (2004). Regulation of early events in chromosome replication. Curr. Biol. *14*, 778-786.

Drake, J.W., Charlesworth, B., Charlesworth, D., and Crow, J.F. (1998). Rates of spontaneous mutation. Genetics *148*, 1667-1686.

Dua, R., Levy, D.L., and Campbell, J.L. (1999). Analysis of the essential functions of the C-terminal protein/protein interaction domain of Saccharomyces cerevisiae pol epsilon and its unexpected ability to support growth in the absence of the DNA polymerase domain. J Biol Chem *274*, 22283-22288.

Dua, R., Levy, D.L., Li, C.M., Snow, P.M., and Campbell, J.L. (2002). In vivo reconstitution of Saccharomyces cerevisiae DNA polymerase epsilon in insect cells. Purification and characterization. J Biol Chem *277*, 7889-7896.

Dudley, A.M., Aach, J., Steffen, M.A., and Church, G.M. (2002). Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. Proc Natl Acad Sci USA *99*, 7554-7559.

Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., Li, H., Stillman, B., and Speck, C. (2009). A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. Proc Natl Acad Sci USA *106*, 20240-20245.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. Nature *421*, 448-453.

Feng, W., Collingwood, D., Boeck, M.E., Fox, L.A., Alvino, G.M., Fangman, W.L., Raghuraman, M.K., and Brewer, B.J. (2006). Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. Nat. Cell Biol. *8*, 148-155.

Feng, W., and D'Urso, G. (2001). Schizosaccharomyces pombe cells lacking the amino-terminal catalytic domains of DNA polymerase epsilon are viable but require the DNA damage checkpoint control. Mol Cell Biol *21*, 4495-4504.

Feng, W., Rodriguez-Menocal, L., Tolun, G., and D'Urso, G. (2003). Schizosacchromyces pombe Dpb2 binds to origin DNA early in S phase and is required for chromosomal DNA replication. Mol Biol Cell *14*, 3427-3436.

Formosa, T., and 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*, 1459-1470.

Forsburg, S.L. (2004). Eukaryotic MCM Proteins: Beyond Replication Initiation. Microbiology and Molecular Biology Reviews *68*, 109-131.

Gambus, A., Jones, R.C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R.D., and Labib, K. (2006). GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. Nat. Cell Biol. *8*, 358-366.

Garg, P., and Burgers, P.M. (2005). DNA polymerases that propagate the eukaryotic DNA replication fork. Crit Rev Biochem Mol Biol *40*, 115-128.

Georgescu, R.E., Kim, S.S., Yurieva, O., Kuriyan, J., Kong, X.P., and O'Donnell, M. (2008). Structure of a sliding clamp on DNA. Cell *132*, 43-54.

Goldsby, R.E., Hays, L.E., Chen, X., Olmsted, E.A., Slayton, W.B., Spangrude, G.J., and Preston, B.D. (2002). High incidence of epithelial cancers in mice deficient for DNA polymerase delta proofreading. Proc Natl Acad Sci USA *99*, 15560-15565.

Haase, S.B., and Lew, D.J. (1997). Flow cytometric analysis of DNA content in budding yeast. Methods Enzymol *283*, 322-332.

Hanna, J.S., Kroll, E.S., Lundblad, V., and Spencer, F.A. (2001). Saccharomyces cerevisiae CTF18 and CTF4 are required for sister chromatid cohesion. Mol. Cell. Biol. *21*, 3144-3158.

Hiraga, S., Hagihara-Hayashi, A., Ohya, T., and Sugino, A. (2005). DNA polymerases alpha, delta, and epsilon localize and function together at replication forks in Saccharomyces cerevisiae. Genes Cells *10*, 297-309.

Hua, X.H., and Newport, J. (1998). Identification of a preinitiation step in DNA replication that is independent of origin recognition complex and cdc6, but dependent on cdk2. J Cell Biol *140*, 271-281.

Hyrien, O., Marheineke, K., and A., G. (2003). Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem. Bioessays *25*, 116-125.

Iida, T., and Araki, H. (2004). Noncompetitive counteractions of DNA polymerase epsilon and ISW2/yCHRAC for epigenetic inheritance of telomere position effect in Saccharomyces cerevisiae. Mol Cell Biol *24*, 217-227.

Ilves, I., Petojevic, T., Pesavento, J.J., and Botchan, M.R. (2010). Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. Mol Cell *37*, 247-258.

Ishimi, Y., Ichinose, S., Omori, A., Sato, K., and Kimura, H. (1996). Binding of human minichromosome maintenance proteins with histone H3. J Biol Chem *271*, 24115-24122.

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

Ishimi, Y., Komamura-Kohno, Y., Arai, K., and Masai, H. (2001). Biochemical activities associated with mouse Mcm2 protein. J Biol Chem *276*, 42744-42752.

Ishimi, Y., Komamura-Kohno, Y., Kwon, H.J., Yamada, K., and Nakanishi, M. (2003). Identification of MCM4 as a target of the DNA replication block checkpoint system. J Biol Chem *278*, 24644-24650.

Ivessa, A.S., Lenzmeier, B.A., Bessler, J.B., Goudsouzian, L.K., Schnakenberg, S.L., and Zakian, V.A. (2003). The Saccharomyces cerevisiae helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. Mol Cell *12*, 1525-1536.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jeruzalmi, D., O'Donnell, M., and Kuriyan, J. (2002). Clamp loaders and sliding clamps. Curr Opin Struct Biol *12*, 217-224.

Johnson, A., and O'Donnell, M. (2005). Cellular DNA replicases: components and dynamics at the replication fork. Annu Rev Biochem. *74*, 283-315.

Kanemaki, M., and Labib, K. (2006). Distinct roles for Sld3 and GINS during establishment and progression of eukaryotic DNA replication forks. EMBO J. *25*.

Kanemaki, M., Sanchez-Diaz, A., Gambus, A., and Labib, K. (2003). Functional proteomic identification of DNA replication proteins by induced proteolysis in vivo. Nature *423*, 720-725.

Kesti, T., Flick, K., Keranen, S., Syvaoja, J.E., and Wittenberg, C. (1999). DNA polymerase epsilon catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. Mol Cell *3*, 679-685.

Kitamura, E., Blow, J.J., and Tanaka, T.U. (2006). Live-cell imaging reveals replication of individual replicons in eukaryotic replication factories. Cell *125*, 1297-1308.

Kouprina, N., Kroll, E., Bannikov, V., Bliskovsky, V., Gizatullin, R., Kirillov, A., Shestopalov, B., Zakharyev, V., Hieter, P., Spencer, F., and et, a.l. (1992). CTF4 (CHL15) mutants exhibit defective DNA metabolism in the yeast Saccharomyces cerevisiae. Mol Cell Biol. *12*, 5736-5747.

Krutchinsky, A.N., Kalkum, M., and Chait, B.T. (2001). Automatic identification of proteins with a MALDI-quadrupole ion trap mass spectrometer. Anal Chem *73*, 5066-5077.

Kubota, Y., Takase, Y., Komori, Y., Hashimoto, Y., Arata, T., Kamimura, Y., Araki, H., and Takisawa, H. (2003). A novel ring-like complex of Xenopus proteins essential for the initiation of DNA replication. Genes Dev. *17*, 1141-1152.

Labib, K., and Gambus, A. (2007). A key role for the GINS complex at DNA replication forks. Trends Cell Biol. *17*, 271-278.

Langston, L.D., and O'Donnell, M. (2008). DNA polymerase delta is highly processive with proliferating cell nuclear antigen and undergoes collision release upon completing DNA. J Biol Chem *283*, 29522-29531.

Liang, D.T., Hodson, J.A., and Forsburg, S.L. (1999). Reduced dosage of a single fission yeast MCM protein causes genetic instability and S phase delay. J Cell Sci *112*, 559-567.

Loeb, L.A. (2001). A mutator phenotype in cancer. Cancer Res *61*, 3230-3239.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

MacAlpine, D.M., and Bell, S.P. (2005). A genomic view of eukaryotic DNA replication. Chromosome Res *13*, 309-326.

Machida, Y.J., Hamlin, J.L., and Dutta, A. (2005). Right place, right time, and only once: replication initiation in metazoans. Cell *123*, 13-24.

MacNeill, S.A. (2010). Structure and function of the GINS complex, a key component of the eukaryotic replisome. Biochem J *425*, 489-500.

Makarova, K.S., Wolf, Y.I., Mekhedov, S.L., Mirkin, B.G., and Koonin, E.V. (2005). Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. Nucl. Acids Res. *33*, 4626-4638.

Marinsek, N., Barry, E.R., Makarova, K.S., Dionne, I., Koonin, E.V., and Bell, S.D. (2006). GINS, a central nexus in the archaeal DNA replication fork. EMBO Rep.

Masai, H., and Arai, K. (2002). Cdc7 kinase complex: a key regulator in the initiation of DNA replication. J Cell Physiol *190*, 287-296.

Masumoto, H., Muramatsu, S., Kamimura, Y., and Araki, H. (2002). S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. Nature. *415*, 651-655.

McCarroll, R.M., and Fangman, W.L. (1988). Time of replication of yeast centromeres and telomeres. Cell *54*, 505-513.

McHenry, C.S. (2003). Chromosomal replicases as asymmetric dimers: studies of subunit arrangement and functional consequences. Mol Microbiol *49*, 1157-1165.

Mendez, J., and Stillman, B. (2003). Perpetuating the double helix: molecular machines at eukaryotic DNA replication origins. BioEssays *25*, 1158-1167.

Mesner, L.D., Crawford, E.L., and Hamlin, J.L. (2006). Isolating apparently pure libraries of replication origins from complex genomes. Mol Cell *21*, 719-726.

Miles, J., and Formosa, T. (1992). Evidence that POB1, a Saccharomyces cerevisiae protein that binds to DNA polymerase alpha, acts in DNA metabolism in vivo. Mol Cell Biol *12*, 5724-5735.

Moyer, S.E., Lewis, P.W., and Botchan, M.R. (2006). Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. Proc. Natl. Acad. Sci. USA *103*, 10236-10241.

Muramatsu, S., Hirai, K., Tak, Y.S., Kamimura, Y., and Araki, H. (2010). CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol-epsilon, and GINS in budding yeast. Genes Dev *24*, 602-612.

Murray, A.W. (1989). The Cell Cycle. Amer. Zool. *29*, 511-522.

Murray, A.W. (2004). Recycling the cell cycle: cyclins revisited. Cell. *116*, 221-234.

Newlon, C.S., and Theis, J.F. (1993). The structure and function of yeast ARS elements. Curr Opin Genet Dev *3*, 752-758.

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

Nick McElhinny, S.A., Gordenin, D.A., Stith, C.M., Burgers, P.M., and Kunkel, T.A. (2008). Division of labor at the eukaryotic replication fork. Mol Cell *30*, 137-144.

Nieduszynski, C.A., Hiraga, S.-I., Ak, P., Benham, C.J., and Donaldson, A.D. (2007). OriDB: a DNA replication origin database. Nucleic Acids Res. *35*, 40-46.

Nishitani, H., and Lygerou, Z. (2004). DNA replication licensing. Front Biosci *9*, 2115-2132.

Nishitani, H., Lygerou, Z., Nishimoto, T., and Nurse, P. (2000). The Cdt1 protein is required to license DNA for replication in fission yeast. Nature *404*, 625-628.

Pacek, M., Tutter, A.V., Kubota, Y., Takisawa, H., and Walter, J.C. (2006). Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. Mol. Cell *21*, 581-587.

Pavlov, Y.I., and Shcherbakova, P.V. (2010). DNA polymerases at the eukaryotic fork-20 years later. Mutat Res *685*, 45-53.

Poloumienko, A., Dershowitz, A., De, J., and Newlon, C.S. (2001). Completion of replication map of Saccharomyces cerevisiae chromosome III. Mol. Biol. Cell *12*, 3317-3327.

Pursell, Z.F., Isoz, I., Lundstrom, E.B., Johansson, E., and Kunkel, T.A. (2007). Yeast DNA polymerase epsilon participates in leading-strand DNA replication. Science *317*, 127-130.

Raghuraman, M.K., Winzeler, E.A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D.J., Davis, R.W., Brewer, B.J., and Fangman, W.L. (2001). Replication dynamics of the yeast genome. Science *294*, 115-121.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O.M., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., *et al.* (2000). Genome-wide location and function of DNA binding proteins. Science *290*, 2306-2309.

Reynolds, N., Watt, A., Fantes, P.A., and MacNeill, S.A. (1998). Cdm1, the smallest subunit of DNA polymerase d in the fission yeast Schizosaccharomyces pombe, is non-essential for growth and division. Curr Genet *34*, 250-258.

Richman, D.D., Cleveland, P.H., Oxman, M.N., and Johnson, K.M. (1982). The binding of staphylococcal protein A by the sera of different animal species. J Immunol *128*, 2300-2305.

Richmond, T.J., and Davey, C.A. (2003). The structure of DNA in the nucleosome core. Nature *423*, 145-150.

Rothstein, R., Michel, B., and Gangloff, S. (2000). Replication fork pausing and recombination or "gimme a break". Genes Dev *14*, 1-10.

Rowles, A., Tada, S., and Blow, J.J. (1999). Changes in association of the Xenopus origin recognition complex with chromatin on licensing of replication origins. J Cell Sci *112 (Pt 12)*, 2011-2018.

Sancar, A., Lindsey-Boltz, L.A., Unsal-Kacmaz, K., and Linn, S. (2004). Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. Annu Rev Biochem *73*, 39-85.

Sclafani, R.A. (2000). Cdc7p-Dbf4p becomes famous in the cell cycle. J Cell Sci *113*, 2111-2117.

Sclafani, R.A., and Holzen, T.M. (2007). Cell cycle regulation of DNA replication. Annu. Rev. Genet. *20*, 0.

Segal, E., and Widom, J. (2009). What controls nucleosome positions? Trends Genet *25*, 335-343.

Sheu, Y.J., and Stillman, B. (2010). The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. Nature *463*, 113-117.

Shirahige, K., Hori, Y., Shiraishi, K., Yamashita, M., Takahashi, K., Obuse, C., Tsurimoto, T., and Yoshikawa, H. (1998). Regulation of DNA-replication origins during cell-cycle progression. Nature *395*, 618-621.

Shiratori, A., Shibata, T., Arisawa, M., Hanaoka, F., Murakami, Y., and Eki, T. (1999). Systematic identification, classification, and characterization of the open reading frames which encode novel helicase-related proteins in Saccharomyces cerevisiae by gene disruption and Northern analysis. Yeast *15*, 219-253.

Spiesser, T.W., Diener, C., Barberis, M., and Klipp, E. (2010). What influences DNA replication rate in budding yeast? PLoS One *5*, e10203.

Stillman, B. (1996). Cell cycle control of DNA replication. Science *274*, 1659-1664.

Stillman, B. (2008). DNA polymerases at the replication fork in eukaryotes. Mol Cell *30*, 259-260.

Subramanian, S. (2008). Nearly neutrality and the evolution of codon usage bias in eukaryotic genomes. Genetics *178*, 2429-2432.

Tackett, A.J., Dilworth, D.J., Davey, M.J., O'Donnell, M., Aitchison, J.D., Rout, M.P., and Chait, B.T. (2005). Proteomic and genomic characterization of chromatin complexes at a boundary. J. Cell Biol. *169*, 35-47.

Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A., and Araki, H. (2003). GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. Genes Dev. *17*, 1153-1165.

Tanaka, S., Umemori, T., Hirai, K., Muramatsu, S., Kamimura, Y., and H., A. (2007). CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. Nature *445*, 328-332.

Tourriere, H., and Pasero, P. (2007). Maintenance of fork integrity at damaged DNA and natural pause sites. DNA Repair (Amst) *6*, 900-913.

Van Kolde, K.E. (1989). Chromatin. New York, Springer-Verlag *1st ed*.

Van Straalen, N.I., and Roelofs, D. (2006). Introduction to Ecological Genetics. New York, Oxford University Press.

Vermaak, D., Ahmad, K., and Henikoff, S. (2003). Maintenance of chromatin states: an open-and-shut case. Curr Opin Cell Biol *15*, 266-274.

Walter, B.E., and Henry, J.J. (2004). Embryonic expression of pre-initiation DNA replication factors in Xenopus laevis. Gene Expression Patterns *5*, 81-89.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737-738.

Wippold, F.J., 2nd, and Perry, A. (2007). Neuropathology for the neuroradiologist: fluorescence in situ hybridization. Am J Neuroradiol *28*, 406-410.

Wyrick, J.J., Aparicio, J.G., Chen, T., and Barnett, J.D. (2001). Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. Science *294*, 2357-2360.

Xu, W., Aparicio, J.G., Aparicio, O.M., and Tavaré, S. (2006). Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae. BMC Genomics *7*.

Yabuki, N., Terashima, H., and Kitada, K. (2002). Mapping of early firing origins on a replication profile of budding yeast. Genes Cells *7*, 781-789.

Yamashita, M., Hori, Y., Shinomiya, T., Obuse, C., Tsurimoto, T., Yoshikawa, H., and Shirahige, K. (1997). The efficiency and timing of initiation of replication of multiple replicons of Saccharomyces cerevisiae chromosome VI. Genes Cells *2*, 655-665.

Yankulov, K., Todorov, I., Romanowski, P., Licatalosi, D., Cilli, K., McCracken, S., Laskey, R., and Bentley, D.L. (1999). MCM proteins are associated with RNA polymerase II holoenzyme. Mol Cell Biol *19*, 6154-6163.

Zegerman, P., and Diffley, J.F. (2007). Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. Nature *445*, 281-285.

Zhang, J.J., Zhao, Y., Chait, B.T., Lathem, W.W., Ritzi, M., Knippers, R., and Darnell, J.E., Jr. (1998). Ser727-dependent recruitment of MCM5 by Stat1alpha in IFN-gamma-induced transcriptional activation. Embo J *17*, 6963-6971.

Zhang, W., and Chait, B.T. (2000). ProFound: An Expert System for Protein Identification Using Mass Spectrometric Peptide Mapping Information. Anal. Chem. *72*, 2482-2489.