The investigation of the function of repair proteins at G-quadruplex structures in *Saccharomyces cerevisiae* revealed that Mms1 promotes genome stability

Die Untersuchung der Funktion von Reparaturproteinen an G-Quadruplex Strukturen in *Saccharomyces cerevisiae* zeigte, dass Mms1 Genomstabilität fördert

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Biomedicine
submitted by

Katharina Wanzek

from

Braunschweig

Würzburg 2016

Reverse page

Submitted on: ……..

Office stamp

Members of the *Promotionskomitee*:

Chairperson: Prof. Alexander Buchberger

Primary Supervisor: Dr. Katrin Paeschke

Supervisor (Second): Prof. Hans-Joachim Lipps

Supervisor (Third): Prof. Caroline Kisker

Date of Public Defence: ……..

Date of Receipt of Certificates: ……..

Index

1 Abstract

G-quadruplex structures are highly stable alternative DNA structures that can, when not properly regulated, impede replication fork progression and cause genome instability (Castillo Bosch et al, 2014; Crabbe et al, 2004; Koole et al, 2014; Kruisselbrink et al, 2008; London et al, 2008; Lopes et al, 2011; Paeschke et al, 2013; Paeschke et al, 2011; Piazza et al, 2015; Piazza et al, 2010; Piazza et al, 2012; Ribeyre et al, 2009; Sabouri et al, 2014; Sarkies et al, 2012; Sarkies et al, 2010; Schiavone et al, 2014; Wu & Spies, 2016; Zimmer et al, 2016). The aim of this thesis was to identify novel G-quadruplex interacting proteins in *Saccharomyces cerevisiae* and to unravel their regulatory function at these structures to maintain genome integrity. Mms1 and Rtt101 were identified as G-quadruplex binding proteins *in vitro* via a pull-down experiment with subsequent mass spectrometry analysis. Rtt101, Mms1 and Mms22, which are all components of an ubiquitin ligase ($Rtt101^{Mms1/Mms22}$), are important for the progression of the replication fork following fork stalling (Luke et al, 2006; Vaisica et al, 2011; Zaidi et al, 2008). The *in vivo* binding of endogenously tagged Mms1 to its target regions was analyzed genome-wide using chromatin-immunoprecipitation followed by deep-sequencing. Interestingly, Mms1 bound independently of Mms22 and Rtt101 to G-rich regions that have the potential to form G-quadruplex structures. *In vitro*, formation of G-quadruplex structures could be shown for the G-rich regions Mms1 bound to. This binding was observed throughout the cell cycle. Furthermore, the deletion of *MMS1* caused replication fork stalling as evidenced by increased association of DNA Polymerase 2 at Mms1 dependent sites. A gross chromosomal rearrangement assay revealed that deletion of *MMS1* results in a significantly increased genome instability at G-quadruplex motifs compared to G-rich or non-G-rich regions. Additionally, binding of the helicase Pif1, which unwinds G4 structures *in vitro* (Paeschke et al, 2013; Ribeyre et al, 2009; Sanders, 2010; Wallgren et al, 2016), to Mms1 binding sites was reduced in *mms1* cells. The data presented in this thesis, together with published data, suggests a novel mechanistic model in which Mms1 binds to G-quadruplex structures and enables Pif1 association. This allows for replication fork progression and genome integrity.

1.1 Zusammenfassung

Bei G-quadruplex Strukturen handelt es sich um stabile Sekundärstrukturen der DNA, welche das Fortschreiten der Replikationsgabel behindern und Genominstabilität verursachen können, falls sie nicht konsequent reguliert werden (Castillo Bosch et al, 2014; Crabbe et al, 2004; Koole et al, 2014; Kruisselbrink et al, 2008; London et al, 2008; Lopes et al, 2011; Paeschke et al, 2013; Paeschke et al, 2011; Piazza et al, 2015; Piazza et al, 2010; Piazza et al, 2012; Ribeyre et al, 2009; Sabouri et al, 2014; Sarkies et al, 2012; Sarkies et al, 2010; Schiavone et al, 2014; Wu & Spies, 2016; Zimmer et al, 2016). Ziel dieser Doktorarbeit war es, neue Proteininteraktionspartner dieser Strukturen in *Saccharomyces cerevisiae* zu identifizieren und zu untersuchen, wie diese Proteine die Strukturen regulieren um Genomstabilität zu gewährleisten. Mit Hilfe eines Pulldown Assays und anschließender massenspektrometrischer Analyse wurden Mms1 und Rtt101 *in vitro* als Interaktionspartner von G-quadruplex Strukturen identifiziert. Rtt101, Mms1 und Mms22, Komponenten der Ubiquitinligase $Rtt101^{Mms1/Mms22}$, spielen eine wichtige Rolle beim Fortschreiten der Replikationsgabel, falls dieses durch Agenzien gehemmt wurde (Luke et al, 2006; Vaisica et al, 2011; Zaidi et al, 2008). Durch Chromatin-Immunpräzipitation mit anschließender Hochdurchsatzsequenzierung wurden die Bindestellen von Mms1 identifiziert. Interessanterweise hat Mms1 genomweit an G-reiche Sequenzen gebunden. Diese G-reichen Sequenzen bildeten G-quadruplex Strukturen *in vitro* aus. Die Bindung von Mms1 erfolgte unabhängig von Rtt101 und Mms22 sowie während des gesamten Zellzyklus. Außerdem kam es zu einer Verlangsamung der Replikationsgabel in *mms1* Zellen, was durch eine verstärkte Bindung der DNA Polymerase 2 nachgewiesen wurde. Ein gross chromsomal rearrangement assay zeigte, dass die Genominstabilität in *mms1* Zellen signifikant erhöht ist, wenn G-quadruplex Motive, im Vergleich zu nicht-G-reichen oder G-reichen Kontrollregionen, vorhanden sind. Zudem war die Bindung der Helikase Pif1, welche G-quadruplex Strukturen *in vitro* entwindet (Paeschke et al, 2013; Ribeyre et al, 2009; Sanders, 2010; Wallgren et al, 2016), stark reduziert, wenn Mms1 fehlte. Mit Hilfe der in dieser Doktorarbeit gewonnenen Ergebnisse, sowie mit Hilfe publizierter Daten, lässt sich ein Model postulieren, in welchem Mms1 an G-

quadruplexe bindet und somit die Bindung von Pif1 ermöglicht. Dadurch werden das Fortschreiten der Replikationsgabel und die Genomstabilität gewährleistet.

2 Introduction

2.1 Maintenance of genome stability

One of the main tasks of a living organism is to ensure the faithful duplication of the genome during mitosis and the correct propagation of the genome during meiosis in order to maintain genome integrity. Under the aspect of evolution, mutations in the genome have positive effects, if these mutations lead to a higher chance of survival. However, mutations can also lead to genome instability, being the cause of inherited diseases, cancer and premature aging (reviewed in (Aguilera & Gomez-Gonzalez, 2008)). These mutations can be of various kinds (point mutations, deletions, insertions, loss of whole chromosomes etc.) and be caused by various endogenous and exogenous sources (e.g. replication defects, failure of repair processes, reactive oxygen species, alkylating agents, ionizing radiation) (reviewed in (Aguilera & Gomez-Gonzalez, 2008; Jeggo et al, 2016)). Replication and repair defects can lead to point mutations (e.g. base substitution, insertion of a few additional nucleotides). Defective separation of chromosomes during mitosis causes chromosomal instability, which is defined by an altered chromosome number (e.g. three instead of two chromosomes). Another kind of genome instability, micro- and minisatellite instability, which can lead to repeat expansion, is for example caused by replication defects. Gross-chromosomal rearrangements (e.g. insertions, duplications, deletions) can result from replication dependent breaks, telomere fusions, *de novo* telomere addition at double-strand break (DSB) sites or failure of repair mechanisms. There are two factors which are involved in the formation of rearrangements. So called *trans* factors prevent rearrangements (e.g. proteins involved in repair and S-phase checkpoint), whereas *cis* factors are regions in the genome with a high frequency of rearrangements as for example fragile sites or highly transcribed genes (reviewed in (Aguilera & Gomez-Gonzalez, 2008)). For more information on problems during replication see 2.1.1.

Cells have developed an ingenious network of mechanisms in order to maintain genome integrity. The replication of eukaryotic DNA is initiated by the DNA polymerase Pol α. After a few nucleotides, the polymerases Pol δ and Pol ε take over. Those two polymerases have various

proofreading activities that secure a mutation rate of only one base per 10^7 bases during one round of replication (reviewed in (Heitzer & Tomlinson, 2014)). Also, a mismatch repair system can remove incorrectly-paired bases (Figure 1). Additionally, two mechanisms exist that repair damage located on one strand of DNA. The base excision repair (BER) mechanism removes defective (e.g. oxidized) bases and repairs single-strand breaks, while the nucleotide excision repair (NER) mechanism removes larger damaged sites, such as dimers or adducts (Figure 1). DNA DSBs can be repaired either by non-homologous end joining (NHEJ) or homologous recombination (HR) (Figure 1) (reviewed in (Jeggo et al, 2016)).

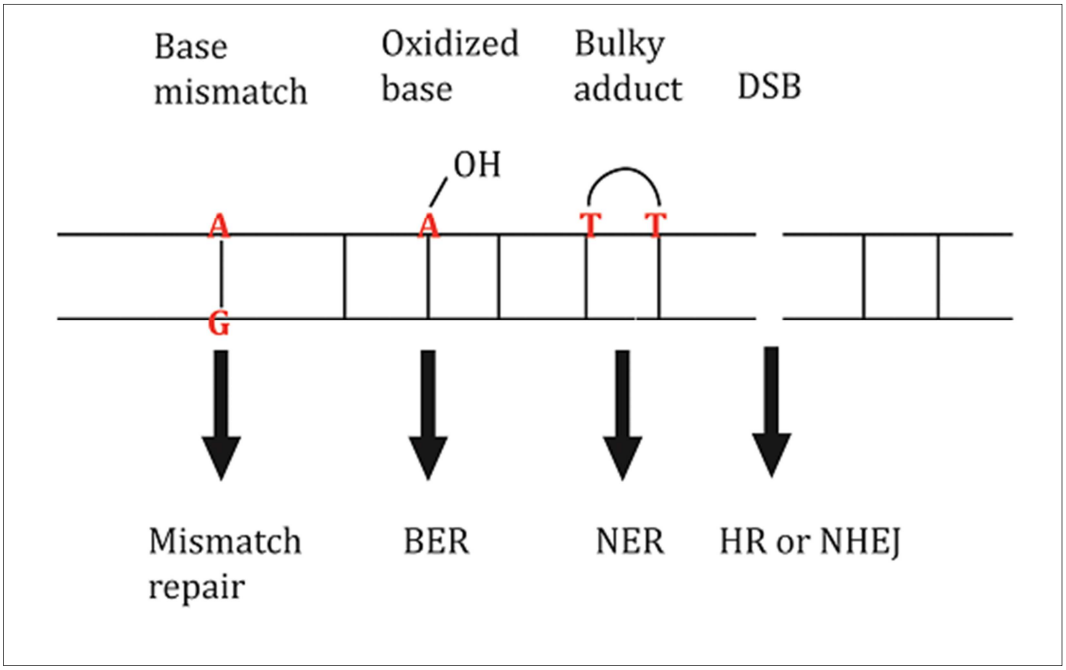


Figure 1: Some DNA repair mechanisms. Different kinds of DNA damage are repaired by different repair mechanisms. DSB (double-strand break), BER (base excision repair), NER (nucleotide excision repair), HR (homologous recombination), NHEJ (non-homologous end joining).

The HR process is the major pathway for DSB repair in yeast; however, this process only takes place in the presence of a sister chromatid. Hence, NHEJ is the alternative mechanism for DSB repair in the G1 phase. In higher eukaryotes NHEJ is the main pathway for DSB repair, although it is more error prone than HR (reviewed in (Hefferin & Tomkinson, 2005)).

Another mechanisms that is important for the maintenance of genome stability is the checkpoint response. In case of DNA damage or replication fork impediments a checkpoint can be activated. Proteins of the DNA damage response (Mec1/Tel1) recognize the damaged site and can induce apoptosis or checkpoint arrest to prevent proliferation of damaged cells or to prevent damaged cells from entering the next phase of the cell cycle. The checkpoint can be activated in G1, S or G2 phase depending on when the DNA damage occurs (reviewed in (Jeggo et al, 2016)). The Mec1 and Tel1 checkpoint kinases play an important role in the phosphorylation of various target genes (e.g. Rad9, Rad17) to induce checkpoint activation (Emili, 1998; Nakada et al, 2003a; Nakada et al, 2003b; Sanchez et al, 1996; Smolka et al, 2007; Sun et al, 1996; Sweeney et al, 2005). Rad9 and Rad53 are additional key proteins, besides Mec1 and Tel1, that play a role in G1, S and G2 checkpoint activation (Allen et al, 1994; Paulovich & Hartwell, 1995; Paulovich et al, 1997; Siede et al, 1993; Weinert & Hartwell, 1988; Weinert et al, 1994). Many checkpoint proteins are encoded by tumor suppressor genes. Hence, mutations in these genes result in defective checkpoints and uncontrolled duplication of tumor cells. All these mechanisms of faithful DNA replication, DNA repair and checkpoint regulation have to be coordinated and tightly regulated to maintain genome stability (reviewed in (Aguilera & Gomez-Gonzalez, 2008)). Cancer development is a multistep process, which is facilitated by various defective cellular processes, resulting in genetic and epigenetic instability. It is discussed that cancer is initiated by the mutation of proto-oncogenes, which then become oncogenes. Oncogenes lead to replication stress and upregulated proliferation of tumor cells. Another reason for the high proliferation of tumor cells is the downregulation of the DNA damage response. Mutations in DNA repair genes or genes that activate the DNA damage response enhance the probability to develop cancer. Also, telomerase is often activated in tumors, which prevents the senescence of tumor cells (reviewed in (Jeggo et al, 2016)). In Figure 2 pathways that can contribute to genome instability and cancer development are summarized.

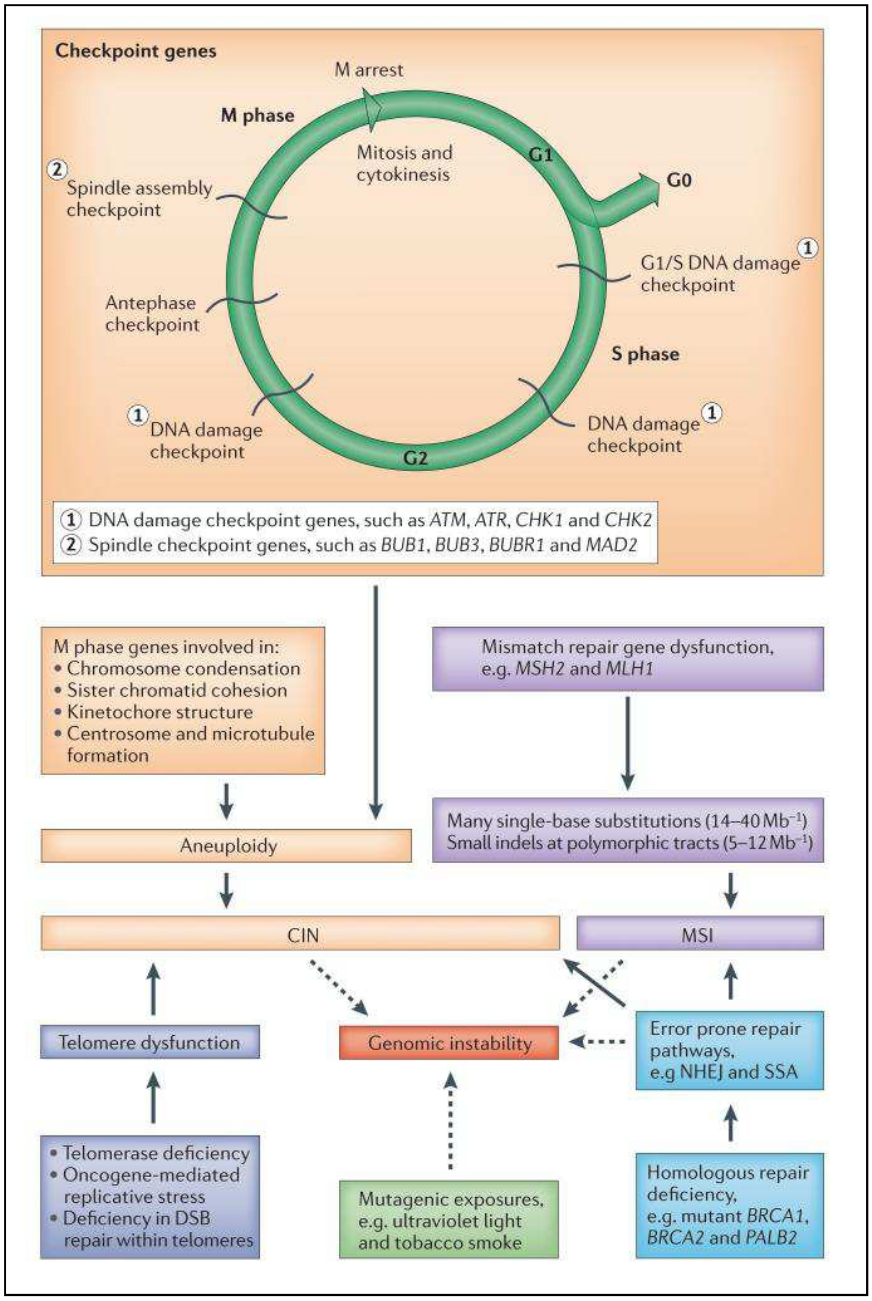


Figure 2: Causes of genome instability and cancer. Mutations in genes involved in cell cycle regulation and checkpoint activation might cause chromosomal instability (CIN). This CIN can be numeric (aneuploidy) or structural and is occurs in all types of cancer. CIN may also arise from other sources as telomere dysfunction or error-prone DNA repair mechanisms. CINs can result in genomic instability as also can microsatellite instability (MSI), which can be caused by mutations in mismatch repair (MMR) genes. Genomic instability can also be caused by exogenous factors (e.g. smoke) or defective/error-prone repair mechanisms (reviewed in (Yates & Campbell, 2012)).

2.1.1 Problems during replication fork progression and their repair

Mistakes during replication and failure of the DNA damage sensing pathway to recognize these mistakes are discussed as the main cause of cancer. Problems during replication can frequently occur at common fragile sites or highly transcribed genes, probably due to the formation of secondary structures or adducts or due to interference between DNA and RNA polymerase. At these regions, S-phase checkpoint proteins are necessary to prevent replication fork collapse and genome instability. Interestingly, common fragile sites are AT-rich and are discussed to be the cause of some chromosome rearrangements found in tumor cells (reviewed in (Aguilera & Gomez-Gonzalez, 2008)).

Replication forks can also stall at specific regions of DNA, such as ribosomal (r) DNA or centromeres (Brewer & Fangman, 1988; Greenfeder & Newlon, 1992; Linskens & Huberman, 1988). At the rDNA, replication forks that move in the opposite direction as the RNA polymerase stall at the transcription termination site (Brewer & Fangman, 1988; Linskens & Huberman, 1988). This locus is called the replication fork barrier (RFB) (Brewer & Fangman, 1988). This stall at the RFB prevents the collision of DNA and RNA polymerases moving in the opposite direction. Hence, the DNA polymerase on the rDNA moves unidirectional and in the same direction as the RNA polymerase (Brewer & Fangman, 1988; Linskens & Huberman, 1988). On both the rDNA and the centromere, replication fork stalling is due to protein-DNA complexes, with rDNA fork stalling being caused by the Fob1 protein and centromere stalling by the Cbf3 protein complex (Brewer et al, 1992; Greenfeder & Newlon, 1992; Kobayashi, 2003; Kobayashi & Horiuchi, 1996). In the case of fork pausing at the RFB it was discovered that the replisome, consisting of Mrc1, Tof1, MCM-Cdc45, GINS, polymerase α and ε, remains intact at a paused fork and recruits the helicase Rrm3 (Calzada et al, 2005). The catalytic activity of Rrm3 was necessary for fork progression at the RFB (Ivessa et al, 2000). Both Tof1 and Csm3 were required for fork pausing. The authors also found that no checkpoint was activated and that Mec1 and Rad53 are dispensable for replisome integrity as well as fork pausing and restart (Calzada et al, 2005).

In addition to that replication pausing can occur at DNA damage sites, e.g. after exposure to UV light as well as after treatment with the ribonucleotide reductase inhibitor hydroxyurea (HU) or the alkylating agent methyl methanesulphonate (MMS) (Lopes et al, 2001; Lopes et al, 2006; Shirahige et al, 1998; Tercero & Diffley, 2001). On the contrary to replication fork stalling at the RFB, fork stalling after treatment with DNA damaging agents caused the activation of the intra-S phase checkpoint as evidenced by Mec1 dependent phosphorylation of Rad53 and delayed S phase progression (Ivessa et al, 2003; Shirahige et al, 1998) (see 2.1 for further information on checkpoint activation). This checkpoint activation by Mec1 and Rad53 was required for replisome integrity and fork restart at HU-stalled forks (Lopes et al, 2001; Lopes et al, 2006) and Tof1 as well as Mrc1 were required for fork pausing and replisome integrity (Katou et al, 2003).

If the replication fork remains attached to the replisome at a stall site, replication can resume after the removal of the obstacle that caused the stall. However, if S-phase checkpoint proteins are defective or if replication is inhibited, the replication fork can collapse, which leads to the disassembly of the replisome and DNA breaks. There are many mechanisms that play a role in ensuring replication fork progression or restart of replication forks. In case of a single-strand break on the leading strand, a DSB site is created and the collapsed replication fork is restarted via break-induced replication (BIR) (Figure 3a). Blocking of replication fork progression or blocking of leading strand synthesis, e.g. through a bound protein, lead to replication fork reversal and formation of a Holliday junction (HJ) or a “chicken-foot” structure. The replication fork is then restarted either by HJ cleavage and induction of BIR or, if the lesion was removed or bypassed, by reversing the HJ and direct restart (Figure 3b). In case of a lesion on the lagging or leading strand the replication fork is not stalled and can bypass the lesion, but it leaves a gap behind. This gap is then repaired either by translesion-synthesis (TLS) or by HR (Figure 3c) (reviewed in (Aguilera & Gomez-Gonzalez, 2008).

In recent years specific DNA secondary structures, called G-quadruplex (G4) structures, have been discussed in various studies as impediments for replication fork progression (Castillo

Bosch et al, 2014; Crabbe et al, 2004; Lopes et al, 2011; Paeschke et al, 2011; Sabouri et al, 2014; Sarkies et al, 2010; Schiavone et al, 2014; Wu & Spies, 2016).

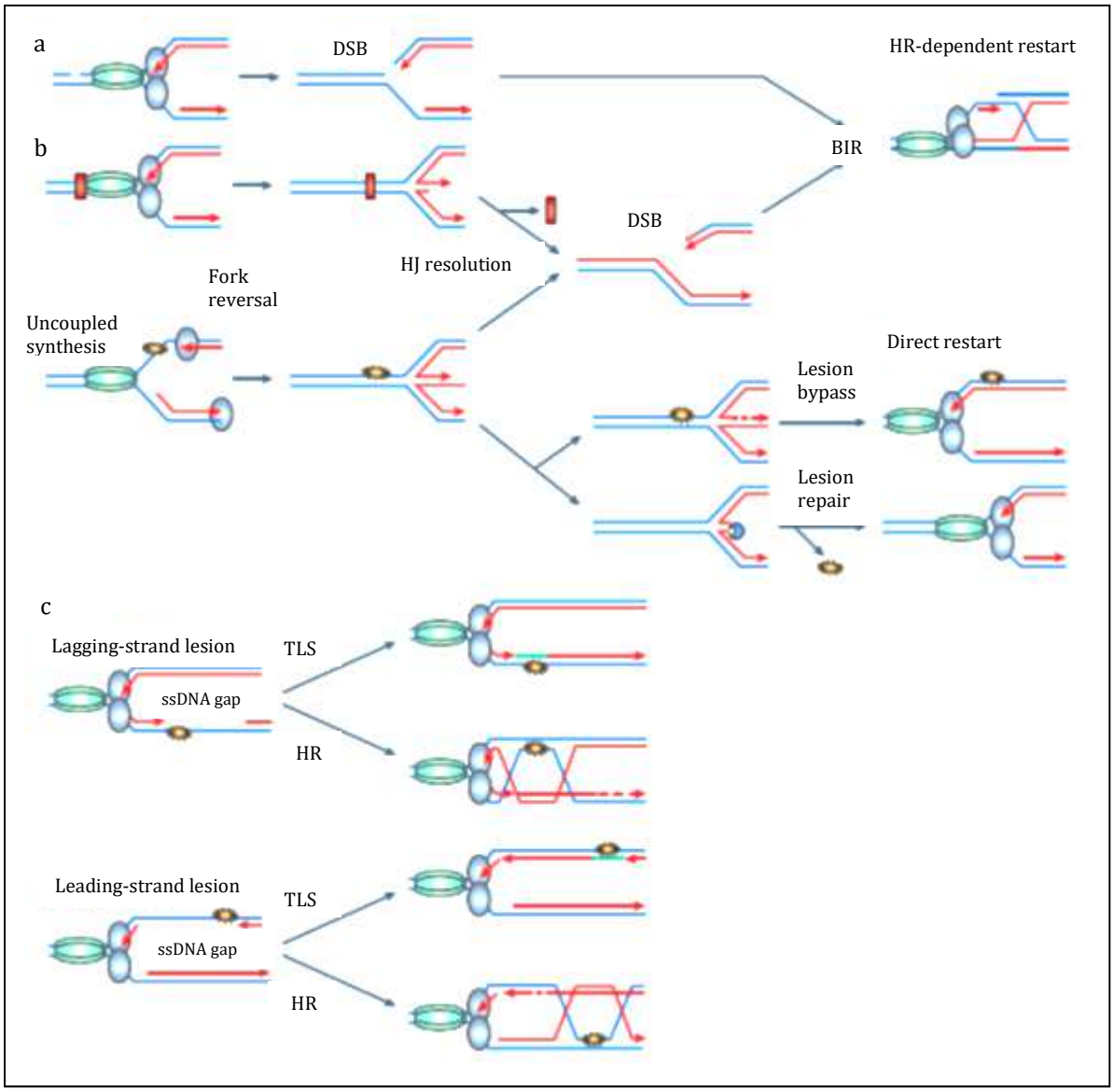


Figure 3: Problems during replication and mechanisms of fork restart. (a) A single-strand break on the leading strand results in a DSB and the replication fork is restarted by break-induced replication (BIR). (b) Blocking of replication fork progression or of leading strand synthesis lead to fork reversal and formation of a Holliday junction (HJ) or a "chicken-foot" structure. The fork is then restarted either by HJ cleavage and induction of BIR or by reversing the HJ and direct restart. (c) A lesion on either strand causes no replication fork stall. The fork can bypass the lesion, but it leaves a gap behind. This gap is the repaired either by TLS or by HR (adapted from (Aguilera & Gomez-Gonzalez, 2008)).

2.2 G-quadruplex structures

Due to the main theme of this thesis, the introduction focuses on DNA G4 structures. For some information on RNA G4 structures see 2.2.6.

2.2.1 Discovery and morphology

The discovery of G4 structures dates back until 1910 when it was demonstrated that a concentrated solution of guanylic acid can form a gel (Bang, 1910). This behavior was also demonstrated 50 years later by Gellert and colleagues (Gellert et al, 1962). Gellert and colleagues studied the characteristics of this gel in more detail and X-ray diffraction analysis suggested that the guanylic acid forms a special helix for which the authors proposed a planar structure in which four guanines are held together via hydrogen bonding (Gellert et al, 1962). This planar structure was also proposed by other authors (Sen & Gilbert, 1988; Sundquist & Klug, 1989; Williamson et al, 1989) and called G-quartet (Figure 4A) (Sundquist & Klug, 1989; Williamson et al, 1989). After that, it was elucidated that single-stranded DNA can form a four stranded structure, called "G4 DNA", under physiological salt concentrations, if the DNA sequence contains guanine-rich tracts (Sen & Gilbert, 1988). It was proposed that the G-quartets are stacked on top of each other to form the G4 structure (Figure 4B) (Sen & Gilbert, 1988; Sundquist & Klug, 1989; Williamson et al, 1989). Such a G4 structure, containing four G-quartets, was formed by the 3´ telomeric overhang (TTGGGG)n of *Tetrahymena* (Sundquist & Klug, 1989; Williamson et al, 1989). The same was observed for telomeric sequences ($(T_4G_4)n$) from *Oxytricha* (Giraldo et al, 1994; Lipps et al, 1982; Williamson et al, 1989) and from yeast (Giraldo et al, 1994). Further analysis of those structures revealed that monovalent cations can promote structure formation, with K^+ being the most and Li^+ being the less efficient, and that those ions are located in the center of the G-quartet (Chaires & Buscaglia, 2013). Crystallization and nuclear magnetic resonance (NMR) experiments confirmed the formation of the G-quartet and the structure of a G-quadruplex (Adrian et al, 2014; Chung et al, 2015; Clark et al, 2012; Kang et al, 1992; Smith & Feigon, 1993). *In vitro* G4 formation was possible in the presence of four tracts with at least two adjacent guanines separated by three loop regions with variable nucleotide

composition and length within one strand of DNA (Hazel et al, 2004; Qin et al, 2015; Williamson et al, 1989), hereafter called "G4 motif" (Figure 4C). It was also elucidated that the formation of a G4 structure is possible in duplex DNA at physiological salt concentrations *in vitro* (Deng & Braunlin, 1995; Kumar et al, 2008; Shirude et al, 2007). The competition between duplex or G4 formation is influenced by pH (Phan & Mergny, 2002), salt type and concentration (Deng & Braunlin, 1995; Phan & Mergny, 2002; Salazar et al, 1996) as well as temperature, DNA concentration (Phan & Mergny, 2002; Salazar et al, 1996) and loop length (Kumar et al, 2008).

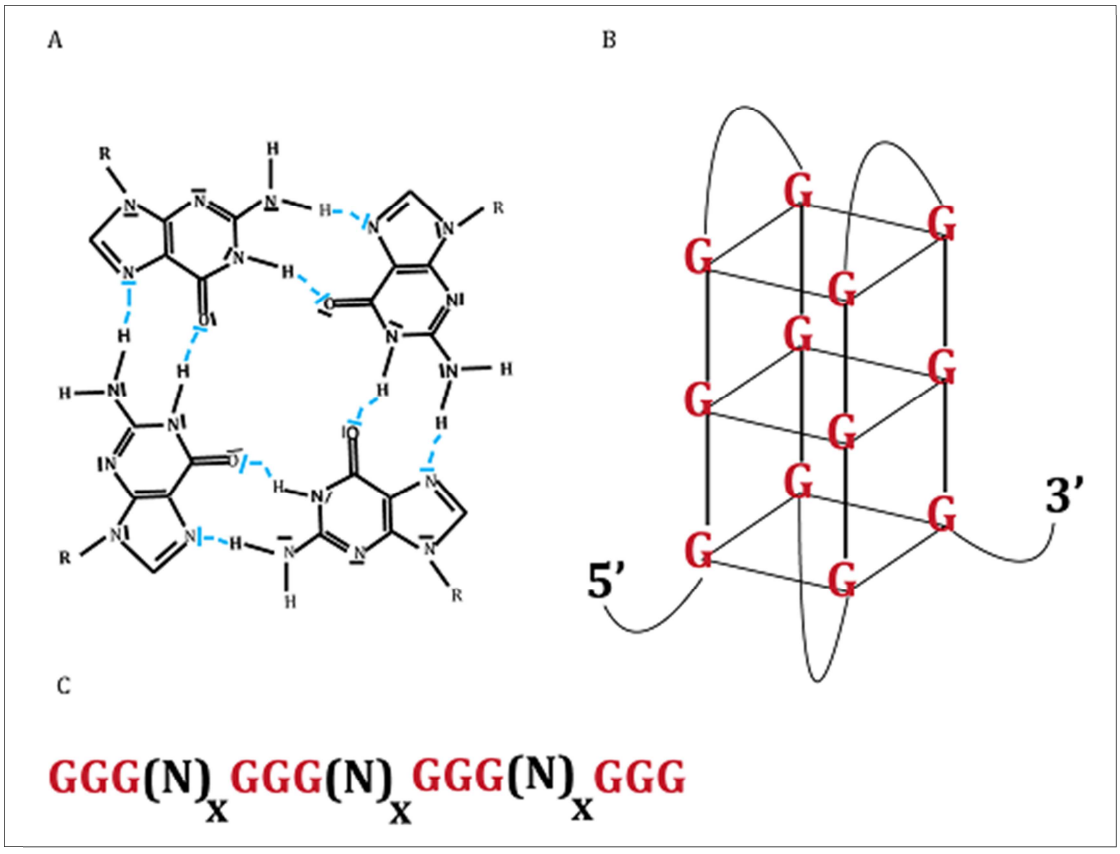


Figure 4: Composition of a G4 structure. (A) Planar G-quartet structure formed by four guanines from four different G-tracts. Hydrogen bonds are marked blue. (B) Anti-parallel intramolecular G4 structure with three G-quartets. Guanines that are located in the G-tract and form the G-quartet are shown as red "G". (C) G4 motif with four G-tracts separated by loop regions (N). X means that a variable number of nucleotides can be in the loops (modified according to (Capra et al, 2010)).

G4 DNA structures can form inter (with more than one DNA strand)- or intramolecular (within one DNA strand), the strands can run parallel, antiparallel or even mixed and the loop can run edgewise, chain-reversal or diagonal (Crnugelj et al, 2002; Kang et al, 1992; Phan et al, 2006;

Phan & Patel, 2003; Qin et al, 2015; Sen & Gilbert, 1988; Smith & Feigon, 1993; Smith et al, 1994; Sundquist & Klug, 1989).

2.2.2 Parameters that influence the stability of G4 structures

G4 structures possess a high kinetic and thermodynamic stability (Jin et al, 1992; Sundquist & Klug, 1989). A G4 structure can even be more stable than duplex DNA (Jin et al, 1992). A G4 structure with three guanines in the G-tract and a mean loop length of < two nucleotides, for example, can have melting temperatures around 82°C (Guedin et al, 2010). The number of guanines in the G-tract has an influence on the stability of the G4 structure. The more guanines in the G-tract the more stable are the G4 structures. For example, a G4 with the same loop composition had a ~20°C higher melting temperature if three instead of two guanines are located in the G-tracts (~65°C to ~46°C) (Smirnov & Shafer, 2000).

Generally, by measuring the melting temperature, it has been demonstrated that G4 structures formed in a potassium solution *in vitro* are more stable than those formed in a sodium solution, while some G4 structures with modified guanines were more stable in a sodium rather than in a potassium solution (Mekmaysy et al, 2008; Sagi et al, 2010; Skolakova et al, 2010). It is difficult to estimate how much more stable G4 structures are in potassium than in sodium solution. In one study the difference in melting temperatures was between 1 and >39°C (Guedin et al, 2010). Also, a recent study identified an unmodified higher-order G4 structure that is less stable in a potassium that in a sodium solution (Saintome et al, 2016).

The stability of a G4 structure in a potassium solution decreases as the loop length increases (Guedin et al, 2010; Hatzakis et al, 2010; Hazel et al, 2004), even though there are some exceptions (Piazza et al, 2015; Smirnov & Shafer, 2000). Interestingly, it was elucidated that, in the presence of sodium, shorter loops do not generally lead to more stable G4 structures. For example a G4 from the consensus sequence `GGGTTTGGGTTTGGGTTTGGG` had a ~8°C higher melting temperature than one from the sequence `GGGTGGGTGGGTTTGGG` (Guedin et al, 2010). Also, the kind of nucleotides in the loops has an influence on G4 stability (Hatzakis et al, 2010; Piazza et al, 2015; Smirnov & Shafer, 2000) as well as the location of the longest loop (Guedin et

al, 2010; Hatzakis et al, 2010; Piazza et al, 2015). Due to a different main theme of this thesis, the effect of loop composition is not described in more detail, but for example a G4 with a central loop of TGT had a ~7°C higher melting temperature than one with the central loop of GTT (Smirnov & Shafer, 2000).

The diverse morphologies of G4 structures make it difficult to study those structures. To my knowledge, there is no rule by which the topology of a G4 structure or its thermodynamic and kinetic stability can be predicted.

2.2.3 Location of G4 motifs in different genomes and proof of their existence *in vivo*

Due to the finding that the telomeric proteins TEBPβ (telomere binding protein in ciliates) and Rap1 (telomere binding protein in yeast) facilitate G4 structure formation *in vitro* (Fang & Cech, 1993; Giraldo & Rhodes, 1994; Paeschke et al, 2005) and that human disease related helicases as BLM and WRN (see 2.2.5.1) unwound G4 structures *in vitro* (Fry & Loeb, 1999; Kamath-Loeb et al, 2001; Sun et al, 1998), it was speculated that G4 structures might also form *in vivo* (reviewed in (Bochman et al, 2012)).

The complete sequencing of the yeast (Goffeau et al, 1996) and human genome (International Human Genome Sequencing, 2004) enabled the examination of G4 structures on a genomic level. The identification of a G4 motif and of general factors that influence the stability of a formed G4 structure (loop length, tract length) made it possible to develop computational algorithms to look for G4 motifs in different genomes (reviewed in (Huppert, 2008)).

Using computational analysis more than 370,000 G4 motifs with at least three guanines in the G-tract have been identified in the human genome (Huppert & Balasubramanian, 2005; Todd et al, 2005). A recent study identified more than 700000 G4 motifs in the human genome applying a combination of the polymerase stop assay and high-resolution sequencing. In this analysis, also G4 motifs with two guanines in the G-tract were included (Chambers et al, 2015). Using computational analysis, more than 500 G4 motifs (excluding those in telomeres and rDNA) with at least three guanines in the G-tract were identified in the *S. cerevisiae* genome (Figure 5) (Capra et al, 2010; Hershman et al, 2008) and around 500 (excluding those in telomeres and

rDNA) in the *Schizosaccharomyces pombe* genome (Sabouri et al, 2014). The numbers of G4 motifs within the *S. cerevisiae* and human genomes were higher than expected from the GC content of those genomes (Hershman et al, 2008; Huppert & Balasubramanian, 2005).

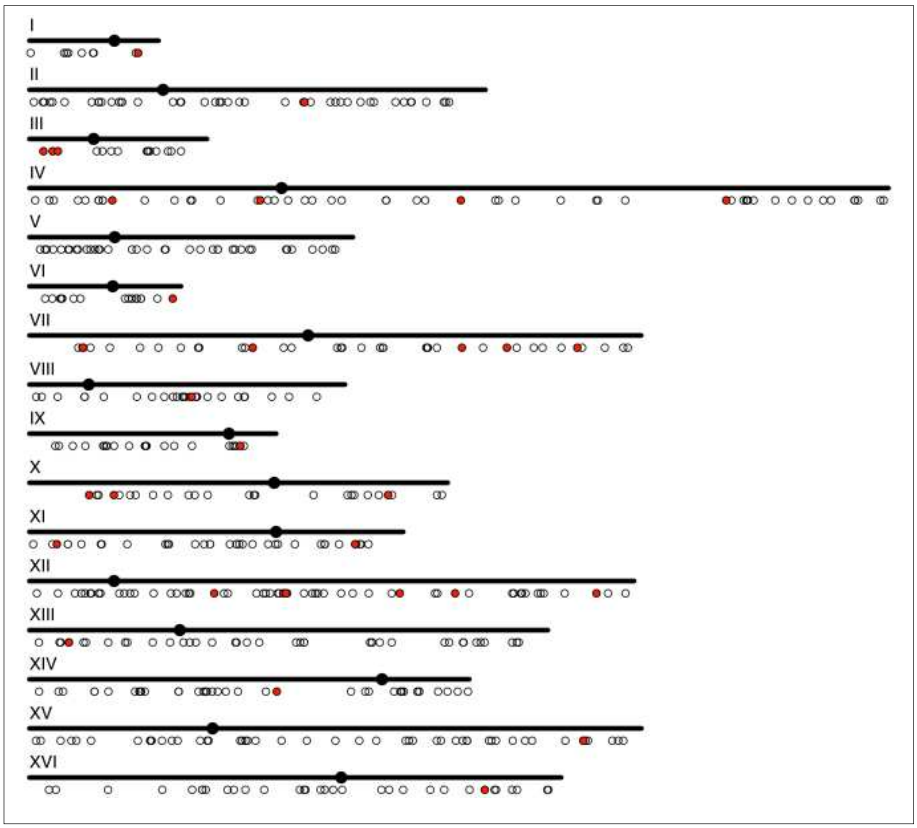


Figure 5: Location of G4 motifs within the *S. cerevisiae* genome. The G4 motifs are represented by a black unfilled circle and red filled circles mark all G4 motifs that are conserved through *sensu stricto* (Capra et al, 2010).

The position in the genome and nucleotide composition of G4 motifs are evolutionarily conserved in *S. cerevisiae* across *sensu stricto* species and among human populations (Capra et al, 2010; Nakken et al, 2009). In both human and yeast G4 motifs are significantly enriched at specific sites in the genome, such as gene promoters, open reading frames (ORF), DNA DSB sites, telomeres and rDNA. It was concluded that G4 motifs are enriched at a specific site, if their density is higher than expected from the GC content of a random sequence (Capra et al, 2010; Hershman et al, 2008; Huppert & Balasubramanian, 2005; Huppert & Balasubramanian, 2007; Sabouri et al, 2014). G4 motifs are also enriched in bacteria and similarly distributed as in yeast and humans (Rawal et al, 2006). To my knowledge the enrichment at those specific sites was

detected in all organisms tested so far. In addition to human, yeast and bacteria, G4 motifs were also highly enriched at promoters in plants, chimpanzee, mouse and rat (Mullen et al, 2010; Yadav et al, 2008). Additionally, in humans, G4 motifs are overrepresented in proto-oncogenes, while they are underrepresented in tumor suppressor genes (Eddy & Maizels, 2006; Huppert & Balasubramanian, 2007). The conservation and specific position of G4 motifs within different genomes suggests that organisms are under a selective pressure to retain the G4 motifs and that the motifs might have positive biological functions (reviewed (Bochman et al, 2012; Rhodes & Lipps, 2015)). However, up to now, it is not known how many G4 motifs also form *in vivo* and if they form, under which circumstances (reviewed in (Bochman et al, 2012)). This dynamic behavior is one reason why the examination of G4 structures *in vivo* is difficult (reviewed in (Maizels & Gray, 2013)).

The formation of G4 structures *in vivo* has been controversially discussed for a long time (reviewed in (Rhodes & Lipps, 2015)). On the one hand, G4 motifs are not randomly distributed in organisms of different species (Capra et al, 2010; Hershman et al, 2008; Huppert & Balasubramanian, 2005; Huppert & Balasubramanian, 2007; Sabouri et al, 2014) and their location is conserved among human populations and in *S. cerevisiae* across *sensu stricto* species (Capra et al, 2010; Nakken et al, 2009). This suggests that organisms are under a selective pressure to retain the G4 motifs and that the motifs might have positive biological functions (reviewed (Bochman et al, 2012; Rhodes & Lipps, 2015)). On the other hand, some scientists are skeptic whether G4 structures have regulatory functions, because their formation and dissociation are too slow for *in vivo* processes and once formed the structures might be an obstacle for replication due to their high thermal stability (reviewed in (Bochman et al, 2012)). Therefore, their formation and unwinding need to be tightly regulated (reviewed in (Rhodes & Lipps, 2015)). Indeed, it was demonstrated that the telomere proteins TEBPβ (ciliates) and Rap1 (yeast) serve as a chaperone in the formation of G4 structures *in vitro* (Fang & Cech, 1993; Giraldo & Rhodes, 1994; Paeschke et al, 2005) and various helicases unwound G4 structures *in vitro* (for more information see 2.2.5.1) (Fry & Loeb, 1999; Huber et al, 2006; Kamath-Loeb et al,

2001; London et al, 2008; Paeschke et al, 2013; Ribeyre et al, 2009; Sun et al, 1999; Sun et al, 1998; Wallgren et al, 2016). In recent years, in addition to computational analyses of G4 motif location, studies using G4 specific antibodies or G4 stabilizing ligands have contributed to validate the occurrence of G4 structures in living cells (reviewed in (Bochman et al, 2012)). The development of G4 specific probes turned out to be a difficult task. In order to specifically detect G4 structures with a probe, this probe has to efficiently discriminate between G4 structures and double-stranded DNA or other biomolecules. In addition to that, the probe has to produce a strong signal when binding to the G4 structure, while producing a low background signal. This is difficult to achieve due to the low abundance of G4 structures in most living cells. Also, it has to be ruled out that the probe induces G4 structure formation (reviewed in (Ma et al, 2015; Vummidi et al, 2013)). Furthermore, the fact that G4 structures are highly polymorphic makes it difficult to develop probes that can specifically discriminate between the different forms of G4 structures (reviewed in (Bhasikuttan & Mohanty, 2015)). This would for example be relevant for the design of anti-tumor agents or for the examination whether proteins bind to specific G4 structures. Due to the high number of G4 specific probes, due to the extensive research on them (reviewed in (Bhasikuttan & Mohanty, 2015; Ma et al, 2015)) and due to a different main theme of this thesis, they are not described in detail.

Some of the data, which supports the presence of G4 structures in living organisms, will be discussed in the next paragraphs. For information on a regulatory function of G4 structures *in vivo* see section 2.2.4.

A breakthrough in the identification of G4 structures in living cells was achieved by three different groups. The first group developed a G4 specific antibody for *Stylonychia lemnae* (Schaffitzel et al, 2001), and later, two other groups developed a G4 specific antibody for human cells (Biffi et al, 2013; Fernando et al, 2008; Henderson et al, 2014). Schaffitzel and colleagues generated two antibodies, Sty3 and Sty49, with high affinity for telomeric G4 structures in *S. lemnae*. *S. lemnae* is a single cell eukaryote that is a model organism utilized for research on telomeres. This organism contains a macronucleus that is transcriptionally active and a

micronucleus that is mostly transcriptionally silent. Around 10^8 nanochromosomes, which all contain telomeres at their ends, are present in the macronucleus. This high number of nanochromosomes results in a high concentration of telomeres in the macronucleus. The DNA of the macronucleus is replicated in a distinct region, called replication band. This distinct region and the high concentration of telomeres in the macronucleus make *S. lemnae* to an organism well suited to perform *in situ* research of telomeres during replication (reviewed in (Lipps & Rhodes, 2009)). The authors discovered that the antibody Sty3 only bound to parallel G4 structures, while the antibody Sty49 bound to parallel and antiparallel structures *in vitro*. Using Sty3, no DNA could be detected *in vivo*, whereas a signal was observed in the macronucleus using Sty49. This was the first direct evidence for G4 structure formation *in vivo*. Due to the fact that only a Sty49 signal, but no Sty3 signal, was observed *in vivo*, it was hypothesized that telomeres in ciliates fold into antiparallel G4 structures *in vivo*. Also, no staining was observed in the replication band, which suggested that G4 structures are unwound during replication and lengthening of telomeres. Although the micronucleus contains an equal concentration of DNA as the macronucleus, no staining was observed here. The authors explained this observation by the fact that the micronucleus contains less telomeres than the macronucleus. This result confirmed the specificity of Sty49 towards G4 structures and that ciliate telomeres fold into G4 structures (Schaffitzel et al, 2001).

The antibodies BG4 and 1H6 are directed against G4 structures formed in human cell lines and preferentially bound to G4 structures *in vitro*. *In vivo* experiments indicated that antibody staining was enhanced following treatment with G4 stabilizing ligands, in the absence of the G4 unwinding helicase FANCJ or after transfecting cells with G4 structures. Thus, G4 structures form *in vivo* in human cell lines (Figure 6) (Biffi et al, 2013; Henderson et al, 2014). Using the single chain antibody hf2 developed by Fernando and colleagues (Fernando et al, 2008), it was demonstrated that G4 structures can even form in duplex DNA *in vivo* (Lam et al, 2013).

Another group developed a small molecule that recognized and stabilized G4 structures, but not double-stranded DNA *in vitro*. Using this small molecule, the authors were able to isolate G4

motifs from human cells, which provided further evidence for the presence of G4 structures *in vivo* (Muller et al, 2010). Additionally, the presence of G4 structures within the human genome was demonstrated by identifying target sites of the G4 interacting small molecule pyridostatin. These binding sites also overlapped with binding sites of the G4 unwinding helicase Pif1 (Rodriguez et al, 2012; Rodriguez et al, 2008).

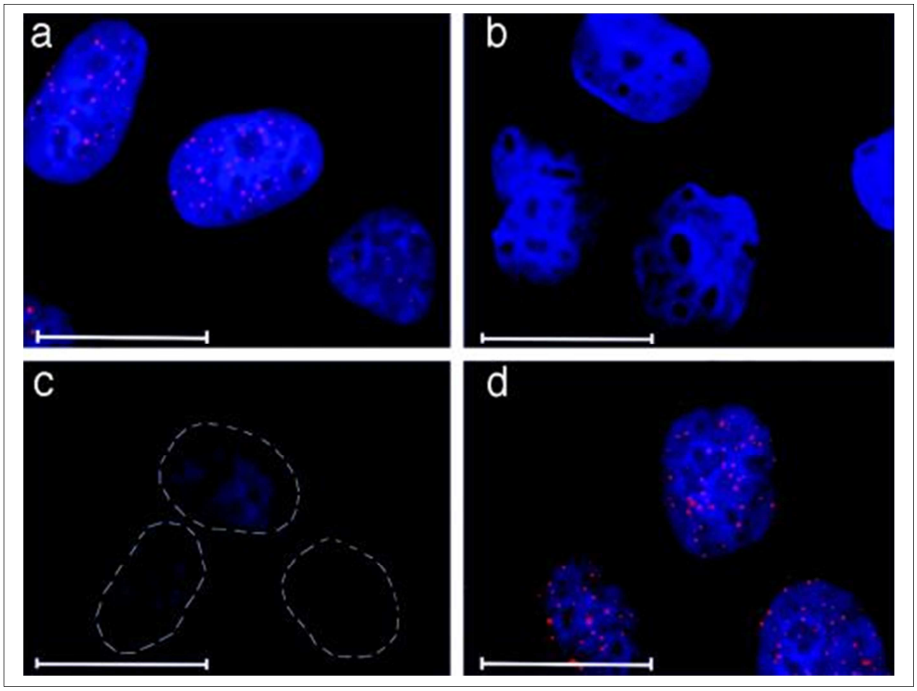


Figure 6: G4 structures stained with BG4 antibody in human cells. (a) BG4 foci in U2OS cells. (b+c) No BG4 foci detection after pre incubation of BG4 with G4 structures or after DNase I treatment respectively. (d) More BG4 foci than in (a) after transfection of G4 structures (Biffi et al, 2013).

2.2.4 Regulatory functions of G4 structures *in vivo*

Because the locations of G4 motifs in the genome are conserved among human populations and in *S. cerevisiae* across *sensu stricto* species (Capra et al, 2010; Nakken et al, 2009) and because the G4 motifs are enriched at specific sites in these genomes, as for example promoters, (Capra et al, 2010; Hershman et al, 2008; Huppert & Balasubramanian, 2005; Huppert & Balasubramanian, 2007), it was suggested that they possess regulatory functions (reviewed in (Rhodes & Lipps, 2015)). It is assumed that G4 structures form in regions where DNA is in the single-stranded conformation as at telomeres or for example during replication as well as

transcription (reviewed in (Rhodes & Lipps, 2015)) (Figure 7). However, a study revealed that G4 structure formation is also possible in duplex DNA (Lam et al, 2013).

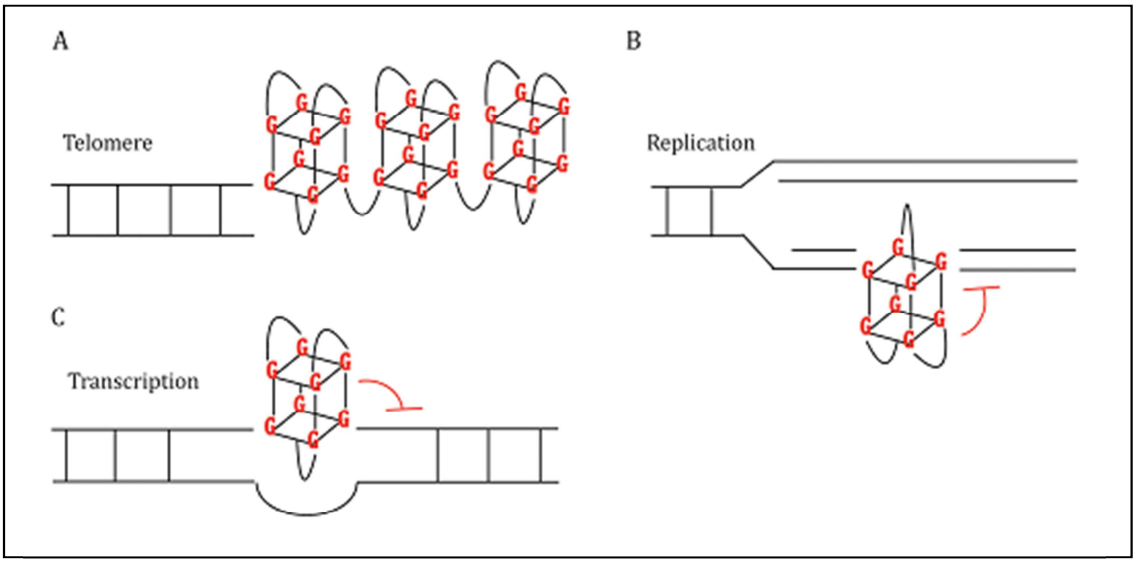


Figure 7: Possible locations of DNA G4 structures *in vivo*. (A) G4 structures could form at the end of telomeres. (B+C) G4 structures might form in single-stranded DNA during replication (B) or transcription (C). Red T-bars show that G4 structures might be an obstacle for replication or transcription. Modified according to (Rhodes & Lipps, 2015).

Indeed, multiple publications showed that G4 structures influence many biological processes such as DNA replication initiation in mammals (Valton et al, 2014), transcription in humans (Johnson et al, 2010; Lam et al, 2013; Siddiqui-Jain et al, 2002) and telomere metabolism in ciliates (Paeschke et al, 2008; Paeschke et al, 2005; Postberg et al, 2012). Some examples of regulatory functions of G4 structures are described in more detail in the next paragraphs.

2.2.4.1 G4 function at telomeres in ciliates

The first direct evidence for a regulatory function of G4 structures was achieved through the use of the *S. lemnae* Sty49 antibody (Paeschke et al, 2008; Paeschke et al, 2005; Schaffitzel et al, 2001). The authors observed that antiparallel G4 structures form at telomeres *in vivo* in a cell cycle dependent manner. During replication no G4 structures could be detected. Hence the authors hypothesized that the G4 structures are unwound during replication. Also, they found that the telomere binding proteins TEBPα and TEBPβ are required for G4 formation at telomeres *in vitro* and *in vivo* (Paeschke et al, 2005). In a further study it was discovered that

phosphorylation of TEBPβ is necessary for the recruitment of telomerase to telomeres during S-phase and that this recruitment leads to the unwinding of G4 structures through the displacement of TEBPβ (Paeschke et al, 2008). A later study specified that a RecQ-like helicase is recruited by telomerase to telomeres in *S. lemnae* and that this helicase is responsible for the unwinding of the G4 structure (Postberg et al, 2012). With the data from Paeschke and colleagues (Paeschke et al, 2008; Paeschke et al, 2005) as well as from Postberg and colleagues (Postberg et al, 2012) the authors proposed a model how telomeric G4 structures are regulated in ciliates (Figure 8). In this model, TEBPα and TEBPβ bind to telomeres in G1 and G2 phase. This leads to the formation of G4 structures (Figure 8A+B). After entry into S-phase TEBPβ is phosphorylated and recruits the telomerase-RecQ-like-helicase complex (Figure 8C+D). The RecQ-like helicase unwinds the G4 structure during replication and by this enables telomerase action (Figure 8E). TEBPβ is dephosphorylated at the end of S phase (Figure 8 F) and again binds to telomeres in G2 phase (Figure 8A) (Postberg et al, 2012).

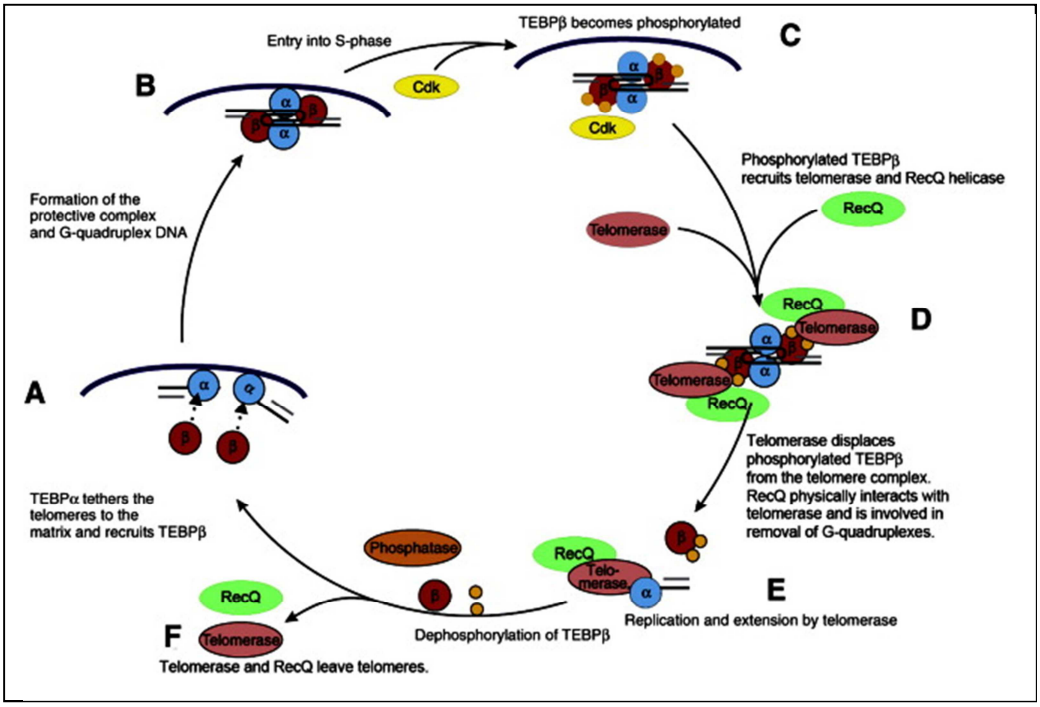


Figure 8: Model for the cell cycle dependent regulation of G4 structures at telomeres in ciliates. (A+B) TEBPα (blue) and TEBPβ (red, round) bind to telomeres in G1 and G2 phase. This leads to the formation of G4 structures. (C+D) After entry into S-phase TEBPβ is phosphorylated and recruits the telomerase-RecQ-like-helicase complex. (E) The RecQ-like helicase unwinds the G4 structure during replication and by this enables telomerase action. (F) TEBPβ is dephosphorylated at the end of S phase and again binds to telomeres in G2 phase (A) (Postberg et al, 2012).

2.2.4.2 G4 function in replication initiation

Regarding a regulatory function of G4 structures it was discovered that more than 90% of human replication origins are associated with G4 motifs. Also, it was found that G4 motifs occur in a particularly high density at early replicating origins. Therefore the authors hypothesized that G4 motifs determine the location of origins and influence their efficiency (Besnard et al, 2012). Another study in DT40 cells revealed that G4 motifs are important for replication initiation at two tested origins. If the G4 motifs were mutated, the origins were not activated. Additionally, the orientation of the G4 structure influenced the position of the actual start site (Valton et al, 2014). However, a recent study in humans contradicts these findings (Bartholdy et al, 2015). They found no correlation between G4 motifs and origin efficiency. They discovered that origins are replete with sequences of asymmetric G/C and A/T nucleotide distributions and with sequences that can form triplexes. Therefore, the authors proposed that origin activity correlates with non-B DNA structure formation in general and not only with G4 structure formation. The authors argued that Valton and colleagues did not sufficiently prove that the observed effects were due to G4 structure formation, especially because they replaced the G4 motif with AT-rich sequences that destroyed the general GC-richness of the origin (Bartholdy et al, 2015).

In mouse and Drosophila it was discovered that replication initiation sites are replete with G-rich repetitive motifs (Cayrou et al, 2011). Further analyses showed that these "origin G-rich repeated elements" (OGREs) can be used to predict origin location in mouse and Drosophila and that replication is preferentially initiated at the 3´end of the OGREs. More than 80% of the mouse and around 70% of the Drosophila replication origins contain an OGRE (Cayrou et al, 2012; Cayrou et al, 2011). Interestingly, the characteristics of OGREs were highly similar in both organisms, which suggested that a conserved mechanism determines replication initiation sites. Additionally, a higher OGRE density correlated with earlier replicating domains. Furthermore, the authors discovered that 50% of mice OGREs contain G4 motifs and that most replication origins in mice contain G4 motifs (only 43% in Drosophila). Interestingly, the G4 motifs were

specifically located on the leading strand in mice and Drosophila. Initiation of replication was observed around 280 bp downstream (160 bp downstream in Drosophila) of the OGRE/G4 motif in mice. The quite high distance between the OGRE and the initiation site suggests that a large protein complex binds in between those sites to influence replication initiation (Cayrou et al, 2012). Another study confirmed the presence of G4 motifs on the leading strand 250 bp upstream of the initiation site in around 78% of mice replication origins (Cayrou et al, 2015). Some models were proposed how G4 structures might regulate origin activation. Only a few of those will be mentioned in this thesis. For example, due to the fact that G4 motifs are located around 250 bp from the initiation site and that the origin recognition complex (ORC) was shown to bind G4 motifs in mice, it was assumed that they rather play a role in the recognition of an origin at the pre-replication complex (pre-RC) site than in the opening of the origin (Cayrou et al, 2015). It might be that binding of pre-RC induces distortion of DNA, which enables G4 structure formation or that binding of other factors induce the distortion of DNA and G4 structure formation, which then recruits the pre-RC (Valton et al, 2014). It is also imaginable that the G4 structure causes the destabilization of close double-stranded DNA sequences, which is required for the access of replication factors (Besnard et al, 2012; Cayrou et al, 2015; Valton et al, 2014). Because G4 structures were shown to impede replication fork progression (Castillo Bosch et al, 2014; Crabbe et al, 2004; Lopes et al, 2011; Paeschke et al, 2011; Sabouri et al, 2014; Sarkies et al, 2010; Schiavone et al, 2014; Wu & Spies, 2016), it was also discussed by the authors how they could at the same time promote replication. Because only 20-30% of the pre-RCs are initiated during S-phase, it is possible that the folded G4 structure inhibits initiation. But, if the G4 structure is unwound, it might enhance replication initiation at those pre-RCs that are activated. This argues for a regulatory function of G4 structures during replication initiation by regulating formation and unwinding of the G4 structure (Cayrou et al, 2015).

Strikingly, another study revealed that the lambda exonuclease (λ-exo), which is used during nascent strand sequencing (method used to discover replication origins in (Cayrou et al, 2012; Cayrou et al, 2011)), cannot efficiently digest GC-rich DNA and G4 motifs. However, efficient and

uniform digestion of DNA sequences is a prerequisite for this experiment. The authors argued that the high occurrence of G4 motifs at replication origins in former studies (Cayrou et al, 2012; Cayrou et al, 2011) might be an overestimation due to the failure of λ-exo to digest G4 motifs efficiently. When taking into account this behavior of λ-exo, the authors discovered that only ~ 36% of human origins overlap with G4 motifs, which indicates that G4 motifs do not determine origin location in general (Foulk et al, 2015).

2.2.4.3 G4 function in transcription

Due to the enrichment of G4 motifs in promoters of yeast, bacteria and different mammals (Capra et al, 2010; Hershman et al, 2008; Huppert & Balasubramanian, 2007; Rawal et al, 2006; Sabouri et al, 2014; Yadav et al, 2008) as well as due to the overrepresentation of G4 motifs in human proto-oncogenes and underrepresentation in tumor suppressor genes (Eddy & Maizels, 2006; Huppert & Balasubramanian, 2007), it was suggested that G4 motifs play a role in transcriptional regulation (reviewed in (Rhodes & Lipps, 2015)). It is assumed that G4 structures form due to negative supercoiling during transcription (Sun & Hurley, 2009). Depending on which strand G4 motifs are located, the formed structures can either enhance or repress transcription. If they are located on the non-template strand, the formation of the structure keeps the template strand in the single-stranded conformation, which promotes transcription. If the structure is formed on the transcribed strand, it can block transcription. Both enhancement and repression of transcription might also occur through interaction of proteins with G4 structures (reviewed in (Bochman et al, 2012)). Some examples of a transcriptional regulation through G4 structures are presented next.

One of the most studied examples for a role of G4 structures in transcription is the *c-MYC* locus (reviewed in (Bochman et al, 2012). This locus is very interesting, because *c-MYC* expression is enhanced in more than 80% of human cancers (Neidle, 2012). In human cell lines, the transcriptional repression of c-*MYC* was dependent on a G4 motif in the nuclease hypersensitivity element III and this repression was further enhanced in the presence of the G4 structure stabilizing ligand TMPyP4. This indicated that the G4 structure, and not the G4 motif, is

responsible for the transcriptional effect (Siddiqui-Jain et al, 2002). An *in vitro* assay confirmed that the G4 motif in the nuclease hypersensitivity element III forms a G4 structure (Simonsson et al, 1998).

Another study revealed a connection between transcriptional regulation through G4 structures and human disorders. The authors observed that genes, which are enriched for G4 motifs, are upregulated in cells from patients with Werner or Blooms syndrome (Johnson et al, 2010). This suggests that the G4 structure unwinding RecQ helicases WRN and BLM (Fry & Loeb, 1999; Huber et al, 2006; Kamath-Loeb et al, 2001; Sun et al, 1998), which are defective in the autosomal recessive disorders Werner and Blooms syndrome respectively (reviewed in (Suhasini & Brosh, 2013)) (see section 2.2.5.1 for more information), play a role in transcriptional regulation by targeting G4 structures (Johnson et al, 2010).

In a further study, expression of G4 motif containing genes changed upon treatment with the G4 structure stabilizing agent pyridostatin. In this analysis, some genes were upregulated and others downregulated upon G4 stabilization. They also confirmed the folding of G4 motifs into G4 structures *in vitro* (Lam et al, 2013).

2.2.5 G4 and genome instability

Due to their high thermal stability, G4 structures can be an obstacle for DNA transactions and hence challenge genome integrity (reviewed in (Bochman et al, 2012)). This challenge for genome integrity was observed if the G4 structures were stabilized by G4 interacting molecules (e.g. G4 stabilizing ligand Phen-DC_3) or not properly regulated (e.g. in the absence of helicases) (Cahoon & Seifert, 2009; Castillo Bosch et al, 2014; Crabbe et al, 2004; Koole et al, 2014; Kruisselbrink et al, 2008; London et al, 2008; Lopes et al, 2011; Paeschke et al, 2013; Paeschke et al, 2011; Piazza et al, 2015; Piazza et al, 2010; Piazza et al, 2012; Ribeyre et al, 2009; Rodriguez et al, 2012; Sabouri et al, 2014; Wu & Spies, 2016; Sarkies et al, 2012; Sarkies et al, 2010; Schiavone et al, 2014; Zimmer et al, 2016).

Among other proteins, two classes that were shown to be important to prevent genome instability at G4 structures are helicases (Wu & Spies, 2016; Castillo Bosch et al, 2014; Cheung et

al, 2002; Crabbe et al, 2004; Kruisselbrink et al, 2008; London et al, 2008; Lopes et al, 2011; Paeschke et al, 2013; Paeschke et al, 2011; Piazza et al, 2015; Piazza et al, 2012; Ribeyre et al, 2009; Sabouri et al, 2014; Sarkies et al, 2012) and specific DNA polymerases (Koole et al, 2014; Sarkies et al, 2012; Sarkies et al, 2010; Schiavone et al, 2014). In the next sections the function of helicases and specific DNA polymerases at G4 structures are described in more detail.

2.2.5.1 Helicases

DNA helicases unwind double-stranded DNA through ATP hydrolysis during different biological processes as for example replication and DNA repair. Various helicases from human and yeast were shown to unwind G4 structures *in vitro*. These were the helicase Pif1, the RecQ helicases Sgs1, WRN and BLM as well as the helicase FANCJ (Fry & Loeb, 1999; Huber et al, 2006; Kamath-Loeb et al, 2001; London et al, 2008; Paeschke et al, 2013; Ribeyre et al, 2009; Sun et al, 1999; Sun et al, 1998; Wallgren et al, 2016), although Sgs1 and WRN were less efficient than Pif1 (Paeschke et al, 2013). Only a few of those helicases were also investigated *in vivo*. In the next paragraphs some *in vivo* data on the helicases will be described.

Pif1 is a 5´-3´ DNA helicase (Lahaye et al, 1991) that unwound G4 structures *in vitro*. This behavior was demonstrated for *S. cerevisiae* Pif1 (Paeschke et al, 2013; Ribeyre et al, 2009), *S. pombe* Pif1 (Pfh1) (Wallgren et al, 2016) and human Pif1 (Sanders, 2010). Work in *S. cerevisiae* and *S. pombe* demonstrated that Pif1 binds to genomic G4 motifs *in vivo*, prevents replication fork stalling and supports genome stability at such sites (Lopes et al, 2011; Paeschke et al, 2013; Paeschke et al, 2011; Piazza et al, 2015; Piazza et al, 2012; Ribeyre et al, 2009; Sabouri et al, 2014). Different experiments determined that increased breaks occur at G4 motifs in the absence of Pif1 helicase in yeast (Lopes et al, 2011; Paeschke et al, 2011; Piazza et al, 2015; Ribeyre et al, 2009; Sabouri et al, 2014). Additionally, in *pif1-m2* yeast cells mutations in the G-tracts occur. These mutations disrupt the G4 motif and prevent the formation of a G4 structure (Paeschke et al, 2013; Paeschke et al, 2011) (Table 1). In *pif1-m2* mutants, only the mitochondrial isoform, not the nuclear isoform, of Pif1 is expressed (Schulz & Zakian, 1994).

Table 1: Sequences of the G4 insert after a gross chromosomal rearrangement (GCR) event in wild type (wt) and *pif1-m2* cells. Shown are the results from one colony of wild type cells and from three colonies of *pif1-m2* cells. In wild type cells, guanines which can be used to form the G4 structure are underlined. In *pif1-m2* cells, guanines that are mutated to different nucleotides are marked red (modified according to (Paeschke et al, 2013)).

Strain	Colony	Sequence (5´-3´)
wild type	1	ATAATGGGTCCTCCAAGCGGTAAAACTTACATGGGATGGTGGGGTCACATGGGTGGTC
pif1-m2	1	GGNTGGAGTCCCCCCTTCAGTAATAGGGGTTCCGAAGGATCCGGTCATTTCGTTGGTT
	2	GGTAGGAGTCTCCCTATCAGTACCAAGGGTTTCGAAGGATCC----------
	3	GGTTGTAGTCACCCCTTCACTAGTAGGGGTTCCGAAAGATCCGGTCATTTCGTCGGTN

Further experiments in yeast revealed that the G4 motif of the CEB1 tandem array only causes genome instability in the absence of Pif1 helicase if it is located on the leading strand of replication (Lopes et al, 2011). In summary the findings have allowed for a mechanistic model in which the replication machinery slows at G4 DNA structures, but it is able to pass them. The Pif1 DNA helicase recognizes G4 structures and unwinds them at the end of S phase in order to support genome stability. The remaining gap at the location of the unwound G4 structure is then repaired via a so far unknown mechanism. Due to the fact that human Pif1 also unwinds G4 structures *in vitro* (Sanders, 2010) and is most abundant at the end of S phase in human cells (Mateyak & Zakian, 2006), this model might also be valid for the situation in humans (Paeschke et al, 2011).

The RecQ helicase WRN is a 3´-5´ DNA helicase in humans (Gray et al, 1997). Mutation in WRN causes the human autosomal recessive disorder Werner syndrome. Werner syndrome is associated with premature aging, chromosomal instability and replication defects (reviewed in (Suhasini & Brosh, 2013)). Crabbe and colleagues discovered in human cells that the helicase activity of WRN is implicated in the maintenance of telomeres. In the absence of WRN, one of the two telomeric sister chromatids is lost. WRN was found to bind to telomeres during S-phase, but not in other phases. Furthermore, it was elucidated that only the lagging strand, but not the leading strand, sister telomere is lost in cells with a helicase-defective WRN gene. The authors

assumed that G4 structures formed on the lagging strand of telomeres can hamper replication, if the RecQ helicase WRN is absent and cannot unwind the structures (Crabbe et al, 2004).

FANCJ is a 5´-3´ DNA helicase in humans (London et al, 2008). This helicase is linked to fanconi anemia, an autosomal recessive disorder. Fanconi anemia is characterized by chromosomal instability, bone marrow failure and cancer (reviewed in (Suhasini & Brosh, 2013)). Deletions at G4 motifs, but not at other regions, could be detected genome-wide in *Caenorhabditis elegans* and human cell lines in the absence of FANCJ (Cheung et al, 2002; Kruisselbrink et al, 2008; London et al, 2008). In *C. elegans* it was elucidated that the 3´end of the breakpoint was located within the G4 motif region and that the deletion extended towards the flanking region 5´ to the G4 motif, but not 3´ (Cheung et al, 2002; Kruisselbrink et al, 2008). In humans, the exact position of the breakpoint could not be determined. However, the G4 motif was located rather 3´ than 5´ of the deletion (London et al, 2008). The polarity of the deletions and the fact that FANCJ unwinds G4 structures in 5´-3´-direction suggest that FANCJ unwinds G4 structures located on the lagging strand (London et al, 2008). Human FANCJ contains a specific site (two lysine residues) for the recognition of G4 structures, which is also the recognition site of the mismatch repair protein MHL1. Therefore, FANCJ cannot interact with both G4 structures and MHL1 at a time. This indicates how FANCJ differentiates between DNA repair and assistance of replication. FANCJ unfolds and refolds a G4 structure under repetitive cycles *in vitro*. Hence, FANCJ can also partially stabilize the G4 structure (Wu & Spies, 2016). Using the primer extension assay it was revealed that G4 structures impede the progression of T7 DNA polymerase *in vitro*. However, in X*enopus* egg extracts, G4 structures were fully replicated, although the replication fork stalled transiently a few nucleotides away from the G4 structure. This stalling at G4 structures was more persistent if the G4 stabilizing ligand Phen-DC_3 was added or if FANCJ-depleted X*enopus* egg extracts were used. In contrary to previous findings in yeast, humans and *C. elegans* (Cheung et al, 2002; Kruisselbrink et al, 2008; London et al, 2008; Paeschke et al, 2013; Ribeyre et al, 2009), G4 motifs were not mutated or deleted after replication, even not in the absence of FANCJ. After addition of recombinant FANCJ to FANCJ-depleted extracts, replication was

resumed, which suggests that FANCJ is required for replication at G4 structures. This function was independent of other proteins of the Fanconi anemia pathway (Castillo Bosch et al, 2014). Interestingly, the position where the replication stalled is similar to the position of deletions observed in *dog-1* (FANCJ orthologue in *C. elegans*) *C. elegans* mutants (Castillo Bosch et al, 2014; Cheung et al, 2002; Kruisselbrink et al, 2008). For more helicases involved in G4 structure processing see review by Mendoza and colleagues (Mendoza et al, 2016).

2.2.5.2 Specific polymerases

It was previously determined that the human DNA polymerase REV1, which is implicated in the TLS pathway, can disrupt G4 structures *in vitro* (Eddy et al, 2014). Also, REV1 was required for replication and epigenetic stability at G4 structures in DT40 cells. In the absence of REV1, replication efficiency of a G4 motif-containing plasmid was markedly reduced compared to wild type cells, if the G4 motif was located on the leading strand template. This was not the case if the G4 motif was located on the lagging strand. Interestingly, both the polymerase-interacting region and the catalytic activity of REV1 were involved in the replication of G4 motifs. They suggested that REV1 might incorporate a cytosine across the G4 structure, which may be destabilized by base pairing between the incorporated cytosines and the guanines in the G4 structure. To study epigenetic stability, the change in enrichment of different histone modifications at a G4 motif containing locus was measured in *rev1* cells. Deletion of REV1 for example caused the derepression of the G4 motif containing ρ-globin gene. This was due to an altered presence of histone modifications and this was dependent on a G4 motif (Sarkies et al, 2010). Furthermore, a recent study in *rev1* DT40 cells revealed that a G4 motif present on the leading strand and +3.5 kb from the transcriptional start site of the BU-1 gene causes epigenetic instability and by this repression of the gene. The authors excluded that the G4 motif as such influences transcription, because in wild type cells addition of the G4 motif +3.5 kb from the transcriptional start site did not influence *BU-1* expression. Also, if the G4 motif was inserted further away from the transcriptional start site (around 4.5 kb) in *rev1* cells, it had no influence on gene expression, but still histone modifications were lost, which indicated that the histone modifications are due

to replication stalls and not due to transcription. The authors suggested a model in which replication fork stalls at G4 structures in *rev1* DT40 cells. This stalling causes the uncoupling of both DNA helicase and DNA polymerase, leading to post-replicative gaps. At these gaps, histones without post-translational modifications will be incorporated. Hence, parental epigenetic marks are lost and gene expression is influenced (Sarkies et al, 2012; Sarkies et al, 2010; Schiavone et al, 2014). The same group further analyzed epigenetic instability at G4 motifs in DT40 cells and found that also the helicase FANCJ and the RecQ helicases WRN/BLM are, similar to REV1, important for epigenetic stability at G4 structures. They proposed a model in which FANCJ, either in cooperation with REV1 or with WRN/BLM, facilitates replication at G4 motifs and by this prevents epigenetic instability (Sarkies et al, 2012). Because FANCJ was also shown to stabilize G4 structures, Wu and coauthors extended the model of a cooperative function of FANCJ and REV1. The authors proposed a model (Figure 9) in which the replication fork is stalled at a G4 structure. FANCJ binds to the G4 structure and unfolds it in 5´-3´ direction. The resulting single-stranded DNA strand can then be used as a template for replication. Due to the finding that REV1 and FANCJ cooperate to maintain epigenetic stability at G4 structures (Sarkies et al, 2012), the authors hypothesize that the translesion synthesis polymerase REV1 might be the polymerase required to incorporate cytosines across from the unwound G4 structure. They suggest that the stabilization of the G4 structure by FANCJ is required for prolonged binding of FANCJ and recruitment of REV1. FANCJ might recruit REV1 either directly or through interaction with PCNA (Figure 9) (Wu & Spies, 2016).

Another protein that functions in the maintenance of genome integrity at G4 motifs is the polymerase Θ. In *C. elegans* this polymerase was shown to repress severe chromosomal damage at G4 motifs, especially in DOG-1 deficient cells, by inducing theta-mediated end joining. However, this process causes small deletions in the range of 50-300 bp (Koole et al, 2014).

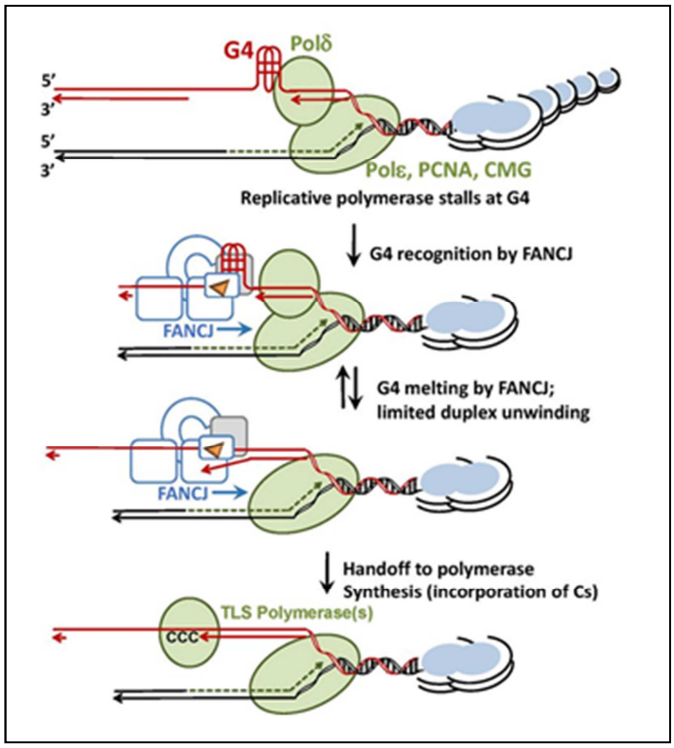


Figure 9: Model of a cooperative function of FANCJ and REV1 at G4 structures. Pol δ stalls at a G4 structure formed on the lagging strand of replication. This G4 structure is bound by FANCJ, which repetitively unfolds and refolds the structure until it recruits the translesion synthesis polymerase REV1. REV1 then incorporates cytosines across the unwound G4 structure. Note, the model is drawn with a G4 motif on the lagging strand. A similar model could be proposed for a G4 structure on the leading strand (Wu & Spies, 2016).

2.2.6 G4 structure formation in RNA

Various *in vitro* studies have confirmed the formation of G4 structures in RNA, in which strands generally run parallel (Arora et al, 2008; Beaudoin & Perreault, 2010; Kumari et al, 2007; Morris et al, 2010; Wieland & Hartig, 2009). Computational analyses in human and yeast revealed that G4 motifs are enriched at 5´ and 3´ untranslated regions (UTRs) of messenger (m) RNAs (Huppert et al, 2008; Sabouri et al, 2014). This suggested that G4 structures have a function in translation (reviewed in (Rhodes & Lipps, 2015)). Indeed, it was shown that G4 motifs influence translation in human cells (Arora et al, 2008; Beaudoin & Perreault, 2010; Morris et al, 2010). For example, it was demonstrated in human cells that the initiation of IRES-dependent translation was due to the presence of a G4 motif. This G4-motif folded into a G4 structure *in vitro* (Morris et al, 2010). In another *in vivo* study with human cells it was observed that a G4-motif located in the 5´ UTR of the mRNA of the zinc-finger protein Zic-1 inhibits translation. This

G4 motif folded into a G4 structure *in vitro* as well (Arora et al, 2008). Also, a G4 motif located in the 5´ UTR of the proto-oncogene *NRAS* folded into a G4 structure *in vitro* and repressed translation *in vitro* (Kumari et al, 2007).

2.3 Repair proteins investigated in this thesis for their function at G4 structures *in vivo*

2.3.1 Mre11

The protein Mre11 is conserved from yeast to humans (Petrini et al, 1995) and has a molecular weight of ~72 kDa (Johzuka & Ogawa, 1995). Yeast Mre11 has 3´-5´exonuclease as well as endonuclease activity (Usui et al, 1998) and is among the first proteins to be recruited to DSB sites in yeast (Lisby et al, 2004). In an immunoprecipitation with yeast lysate it was found that Mre11 bridges the interaction between Rad50 and Xrs2, forming a complex (Usui et al, 1998), called MRX complex in yeast and MRN in humans. During DSB repair the MRX complex functions in HR in human and yeast and in yeast it also functions in NHEJ (see 2.3.2) (reviewed in (Hefferin & Tomkinson, 2005; Heyer et al, 2010)).

The process of HR is schematically shown in Figure 10. The process was simplified, especially the proteins involved in the process, for this thesis, because mostly Mre11 is important for this thesis. For detailed information see review (Heyer et al, 2010). In the first step of HR, the 3´-5´ strand is resected forming a 3´ single-stranded DNA overhang (reviewed in (Symington & Gautier, 2011)). This process requires nuclease and helicase activity. In yeast, the MRX complex, Exo1, Dna2, Sae2 and Sgs1 play a role in this process (reviewed in (Heyer et al, 2010)) (Figure 10). However, studies in yeast cells revealed that the nuclease activity of Mre11 is necessary for the resection of “dirty” DSB sites (e.g. induced by irradiation) (Lisby et al, 2004; Moreau et al, 2001; Westmoreland & Resnick, 2013), but dispensable for the resection at HO-induced DSBs (“clean” ends) (Lisby et al, 2004; Llorente & Symington, 2004; Moreau et al, 1999; Moreau et al, 2001; Westmoreland & Resnick, 2013). The formed single-stranded DNA overhang is then bound by the protein RPA, which removes secondary structures on this overhang if necessary (Figure 10). After that, Rad51 filaments assemble to the overhang. These filaments are

responsible for homology search and invasion of the homologous double-stranded DNA sequence. By this, a so called D-loop is formed. After D-loop formation, three different pathways of HR are possible, depending on the exact *in vivo* situation. These pathways include BIR, synthesis-dependent strand annealing (SDSA) and double Holliday junction (dHJ). So far, not all proteins that are involved in these processes have been identified. The BIR process occurs if the other end of the DSB is missing and the D-loop becomes a replication fork. In this process heterozygosity might partially be lost. SDSA is the main pathway if both ends after a DSB are available. In this pathway, the D-loop is reversed after strand synthesis and this newly synthesized strand is annealed with the overhang from the other end. Therefore, no crossovers are produced. During the process of dHJ, a second D-loop is formed. This process might result in crossovers, if the D-loops are not resolved properly (Figure 10) (reviewed in (Heyer et al, 2010)).

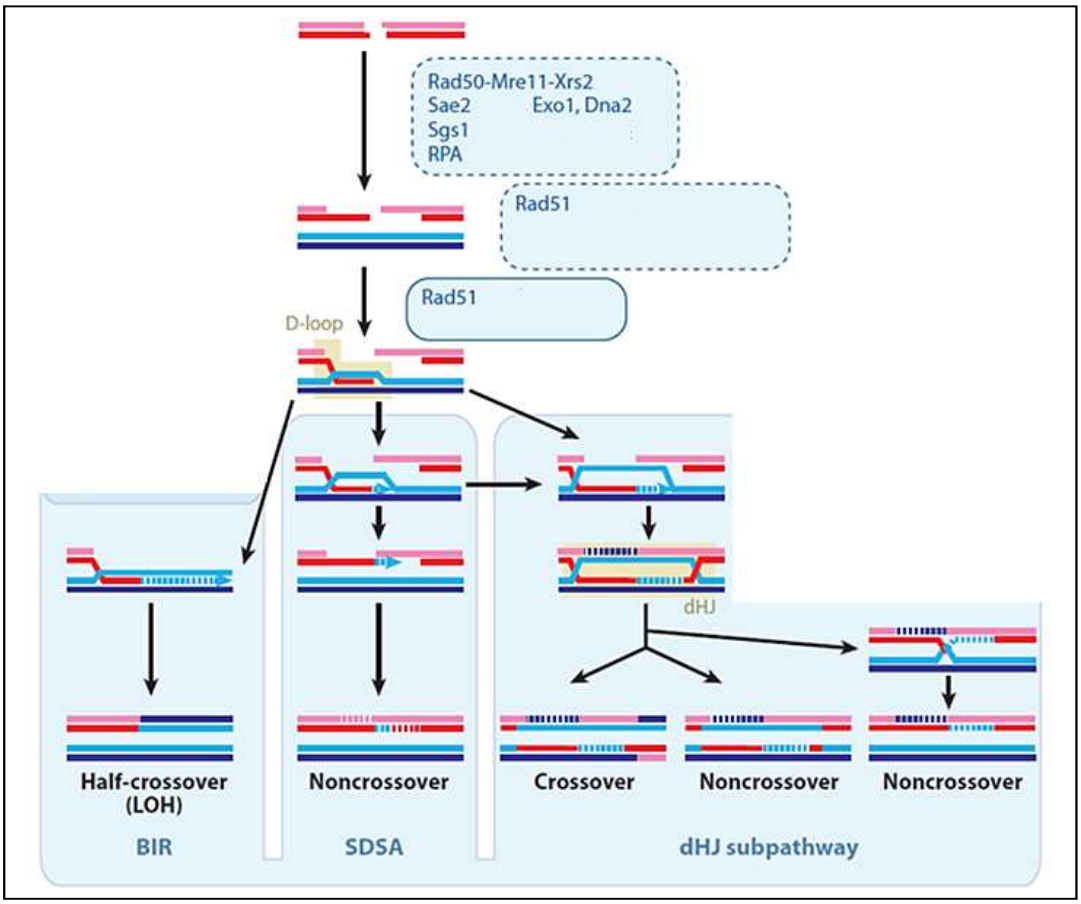


Figure 10: HR pathway in yeast. Only proteins, mentioned in the text are displayed in the figure. Of course, more proteins are involved in the process. Dotted lines resemble newly synthesized DNA strands. BIR (break-induced replication), SDSA (synthesis-dependent strand annealing), dHJ (double Holliday junction). Modified according to (Heyer et al, 2010).

All three components of the MRX complex bound preferentially to telomeric G4 structures compared to double-stranded and single-stranded telomeric DNA *in vitro* (Ghosal & Muniyappa, 2005; Ghosal & Muniyappa, 2007). Binding of the complete MRX complex was more pronounced than binding of the individual complex components (Ghosal & Muniyappa, 2005; Ghosal & Muniyappa, 2007). Additionally, Mre11 as well as the MRX complex cleaved telomeric G4 structures, single-stranded telomeric DNA and double-stranded telomeric DNA *in vitro* (Ghosal & Muniyappa, 2005; Ghosal & Muniyappa, 2007).

Due to the fact that G4 motifs are enriched at telomeres in human and yeast (Capra et al, 2010; Hershman et al, 2008; Huppert & Balasubramanian, 2005; Sabouri et al, 2014), that ciliate, human and yeast telomeres fold into G4 structures *in vitro* (Ghosal & Muniyappa, 2005; Ghosal & Muniyappa, 2007; Giraldo et al, 1994; Lipps et al, 1982; Sundquist & Klug, 1989; Williamson et al, 1989; Zaug et al, 2005) and that ciliate, human as well as yeast telomere binding proteins bind to G4 structures *in vitro* (Fang & Cech, 1993; Giraldo & Rhodes, 1994; Paeschke et al, 2005; Zaug et al, 2005), it can be envisaged that G4 structures form at telomeres in different organisms (reviewed in (Bochman et al, 2012)). In ciliates it was directly demonstrated that G4 structures form at telomeres (Paeschke et al, 2008; Paeschke et al, 2005; Postberg et al, 2012; Schaffitzel et al, 2001), while data in human cells so far only indicates the formation of G4 structures at telomeres, because of insufficient resolution during microscopy (reviewed in (Rhodes & Lipps, 2015)). However, the fact that the G4 unwinding helicase WRN (Fry & Loeb, 1999; Huber et al, 2006; Kamath-Loeb et al, 2001; Paeschke et al, 2013) is required for telomere maintenance in human cells (Crabbe et al, 2004) is one finding that highly suggests that G4 structures form at human telomeres *in vivo* (reviewed in (Rhodes & Lipps, 2015)). To my knowledge, the formation of G4 structures at yeast telomeres *in vivo* has not been demonstrated so far. However, a study revealed that the *in vitro* G4 structure cleaving protein Kem1 is required for telomere maintenance in yeast (Liu et al, 1995). Additionally, it was shown in yeast that the telomerase subunit Est1 promotes the formation of telomeric G4 structures *in vitro* and that this activity of Est1 is required for telomere maintenance (Zhang et al, 2010). Therefore, the authors assumed

that G4 structures are implicated in telomere function (Liu et al, 1995; Zhang et al, 2010). Hence, it is assumed in this thesis that G4 structures can form at telomeres in human and yeast.

In yeast, it was elucidated by chromatin-immunoprecipitation (ChIP) analysis that the MRX complex binds to telomeres in the late S phase (McGee et al, 2010; Takata et al, 2005) and that this binding is required for the association of telomere regulatory proteins (e.g. Cdc13, Est1, Mec1) at telomeres (Takata et al, 2005). Binding of MRX was enhanced at short telomeres compared to wild type length telomeres in yeast (McGee et al, 2010). In yeast, association of the telomere binding protein Cdc13 to telomeres and the catalytic activity of telomerase were independent of Mre11 (Bourns et al, 1998; Tsukamoto et al, 2001), but Mre11 was necessary for the recruitment of the telomerase subunits Est1 and Est2 to telomeric sequences (Goudsouzian et al, 2006). Knock down of either component of the MRX complex caused shortening of the G-rich telomeric overhang in human cells with active telomerase (Chai et al, 2006).

2.3.2 Ku70

Human and yeast Ku70 have a molecular weight of 70 kDa (Feldmann & Winnacker, 1993; Mimori et al, 1986). Yeast Ku70 forms a heterodimer with Ku80 (standard names in yeast are Yku70 and Yku80, but for simplification are referred to as Ku70 and Ku80 in this thesis) *in vitro* (Feldmann & Winnacker, 1993). Human Ku70 and Ku80 interacted in a yeast two hybrid assay and it was further elucidated that the C-terminal region of both proteins is required for dimerization (yeast two hybrid) and for DNA binding (*in vitro*) (Wu & Lieber, 1996). The C-terminal regions of Ku70 that are required for dimerization and DNA binding are conserved between *S. cerevisiae* and human (Wu & Lieber, 1996). The human Ku70/80 complex showed ATPase and helicase activity *in vitro* (Tuteja et al, 1994). However, structures of the Ku70 and Ku80 proteins suggest that they do not contain a helicase domain and it is not known whether the complex also possesses helicase activities *in vivo* (reviewed in (Gullo et al, 2006)). The Ku70/80 complex functions in DSB repair via NHEJ in human and yeast (reviewed in (Hefferin & Tomkinson, 2005)), but it has no function in HR as demonstrated in yeast (Tsukamoto et al,

1996). After DSB induction, MRX and Ku70/80 complexes bind immediately to DSB sites in yeast (Wu et al, 2008).

To induce NHEJ in yeast, first the Ku70/80 complex binds to the DSB site (Figure 11b). Then, the MRX complex is recruited and associates with the Ku70/80 complex (Figure 11c). The MRX and Ku70/80 complex form the end bridging complex, which bridges the two ends of the DSB (reviewed in (Hefferin & Tomkinson, 2005)). Then, the Dnl4-Lif1 ligase complex is recruited and stabilizes Ku70/80 binding (Zhang et al, 2007) (Figure 11d). After that, the endonuclease Rad27 and the polymerase Pol4 are recruited (Figure 11e). Rad27 is required for end processing and Pol4 for gap filling. In the last step, the ligase component Dnl4 ligates the broken DNA ends (Figure 11f) (reviewed in (Hefferin & Tomkinson, 2005)). Note, in Figure 11 Rad50 is drawn binding to the Ku70/80 complex, while Mre11 does not seem to bind. However, a later study demonstrated with a yeast two hybrid assay that Ku80 interacts with Dnl4 as well as Mre11 and that Xrs2 interacts with Lif1. These interactions were essential for NHEJ, but not for telomere maintenance and recombination exerted by Ku70/80 and MRX (Palmbos et al, 2005).

Ku70/80 and Dnl4-Lif1 prevent end resection and therefore HR, as shown in yeast (Zhang et al, 2007). In yeast, the absence of the Ku70/80 complex, Lif1 or Dnl4 in G1 phase promotes 5´ end resection by the MRX complex, while the absence of the Ku70/80 complex during G2 phase has no effect on the loading of the MRX complex and on the end resection (Clerici et al, 2008). This is in line with the preference of yeast cells to repair DSBs in G1 phase by NHEJ and in G2 phase by HR (reviewed in (Hefferin & Tomkinson, 2005)).

Through MS analysis human Ku70 was found to bind to the G4 structure formed in the human *KRAS* proto-oncogene, although it also bound to the double-stranded region of *KRAS* (Cogoi et al, 2008). As explained in 2.3.1 it is assumed that G4 structures form at telomeres in yeast. In yeast, it was demonstrated by ChIP analysis that Ku80 associates with telomeres throughout the cell cycle (Fisher et al, 2004) and that it binds wild type length telomeres as well as short telomeres (Sabourin et al, 2007). Yeast Ku70/80 bound to TLC1 (RNA template of telomerase) *in vitro* and

this interaction was important for telomere maintenance and de novo telomere addition *in vivo* (yeast) (Stellwagen et al, 2003).

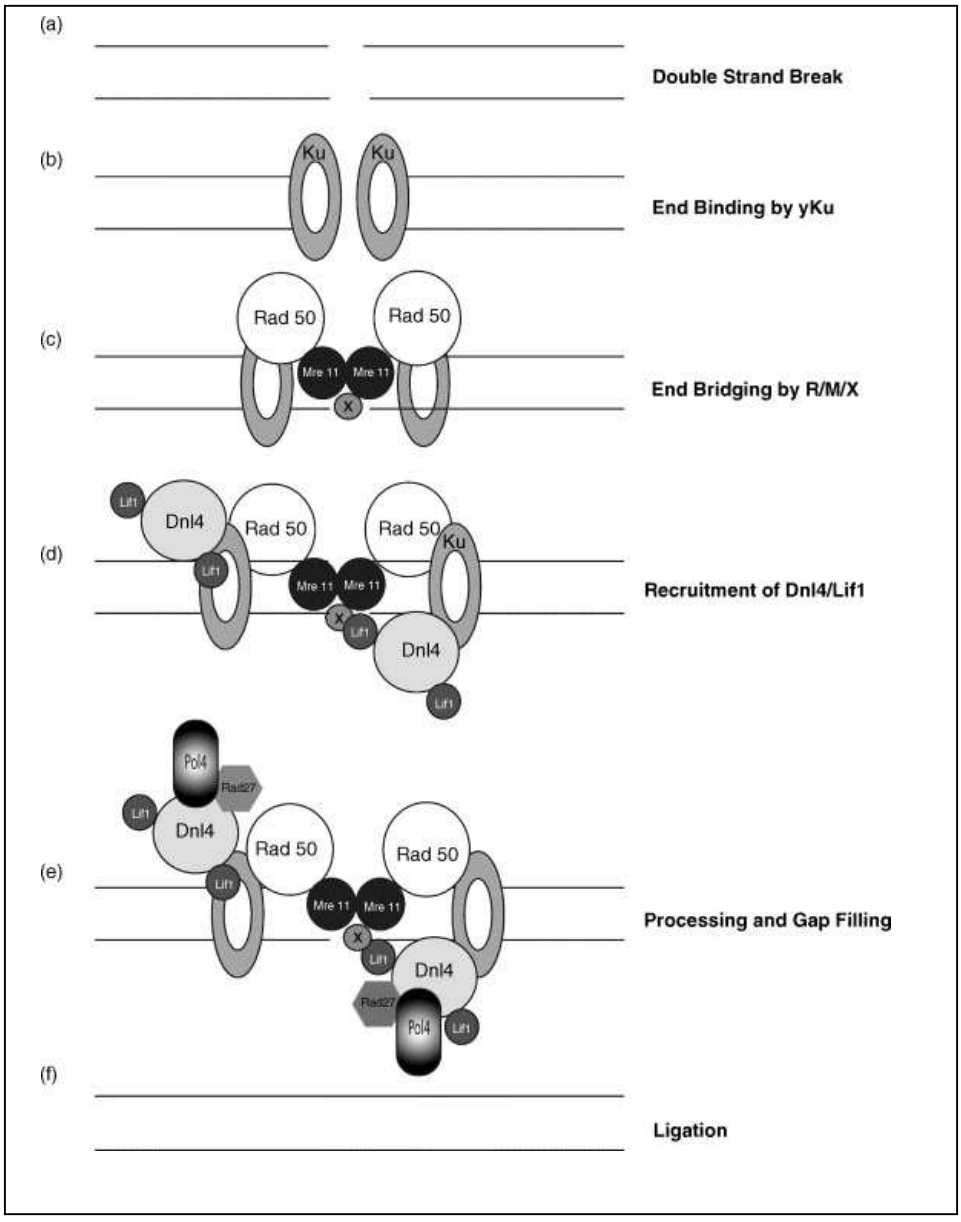


Figure 11: NHEJ in *S. cerevisiae*. (a) DSB induction. (b) Ku binds to the DSB site. (c) The MRX complex binds to Ku. Both complexes form the end-bridging complex. (d) The Dnl4-Lif1 ligase complex is recruited. (e) Processing by Rad27 endonuclease and gap-filling by polymerase Pol4. (f) Dnl4 ligase repairs the break. (x: Xrs2; black and round: Mre11; dark grey and round: Lif1; grey and hexagonial: Rad27 (adapted from in (Hefferin & Tomkinson, 2005)).

This interaction was also necessary for the binding of the telomerase components Est1 and Est2 to telomeres in yeast (Fisher et al, 2004), which suggests that Ku70/80 is required for both the recruitment of telomerase to telomeres and for their lengthening by telomerase (Fisher et al,

2004; Stellwagen et al, 2003). The Ku70/80 complex is also involved in telomere-mediated transcriptional silencing in yeast (Boulton & Jackson, 1998).

2.3.3 Tel1

The kinase Tel1 is the yeast homolog of human ATR and has a molecular weight of 322 kDa (Greenwell et al, 1995). Tel1 is one of the first proteins to bind to DSB sites in response to DNA damage (Lisby et al, 2004), where it acts as a checkpoint kinase. It induces the phosphorylation of various target proteins (e.g. H2A, Rad9, Rad53), which in turn induce checkpoint activation (Javaheri et al, 2006; Nakada et al, 2003b; Smolka et al, 2007). Studies revealed that the interaction of Tel1 with the C-terminal region of Xrs2 (MRX complex) is necessary for the activation of Tel1 and for its recruitment to DSB sites (Nakada et al, 2003a).

As mentioned in 2.3.1, it is assumed that G4 structures form at telomeres in yeast. Tel1 associated with telomeres *in vivo* (yeast) (Hector et al, 2007; McGee et al, 2010; Sabourin et al, 2007; Takata et al, 2004). This binding was preferred at short telomeres (Hector et al, 2007; McGee et al, 2010; Sabourin et al, 2007) and leads to the elongation of telomeres (Hector et al, 2007). In *TEL1* deleted yeast cells telomeres are short, but stable (Lustig & Petes, 1986; Ritchie et al, 1999). Genetic analyses in yeast of *TEL1* and *TLC1* suggest that Tel1 probably does not directly activate telomerase (Ritchie et al, 1999). Goudsouzian and colleagues established in yeast that Tel1 is required for the recruitment of the telomerase subunits Est1 and Est2 to telomeric sequences (Goudsouzian et al, 2006). In *tel1* yeast cells, Est1 and Est2 association with short telomeres in both S and G2 phase was not favored (Sabourin et al, 2007). It was shown in yeast that the binding of Cdc13 to telomeres and the catalytic activity of telomerase are independent of Tel1 (Tsukamoto et al, 2001).

2.3.4 Dot1

Dot1 has a molecular weight of ~ 66 kDa in yeast (Martino et al, 2009) and is a methyltransferase that methylates histone H3 on lysine 79 (H3K79me) in yeast and humans (Feng et al, 2002; Lacoste et al, 2002; Ng et al, 2002; van Leeuwen et al, 2002). Dot1 and the methylation are conserved from yeast to humans (Feng et al, 2002; Ng et al, 2002). As demonstrated in yeast, Dot1 is the only protein involved in this methylation i*n vivo* (van Leeuwen et al, 2002). In the presence of H3K79me, the association of Sir proteins with yeast telomeres is reduced, which suggests that Dot1 plays a role in telomeric silencing (Ng et al, 2002; van Leeuwen et al, 2002). However, another study in yeast revealed that Dot1 did not influence the binding of Sir proteins to all telomeres genome-wide (Takahashi et al, 2011). Dot1 plays a role in promoting the arrest of yeast cells in G1 and intra-S phase. Its methyltransferase activity is required for this action. Dot1 is necessary for phosphorylation of Rad9 and Rad53 to induce checkpoint activation (Giannattasio et al, 2005; Wysocki et al, 2005). The deletion of *DOT1* enhances the resistance of yeast cells towards MMS and reduces the sensitivity of repair mutants (HR, NHEJ, NER, BER) towards MMS. Just the contrary, *dot1 rev1* and *dot1 rev3* yeast double mutants are more sensitive towards MMS compared to the single mutants, which suggests that Dot1 negatively regulates TLS. The authors concluded that the decreased sensitivity of *dot1* yeast cells towards MMS is probably the consequence of increased TLS activity (Conde & San-Segundo, 2008). The hypothesis that Dot1 negatively regulates the TLS pathway is supported by the finding that *dot1 rad6* as well as *dot1 rev1* yeast double mutants were less viable following UV treatment than the single mutants. However, *dot1 rad1* double mutants were not more sensitive towards UV than the single mutants, which indicates that Dot1 plays a role in NER. This is in line with the observation that Dot1 and methylation of histone H3 on lysine 79 are required for repair of UV damage in yeast cells (Bostelman et al, 2007). Dot1 is also required for DSB repair, because H3K79me is needed for Rad9 binding to chromatin (Toh et al, 2006) and Rad9 is necessary for sister chromatid recombination (SCR) in yeast (Conde et al, 2009). The above and further genetic assays suggest that Dot1 plays an important role in coordinating different repair

mechanisms, while methylation of Histone H3 on lysine 79 seems only required for a subset of these functions (Bostelman et al, 2007; Conde & San-Segundo, 2008). Due to the fact that Rev1 is involved in G4 structure regulation (Eddy et al, 2014; Sarkies et al, 2010; Schiavone et al, 2014; Wu & Spies, 2016) and that Dot1 negatively regulates Rev1 (Bostelman et al, 2007; Conde & San-Segundo, 2008), there could be a connection of Dot1 to G4 structures, although, to my knowledge, Rev1 function at G4 structures has so far not been demonstrated in yeast, while the connection of Dot1 with Rev1 has only been shown in yeast.

2.3.5 Mgs1

Yeast Mgs1 is an AAA+ ATPase with a molecular weight of 67 kDa (Hishida et al, 2001). Mgs1 is the yeast homolog of human WHIP. WHIP is conserved from *E.coli* to human (Kawabe et al, 2001). Mgs1 interacted genetically with the helicase Sgs1, demonstrated by the fact that the *mgs1 sgs1* yeast double mutant grew slower compared to the single mutants in undamaged cells (Branzei et al, 2002b; Hishida et al, 2001; Kawabe et al, 2001) and that the double mutant arrested in G2/M phase (Branzei et al, 2002b). Additionally, the yeast *mgs1 sgs1* double mutant showed increased rates of spontaneous SCR compared to the single mutants (Branzei et al, 2002b). This and the slow growth were also observed in DT40 cells for *whip blm* (Sgs1 homolog in humans) double mutants (Hayashi et al, 2008). The two helicases, BLM and Sgs1 unwound G4 structures *in vitro* (Huber et al, 2006; Sun et al, 1999; Sun et al, 1998). Because Mgs1/WHIP seem to function in a parallel pathway to the G4 unwinding helicases Sgs1/BLM (Branzei et al, 2002b; Hayashi et al, 2008; Hishida et al, 2001; Kawabe et al, 2001), it might be possible that Mgs1 also functions at G4 structures. Further genetic analyses in yeast suggest that Mgs1 promotes progression of polymerase δ in cells with defective Rad6 pathway (TLS pathway) (Branzei et al, 2002a; Hishida et al, 2002) and that it negatively regulates the TLS pathway through its interaction with PCNA (Branzei et al, 2002a; Hishida et al, 2006).

2.4 Ubiquitin E3 ligase

Protein degradation regulates the life span of proteins. There are two ways of protein degradation in eukaryotes; lysosomal and proteasomal. The lysosome is responsible for the digestion of extracellular proteins, while the proteasome degrades intracellular proteins. It was discovered that many biological processes, as for example cell division, apoptosis, transcription and protein quality control, are regulated by protein degradation through the proteasome. The proteasome is a multiprotein complex. For proteins to be recognized by the proteasome they have to be polyubiquitinated. This is a multistep process (Figure 12). In the first step ubiquitin is activated by binding to the ubiquitin-activating enzyme (E1) under ATP hydrolysis (step 1). Then, ubiquitin is transferred to the ubiquitin-conjugating enzyme (E2) (step 2). In step three, E2 and the protein substrate bind to a specific ubiquitin-protein ligase (E3) and the activated ubiquitin is transferred to the target protein. Step 1-3 are repeated to create a polyubiquitin chain on the target protein. The polyubiquitinated protein is recognized by the proteasome (step 4). In the proteasome, the protein is degraded to peptides under ATP hydrolysis (step 5), after which the peptides are deubiquitinated (step 6) (Figure 12). There are hundreds of different E3 ligases in human cells. These E3 ligases are substrate specific, which allows for a specific and proper regulation of proteasomal degradation (reviewed in (Ciechanover, 2005)).

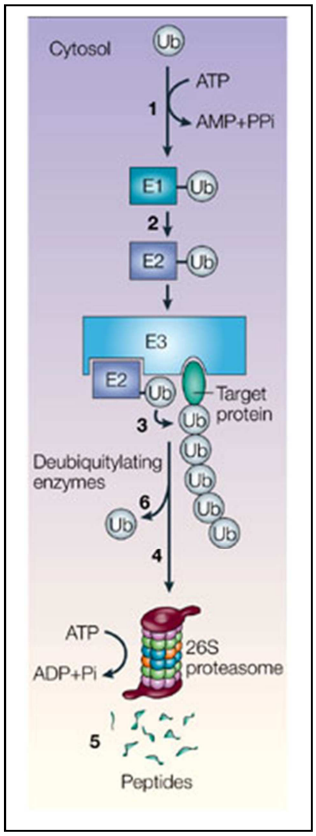


Figure 12: Proteasomal degradation. In the first step ubiquitin is activated by binding to the ubiquitin-activating enzyme (E1) under ATP hydrolysis (step 1). Then, ubiquitin is transferred to the ubiquitin-conjugating enzyme (E2) (step 2). In step three, E2 and the protein substrate bind to a specific ubiquitin-protein ligase (E3) and the activated ubiquitin is transferred to the target protein. Step 1-3 are repeated to create a polyubiquitin chain on the target protein. The polyubiquitinated protein is recognized by the proteasome (step 4). In the proteasome, the protein is degraded to peptides under ATP hydrolysis (step 5), after which the peptides are deubiquitinated (step 6) (adapted from (Ciechanover, 2005)).

2.4.1 Function of human DDB1-CUL4 ubiquitin E3 ligase

All *in vivo* experiments described in this section were performed with human cells lines. Different studies revealed an interaction between DDB1, DDB2, CUL4 and ROC1 (Angers et al, 2006; Groisman et al, 2003; Higa et al, 2006). This DDB1-DDB2-CUL4-ROC1 complex possesses E3 ubiquitin ligase activity and plays a role in global genome (GG) NER *in vivo* (Groisman et al, 2003). Crystallization experiments solved the structure of the DDB1-CUL4-ROC1 ubiquitin ligase and demonstrated that CUL4 presents the linker between DDB1 and ROC1 (Angers et al, 2006). From former studies it can be concluded that DDB2 directly binds to DDB1, while DDB1 directly binds to CUL4. Hence, DDB1 is the linker between DDB2 and CUL4 (Angers et al, 2006; Groisman et al, 2003; Higa et al, 2006). The observed interaction of DDB2 with XPA *in vitro* and *in vivo*

directly links the DDB complex (DDB1-DDB2) to NER repair. This interaction was necessary for cyclobutane pyrimidine dimer (CPD) repair (NER) *in vitro* and *in vivo* (Wakasugi et al, 2009). The DDB complex was shown to associate with various damaged DNA, such as UV induced CPDs and (6-4) photoproducts as well as abasic sites and mismatches *in vitro*. This binding capacity was decreased if DDB2 contained the same mutations that also cause the phenotype in patients suffering from xeroderma pigmentosum (XP) type E (Wittschieben et al, 2005). In patients suffering from XP, the NER pathway is disrupted due to mutations (reviewed in (Suhasini & Brosh, 2013)). Another study demonstrated that DDB1 associates with sites of UV damage *in vivo* and that this binding depends on DDB2, but not on CUL4. Due to the fact that the association of DDB1 with damaged sites fluctuates and that the amount of DDB1 binding to DNA depends on DDB2, the authors hypothesized that DDB2 plays a role in removing DDB1 from DNA. The authors also hypothesized that the whole DDB1-DDB2-CUL4-ROC1 ubiquitin ligase complex is recruited, because the association kinetics of DDB1, DDB2 and CUL4 were similar at sites of damage (Alekseev et al, 2008). Additionally depletion of DDB1 induced checkpoint activation and an enhanced occurrence of DSBs *in vivo*. Depletion of CUL4 had the same effect on checkpoint activation while depletion of DDB2 or XPA did not. This indicates that a dysfunctional NER pathway is not responsible for the arrest and that the function of DDB1 requires CUL4 as part of the ubiquitin ligase complex. The observed effects are in part probably due to false replication mechanisms, as DDB1 prevents re-replication through regulation of CDT1 degradation (Lovejoy et al, 2006). *In vitro*, the CUL4-DDB1 complex also bound to other various WD40-repeat proteins, not just DDB2. In these complexes, DDB1 presents the linker between CUL4 and WD40-repeat proteins as well. It was suggested that the WD40-repeat proteins serve as substrate specific adapters (Angers et al, 2006; Higa et al, 2006). The interaction of CUL4-DDB1 and the WD40-repeat protein L2DTL was necessary for the degradation of CDT1 *in vivo*. For CDT1 to be ubiqutinated by DDB1-CUL4 it has to be bound to PCNA (Lovejoy et al, 2006). Similar results were obtained for the cyclin-dependent kinase (CDK) inhibitor p21. Degradation of p21 in S phase and after UV treatment was abolished if DDB1 was silenced or if a mutation in

PCNA prevented it from binding to p21. Also, *in vitro* analysis with a purified CUL4-DDB1^{Cdt2} complex revealed that is ubiquitinates p21 (Nishitani et al, 2008). In addition to that the CUL4-DDB1^{Cdt2} complex was required for the degradation of the p12 subunit of Polδ in response to DNA damage *in vivo*. This degradation was even observed in S phase (Lee et al, 2014; Zhang et al, 2013). For more information on the CUL4-DDB1 ubiquitin ligase see data reviewed in (Sang et al, 2015). Additionally, CUL4-DDB1 was elucidated to function in H3 methylation *in vivo*. This function was also dependent on some WD40-repeat proteins, but not on DDB2 (Higa et al, 2006). To my knowledge it has not been unraveled if the CUL4-DDB1 ligase binds directly to DNA or what its direct binding targets are. Also, as far as I know, it is not known up to now by which mechanism the CUL4-DDB1 ligase regulates NER processes to prevent human diseases e.g. XP.

2.4.2 Function of yeast Rtt101-Mms1 ubiquitin E3 ligase

Yeast Mms1 is a homolog of the mammalian DDB1 protein and it was suggested that the yeast Rtt101^{Mms1} complex resembles the human CUL4^{DDB1} ubiquitin ligase complex (Zaidi et al, 2008). Zaidi and colleagues revealed that Mms1 interacts with Rtt101 and by this allows the interaction between Rtt101 and Ctr10 or Mms22, forming the ubiquitin ligase Rtt101$^{Mms1/Mms22}$ or Rtt101$^{Mms1/Crt10}$ (Figure 13) (note, the same was observed for the human homologues (Angers et al, 2006; Groisman et al, 2003; Higa et al, 2006)). Ctr10 and Mms22 are hypothesized to determine the function of the complex (Zaidi et al, 2008), similar to the WD40-repeat proteins (e.g. DDB2) in humans (Angers et al, 2006; Higa et al, 2006). Ctr10 was shown to repress the transcription of ribonucleotide reductase (RNR) genes RNR2 and RNR3 (Fu & Xiao, 2006). In line with this, it was elucidated that RNR3 gene expression was enhanced in *mms1* cells (Hryciw et al, 2002). The interactions between Rtt101, Mms1 and Mms22 as well as the dependency of Mms22 association with Rtt101 on Mms1 were confirmed by another study. This study also demonstrated that full-length Mms1 is necessary for Rtt101 binding, while the N-terminal region is sufficient for Mms22 binding. Mms22 binds to Mms1 via its C-terminus (Mimura et al, 2010). The interaction of Mms22 with Rtt101 only occurred in the presence of the MMS, which

causes replicative stress. It was suggested that Mms22 is the ubiquitin ligase component that associates with DNA (Figure 13) (Zaidi et al, 2008).

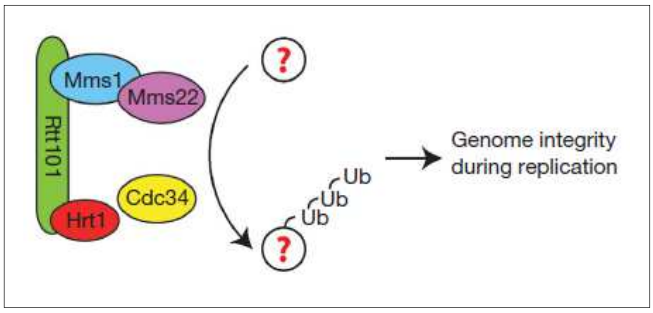


Figure 13: Mms22 interacts with Rtt101 via Mms1. Mms22 probably is the DNA binding component. The Rtt101$^{Mm1/Mms22}$ complex is involved in replication fork progression, probably leading to the ubiquitination of specific targets. Hrt1 is another ubiquitin ligase component and Cdc34 interacts with Rtt101, Mms1 and Mms22 and is likely implicated in the function (Zaidi et al, 2008).

All three components of the ligase Rtt101$^{Mms1/Mms22}$ were required for replication fork progression in the presence of replication stress (Luke et al, 2006; Vaisica et al, 2011; Zaidi et al, 2008). In *rtt101, mms1* or *mms22* cells replication was severely slowed down and more un-replicated gaps could be detected compared to wild type cells following treatment with MMS (Luke et al, 2006; Zaidi et al, 2008). Additionally, genetic analyses demonstrated a synthetic lethality of *mms1* and *mms22* mutants with *mcm10*, which is a proteins implicated in replication fork progression (Araki et al, 2003). *rtt101, mms1*, and *mms22* cells were highly sensitive to MMS, HU, or the alkaloid camptothecin (CPT), but only slightly sensitive to UV radiation (Araki et al, 2003; Hryciw et al, 2002; Luke et al, 2006). In line with this Mms1 and Mms22 were required for homologous recombination (HR) at stalled forks using either sister chromatids (unequal SCR) or homologous chromosomes (Duro et al, 2008). Also, the *mms1 sgs1* double mutant showed a reduced level of unequal SCR compared to *sgs1* cells, which indicates that Mms1 is involved in unequal SCR in *sgs1* cells (Ui et al, 2007). However, neither Rtt101 nor Mms1 nor Mms22 were involved in HR at HO-induced DSB sites or in NHEJ (Araki et al, 2003; Duro et al, 2008; Luke et al, 2006; Zaidi et al, 2008). Genetic assays indicated that Mms1 probably does not play a role in BER, NER or post replication repair (PRR) (Araki et al, 2003; Hryciw et al, 2002). The fact that Mms1 is not implicated in PRR is further supported by the

results from Zaidi and colleagues, who demonstrated that neither Rtt101 nor Mms1 were involved in the ubiquitination of PCNA (Zaidi et al, 2008), which is necessary to induce TLS (PRR) (Griffiths, 2012). Vaisica and colleagues revealed that Mms1 binds near early origins of replication if the forks are stalled by HU and that this binding is dependent on the other two ligase components, Rtt101 and Mms22 (Vaisica et al, 2011). Additionally, Rtt101 and Mms1, as well as their homologues in humans, are necessary for nucleosome assembly during replication by ubiquitination of H3K56ac (Han et al, 2013). So far it is unclear whether the ligase components bind to specific sequences or structures in the DNA.

2.5 Yeast as a model organism

In this thesis, the baker´s yeast *Saccharomyces cerevisiae* was used as a model organism. *S. cerevisiae* is often used as a model organism, because it is a quite simple single-celled eukaryote, in which genes can be easily manipulated. *S. cerevisiae* is highly used to study for example the cell cycle, aging, gene regulation and chromosome structure. One of the facts that makes *S. cerevisiae* to an extremely useful model organism is that around 6000 *S. cerevisiae* proteins, for example proteins that control cell division, also exist in most if not all eukaryotes (Lodish, 2013). Also, metabolic and regulatory mechanisms are highly conserved from yeast to humans (Sherman, 1991). Additionally, yeast cells grow easily and fast, proteins can be purified from yeast cells and meiosis can easily be studied in yeast because diploids undergo meiosis under starvation condition. Furthermore, because yeast cells can not only grow as haploids, but also as diploids it is possible to identify essential genes (Lodish, 2013; Sherman, 1991). The development of genetic assays in yeast was an important step for the examination of genes involved in genome instability (reviewed in (Aguilera & Gomez-Gonzalez, 2008)). HR is used in yeast to integrate transforming DNA into the genome, which allows for a rapid and specific replacement or modification of genes. Also, the genome of *S. cerevisiae* was completely sequenced before those of other eukaryotes, which made it a highly important organism for genomic studies. The genome of *S. cerevisiae* contains 16 chromosomes with a size from 200-2.2 Kbp and is very compact. Yeast in generally grown at 30°C and it normally has a doubling time of

90 min. Yeast haploids exist as mating type a (MATa) and mating type α (MATα). Diploid yeast cells can be created mixing MATa and MATα strains (Sherman, 1991). However, there are also limitations for the use of *S. cerevisiae* in science. Due to its simple unicellular structure it cannot be used to examine for example development or organ systems (Lodish, 2013).

3 Aim

Even if some proteins have already been shown to be important for replication and genome stability at G4 motifs *in vivo* (Castillo Bosch et al, 2014; Cheung et al, 2002; Crabbe et al, 2004; Koole et al, 2014; Kruisselbrink et al, 2008; London et al, 2008; Lopes et al, 2011; Paeschke et al, 2013; Paeschke et al, 2011; Piazza et al, 2015; Piazza et al, 2012; Ribeyre et al, 2009; Sabouri et al, 2014; Sarkies et al, 2012; Sarkies et al, 2010; Schiavone et al, 2014; Wu & Spies, 2016; Zimmer et al, 2016), not all mechanisms of G4 regulation have been unraveled. Besides that, to my knowledge, it is not known so far how G4 induced breaks are repaired, which mechanism is involved in mutating guanines in the absence of Pif1 and what happens to the mutations during the next round of replication.

The aim of this thesis was to identify novel G4 interacting proteins and to unravel their biological function at G4 motifs *in vivo*. *S. cerevisiae* was used as a model organism. Due to the fact that the goal of this thesis was to elucidate how G4-induced damage is repaired and which proteins are required for overall genome stability at G4 motifs *in vivo*, proteins that play a role in DNA repair processes were of particular interest.

In order to identify new G4 interacting proteins, a pull-down experiment with subsequent mass-spectrometry (MS) analysis was performed. ChIP was then used in order to determine binding of specific proteins to G4 motifs *in vivo*. A gross-chromosomal-rearrangement (GCR) assay was utilized to evaluate the influence of G4 interacting proteins on genome integrity. Additionally, DNA Pol2 ChIP was performed to investigate replication fork progression and γ-H2A ChIP was used to study the occurrence of DSBs at G4 motifs.

4 Methods

4.1 Strains, constructs, and media

Standard experiments with minor modifications were performed to produce all *S. cerevisiae* strains used in this thesis (Annex Table 1) (Gietz & Woods, 2002; Sikorski & Hieter, 1989; Van Driessche et al, 2005). Yeast transformation was performed according to (Gietz & Woods, 2002) with minor modifications. An overnight (O/N) culture of yeast was used to inoculate 30 ml of yeast extract peptone dextrose (YPD) medium enriched with 2% glucose. Cells were grown from an optical density (OD_{600}) of 0.2 until ~0.7 at 30°C and 200 rounds per minute (rpm). Then cells were centrifuged at 1811 *g* for 3 minutes (min) at room temperature (RT). The supernatant was removed, the cell pellet was washed with 25 ml sterile double distilled water (ddH_2O) and the suspension was again centrifuged. Washing and centrifugation were repeated one more time. In between salmon sperm DNA was boiled for 5 min and put on ice. After centrifugation, cells were resuspended in 1 ml of 100 mM lithium acetate (LiAc) and transferred to a 1.5 ml tube. After centrifugation at 20238 *g* for 15 seconds (sec) and removal of the supernatant, cells were resuspended in 400 µl 100 mM LiAc. For each transformation 50 µl of the cell suspension was transferred into a fresh tube and centrifuged at 20238 *g* for 15 sec. The supernatant was removed and the transformation mix (240 µl 50% polyethylene glycol, 36 µl 1 M LiAc, 25 µl salmon sperm DNA (10 g/l), 50 µl transforming DNA+ ddH_2O) was added to the cells. The suspension was mixed vigorously for 1 min using a VORTEX-GENIE 2. Next the cells were incubated at 30°C and 320 rpm. After 30 min the suspension was mixed with 40 µl dimethyl sulfoxide and incubated at 42°C for exactly 20 min. Then the cells were centrifuged at 20238 *g* for 15 sec and resuspended in 200 µl Tris-ethylenediaminetetraacetic acid (EDTA) (TE) (10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.0) buffer. Cells were plated on selective media.

The strains are derivatives of the W303 background (created by R. Rothstein) or of the YPH background (Sikorski & Hieter, 1989). Successful transformants were confirmed by single-colony PCR and in case of endogenous tags also by western blots.

4.2 Folding of G4 and confirmation of folding

4.2.1 Folding

The folding was performed according to (Bachrati & Hickson, 2006) with minor modifications. Oligodeoxynucleotides (Sigma) (for sequences see Annex Table 2) were solved in 1 M sodium chloride (NaCl) in a final concentration of 2 g/l. The solution was incubated at 100°C for 5 min and at 60°C for 48 h. After that the DNA was precipitated with 0.3 M sodium acetate (pH 5.2) and one volume isopropanol. The DNA was solved in the original volume ddH_2O.

4.2.2 Confirmation of folding

4.2.2.1 Circular dichroism

15-20 µg DNA in ddH_2O (total volume of 200 µl) was submitted to circular dichroism (CD) measurements at 25°C using Jasco J-810 spectropolarimeter (Jasco). The parameters used for the pull-down experiments were continuous scanning mode (200-350 nm), accumulation 10, scanning speed 100 nm/min, response 0.5 sec, band width 2 nm, data pitch 0.1 nm and later for the confirmation of folded binding motifs response 0.25 sec and data pitch 0.2 nm was used.

4.2.2.2 Native polyacrylamide gel

5.25 ml acrylamide (AA) (40%, 19:1), 0.3 ml 10% ammonium persulfate (APS), 0.03 ml tetramethylethylenediamine (TEMED), 1x Tris-borate-EDTA (TBE) buffer (pH 7.0) (89 mM Tris, 89 mM boric acid, 2 mM EDTA) and 18.42 ml ddH_2O were mixed. Combs with 1 cm wide pockets were used to avoid disruption of the G4 structure. The gel was run in 1x TBE buffer (pH 7.0) at 100 V and at RT for 1 h 45 min. The DNA was blotted on a Hybond™-N^+ (GE healthcare) membrane applying 2 mA/cm^2 for 60 min. The DNA was then stained with methylene blue (0.04% methylene blue hydrate and 0.5 M sodium acetate trihydrate).

4.3 Pull-down experiment

7000 pmol pre-folded oligodeoxynucleotides (G4 and mutated G4) were biotinylated by incubation with 0.1 mM biotin-14-dATP (Invitrogen), 1x terminal deoxynucleotidyl transferase (TdT) reaction buffer and 15 units TdT (Invitrogen) for 4 h. To remove unincorporated biotin,

the DNA was precipitated with 2.5 volumes 100% ethanol and 100 mM NaCl for 30 min at 4°C. After centrifugation (17949 *g*, 30 min, 4°C) the pellet was washed with 1 ml 70% ethanol and solved in 150 µl ddH_2O. A YM-30 column (Microcon) was used to desalt the DNA. The DNA concentration was determined on a NanoVue Plus (GE Healthcare).

The amounts described below are for each DNA sequence analyzed. Yeast proteins were isolated using the following protocol. An O/N culture was used to inoculate 50 ml YPD + 2% glucose. For this experiment, 16 x 50 ml cells were used (400 ml cells per G4/mutated G4). Cells were grown at 30°C and 200 rpm to an OD_{600} of 0.5 (approximately two doubling times). Cells were collected (1811 *g*, 5 min, 4°C) and washed with 30 ml ice-cold ddH_2O. Then cells were washed with 20 ml ice-cold lysis buffer (0.1 M HEPES pH 7.5, 0.01 M potassium acetate (KAc), 10% glycerin, 0.5% Nonidet P-40, 1 mM EDTA pH 8.0, 1 mM dithiothreitol (DTT), 1 µg/ml pepstatin A/leupeptin (LP) (Roth), 1 µg/ml aprotinin (AP) (Roth) and 0.5 mM 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF) (Roth)). Cells were resuspended in 400 µl ice-cold lysis buffer and quick-frozen in liquid nitrogen. Cells were thawed on ice and glass beads were added to 2 mm below the meniscus. Lysis was performed using Fastprep (S 6.5, TH MP, 3 x 20 sec, put on ice between runs for 1 min, MP Biomedicals). A hole was poked into the bottom of the tubes using a 26-gauge needle and the lysate was centrifuged (106 *g*, 1 min, 4°C) into a new tube. The lysates (16) were mixed, 480 µl avidin (5 g/l) (Calbiochem) was added and incubated for 30 min at 4°C on a head-over-tail rotator wheel. All subsequent steps until the dialysis were performed with modifications according to a published protocol (Jutras et al, 2012). 1600 µl Dynabeads M-270 streptavidin (Invitrogen) were washed three times with 4000 µl BS/THES buffer (22 mM Tris-HCl pH 7.5, 10 mM HEPES pH 7.5, 8.9% saccharose, 62 mM NaCl, 5 mM calcium chloride, 50 mM potassium chloride (KCl), 1 mM EDTA pH 8.0, 12% glycerin, 1 mM DTT, 1 µg/ml LP (Roth), 1 µg/ml AP (Roth) and 0.5 mM AEBSF (Roth)). The lysate was cleared by pre-incubation with the washed beads for 60 min at 4°C on the head-over-tail rotator wheel. 2 µl of the supernatant were kept as input control. During the incubation 1500 µl Dynabeads M-280 streptavidin (Invitrogen) were washed with 3750 µl 2xBW buffer (10 mM

Tris-HCl pH 7.5, 1 mM EDTA pH 8.0, 2 M NaCl, 1% 0.1 M phenylmethylsulfonyl fluoride (PMSF) (Serva)) for three times. Beads were resuspended in 1490 µl 2xBW buffer, mixed with 1500 µl biotinylated DNA (7000 pmol) and incubated for 60 min at RT on a rotator. After 60 min the DNA concentration in the supernatant was measured. If the concentration of unbound DNA in all samples was not equal, the incubation was extended for 60 min. The immobilized DNA (~3000 pmol) was washed three times with 3000 µl TE buffer. For controls, supernatants were kept and DNA concentrations were measured using the NanoVue Plus (GE Healthcare). The beads were then blocked with 1500 µl 0.1% (w/v) bovine serum albumin (BSA) in 2xBW buffer for 15 min at 4°C on a rotator. Beads were washed two times with 3750 µl BS/THES buffer and once with 3750 µl BS/THES buffer containing 5 µg DNA (random oligodeoxynucleotides). Beads were then resuspended in 1500 µl BS/THES buffer and incubated with cleared protein lysate, 50 mM KAc and 100 fold excess DNA (compared to bound bait DNA, random oligodeoxynucleotides) for 12 h at 4°C on a head-over-tail rotator wheel The supernatant was kept as a control and the beads were washed twice with 3750 µl BS/THES buffer supplemented with 5 µg PCR product containing G4 motifs. Then the beads were washed five times with 3750 µl BS/THES buffer. To elute the proteins the beads were incubated with increasing concentrations of NaCl. Per elution 600 µl elution buffer (200 mM, 400 mM, 600 mM, and 800 mM NaCl in 25 mM Tris-HCl (pH 7.5)) was added and beads were incubated at RT for 4 min on a rotator. Elution fractions were dialyzed and concentrated in a speed vac or proteins were precipitated with acetone (one volume sample: four volumes acetone, O/N -20°C). Samples were mixed with 4x LDS loading buffer (NuPAGE, Invitrogen) to a final concentration of 1x in 1x TE buffer and treated with 50 mM DTT at 95°C for 10 min. After cooling down to RT 120 mM iodacetamide was added to the sample and incubated at RT for 20 min in the dark. Next the samples were separated in 10% sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) gels and proteins were visualized by staining with SimplyBlue SafeStain (Invitrogen). Proteins were identified via MS.

4.4 Gross-chromosomal-rearrangement (GCR) assay

The GCR assay was performed according to (Putnam & Kolodner, 2010) with minor modifications. Per GCR assay seven biological replicates were inoculated in 5 ml YPD + 2% glucose and grown for 48 h at 30°C. 1.5 ml cell suspension was transferred to 1.5 ml tubes. 10 µl cell suspension from each tube was serially diluted in ddH_2O and cells were plated on YPD plates with a final dilution of $1:10^7$. The remaining 1.49 ml suspension were collected by centrifugation (20238 *g*, 30 sec, RT), resuspended in 200 µl ddH_2O, and plated on FOA/CAN selective plates. YPD plates were grown for two days at 30°C and FOA/CAN plates for seven days at 30°C. After incubation colonies were counted and the GCR rate was determined via fluctuation analysis with the program FALCOR (Hall et al, 2009). The parameters used were "calculate mutation rate: MSS maximum likelihood method", "group data by 7", "calculate 95% confidence interval", "rate: per 10^-9 mutations/generation".

4.4.1 Multiplex PCR

For one PCR reaction 1 µg genomic DNA was mixed with 0.3 µM primers KW 146+147, 0.18 µM primers KW 156+157, 0.77 µM primers KW 151+152+153+154, 0.62 µM primers KW 148+155+149/150 (see Annex Table 3 for primer specificities), 0.4 mM dNTPs, 1x polymerase buffer and 2.5 units Ex Taq polymerase (TaKaRa). The PCR was performed with these parameters: initial denaturation (95°C, 5 min), 35 cycles of denaturation (95°C, 30 sec), annealing (56°C, 30 sec), elongation (72°C, 1 min) and a final elongation (72°C, 10 min).

4.5 ChIP

4.5.1 Of asynchronous samples

For conventional ChIP experiments, 50 ml of cells were used. DNA binding proteins were crosslinked to the DNA with formaldehyde (final concentration 1%) for exactly 5 min with 160 rpm at 25°C. Immediately glycine was added to a final concentration of 125 mM. The cells were incubated for 5 min (160 rpm, 25°C) and transferred on ice. Cells were centrifuged (1811 *g*, 3 min, 4°C) and washed with 30 ml ice-cold HBS buffer (50 mM HEPES, pH 7.5, 140 mM

NaCl). The cells were washed in 20 ml ice-cold lysis buffer (50 mM HEPES, pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 1% IGEPAL CA-630 (Sigma), 0.1% sodium deoxycholate (Sigma), 1 mM PMSF (Serva)), resuspended in 400 µl lysis buffer, and quick-frozen in liquid nitrogen. Then the cells were thawed, glass beads were added to 2 mm below the meniscus and cells were lysed in the Fastprep (S 6.5, TH MP, 1.5 min, placed on ice after 1 min for 5 min, Biomedicals). A hole was poked into the tubes with a 26-gauge needle and the lysed cells were centrifuged (1 min, 106 *g*, 4°C) into a new tube. The tube with the glass beads was removed and the samples were centrifuged (20 min, 20817 *g*, 4°C). The supernatant was discarded and the pellet (inclusive crosslinked DNA layer on top of the pellet) was resuspended in 500 µl ice-cold lysis buffer. The sample was sonicated at 50% amplitude and 50% duty cycle (1 sec pulse, 1 sec pause) for 5x5 pulses to shear the DNA (Branson sonifier W250-D). The sample was put on ice after every five pulses for 4 min. The sonicated sample was centrifuged (5 min, 10000 *g*, for 4°C) and the supernatant was transferred to a low retention tube. 5 µl were put into a separate tube for input sample, 10 µl for western control input sample and 20 µl for DNA sample. The DNA sample was mixed with 80 µl TE + 1% SDS and incubated for O/N at 65°C. The DNA was extracted with phenol-chloroform, precipitated and loaded on an agarose gel to determine the shearing size.

The remaining supernatant was incubated with 8 µl c-Myc antibody (1 mg/ml, Clontech), with 4 µl HA-antibody (2 mg/ml, Santa Cruz Biotechnology), with 6 µl γ-H2A antibody (Abcam) or 6 µl H3 antibody (Abcam) for 1 h at 4°C on the head-over-tail rotator wheel. During the incubation Dynabeads Protein G (Invitrogen) were washed two times with 1 ml lysis buffer without PMSF and resuspended in the original volume in lysis buffer without PMSF. After the incubation 80 µl equilibrated beads were added to the sample and incubated for 2 h at 4°C on the head-over-tail rotator wheel. The beads were collected on a magnetic rack and the supernatant was removed. The beads were washed for exactly 4 min with 2 x 1 ml SDS buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 0.025% SDS), 1 x 1 ml Hi salt buffer (50 mM HEPES pH 7.5, 1 M NaCl, 1 mM EDTA pH 8.0), and 1 x 1 ml T/L buffer (20 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.0, 1.06% LiCl, 0.5% sodium deoxycholate (Sigma), 0.5% IGEPAL-CA-

630 (Sigma)). Finally, samples were washed two times with 1 ml T/E buffer (20 mM Tris-HCl pH 7.5, 0.1 mM EDTA pH 8.0). After removal of the last washing solution the beads were resuspended in 135 µl T/E + 1% SDS by vortexing, incubated 2 min at 65°C, vortexed again, and placed on the magnetic rack. 120 µl of the supernatant (ChIP sample) were transferred to a fresh tube and put at 65°C for O/N. 10 µl were taken as western eluate control sample. The western control samples (input + eluate) were mixed with 6x SDS loading buffer (0.3 M Tris-HCl pH 6.8, 1.2 mg bromphenol blue, 15% glycerin, 6% SDS and 168 mM DTT) to a final concentration of 1x and incubated at 95°C for 30 min. Then SDS-PAGE was performed and the proteins were blotted on a nitrocellulose membrane (GE healthcare) to control if the tagged proteins were stable during the analysis. The 5 µl input sample (see above) was mixed with 115 µl T/E + 1% SDS and incubated in the same way as the ChIP sample. After the incubation the DNA was purified using the QIAquick PCR Purification Kit (Qiagen). Immunoprecipitated DNA was analyzed by qPCR (CFX96™ and SYBR Green (172-5274) from Biorad) using the input sample as a control. The following program was used: initial denaturation (95°C, 3 min), 40 cycles of denaturation (95°C, 30 sec), annealing (52.5°C (except primers KP 230, rDNA, ARO and tRNA: 52°C), 30 sec), elongation (72°C, 30 sec). The melt curve was measured using the program 65-95°C, increment 0.5°C for 5 sec.

4.5.2 Of synchronous samples

The time point 0 min sample (G1 arrested cells, 50 ml, see 4.6.1) was crosslinked as described in 4.5.1. The remaining 450 ml synchronous cells were released from their arrest. The sample was concentrated to a few milliliters with a bottle-top filter using a water pump jet to remove the α-factor. Cells were resuspended in ~ 20 ml of the remaining media, centrifuged (2465 *g*, 2 min, RT), washed twice with 25 ml ddH_2O each and resuspended in 15.5 ml fresh YPD + 2% glucose. The cells were transferred into a plastic beaker and sonicated at 40% amplitude and 50% duty cycle for 2 x 5 pulses (Branson sonifier W250-D). For each time point 3 ml cells were pipetted into 34 ml YPD + 2% glucose + pronase (0.2 mg/ml, Sigma). The samples were incubated at 24°C and 160 rpm. Every 15 min after the addition of the pronase-containing media a sample was

crosslinked as described in 4.5.1. The last sample was taken at 75 min. Just before every sample was crosslinked, 7 ml of cells were taken for Fluorescence-activated cell sorting (FACS) analysis (see 4.6.1). After the crosslink all samples were washed with HBS and lysis buffer and quick-frozen in liquid nitrogen as described in 4.5.1. The remaining steps were performed as described in 4.5.1 with minor changes. Before addition of the antibodies, the sample was divided into 1/3 and 2/3 and filled with lysis buffer containing PMSF up to the original volume. 5 µl of both samples was kept as input sample. The 2/3 sample was incubated with 3 µl of c-Myc monoclonal antibody (1 mg/ml, Clontech) and the 1/3 sample with 1.5 µl HA-antibody (2 mg/ml, Santa Cruz Biotechnology). After incubation with the antibody, 30 µl equilibrated beads were added to the sample.

4.5.3 ChIP seq

ChIP-seq was performed as described in 4.5.1 with minor modifications. Per strain 700 ml YPD + 2% glucose were inoculated and crosslinked. For the washing steps 700 ml HBS and 200 ml lysis buffer were used. Cells were quick-frozen in liquid nitrogen in a volume of 0.5 ml. After cell lysis the pellet was resolved in 900 µl lysis buffer supplemented with protease inhibitors 1 µg/ml LP (Roth), 1 µg/ml AP (Roth), 0.5 mM AEBSF (Roth) and 1% 0.1 M PMSF (Serva). Instead of sonication a M220 Focused-ultrasonicator (Covaris) was used to fragment the DNA. The applied parameters were 75 watt, 25 duty, and 200 cycles/burst for 20 min. The sheared samples were filled up to 1.4 ml with lysis buffer and centrifuged at 10000 *g* for 15 min. After centrifugation 15 µl of the supernatant was taken as input sample, 10 µl as western input control sample and 20 µl as DNA sample. The DNA sample was used to determine the shearing size (see 4.5.1). The remaining sample was incubated with 60 µl c-Myc monoclonal antibody (1mg/ml, Clontech) for 12 h at 4°C on a head-over-tail rotator wheel. The input sample (15 µl sample + 345 µl TE + 1% SDS) was incubated for O/N at 65°C. After that, the input DNA was purified using a QlAquick PCR Purification Kit applying some minor changes. The sample was divided into three parts, mixed with five times volume of binding buffer, loaded on the column and centrifuged at RT and 2348 *g* for 1 min. 500 µl wash buffer was added to the column and the samples were centrifuged at

18407 *g* for 30 sec. Washing was repeated with 750 µl buffer. After 5 min of incubation the samples were centrifuged (18407 *g*, 30 sec). The flow through was removed and centrifugation was continued for 2 min. DNA was eluted with 40 µl ddH_2O (1 min incubation, 18407 *g*, 1 min). The elution step was repeated once. 400 µl equilibrated Dynabeads Protein G (Invitrogen) were added to the ChIP sample and the sample was incubated for 4 h on the head-over-tail rotator wheel. After incubation the beads were divided into four parts and washed according to 4.5.1. After removal of the last washing solution the beads were resuspended in 145 µl T/E + 1% SDS buffer by vortexing, incubated 2 min at 65°C, vortexed again and placed on the magnetic rack. 120 µl of the supernatant (ChIP sample) was transferred to a fresh tube and put for O/N at 65°C. 20 µl were taken as western eluate control sample. The western samples (input and eluate) were used to determine the stability of the tagged protein (see 4.5.1). The ChIP sample was purified using the QIAquick PCR Purification Kit as the input sample (see above), except the DNA was eluted with 2x30 µl ddH_2O. The concentration of input and ChIP DNA was determined via Qubit (Thermo scientific) measurements. Of the input samples 8 ng and of the ChIP samples the total volume (2-8 ng) was used for the library preparation. The library was prepared using the NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina (NEB). The adapter-ligated sample was loaded on a 2% agarose gel and DNA with a fragment size of 175-225 bp was cut out (DNA was not detectable) and purified with the MinElute Gel Extraction Kit (Qiagen). The eluted DNA was subjected to PCR enrichment. 20 µl DNA was mixed with 25 µl NEBNext High-Fidelity 2x PCR Master Mix and 2.5 µl of each primer (from the NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina). The PCR was performed with these parameters: initial denaturation (98°C, 30 sec), 17 cycles of denaturation (98°C, 10 sec), annealing (65°C, 30 sec), elongation (72°C, 30 sec) and a final elongation (72°C, 5 min). After the PCR the DNA was purified using the MinElute PCR Purification Kit (Qiagen). The following steps were performed by Elmar Wolf. The eluted DNA was analyzed with an Experion system (Bio-Rad). Samples were mixed equimolar and sequenced using an Illumina Nextseq500 sequencer. Only reads passing the internal

Illumina raw data-filter were aligned to the yeast reference genome (SACCER3) (https://genome.ucsc.edu/) with BOWTIE (Langmead et al, 2009).

4.5.3.1 Biocomputational analysis

After alignment, the number of reads was normalized to the sample with the lowest number of reads, which was performed by Elmar Wolf. Binding regions were identified by Peakcall. Peaks were called using the program "Model-based Analysis for ChIP-Seq (MACS 2.0)" with default settings, nomodel option and 180 extsize, which correlates to the minimal fragment size, determined by Experion measurements (Zhang et al, 2008). The data from input sample was used as a control. In order to define a consensus binding motif of Mms1 a MEME-based motif elicitation was performed (Bailey et al, 2009). To evaluate overlap of binding sites with G4 motifs, the published data of G4 motif distribution (at least three guanines in the G-tract (G4tract3)) within the yeast genome (Capra et al, 2010) was used first. Overlaps with G4 motifs (and later genome features) were determined by the command: bedtools window –a (file a) –b (file b) –w X. X corresponds to 400 when evaluating overlap of binding regions and to 500 when evaluating qPCR regions. This difference in windows is due to different fragment sizes of samples during ChIP-seq and conventional ChIP (see Annex Figure 1). Because the MEME-based binding motif suggests binding of Mms1 to G4 motifs with two guanines in the G-tract (G4tract2 motifs), the script developed to identify G4 motifs within the genome (Capra et al, 2010) was applied. G4 motifs with a minimal number of two guanines in the tract and maximal loop size of seven nucleotides were identified in the genome. The used reference genome sequence was from S288C that was released Jan. 2015 (http://downloads.yeastgenome.org/sequence/S288C_reference/genome_releases/?C=M;O=D). Then the bedtools window command was utilized to evaluate overlapping of binding regions and qPCR regions with the genome-wide G4 motifs. To study the overlap with genome features, the file saccharomyces_cerevisiae_R64-2-1_20150113.gff that was supplied within the genome release of S288C in Jan. 2015 was used and the locations of all genome features were grepped. Genes include all ORFs and non-protein coding RNAs. For overlap with γ-H2A binding sites,

genome-wide published data was used (Capra et al, 2010). To examine if the qPCR regions contain G4 motifs replicated on the leading and/or lagging strand, the closest ARS (bedtools closest –d –t all –a (G4 motifs) –b (ARS)) of those G4 motifs was identified and by this and strand location (Watson/Crick) it was determined whether the G4 motifs are replicated on the leading or lagging strand. The examination which Mms1 binding sites overlap with Pif1 binding sites as well as all tests for significance were performed by John A. Capra as part of collaboration.

4.6 Synchronization of yeast cells in G1, S and G2 phase

4.6.1 Synchronization in G1 phase

The synchronization was performed according to (Azvolinsky et al, 2006) with minor modifications. *bar1* deleted cells were grown in YPD + 2% glucose at 30°C and 200 rpm to an OD_{600} of 0.15 allowing at least two doublings. α-factor (Genscript) was added to a final concentration of 10 ng/ml. Cultures were put on the shaker at 24°C and 160 rpm for 3-4 h. After 3 h the formation of "shmoos" was controlled with the microscope. If at least 80% of the cells had formed "shmoos", part of the culture was taken for FACS analysis to confirm G1 arrest of cells (e.g. 7 ml culture+45.5 µl 10% NaN_3+1456 µl 0.5 M EDTA, pH 8.0). The other part was either used for ChIP analysis or protein isolation.

4.6.2 Synchronization in S and G2 phase

These synchronizations were performed as published with minor changes (Amberg et al, 2005). Yeast cells were grown in YPD + 2% glucose at 30°C and 200 rpm until OD_{600} 0.4 was reached, allowing at least two doublings. For S phase arrest HU (Sigma, final concentration 0.2 M) and for G2 arrest nocodazole (Sigma, final concentration 15 µg/ml in 1% DMSO) was added to the sample. Cells were grown at 30°C and 200 rpm for 3.5 h. After that, part of the sample was taken for FACS analysis to confirm arrest of cells (e.g. 7 ml culture+45.5 µl 10% NaN_3+1456 µl 0.5 M EDTA, pH 8.0). The other part was either used for ChIP analysis or protein isolation.

4.6.3 FACS analysis to confirm arrests (adapted from Boris Pfander)

1-2 x 10^7 cells (see 4.6.1+4.6.2) were centrifuged (RT, 20238 *g*, 3 min) and the supernatant was removed. The cells were resuspended in 1 ml 70% ethanol and 50 mM Tris-HCl pH 7.8. Then, the samples were centrifuged (RT, 20238 *g*, 3 min) and the pellet was resuspended in 1 ml 50 mM Tris-HCl pH 7.8. The centrifugation was repeated, the pellet was resuspended in 520 µl RNase solution (500 µl 50 mM Tris-HCl pH 7.8+20 µl RNase A (10 mg/ml)) and incubated for 4 h at 37°C. After centrifugation (RT, 20238 *g*, 3 min) and removal of the supernatant, the sample was resuspended in 220 µl Proteinase K solution (200 µl 50 mM Tris-HCl pH 7.8+20 µl Proteinase K (10 mg/ml)). Next, the sample was incubated for 1 h at 50°C and centrifuged (RT, 20238 *g*, 3 min). After removal of the supernatant, the pellet was resuspended in 500 µl 50 mM Tris-HCl pH 7.8. Then, the samples were sonicated at 30% amplitude and 50% duty cycle (1 sec pulse, 1 sec pause) for five pulses (Branson sonifier W250-D). After sonication, 200 µl of the sample was mixed with 1000 µl SYTOX solution (1.5 ml 50 mM Tris-HCl pH 7.8+0.5 µl SYTOX green (Thermo Fisher Scientific)). The sample was then measured by FACS using FACS Canto II (Becton Dickinson) using the parameters FSC Log 200V, SSC Log 300V, GFP Lin 416V, FSC threshold 5000, flowrate high and record 10000 events.

4.7 Detection of proteins that bind to G4 *in vitro*

4.7.1 Native PAGE

Yeast protein lysate was prepared according to ChIP protocol (4.5.1), except Fastprep was only performed for 1 min. After centrifugation lysates were mixed with 10x ficoll loading buffer (25% ficoll, 0.1 M Tris-HCl, pH 7.4, 0.1 M EDTA, pH 8.0, 0.25-0.5% orange G and xylene cyanol) to a final concentration of 1x and loaded on a 10% native polyacrylamide gel (3.33 ml AA (30%, 37.5:1), 0.1 ml APS, 0.004 ml TEMED, 2 ml 5x TBE pH 7.5, 4.57 ml ddH_2O). One gel was prepared for basic and one for acidic proteins. For the electrophoresis of basic proteins the power was reversed. The gels were run in 1x TBE pH 7.5 at 4°C applying 80 V for 2 h. Proteins were then blotted onto a PVDF membrane at RT. For basic proteins, the membrane was put on top of the gel, while for acidic proteins the gel was on top of the membrane. The blot was performed using

a native blotting buffer (48 mM Tris, 39 mM glycine, 0.0375% SDS (Dunn & International Electrophoresis Society, 1986)) for 1 h at 0.8 mA/cm^2. Membranes were cut lane wise and blocked with random oligodeoxynucleotides (100 µg/ml) in 1x TBE pH 7.5 for 45 min 4°C. One membrane was incubated with biotinylated G4, biotinylated mutated G4 (each ca. 500 pmol) or no DNA in 1x TBE pH 7.5 for O/N at 4°C. For sequences of used DNA see Annex Table 2. For the biotinylation 1000 pmol DNA, 1x TdT reaction buffer, 0.1 mM biotin-14-dATP and 2 units TdT were mixed, incubated at 37°C for 4 h and purified with a G25 column according to manufactures instructions. After incubation the membranes were washed three times for 10 min with 1x TBE pH 7.5, blocked with 2% milk powder in 1x TBE pH 7.5 and then incubated with anti-biotin (anti-biotin peroxidase conjugate, Sigma, 1:4000) antibody in 1x TBE pH 7.5 supplemented with 2% milk powder for 3 h at 4°C. The membrane was washed again three times for 10 min with 1x TBE pH 7.5. DNA was detected via chemiluminiscence.

4.7.2 SDS PAGE and renaturing

Yeast protein lysate was prepared according to ChIP protocol (see 4.5.1), except Fastprep was only performed for 1 min. The lysate was mixed with SDS loading buffer and incubated for 5 min at 95°C. A 10% SDS-polyacrylamide gel was loaded with the proteins and run for 2.5 h at RT first applying 80 V and later 130 V. Separated proteins were blotted on a nitrocellulose membrane. The membrane was cut lanes wise. On two lanes proteins were refolded with renaturing buffer 1, on other two lanes with renaturing buffer 2 and another two lanes were not renatured. Those lanes that were not renatured were blocked with 100 µg/ml random oligodeoxynucleotides in 1x TBE pH 7.5 for 45 min at 4°C. All subsequent steps after this blocking, starting with washing, were performed as described below. The membranes that were refolded using buffer 1 (4M urea) were incubated for O/N at 4°C. Then the urea concentration was reduced stepwise to 150 mM. The membranes that were refolded using buffer 2 (50 mM HEPES pH 7.0, 100 mM KAc, 5 mM magnesium acetate, 3 mM DTT, 0.3% Tween-20, 1% BSA) were incubated for 48h at 4°C as described in (Coutavas et al, 1993). After refolding, the membranes were blocked with 100 µg/ml random oligodeoxynucleotides in 1x TBE pH 7.5 for

45 min at 4°C prior to incubation with biotinylated bait (G4 or mutated G4, 400 pmol each, see 4.7.1 for biotinylation and Annex Table 2 for sequences) for O/N at 4°C in 150 mM urea or buffer 2 respectively. Washing and blocking with milk powder was performed according to section 4.7.1. Incubation with anti-biotin (anti-biotin peroxidase conjugate, Sigma, 1:2000) antibody was performed at 4°C for 4 h. DNA was detected via chemiluminiscence.

4.8 Western analysis

Proteins for western analysis were isolated according to a protocol by Marco Foiani (Cotterill, 1999). Western analysis was performed according to standard protocols. As primary antibodies c-Myc (Clontech), HA (Santa Cruz Biotechnology), γ-H2A (Abcam), H3 (Abcam), anti-biotin peroxidase conjugate (Sigma) or Hsp60 (Abcam) were used according to manufactures protocol. Hsp60 served as a reference protein. As secondary antibody HRP-coupled secondary antibody (Santa Cruz Biotechnology) were used. Proteins were detected by chemiluminescence. Quantification was performed using Image Lab (Biorad).

4.9 Co-immunoprecipitation of Mms1

Protein lysate was prepared from Mms1-Myc, Mms1-Myc *pif1-m2* and *pif1-m2* cells according to 4.3 from 14 x 50 ml culture each. Cells were lysed using Fastprep (S 6.5, TH MP, 4 x 20 sec, put on ice between runs for 1 min, Biomedicals). A hole was poked into the bottom of the tubes using a 26-gauge needle and the lysate was centrifuged (106 *g*, 1 min, 4°C) into a new tube. The supernatant was then incubated with each 7 µl DNase (20 g/l) and RNase (20 g/l) at 30°C for 15 min. The sample was centrifuged (20000 *g*, 15 min, 4°C) and all the supernatants of each strain were mixed. The samples were diluted 1:2 with lysis buffer without DTT to reduce the amount of DTT for incubation with antibody and beads. Then, 40 µl c-Myc monoclonal antibody (1 mg/ml, Clontech) was added to each lysate and incubated for 1 h at 4°C on a head-over-tail rotator wheel. After the incubation 300 µl equilibrated (equilibration according to 4.5.1) Dynabeads Protein G (Invitrogen) were added to each sample and incubated for 2 h at 4°C on the head-over-tail rotator wheel. Next, the beads were washed for four times with each 3 ml ice-cold

phosphate buffered saline (PBS) buffer (137 mM NaCl, 2.7 mM KCl, 10 mM sodium hydrogen phosphate, 1.8 mM potassium dihydrogen phosphate). Then, 24 µl TE buffer and 6 µl 6x SDS loading buffer (0.3 M Tris-HCl, pH 6.8, 1.2 mg bromphenol blue, 15% glycerin, 6% SDS and 168 mM DTT) were added to the beads. The samples were incubated at 95°C for 5 min and submitted to MS analysis.

4.10 Mass spectrometry

MS experiments and quantitative evaluations were performed by the group of Andreas Schlosser. I performed analysis according to peptide counts and go term analysis (section 5.2).

4.10.1 Pull-down experiment

The excised gel bands were destained with 30% acetonitrile in 0.1 M ammonium bicarbonate (NH_4HCO_3), pH 8, shrunk with 100% acetonitrile and dried in a vacuum concentrator. Digests were performed with 0.1 µg trypsin per gel band overnight at 37°C in 0.1 M NH_4HCO_3, pH 8. After removing the supernatant, peptides were extracted from the gel slices with 5% formic acid, and extracted peptides were pooled with the supernatant. NanoLC-MS/MS analyses were performed on an LTQ-Orbitrap Velos Pro (Thermo Scientific) equipped with an EASY-Spray Ion Source and coupled to an EASY-nLC 1000 (Thermo Scientific). Peptides were loaded on a trapping column (2 cm x 75 µm ID. PepMap C18, 3 µm particles, 100 Å pore size) and separated on an EASY-Spray column (25 cm x 75 µm ID, PepMap C18, 2 µm particles, 100 Å pore size) with a 30-minute linear gradient from 3% to 30% acetonitrile and 0.1% formic acid. MS and MS/MS scans were acquired in the Orbitrap analyzer with a resolution of 30,000 or 7500 respectively at m/z 400 using HCD fragmentation with 30% normalized collision energy. A TOP5 data-dependent MS/MS method was used. Dynamic exclusion was applied with a repeat count of one and an exclusion duration of 30 seconds. Singly charged precursors were excluded from selection. Minimum signal threshold for precursor selection was set to 50,000. Predictive AGC was used with AGC target a value of $1x10^6$ for MS scans and $5x10^4$ for MS/MS scans. Lock mass option was applied for internal calibration in all runs using background ions from protonated

decamethylcyclopentasiloxane (m/z 371.10124). Mascot Distiller 2.5 was used for raw data processing and for generating peak lists, essentially with standard settings for the Orbitrap Velos (high/high settings). Mascot Server 2.5 was used for database searching with the following parameters: peptide mass tolerance: 10 ppm, MS/MS mass tolerance: 0.02 Da, enzyme: "trypsin"; fixed modifications: carbamidomethyl (C); variable modifications: Gln->pyroGlu (N-term Q), oxidation (M) and acetyl (protein N-term). Database searching was performed against UniProt Yeast (Saccharomyces cerevisiae S288c) database and the contaminants database from Max Planck Institute of Biochemistry (Martinsried).

4.10.2 Co-immunoprecipitation

Proteins were reduced in 50 mM DTT for 10 min at 90 °C, and alkylated with 120 mM iodoactamide for 20 min at RT in the dark. Protein precipitation was performed according to Wessel and Fluegge with chloroform/methanol (Wessel & Flugge, 1984). Precipitated proteins were dissolved in 0.5% sodium deoxycholate (Sigma) in 0.1 M NH_4HCO_3, pH 8. Digests were performed with trypsin (trypsin-to-protein ratio: 1:100) overnight at 37°C. Sodium deoxycholate was removed by extraction with ethylacetate (Masuda et al, 2008). Peptides were dried in a vacuum concentrator to remove remaining ethylacetate. Peptides were desalted using C18 stage tips (Rappsilber et al, 2003). Each stage tip was prepared with three disks of C18 Empore SPE Disks (3M) in a 200 µl pipet tip. Peptides were eluted with 60% acetonitrile in 0.1% formic acid, dried in a vacuum concentrator, and stored at -20°C. Peptides were dissolved in 2% acetonitrile/0.1% formic acid prior to NanoLC-MS/MS analysis. NanoLC-MS/MS analyses were performed on an Orbitrap Fusion (Thermo Scientific) equipped with an EASY-Spray Ion Source and coupled to an EASY-nLC 1000 (Thermo Scientific). Peptides were loaded on a trapping column (2 cm x 75 µm ID, PepMap C18, 3 µm particles, 100 Å pore size) and separated on an EASY-Spray column (50 cm x 75 µm ID, PepMap C18, 2 µm particles, 100 Å pore size) with a 180-minute linear gradient from 3% to 40% acetonitrile and 0.1% formic acid. Both MS and MS/MS scans were acquired in the Orbitrap analyzer with a resolution of 60,000 for MS scans and 15,000 for MS/MS scans. HCD fragmentation with 35% normalized collision energy was

applied. A Top Speed data-dependent MS/MS method with a fixed cycle time of 3 sec was used. Dynamic exclusion was applied with a repeat count of one and an exclusion duration of 120 seconds. Singly charged precursors were excluded from selection. Minimum signal threshold for precursor selection was set to 50,000. Predictive AGC was used with AGC a target value of $5x10^5$ for MS scans and $5x10^4$ for MS/MS scans. EASY-IC was used for internal calibration. Raw MS data files were analyzed with MaxQuant version 1.5.3.12 (Cox & Mann, 2008). Database search was performed with Andromeda, which is integrated in the utilized version of MaxQuant. The search was performed against the UniProt Saccharomyces cerevisiae (strain ATCC 204508 / S288c, taxon identifier: 559292) reference proteome database. Additionally, a database containing common contaminants was used. The search was performed with tryptic cleavage specificity with two allowed miscleavages. Protein identification was under control of the false-discovery rate (<1% FDR on protein and peptide level). In addition to MaxQuant default settings (e.g. at least one razor/unique peptide for identification, two allowed miscleavages), the search was performed against following variable modifications: Protein N-terminal acetylation, Gln to pyro-Glu formation (N-term. Q) and oxidation (on Met). For protein quantitation, the LFQ intensities were used (Cox et al, 2014). Proteins with less than two identified razor/unique peptides were dismissed. Missing LFQ intensities in the control samples were imputed with values close to the baseline if intensities in the corresponding IP samples were present. Data imputation was performed with values from a standard normal distribution with a mean of the 5% quantile of the combined LFQ intensities and a standard deviation of 0.1.

5 Results

5.1 G4 structures bind to specific proteins

The position in the genome and nucleotide composition of G4 motifs are evolutionarily conserved in *S. cerevisiae* across *sensu stricto* species and among human populations (Capra et al, 2010; Nakken et al, 2009). This suggests that organisms are under a selective pressure to retain the G4 motifs and that the motifs might have positive biological functions (reviewed in (Bochman et al, 2012; Rhodes & Lipps, 2015)). However, because their formation and dissociation are too slow for *in vivo* processes and once formed the structures might be an obstacle for replication due to their high thermal stability (reviewed in (Bochman et al, 2012)), their formation and unwinding need to be tightly regulated by proteins (reviewed in (Rhodes & Lipps, 2015)). Indeed, it was demonstrated that TEBPβ, for example, promotes the formation of G4 structures *in vitro* (Fang & Cech, 1993; Giraldo & Rhodes, 1994; Paeschke et al, 2005) as well as *in vivo* (Paeschke et al, 2008; Paeschke et al, 2005) and various helicases unwound G4 structures *in vitro* and *in vivo* (for more information see 2.2.4.1 and 2.2.5.1) (Wu & Spies, 2016; Castillo Bosch et al, 2014; Cheung et al, 2002; Crabbe et al, 2004; Kruisselbrink et al, 2008; London et al, 2008; Lopes et al, 2011; Paeschke et al, 2013; Paeschke et al, 2011; Piazza et al, 2015; Piazza et al, 2012; Ribeyre et al, 2009; Sabouri et al, 2014; Sarkies et al, 2012). In *S. lemnae* it was discovered that TEBPα and TEBPβ are required for G4 structure formation and RecQ for their unwinding at telomeres. By this the proteins regulate telomere metabolism (Paeschke et al, 2008; Paeschke et al, 2005; Postberg et al, 2012). Apart from that, even more proteins were discovered to bind to G4 structures and to be required for genome stability at G4 structures in different organisms (for more information see introduction and data reviewed in (van Kregten & Tijsterman, 2014)). However, not all mechanisms of G4 regulation have been unraveled and not all the proteins that are involved in this process have been identified. In order to understand how G4 structures are regulated in living cells to control biological processes and to completely understand under which circumstances G4 structures cause genome instability, all proteins that

play a role in the specific regulation of G4 structures need to be identified and their exact function at G4 structures has to be unraveled.

If proteins influence biological processes through regulation of G4 structures, they need to be specific for G4 structures. Hence, it was assumed that specific proteins bind to G4 structures compared to linear G-rich DNA.

To support this hypothesis it was first investigated *in vitro* if G4 structures bind specifically to other proteins than linear G-rich DNA. The G4 motif (Chr XI_highγH2A_G4, see Annex Table 2 for sequence) was folded as published (Bachrati & Hickson, 2006). As previously shown, this G4 motif is located at a γ-H2A binding site (Capra et al, 2010), which was considered an interesting feature due to the fact that the aim was to search for repair proteins. γ-H2A is formed in response to DNA DSBs (Downs et al, 2000; Shroff et al, 2004). The folding of the G4 structure was confirmed by native PAGE (Figure 14A). The mutated G4 was not folded *in vitro*, but the non-folding was not tested. In the mutated G4 motif, some of the guanines in the G-tracts were replaced by other nucleotides to disrupt the G4 motif and to prevent folding of this sequence (Chr XI_highγH2A_G4mut, see Annex Table 2 for sequence). The mutated G4 motif was used to distinguish proteins that bind to the G4 structure from proteins that bind to linear DNA.

In one experiment yeast proteins were run on a native polyacrylamide gel and blotted (Figure 14B+C) and in the other experiment yeast proteins were denatured, separated by SDS PAGE and renatured on a membrane (Figure 14D-F). The second experiment was performed, because proteins can only be efficiently separated by size on a gel if they are denatured. The renaturing was performed, because it was the interest in this thesis to investigate binding of G4 structures to native proteins, which is their functional form *in vivo*. If proteins bind to G4 structures *in vivo* to regulate biological processes, they need to be in the native and active conformation. The membranes were incubated either with the biotinylated folded G4 or with biotinylated mutated G4. In the native experiment the membranes were also incubated with no DNA sample as a control (Figure 14B+C, lane 3). In the SDS PAGE experiment the binding of G4 or mutated G4 to denatured proteins without refolding was performed as a control (Figure 14D). The DNA was

detected via chemiluminiscence. This means, proteins were detected via the DNA that bound to them. In both experiments more proteins were detected if the proteins had been incubated with the G4 structure compared to linear G-rich DNA (Figure 14B+C+E+F), although the same amounts of DNA (determined by nanodrop measurements) and proteins (same volume of the same protein lysate) were used.

In the experiment with the native PAGE, only one big band was detected, which is due to the fact that the proteins were not properly separated (Figure 14B+C). Here, much more proteins were detected if the membranes were incubated with the G4 structure compared to linear DNA (Figure 14B+C). After incubating the membranes with no DNA, no signal was detected as expected (Figure 14B+C, lane 3).

In the experiment with SDS-PAGE and renaturing, proteins were separated according to their molecular weight (Figure 14D-F). After refolding with buffer 1, more DNA signal could be detected, especially in the molecular range of 25-70 kDa, if the refolded proteins were incubated with the G4 structure compared to the mutated control (Figure 14E). After refolding with buffer 2, the mutated control bound to more proteins than after refolding with buffer 1, especially in the molecular range of 25-70 kDa, (Figure 14E+F), but even more bands could be detected if the refolded proteins were incubated with the G4 structure compared to the mutated control (Figure 14F). After incubating the denatured proteins without refolding with G4 or mutated G4, only four bands (25 kDa, 35 kDa, 130 kDa, >170 kDa) were detected after incubation with the G4 structure, while no proteins were detected after incubation with the mutated control. This result was expected, because it was assumed that DNA sequences preferentially and specifically bind to native proteins compared to denatured proteins, because the native form is the form that is active *in vivo*. Also in this experiment (Figure 14D-F), the same amounts of DNA (determined by nanodrop measurements) and proteins (same volume of the same protein lysate) were used. From these experiments (Figure 14) it can be concluded that there are many proteins to which the G4 structure binds specifically compared to G-rich linear DNA.

In order to identify the proteins that specifically bind to G4 structures, a pull-down experiment with subsequent MS analysis was performed next.

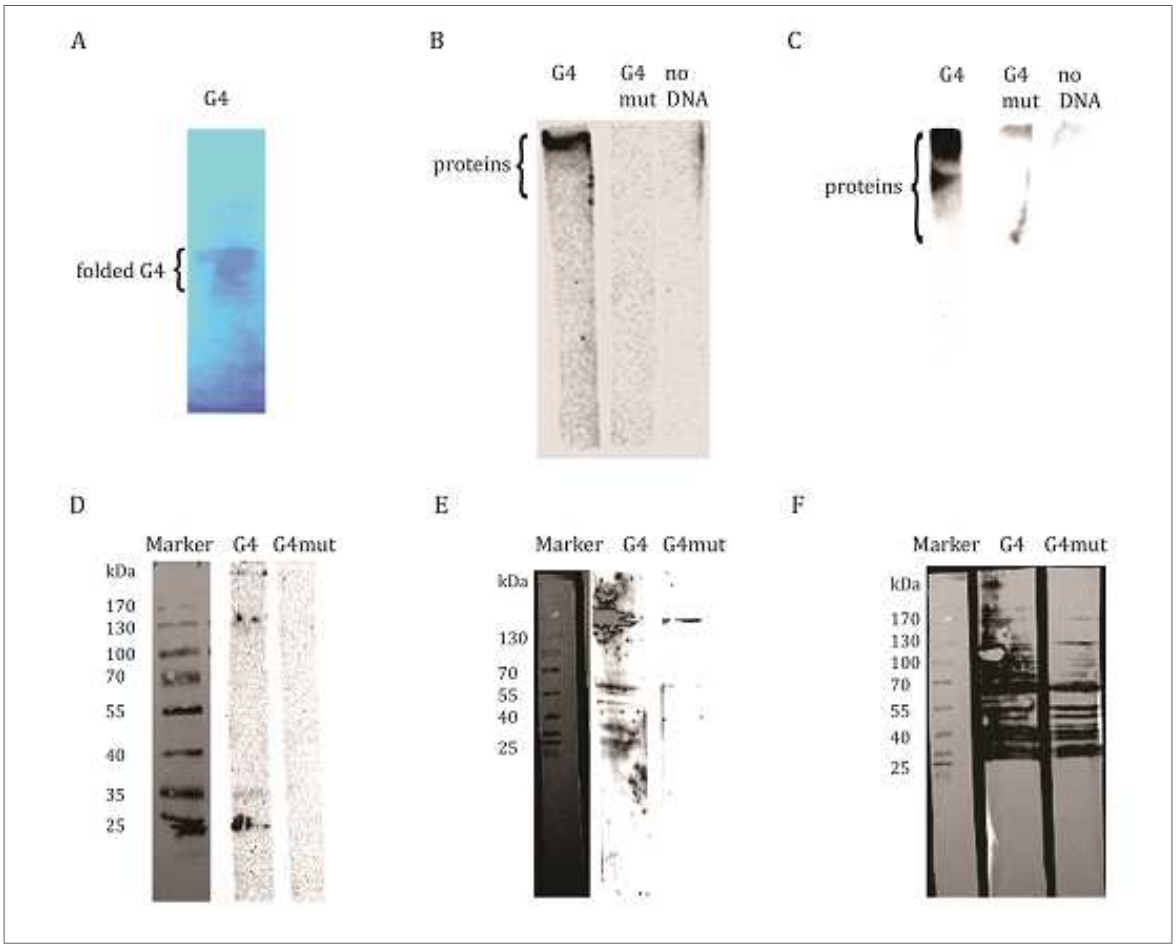


Figure 14: G4 structure binds specifically to other proteins compared to linear DNA. (A) Methylene blue staining of G4 oligodeoxynucleotide after folding, native PAGE and blotting. (B+C) Detection of biotinylated G4 (G4), mutated G4 (G4mut) or no DNA that bound to acidic (B) or basic (C) proteins separated by native PAGE. (D) Detection of biotinylated G4 (G4) or mutated G4 (G4mut) that associated with denatured proteins separated by SDS-PAGE without refolding. Marker proteins are separated in lane 1 (Thermo fisher scientific, 26616). (E+F) Detection of biotinylated G4 (G4) or mutated G4 (G4mut) that bound to denatured proteins separated by SDS PAGE after refolding with buffer 1 (E) or 2 (F). Marker proteins are separated in lane 1 of each Figure 14E+F (Thermo fisher scientific, 26616).

5.2 Many proteins specifically bind to G4 structures *in vitro*

Using a G4 structure or mutated G4 as bait, a pull-down assay and subsequent MS analysis was performed to identify proteins that specifically bind to G4 structures. Biotinylated oligodeoxynucleotides containing either a G4 motif or a mutated G4 motif were folded according to the protocol described in (Bachrati & Hickson, 2006) (Chr XI_highγH2A_G4 and Chr

XI_highγH2A_G4mut or Chr VI_G4 and Chr VI_G4mut, see Annex Table 2 for sequences of motifs). Pull-down experiment was performed four times under different experimental conditions. In the first experiment (Chr XI_highγH2A_G4 and Chr XI_highγH2A_G4mut, wild type cells, lysis and washing according to section 4.5.1, elution via boiling, see Annex Figure 2A for SDS-PAGE prior to MS analysis), only 13 proteins were identified if the parameters 20% protein threshold, 20% peptide threshold and minimum of two unique peptide counts were applied (Annex Table 4). The correct folding of the G4 sample and no folding of the mutated G4 was not confirmed. Because of the small number of identified proteins, these results were not further analyzed and the protocol was optimized (section 4.3). In the first pull-down experiment, lysis and washing was performed according to section 4.5.1 and in the following experiments (second, third and fourth pull-down experiment) according to section 4.3. It was assumed that washing with ChIP buffer (first pull-down experiment) would be too harsh, because DNA and proteins were not crosslinked in the pull-down experiment and that proteins might be more stable in their native form if the lysis buffer from section 4.3 was used, because it contained more protease inhibitors and DTT. In addition to that, in the first pull-down experiment proteins were eluted via boiling and in the following experiments via increasing salt concentrations. Elution with boiling is quite harsh and all proteins are eluted at once, while eluting with increasing salt concentrations leads to different protein fractions. By this the stability of DNA-protein interactions can be estimated, because if a protein is eluted at low salt concentration, its binding to the DNA is weaker than the binding of a protein, which is eluted with higher salt concentration. Also, in the first experiment the G4 motif and mutated G4 motif from section 5.1 were used (Chr XI_highγH2A_G4 + Chr XI_highγH2A_G4mut, see Annex Table 2 for sequences), while in the following experiments other motifs (Chr VI_G4 and Chr VI_G4mut, see Annex Table 2 for sequences of both motifs) were used, because it was assumed that the Chr VI_G4 forms a more stable G4 structure than Chr XI_highγH2A_G4 due to its smaller loop lengths. This Chr VI_G4 motif is also located at a γ-H2A binding site (Capra et al, 2010). It is known that the *S. cerevisiae* helicase Pif1 unwinds G4 structures *in vitro* (Paeschke et al, 2013; Ribeyre et al, 2009; Paeschke et al, 2011). Hence, in

order to prevent unfolding of G4 structures during this analysis, yeast protein lysate from *pif1-m2* cells was used in some experiments (pull-down experiment 2+3). The highest amount and most promising proteins were identified in the MS analysis shown in Figure 15 (pull-down experiment 2). Here, the protocol from section 4.3, the motifs Chr VI_G4 and Chr VI_G4mut as well as *pif1-m2* cells were used. The correct folding of a G4 and non-folding of the mutated control was confirmed by CD (Annex Figure 2B). Proteins were eluted via increasing salt concentrations. Bound proteins were analyzed initially by SDS PAGE (Figure 15A (elution 400 mM + 600 mM NaCl), Annex Figure 2C+D (elution 200 mM + 800 mM NaCl)) and subsequently by MS (Annex Table 5+6). MS analysis was performed by the group of Andreas Schlosser. Due to the high amount of detected protein bands in the elution fraction 400 mM and 600 mM NaCl and because mostly proteins that bind strongly to G4 structures were interesting candidates, only the elution fractions 400 mM and 600 mM NaCl were analyzed by MS. The group of Prof. Schlosser aimed to analyze the identified proteins of these two fractions quantitatively (see Annex Figure 3), but they claimed that there were too less background proteins (personal communication, July 2016). They said, normally in pull-down experiments there are many proteins, which unspecifically bind to the beads and that those proteins can be used for normalizing. They found that in the pull-down experiment performed in this thesis there is a wide distribution of proteins around zero protein intesities (grey dots) (Annex Figure 3), which normally represent unspecific interaction partners. However, when they took a closer look at these grey dotted proteins, it was revealed that also specific interaction partners might be among them. Therefore, the group of Prof. Schlosser claimed that the method normally used for quantification does not work for this pull-down experiment (personal communication, July 2016).

Hence, I evaluated the data using exclusive unique peptide counts. It was assumed that the higher the amount of identified exclusive unique peptides of a protein, the higher is the specificity of its binding. Also, the higher the number of exclusive unique peptide counts, the higher is the probability, that the protein was correctly identified. Because both elution fractions

resulted from the same pull-down experiment and analyses would be too complicated otherwise, both fractions were evaluated together and not individually (see Annex Table 5+6 for individual lists). Applying the parameters 20% protein threshold, 20% peptide threshold and minimum of two exclusive unique peptide counts, 896 yeast proteins were identified in both fractions (proteins from other organisms e.g. keratin as well as decoys were skipped from the analysis). 288 yeast proteins were identified that uniquely bound to the G4 sample (Figure 15B). Another 525 yeast proteins showed no preferred binding, and 83 yeast proteins only bound to the mutated G4 (Figure 15B). Go term analysis revealed that many of the 288 G4 binding proteins are implicated in similar biological processes (Figure 15C, Annex Table 7): 34% are involved in metabolic processes of nucleic acids, of which, 82% play a role in gene expression, 23% in ribosome biogenesis, 24% in the DNA damage response, 19% in mRNA processing and 12% in replication.

With a focus on the role and consequences of G4 structures formed during DNA replication and their influence on DNA damage, many proteins involved in repair processes were identified. Next, data on some repair proteins identified in the pull-down experiment (for more information see introduction section 2.3) will be presented and it will be demonstrated that many of them are connected through a function in the same biological process. For example, Mec1 and Tel1 are checkpoint kinases responsible for the phosphorylation of various targets (e.g. Rad9, Rad17, Rpa1, H2A) in response to DNA damage (Downs et al, 2000; Emili, 1998; Nakada et al, 2003a; Nakada et al, 2003b; Sanchez et al, 1996; Shroff et al, 2004; Smolka et al, 2007; Sun et al, 1996; Sweeney et al, 2005). Smc1 and Smc6 were shown to be involved in sister chromatid cohesion after DNA damage (Strom et al, 2007; Strom et al, 2004), Smc3 and Dot1 play a role in SCR (Conde et al, 2009; Cortes-Ledesma & Aguilera, 2006) and Smc5 was observed to be recruited to DSB sites and to be required for the recruitment of Smc1 (Potts et al, 2006). Mec1, Tel1, Smc1, Smc3, Smc5 and Dot1 are connected functionally as it was demonstrated that Mec1 and Tel1 can phosphorylate H2A (γ-H2A) (Downs et al, 2000; Shroff et al, 2004) and Dot1 methylates histone H3 on lysine 79 (H3K79me) (Feng et al, 2002; Lacoste et al, 2002; Ng et al, 2002; van Leeuwen et

al, 2002). Both γ-H2A and H3K79me are needed for Rad9 binding to chromatin (Toh et al, 2006) and Rad9 is necessary for SCR (Conde et al, 2009) as also are the cohesion complex (Smc1-Smc3) and the Smc5-Smc6 complex (Cortes-Ledesma & Aguilera, 2006; De Piccoli et al, 2006; Potts et al, 2006).

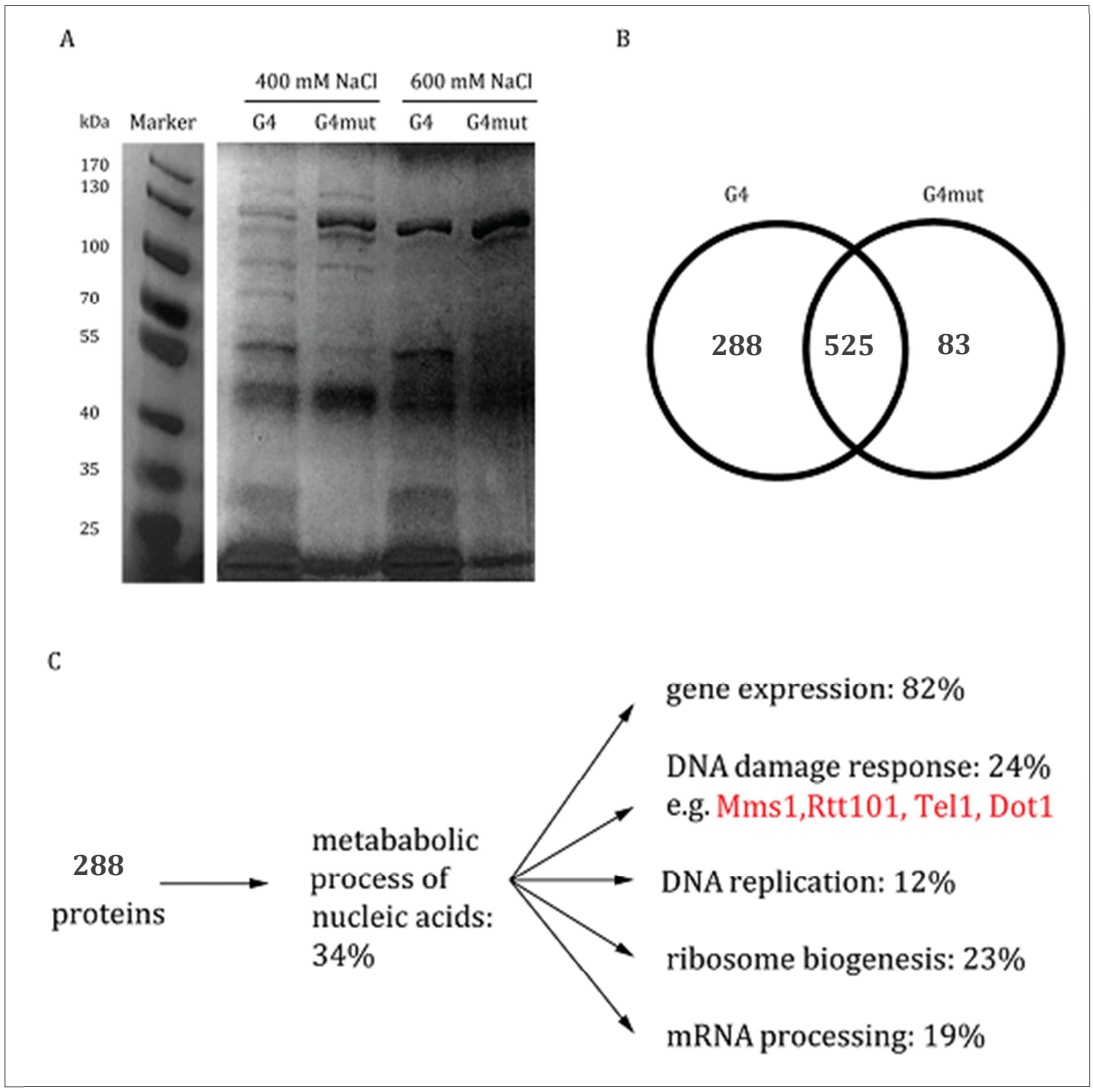


Figure 15: Repair proteins bind to G4 structures *in vitro*. (A) Coomassie staining of separated proteins eluted from G4 or mutated G4 (G4mut) DNA with 400 mM or 600 mM NaCl. Protein ladder (marker, Thermo fisher scientific, 26616) was separated on lane 1. (B) Number of identified proteins by MS that bound to the G4, mutated G4 (G4mut) or both. MS analysis was performed by the group of Andreas Schlosser. (C) Go term analysis of G4 specific proteins.

Conde and colleagues hence proposed a model in which a DSB induces the phosphorylation of H2A by Mec1 and/or Tel1 and the methylation of Histone H3 on lysine 79 by Dot1. Both γ-H2A and H3K79me are required for the recruitment of Rad9, which then enables the loading of cohesion and repair via SCR (Conde et al, 2009). For another protein identified, Mgs1, different

studies suggest that it promotes replication fork progression (Branzei et al, 2002a; Hishida et al, 2002), which was also the case for Mms1 and Rtt101 (Luke et al, 2006; Vaisica et al, 2011; Zaidi et al, 2008).

Evaluated were the exclusive unique peptide counts of the fraction 400 mM and 600 mM NaCl together (see Annex Table 5+6). For the G4 sample, the exclusive unique peptide counts for Mec1 were 40, for Tel1 and Smc1 seven each, for Smc3 eleven, for Smc5 three, for Dot1 six, for Mgs1 32, for Mms1 twelve and for Rtt101 seven. The protein identification probability was over 95% for all proteins. Mec1, Tel1, Smc1, Smc3, Smc5, Dot1 and Mgs1 were not found in the mutated control. Mms1 and Rtt101 were found with one peptide count each in the mutated control (protein identification probability was over 95%), but due to the applied parameters (at least two exclusive unique peptides) it was considered that they did not bind to the control.

Then, it was aimed to reproduce the results. Hence, two more pull-down experiments and MS analyses were performed. One pull-down experiment (third) was performed as the second pull-down experiment. The folding of the G4 and no folding of the mutated G4 was confirmed by CD measurements (Annex Figure 2E). Proteins were first separated by SDS-PAGE and stained with coomassie (Annex Figure 2F). Because no protein bands could be detected with coomassie staining, the 800 mM elution and the input sample were stained with silver staining (Annex Figure 2G). With this staining, bands were detected in the 800 mM elution and in the input sample. Therefore it was assumed, that the coomassie was not sensitive enough for a detection of proteins in the other fractions. Only the 400 mM NaCl elution of the G4 sample was analyzed by MS. 223 proteins were identified applying the parameters 20% protein threshold, 20% peptide threshold and minimum of two exclusive unique peptide counts (proteins from other organisms e.g. keratin as well as decoys were skipped from the analysis) (Annex Table 8). Because none of the proteins discussed above (Mec1, Tel1, Smc1, Smc3, Smc5, Dot1, Mgs1, Mms1 and Rtt101), were identified in this analysis and none of the identified proteins were considered interesting candidates for this thesis, MS analysis of the mutated control was not performed. In the last (fourth) pull-down experiment, yeast protein lysate from wild-type cells was used to

examine whether Pif1 can be reproduced as a G4 structure interacting protein. The other experimental conditions were the same as for pull-down experiment two and three. Proteins were first separated by SDS-PAGE and stained with coomassie (Annex Figure 2H, performed by the group of Andreas Schlosser). Due to the highest number of detected protein bands, the 400 mM NaCl elution fraction of the G4 and mutated G4 sample were analyzed by MS and 490 proteins were identified in the G4 and mutated G4 sample applying the parameters 20% protein threshold, 20% peptide threshold and minimum of two exclusive unique peptide counts (proteins from other organisms e.g. keratin as well as decoys were skipped from the analysis) (Annex Table 9). 286 proteins uniquely bound to the G4 sample, 37 only bound to the mutated G4 sample and 167 proteins bound to both samples. The G4 and mutated G4 samples were the same as for the third pull-down experiment (see Annex Figure 2E for CD). In this analysis, from the discussed proteins above (Mec1, Tel1, Smc1, Smc3, Smc5, Dot1, Mgs1, Mms1 and Rtt101), only Mgs1 was identified. Mgs1 was found with six peptide counts in the G4 sample, but no peptides were found in the mutated control. Go term analysis revealed, that thirteen proteins (Hta1, Rad7, Dna2, Ckb1, Cka2, Ckb2, Ctr9, Rfc4, Rfc3, Mgs1, Bdf1, Rfc2, Msh3) (see Annex Table 9 for peptide counts) uniquely identified in the G4 sample, are implicated in the response to DNA damage. Except for Mgs1, Bdf1 and Ckb1 none of these proteins were also identified as unique G4 interacting proteins in another analysis (Annex Table 4+5+6+8+9). Mgs1 was identified in the second pull-down experiment with 32 (0 peptides in mutated G4) and in the fourth pull-down experiment with six (0 peptides in mutated G4) peptides (Annex Table 5+6+9). Bdf1 was identified with six peptides in the second and with eleven peptides in the fourth pull-down experiment (one and zero peptides in mutated G4 respectively) (Annex Table 5+9). Ckb1 was identified with six peptides in the third pull-down experiment and with twelve peptides in the fourth pull-down experiment (one peptide in the mutated control). In the second pull-down experiment Ckb1 was also identified in the mutated control (Annex Table 5+6+8+9).

Surprisingly, the G4 unwinding helicase Sgs1 (Paeschke et al, 2013; Sun et al, 1999) was identified in two of four pull-down experiments (Annex Table 5+6+9) (39 (*pif1-m2* cells) and 16

(wild type cells) peptide counts), but only in the mutated G4. Also the 5´-3´ DNA helicase Rrm3 (Ivessa et al, 2002) was found in the fourth pull-down experiment (wild-type cells, Annex Table 9) with 13 peptide counts in the mutated control, but with only one peptide count in the G4 sample. Rrm3 was shown bind to telomeres and to promote telomere replication in yeast (Ivessa et al, 2002). Because it was assumed for this thesis that G4 structures can form at telomeres in yeast (see 2.3.1), it would have been expected that Rrm3 rather binds to the G4 structure than to the mutated G4 during the pull-down experiment. Even more surprisingly, the *in vitro* G4 unwinding helicase Pif1 (Paeschke et al, 2013; Ribeyre et al, 2009; Sanders, 2010; Wallgren et al, 2016) was identified in the second pull-down experiment with *pif1-m2* cells (Annex Table 5+6) only in the mutated control (16 peptide counts, only one peptide count for the G4 sample in each elution), while in the pull-down experiment with wild type cells (Annex Table 9) Pif1 was identified with 13 peptide counts in the G4 and with 24 peptide counts in the mutated control. Therefore, in this thesis, it could not be confirmed that Sgs1 and Pif1 preferentially bind to G4 structures.

Interestingly, the mismatch repair protein Msh1 was predominantly found in the mutated G4 sample (Annex Table 5+6+9). In the second pull-down experiment with *pif1-m2* cells Msh1 was identified with 4/44 (G4/G4mut) (Annex Table 5+6) and in the fourth pull-down experiment with wild-type cells with 4/28 (G4/G4mut) peptide counts (Annex Table 9). In the second pull-down experiment with *pif1-m2* cells also the mismatch proteins Msh2 (6/50=G4/G4mut), Msh3 (7/45=G4/G4mut) and Msh6 (3/37=G4/G4mut) were predominantly found in the mutated G4 sample (Annex Table 5+6). In the fourth pull-down experiment with wild type cells, Msh2 was identified with seven peptide counts in each sample, Msh3 with six peptides in the G4, but with no peptide in the mutated G4 and Msh6 with two peptides in the G4 and three peptides in the mutated G4 sample (Annex Table 9).

Interestingly, 224 unique G4 binding proteins were identified in at least two independent pull-down experiments (thanks to Stefan Juranek for this evaluation) (Annex Table 10).

To get more insights into the function of the identified G4-binding proteins, their *in vivo* binding to G4 structures as well as their *in vivo* function at G4 structures was examined. Additionally, some proteins already identified as G4 structure binding partners were investigated.

5.3 ChIP of G4 binding proteins

G4 motifs are enriched at DSB sites and are connected to DSB occurrence (Capra et al, 2010; Crabbe et al, 2004; Hershman et al, 2008; Koole et al, 2014; London et al, 2008; Lopes et al, 2011; Paeschke et al, 2011; Piazza et al, 2015; Piazza et al, 2012; Ribeyre et al, 2009; Rodriguez et al, 2012; Sabouri et al, 2014; Zimmer et al, 2016). Furthermore, previous studies revealed a connection between G4 induced DNA damage and HR (Lopes et al, 2011; Piazza et al, 2012; Ribeyre et al, 2009; Zimmer et al, 2016). Repair proteins that were identified to bind to G4 structures *in vitro* were Ku70 and Mre11 (Cogoi et al, 2008 ; Ghosal & Muniyappa, 2005). Therefore it was investigated in this thesis whether Mre11 and Ku70 specifically bind to non-telomeric G4 motifs *in vivo* to enable the repair of G4 induced damage. Ku70 was identified in one pull-down experiment (Annex Table 4), but with only one peptide count in the G4 and no peptide count in the mutated G4 and the protein identification probability was below 50%. Hence, this result is negligible. Mre11 was not found in any of the pull-down experiments performed in this thesis (Annex Table 4+5+6+8+9). Also, the *in vivo* binding of three proteins (Tel1, Dot1, Mgs1) identified in a pull-down experiment (Tel1 and Dot 1 once, Mgs1 twice (Annex Table 5+6+9)) was determined. Tel1 (seven peptide counts, only G4 sample) and Dot1 (six peptide counts, only G4 sample) are implicated in DSB repair (Conde et al, 2009; Shroff et al, 2004) and Mgs1 (32+6 peptide counts, only G4 sample) is important for replication fork progression (Branzei et al, 2002a; Branzei et al, 2002b; Hishida et al, 2002), which was also impeded by G4 structures in previous studies (Castillo Bosch et al, 2014; Crabbe et al, 2004; Lopes et al, 2011; Paeschke et al, 2011; Sabouri et al, 2014; Sarkies et al, 2010; Schiavone et al, 2014; Wu & Spies, 2016).

To determine the binding of those proteins to G4 motifs *in vivo* ChIP of endogenously Myc-tagged Ku70, Mre11, Tel1, Dot1 and Mgs1 followed by qPCR (see Annex Table 11 for sequences

of all primers used in this thesis) was performed. In these and all subsequent ChIP experiments (except for ChIP-seq) at least three biological replicates were analyzed independently and qPCR was performed with two technical replicates. Also, every protein that was tagged in this thesis was tagged endogenously at its C-terminus. The expression of the tagged proteins was confirmed by western blot analysis (Annex Figure 4). All five tagged proteins investigated in this section (Ku70, Mre11, Tel1, Dot1 and Mgs1) were functionally expressed (Annex Figure 4 A+B). In western blot analysis a single band was detected at a similar size (Myc-Tag accounts for ~16 kDa and HA-Tag for ~3 kDa). Protein bands were observed for Ku70-Myc at around 90 kDa (86 kDa expected), for Mre11-Myc at 120 kDa (88 kDa expected), for Tel1-Myc at above 170 kDa (338 kDa expected), for Dot1 at 90 kDa (82 kDa expected) and for Mgs1-Myc at 110 kDa (83 kDa expected) (Annex Figure 4A+B). A reason for the fact that the proteins sometimes run higher than expected might be that the separation conditions were not ideal. Western blot analysis was also performed with the western input and eluate controls for conventional ChIP (see 4.5.1). However, for some proteins no bands could be detected in the input and/or eluate controls. Due to the fact that the ChIP procedures where the same for all tested proteins, the western analysis of Ku70 is shown exemplarily. In western input and eluate conventional ChIP control samples of Myc-tagged Ku70 protein bands were detected at around 110 kDa (86 kDa expected) and the protein was stable during the analysis (Annex Figure 4C+D).

To determine whether the examined proteins show binding preference towards G4 motifs compared to other DNA, five G4 motifs and two control regions were tested. The two control regions used were one non-G-rich and one G-rich region that were previously used in a publication (Paeschke et al, 2011). The G4 motifs that were tested were chosen from published G4tract3 motifs (Capra et al, 2010). In this thesis, G4 motifs are considered G4tract3 motifs, if they contain at least three guanines in the G-tract. The control regions do not contain a G4tract3 motif within a window of ± 500 bp, which was the maximal shearing size during conventional ChIP (Annex Figure 1A). It was assumed that repair pathways and therefore binding of repair proteins are required at G4 motifs with high γ-H2A accumulation. The histone variant γ-H2A is

formed by phosphorylation of histone H2A as a consequence of DSBs and it is part of the signal transduction pathway after DNA damage (Downs et al, 2000; Shroff et al, 2004). Hence, three of the tested G4 motifs (Chr $VI_{G4tract3}$, $XI_{G4tract3}$, $XII_{G4tract3}$) and also the non-G-rich (Chr XI_{NG}) and G-rich control region (Chr $XIII_{GR}$) were chosen, because they are located at a previously identified γ-H2A binding site (Capra et al, 2010) within a window of ± 500 bp, which was the maximal shearing size during conventional ChIP (Annex Figure 1A).

None of the five tested proteins (Ku70, Mre11, Tel1, Dot1, and Mgs1) associated significantly more with the tested G4 motifs (Chr $VI_{G4tract3}$, $IX_{G4tract3}$, $XI_{G4tract3}$, $XII_{G4tract3}$ and $XVI_{G4tract3}$) compared to a G-rich region (Chr $XIII_{GR}$) (Figure 16). Tel1 preferentially bound to the five G4 motifs tested (Chr $VI_{G4tract3}$, $IX_{G4tract3}$, $XI_{G4tract3}$, $XII_{G4tract3}$ and $XVI_{G4tract3}$) compared to the non-G-rich (Chr XI_{NG}) region (3.3-18.5-fold), but this was also the case for the G-rich region (Chr $XIII_{GR}$) (8.6-fold) (Figure 16A). Mgs1 associated more with G4 motifs of Chr $VI_{G4tract3}$ (1.5-fold) and $XII_{G4tract3}$ (three-fold) than with the non-G-rich control region (Chr XI_{NG}) (Figure 16B). Ku70, Mre11 and Dot1 did not bind preferentially to any of the five G4 motifs tested (Chr $VI_{G4tract3}$, $IX_{G4tract3}$, $XI_{G4tract3}$, $XII_{G4tract3}$ and $XVI_{G4tract3}$) compared to the non-G-rich control region (Chr XI_{NG}) (Figure 16).

Because the examined proteins were identified as G4 structure binding proteins *in vitro* (this thesis, Annex Table 5+6 or as published (Cogoi et al, 2008 ; Ghosal & Muniyappa, 2005)), but their binding was not observed at non-telomeric G4 motifs *in vivo* (Figure 16), conventional ChIP experiments were performed in *pif1-m2* mutants. It was assumed that association of G4 structure binding proteins is increased in the absence of G4 unwinding helicases. Again, binding of Ku70, Mre11, Tel1, Dot1 and Mgs1 was determined.

Tested regions were the same as in Figure 16, except also two telomeric regions (Chr $VI_{Tel\_VI\text{-}R}$+ Chr $VII_{Tel\_VII\text{-}L}$) as in a publication (McGee et al, 2010) were examined for Ku70, Mre11 and Tel1 binding, because those proteins associated with telomeres in *S. cerevisiae* (Fisher et al, 2004; Hector et al, 2007; McGee et al, 2010; Sabourin et al, 2007; Takata et al, 2005).

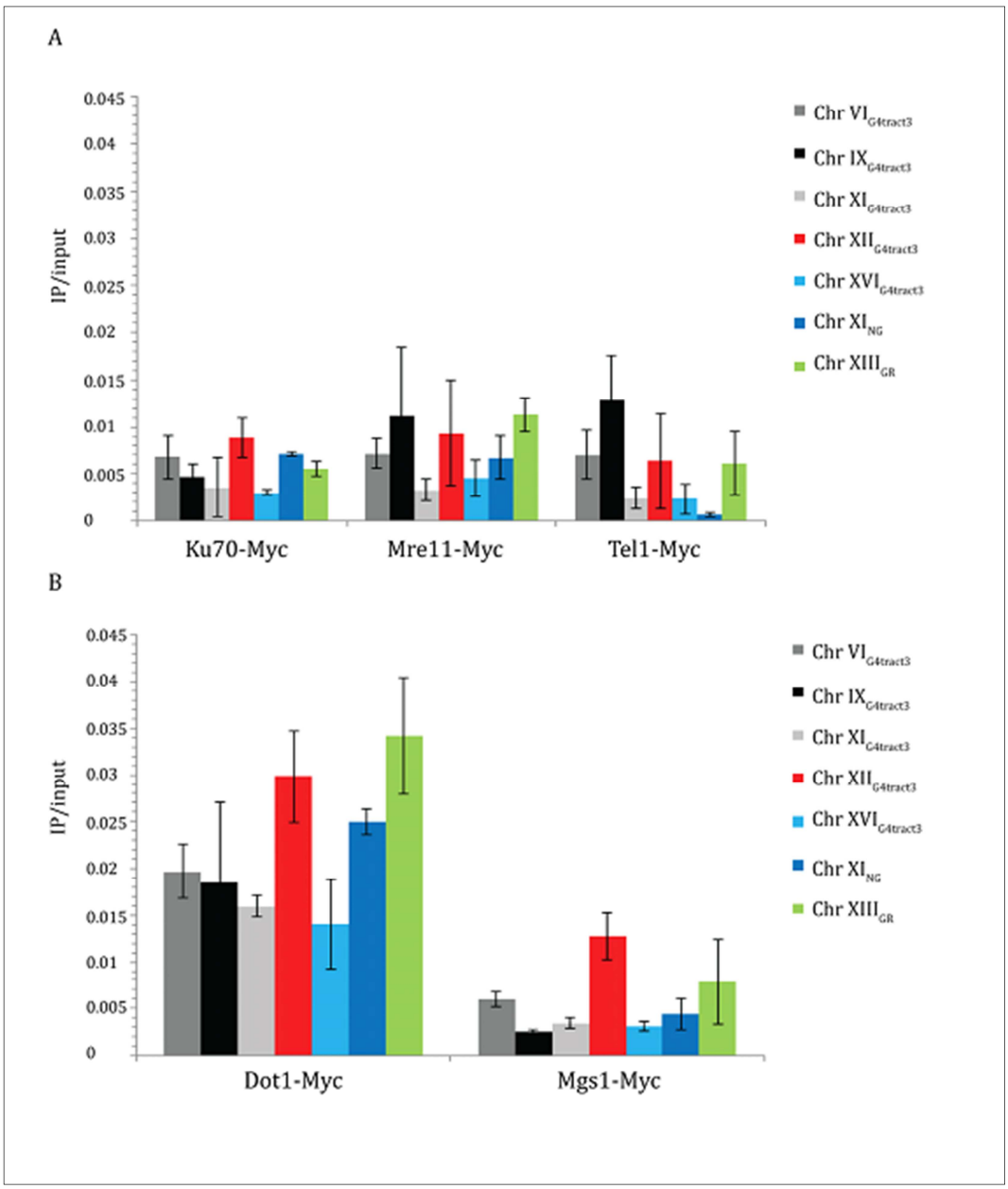


Figure 16: None of the examined proteins is enriched at the tested G4 motifs compared to a G-rich region. (A+B) ChIP was performed and the binding of Myc-tagged Ku70, Mre11 and Tel1 (A) as well as Dot1 and Mgs1 (B) was analyzed by qPCR using primer pairs for the shown regions. As control regions, the enrichment at non-G-rich (Chr XI_{NG}) and G-rich (Chr $XIII_{GR}$) regions was determined. Plotted are the IP/input values as means ± standard deviation (SD) from at least three independent experiments.

The G4 motifs Chr $IX_{G4tract3}$, $XI_{G4tract3}$ and $XVI_{G4tract3}$ and the control regions (ChrXI_{NG} and Chr$XIII_{GR}$) are located at a previously identified Pif1 binding site (Paeschke et al, 2011) within a window of ± 500 bp, which was the maximal shearing size during conventional ChIP (Annex Figure 1A). Also the telomeres tested (Chr $VI_{Tel\_VI\text{-}R}$ and $VII_{Tel\_VII\text{-}L}$) were previously identified as Pif1 binding sites (Paeschke et al, 2011; Phillips et al, 2015), although $VII_{Tel\_VII\text{-}L}$ was modified and other

primers were used in (Phillips et al, 2015). The G4 motifs Chr $VI_{G4tract3}$ and $XII_{G4tract3}$ are not located at a previously identified Pif1 binding site (Paeschke et al, 2011) within a window of ± 500 bp, but within ± 1500 bp. At the time when these ChIP-qPCR analyses were performed, it was thought that the shearing size during conventional ChIP was around 3000 bp. At this time point the crosslink was not reversed before analyzing the shearing size and protein-DNA complexes run higher than unbound DNA. Therefore, taken into account the correct shearing size of 500 bp, the G4 motifs Chr $VI_{G4tract3}$ and $XII_{G4tract3}$ are no Pif1 binding sites.

Unexpectedly, absence of nuclear Pif1 tended to result in less binding of Ku70 to all non-telomeric G4 motifs tested (Chr $VI_{G4tract3}$, $IX_{G4tract3}$, $XI_{G4tract3}$ and $XVI_{G4tract3}$) and to the telomeres tested (Chr $VI_{Tel\_VI\text{-}R}$ and $VII_{Tel\_VII\text{-}L}$), even though the effect was mostly not significant due to high SD (Figure 17A+B). Only at Chr $XII_{G4tract3}$, no Pif1 dependent binding was detected (Figure 17A). This is in line with the fact that this site was not identified as a Pif1 binding site in a previous study (Paeschke et al, 2011). A significant effect was observed at the G4 motifs of Chr $VI_{G4tract3}$ (although this is not a Pif1 binding site (Paeschke et al, 2011)), $IX_{G4tract3}$ and $XVI_{G4tract3}$ (2.1-3.5-fold) and the telomeres Chr $VI_{Tel\_VI\text{-}R}$ and $VII_{Tel\_VII\text{-}L}$ (2.4-16.7-fold), but not at the non-G-rich (Chr XI_{NG}) and G-rich (Chr $XIII_{GR}$) control regions (Figure 17A+B). The 4.3-fold decrease of Ku70 binding to Chr $XI_{G4tract3}$ was not significant due to high SD (Figure 17A).

Binding of Mre11 (Figure 17C) and Tel1 (Figure 17D) was Pif1 independent at all tested non-telomeric G4 motifs (Chr $VI_{G4tract3}$, $IX_{G4tract3}$, $XI_{G4tract3}$, $XII_{G4tract3}$ and $XVI_{G4tract3}$) and at the non-G-rich (Chr XI_{NG}) control region. The decrease of Mre11 binding to Chr $VI_{G4tract3}$, $XI_{G4tract3}$ and $XVI_{G4tract3}$ (1.6-2.2-fold) (Figure 17C) was not significant due to high SD. Binding of Mre11 to the telomeres Chr $VI_{Tel\_VI\text{-}R}$ and $VII_{Tel\_VII\text{-}L}$ as well as the G-rich control region (Chr $XIII_{GR}$) significantly decreased (2.8-15.7-fold) in *pif1-m2* compared to wild type cells (Figure 17C). Tel1 associated significantly less (3.5-fold) with the telomere Chr $VI_{Tel\_VI\text{-}R}$ in *pif1-m2* compared to wild type cells, but not with telomere Chr $VII_{Tel\_VII\text{-}L}$ (Figure 17D). The decreased binding of Tel1 to the G-rich region (Chr $XIII_{GR}$) (2.2-fold) (Figure 17D) was not significant due to high SD.

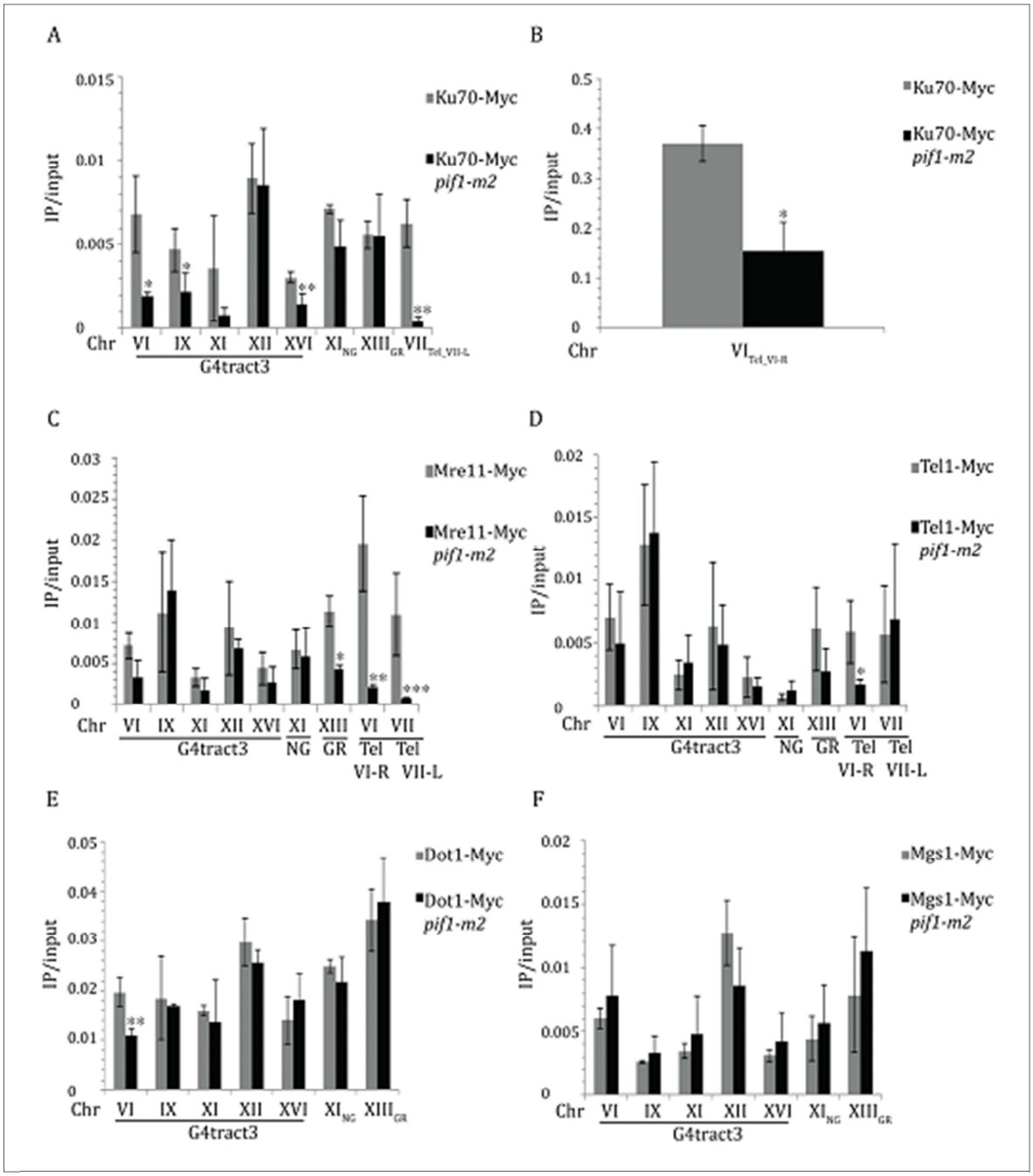


Figure 17: Binding of Ku70 to most G4 motifs was Pif1 dependent. (A+B) ChIP was performed and the binding of Myc-tagged Ku70 in wild type (grey) and *pif1-m2* (black) cells was analyzed by qPCR using primer pairs for the shown regions. As control regions, the enrichment at non-G-rich (Chr XI_{NG}) and G-rich (Chr $XIII_{GR}$) regions was determined. Plotted are the IP/input values as means ± SD from at least three independent experiments. (C-F) ChIP-qPCR of Myc-tagged Mre11 (C), Tel1 (D), Dot1 (E) and Mgs1 (F) was performed as in A+B. Statistical significance compared to wild type Myc-tagged cells was determined by student's T-test. *: p≤0.05, **: p≤0.01, ***: p≤0.001.

Mgs1 and Dot1 (Figure 17E+F) showed Pif1 independent binding to all tested regions except for the G4 motif on Chr $VI_{G4tract3}$ (although this is not a Pif1 binding site (Paeschke et al, 2011)) where less binding of Dot1 (1.8-fold) was seen in *pif1-m2* compared to wild type cells (Figure 17E).

Due to the *in vitro* unwinding activity of Pif1 at G4 structures (Paeschke et al, 2013; Ribeyre et al, 2009; Sanders, 2010; Wallgren et al, 2016) it was assumed that binding of G4 interacting proteins to G4 motifs would increase in the absence of the helicase. An additional reason for this hypothesis was the detection of increased breaks at G4 motifs in the absence of Pif1 helicase in yeast in former studies (Lopes et al, 2011; Paeschke et al, 2011; Piazza et al, 2015; Ribeyre et al, 2009; Sabouri et al, 2014). Unexpectedly, binding of the tested proteins, especially Ku70, rather decreased than increased in the absence of Pif1. Neither in wild type nor in *pif1-m2* cells did the tested proteins specifically bind to non-telomeric G4 motifs (Figure 16+17) compared to a G-rich control region (Chr $XIII_{GR}$). It is possible that the tested G4 motifs are no specific binding sites of the examined proteins, but it does not exclude that the proteins preferentially bind to other G4 motifs. Hence the following experiments were performed to clarify the role of the proteins.

5.4 Deletion of *MRE11* and *MMS1* causes severe chromosomal damage at G4 motifs

To determine if proteins that bind to G4 structures *in vitro* are important for genome integrity at G4 motifs *in vivo*, a previously described GCR assay was performed (Paeschke et al, 2013; Schmidt et al, 2006). For this, a G4 motif from Chr I (G4-*LEU2*) was inserted into the yeast genome on the left arm of Chr V, replacing the non-essential *PRB1* gene (Figure 18A, see Annex Table 12 for detailed information of regions and sequences and for sequencing results that confirm the correct sequences; first all inserts were inserted, then deletions were created. Hence, only one sequencing analysis per insert is shown). Two counter selectable markers (*URA3* and *CAN1*) are located downstream of the *PRB1* locus (Figure 18A). If the inserted sequence induces genomic instability, the markers are lost, and cells can grow on selective media. By counting the colonies on selective media compared to those on rich media plates, the GCR rate can be determined via fluctuation analysis (Hall et al, 2009). The GCR rates here and in

subsequent experiments were determined at least three times with seven biological replicates each.

The GCR rate of the wild type cells without an insert was approximately $4x10^{-9}$ events per generation, which is ~10-fold higher than published ($3.5x10^{-10}$) (Chen & Kolodner, 1999). The fold-enrichment of GCR rates plotted over wild type without G4 motif is depicted in Figure 18B. Most of the GCR assays depicted in Figure 18B were performed by Julia Wille under my supervision. Investigated proteins were the *in vitro* G4 binding proteins Ku70 and Mre11 (Cogoi et al, 2008 ; Ghosal & Muniyappa, 2005) as well as Dot1, Mgs1 and Mms1 (identified in this thesis) (Annex Table 5+6+9). Mms1 is part of two ubiquitin ligases formed with Rtt101 (Zaidi et al, 2008). Rtt101 was also identified in the pull-down experiment and MS as a G4 interacting protein in this thesis (Figure 15, Annex Table 5), but due to the better peptide counts of Mms1 compared to Rtt101 (Annex Table 5), it was decided to examine Mms1 further.

The GCR rate increased significantly in *mre11* (18-fold), *dot1* (1.7-fold) and *mms1* (3.3-fold) cells compared to wild type cells (Figure 18B), but not in *ku70* or *mgs1* cells. Insertion of G4-*LEU2* did not change the GCR rate significantly in wild type, *ku70* and *mgs1* cells compared to wild type cells without G4 motif (Figure 18B). However, insertion of G4-*LEU2* caused a significant enhancement of the GCR rate in *mre11* (~47-fold), *dot1* (1.7-fold) and *mms1* (~20-fold) cells compared to wild type cells without G4 motif (Figure 18B). In *mre11* cells the GCR rate was 2.6–fold higher in the strain with the inserted G4-*LEU2* compared to *mre11* cells without the G4 motif (Figure 18B). This enhancement was ~six-fold in *mms1* cells (Figure 18B). In *dot1* cells no significant difference in the GCR rate was observed between the strain with inserted G4-*LEU2* compared to the strain without the G4 motif (Figure 18B).

From this data it can be concluded that Mre11 and Mms1 are especially important for genome integrity in the presence of a G4 motif. To confirm their role in the specific regulation of G4 motifs compared to other DNA sequences, control regions must be inserted at the *PRB1* locus, and their outcome on genome stability needs to be analyzed (see Figure 28 for results on *mms1*).

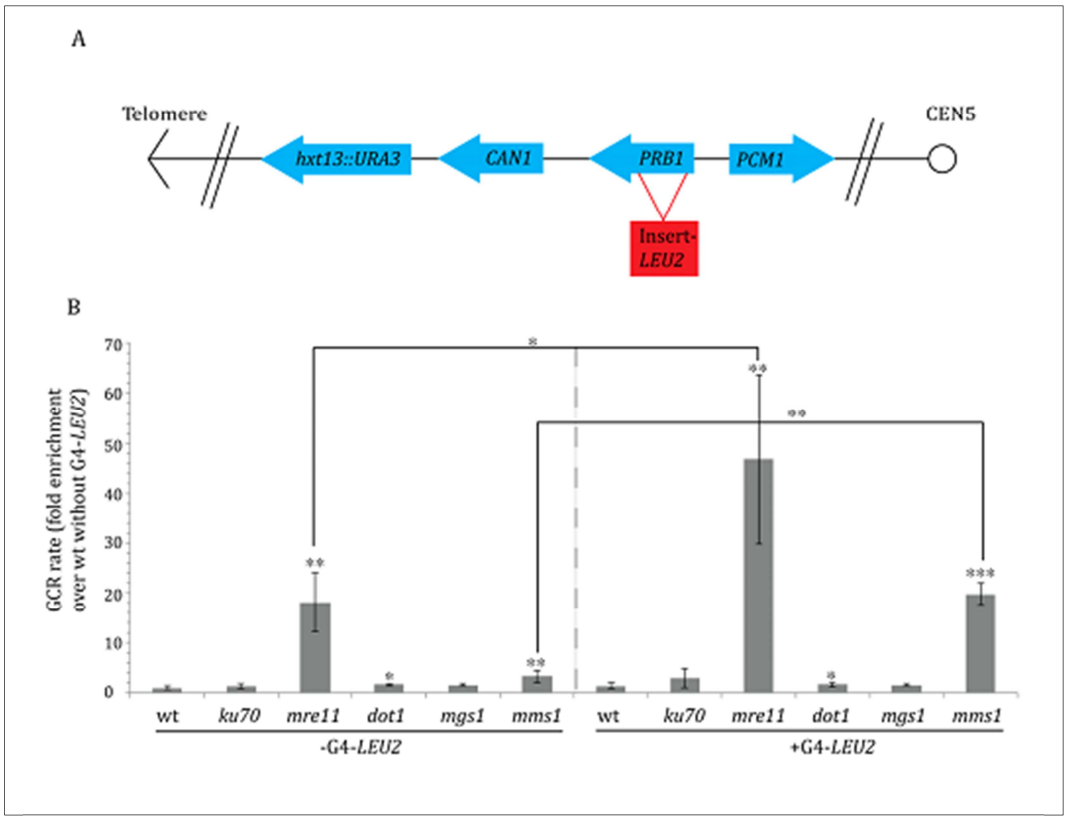


Figure 18: Mre11 and Mms1 are necessary for genome stability, especially in the presence of G4 motifs. (A) Schematic of the genomic region used in the GCR assay. *PCM1* is the last essential gene on the left arm of chromosome V. *PRB*1 is a non-essential gene. If a sequence is inserted at the *PRB1* locus and this sequence induces genome instability, the counter selectable markers *URA3* and *CAN1* are lost. Cells can then grow on selective media. The GCR rate was calculated by fluctuation analysis. (B) The GCR rate was determined for wild type (wt), *ku70*, *mre11*, *dot1*, *mgs1* and *mms1* cells without or with an inserted G4 motif from ChrI (G4-*LEU2*) (Annex Table 12). Shown are mean values from at least three independent experiments ± SD as fold enrichment over wild type without G4 motif. Statistical significance compared to wild type no insert strain was determined by student's T-test. *: p≤0.05, **:p≤0.01, ***:p≤0.001. Further statistical significance is noted in the figure. This was also determined by student's T-test. **:p≤0.01. Most of the GCR assays depicted here were performed by Julia Wille under my supervision.

5.5 γ-H2A-ChIP

To support the observations of the GCR assay or to get more insights into a possible function of the tested proteins at G4 structures, ChIP of endogenous γ-H2A followed by qPCR was performed. As in former studies (McGee et al, 2010; Paeschke et al, 2011) γ-H2A was used as a marker for the occurrence of DSBs. Previously it was shown that G4 structures can result in strand breaks, as evidenced by increased γ-H2A signal (Paeschke et al, 2011; Rodriguez et al, 2012; Sabouri et al, 2014; Zimmer et al, 2016). In addition to that, G4 structures can impede

replication fork progression (Castillo Bosch et al, 2014; Crabbe et al, 2004; Lopes et al, 2011; Paeschke et al, 2011; Sabouri et al, 2014; Sarkies et al, 2010; Schiavone et al, 2014; Wu & Spies, 2016), which might also result in DSBs (reviewed in (Aguilera & Gomez-Gonzalez, 2008)).

Therefore, it was examined in this thesis, whether the repair proteins Ku70 and Mre11 (G4 binding proteins *in vitro* (Cogoi et al, 2008 ; Ghosal & Muniyappa, 2005)) as well as Tel1, Dot1, Mgs1 and Mms1 (*in vitro* G4 binding proteins identified in this thesis, Annex Table 5+6+9) are important for the regulation of G4 structures to prevent DNA DSBs or if they are recruited for the repair of G4-induced DSBs.

Western analysis was performed with the western eluate controls for ChIP (see 4.5.1). Input samples were not taken, because γ-H2A is an endogenous protein. Due to the fact that ChIP procedures where the same for all tested strains, the western analyses of two ChIPs are shown exemplarily in Annex Figure 4E. In western eluate control samples of γ-H2A in wild type and *mre11* cells protein bands were detected at around 20 kDa (14 kDa expected according to manufactures protocol (Abcam)) and the protein was stable during the analysis (Annex Figure 4E). The observed protein band around 55 kDa is due to the heavy chain of the antibody.

Tested were almost the same G4tract3 motifs as in Figure 16+17 (in total five) (Chr $VI_{G4tract3}$, $XI_{G4tract3}$, $XII_{G4tract3}$, $XIII_{G4tract3}$ and $XVI_{G4tract3}$) as wells as two positive control regions (Chr $XIII_{PC\_\gamma\text{-}H2A}$ and $XV_{PC\_\gamma\text{-}H2A}$) and one negative control region (Chr $XIII_{NC\_\gamma\text{-}H2A}$). The G4 motif Chr $XIII_{G4tract3}$ was chosen instead of Chr $IX_{G4tract3}$ (Figure 16+17), because it is located at a previously identified γ-H2A binding site (Capra et al, 2010). Also the G4 motifs Chr $VI_{G4tract3}$, $XI_{G4tract3}$ and $XII_{G4tract3}$, but not the G4 motif Chr $XVI_{G4tract3}$ overlap with a γ-H2A binding site (window ± 500 corresponding to the maximal shearing size, Annex Figure 1A) identified by (Capra et al, 2010). The G4 motif Chr $XVI_{G4tract3}$ was also investigated, because it is possible that G4 motifs lead to a γ-H2A binding at this region in the absence of repair proteins. The positive and negative control regions were chosen from published genome-wide γ-H2A binding sites (Capra et al, 2010). Those control regions do not contain a G4tract3 motif within a window of ± 500 bp (maximal shearing size, Annex Figure 1A).

In *tel1* cells γ-H2A signals were significantly enriched compared to wild type cells at the G4 motifs Chr $XI_{G4tract3}$, $XIII_{G4tract3}$ and $XVI_{G4tract3}$ (1.7-2.0-fold) as well as the positive control site Chr $XIII_{PC\_\gamma\text{-}H2A}$ (1.5-fold) (Figure 19A). At the G4 motif Chr $VI_{G4tract3}$, the positive control region Chr $XV_{PC\_\gamma\text{-}H2A}$ and the negative control region Chr $XIII_{NC\_\gamma\text{-}H2A}$ γ-H2A signal was 1.8-2.1-fold increased, but this effect was not significant due to the high SD (Figure 19A).

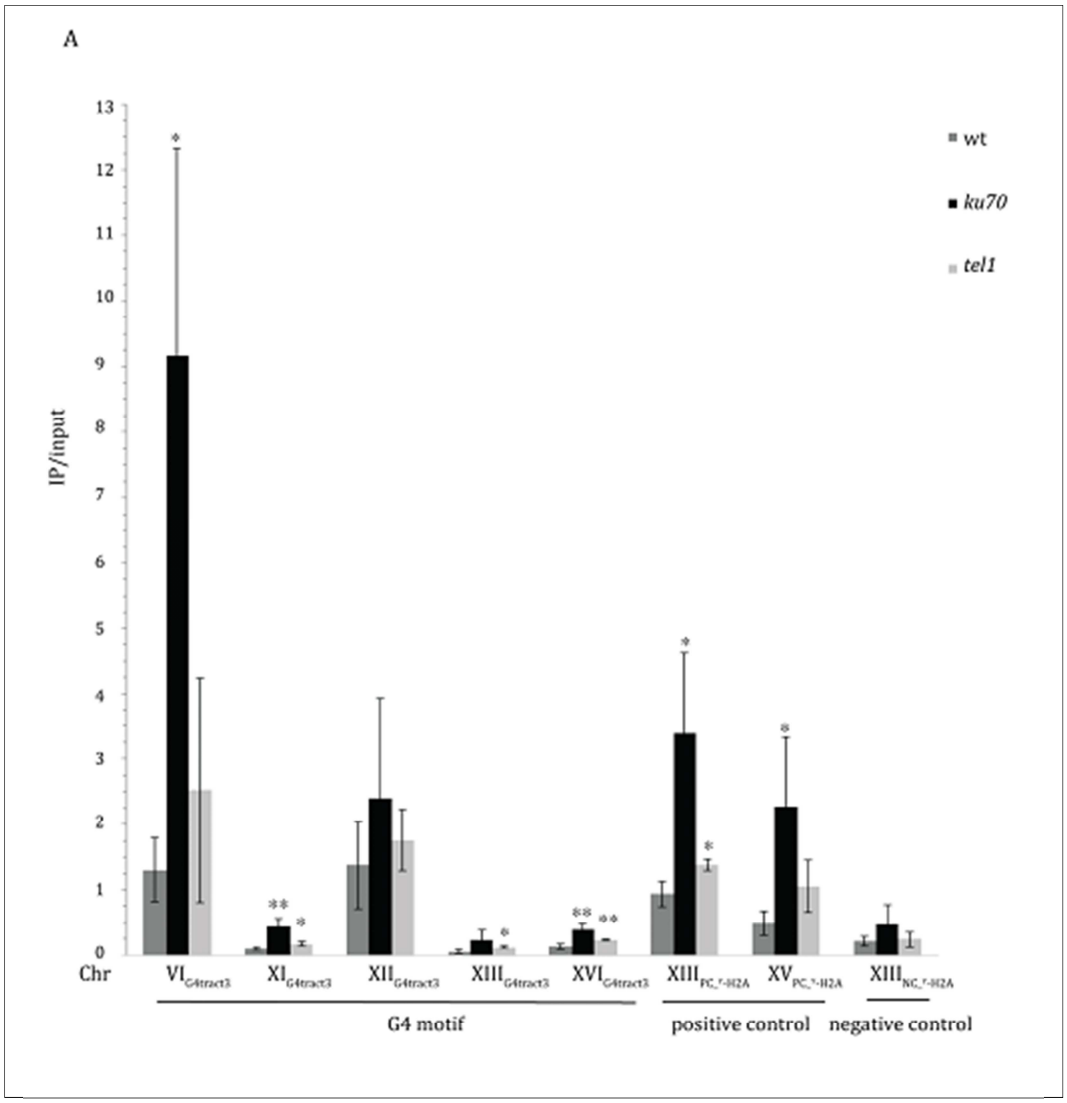
Deletion of *ku70* caused a significant increase of γ-H2A accumulation at the G4 motifs of Chr $VI_{G4tract3}$, $XI_{G4tract3}$ and $XVI_{G4tract3}$ (three-seven-fold) and at both positive control regions (Chr $XIII_{PC\_\gamma\text{-}H2A}$ and $XV_{PC\_\gamma\text{-}H2A}$) (3.6-4.6-fold) (Figure 19A). Also here, the 1.7-3.9-fold increase at the G4 motifs Chr $XII_{G4tract3}$ and $XIII_{G4tract3}$ as well as at the negative control region Chr $XIII_{NC\_\gamma\text{-}H2A}$ was not significant (Figure 19A).

In *mre11* cells a significant increase of γ-H2A signal (2.3-fold) could be detected at the G4 motif Chr $XVI_{G4tract3}$ compared to wild type cells (Figure 19B). At all other tested regions γ-H2A signal tended to increase (1.5-3.3-fold) in *mre11* compared to wild type cells (except for G4 motifs of Chr $VI_{G4tract3}$ and $XIII_{G4tract3}$) (Figure 19B).

In the case of Dot1 and Mms1 γ-H2A accumulation tended to decrease at most G4 motifs and at the positive control regions (Chr $XIII_{PC\_\gamma\text{-}H2A}$ and $XV_{PC\_\gamma\text{-}H2A}$) in the strains with the deleted protein compared to wild type cells (Figure 19B). Significant effects were observed at the G4 motif of Chr $XI_{G4tract3}$ (*dot1*, two-fold) and Chr $VI_{G4tract3}$ (*mms1*, 4.4-fold) and at the positive control region Chr $XIII_{PC\_\gamma\text{-}H2A}$ (*dot1* and *mms1*, 1.6-2-fold) (Figure 19B). In *dot1* cells γ-H2A decreased at the G4 motifs Chr $XII_{G4tract3}$ and $XIII_{G4tract3}$ 1.8-2-fold even though these effects were not significant (Figure 19B). In *mms1* cells γ-H2A signal was reduced at the G4 motifs Chr $XII_{G4tract3}$ and $XVI_{G4tract3}$ 1.8-3.6-fold and at the positive control region Chr $XV_{PC\_\gamma\text{-}H2A}$ 1.9-fold (Figure 19B). Also here, the effects were not significant.

Deletion of *MGS1* did not lead to a consistent effect on γ-H2A signal at the tested G4 motifs (Chr $VI_{G4tract3}$, $XI_{G4tract3}$, $XII_{G4tract3}$, $XIII_{G4tract3}$ and $XVI_{G4tract3}$) (Figure 19B). Only at both positive control regions (Chr $XIII_{PC\_\gamma\text{-}H2A}$ and $XV_{PC\_\gamma\text{-}H2A}$) a significant increase of γ-H2A accumulation (2.3-2.6-fold) was detected in *mgs1* compared to wild type cells (Figure 19B).

Although γ-H2A binding was observed at the G4 motifs Chr $XI_{G4tract3}$ and $XIII_{G4tract3}$ in a former study (Capra et al, 2010), this could not be confirmed in this thesis, because the IP/input values at these regions were in the same range as the IP/input values of the negative control site Chr $XIII_{NC\_\gamma\text{-}H2A}$ (Figure 19).



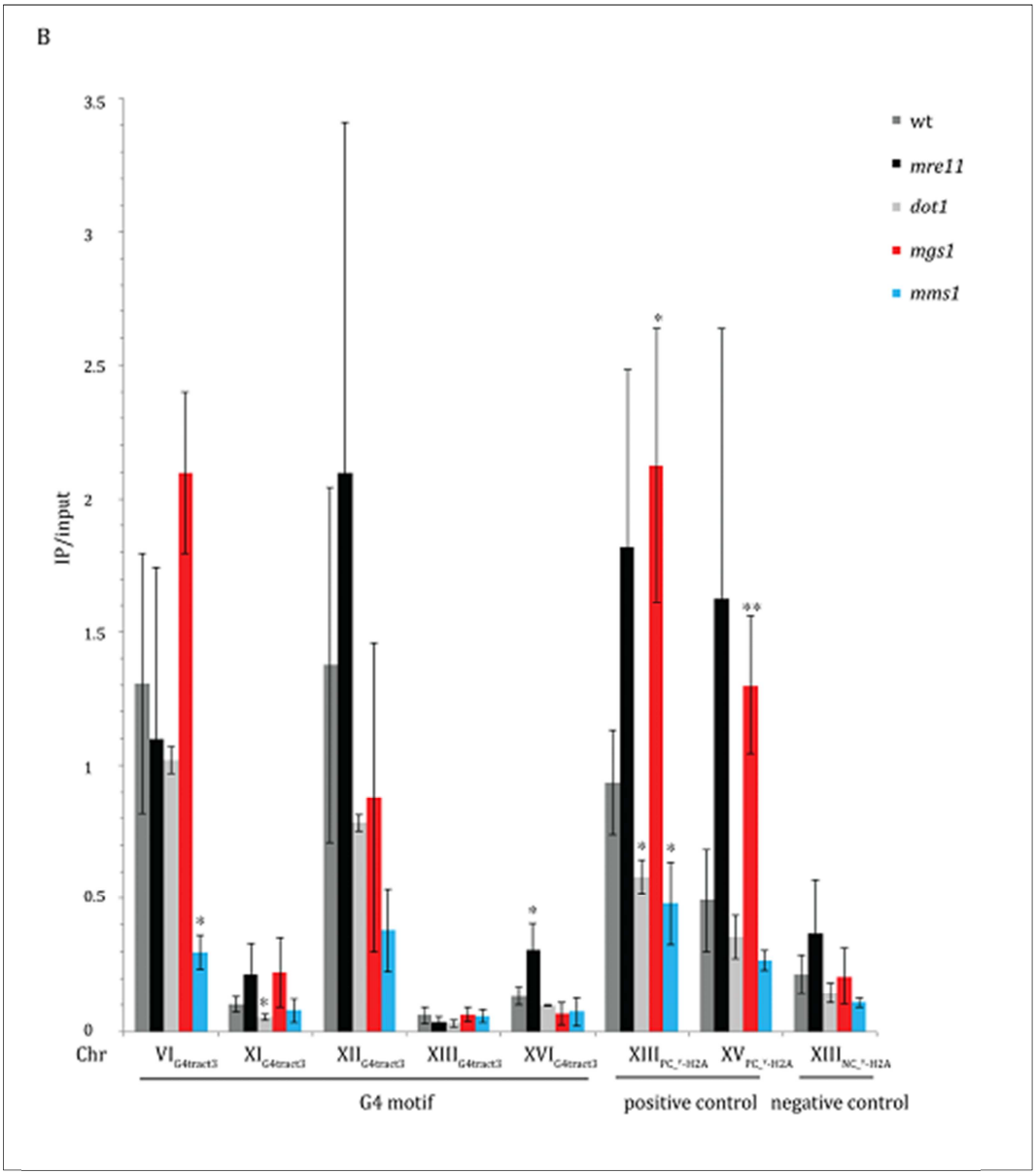


Figure 19: γ-H2A levels change upon deletion of different proteins. (A) ChIP was performed and the binding of endogenous γ-H2A in wild type (wt) (grey), *ku70* (black) and *tel1* (light grey) cells was analyzed by qPCR using primer pairs for the shown regions. As control regions, the enrichment at high ($ChrXIII_{PC\_\gamma\text{-}H2A}$ and $XV_{PC\_\gamma\text{-}H2A}$) or low ($ChrXIII_{NC\_\gamma\text{-}H2A}$) γ-H2A binding sites was determined. Plotted are the IP/input values as means ± SD from three independent experiments. (B) Same as in A only for the strains wt (grey), *mre11* (black), *dot1* (light grey), *mgs1* (red) and *mms1* (blue). Statistical significance compared to wild type cells was determined by student's T-test. *: p≤0.05, **: p≤0.01.

The *in vitro* G4 binding proteins Ku70, Mre11, Tel1, Dot1 and Mgs1 did not show consistent *in vivo* effects at G4 structures (Figure 16-19). Thus, those proteins were not examined further in this thesis. Because of the identification of Mms1 and its binding partner Rtt101 (Zaidi et al, 2008) as G4 binding proteins *in vitro* (Figure 15, Annex Table 5) and because of the observed

in vivo effects of *mms1* cells (Figure 18 and Figure 19B), the effect of Rtt101 and Mms1 at G4 structures *in vivo* was investigated further. Rtt101 and Mms1 form ubiquitin ligase complexes and so far two ubiquitin ligases have been identified that contain Rtt101 and Mms1 (Zaidi et al, 2008). One contains Mms22 ($Rtt101^{Mms1/Mms22}$) as a third component and the other one Crt10 ($Rtt101^{Mms1/Crt10}$). Ctr10 and Mms22 are hypothesized to determine the function of the complex (Zaidi et al, 2008). Ctr10 was demonstrated to be involved in the transcription of RNR genes (Fu & Xiao, 2006). Rtt101, Mms1 and Mms22 promoted replication fork progression and induced HR at stalled replication forks (Duro et al, 2008; Luke et al, 2006; Vaisica et al, 2011; Zaidi et al, 2008). Because of the interest of this thesis to study replication and repair at G4 structures, the ubiquitin ligase $Rtt101^{Mms1/Mms22}$ rather than $Rtt101^{Mms1/Crt10}$ was examined further.

5.6 Function of Mms1 at G4 structures

5.6.1 Mms1 associates genome-wide with G-rich regions

Due to the higher peptide counts of Mms1 compared to Rtt101 in the MS analysis (Annex Table 5), it was assumed that rather Mms1 binds to G4 structures and exerts functions at G4 structures *in vivo*. Also, to my knowledge, Rtt101 is only functional if tagged on its N-terminus, but has not been chipped. Therefore, the genome-wide *in vivo* binding sites of Mms1 were investigated via ChIP followed by genome-wide deep-sequencing analysis (ChIP-seq) using endogenous Myc-tagged Mms1. As in a former publication, Mms1 was tagged at its C-terminus (Vaisica et al, 2011). The functionality of Myc-tagged Mms1 was confirmed utilizing the previous observation that *mms1* cells are highly sensitive towards 0.01% MMS (Hryciw et al, 2002). Performing a spot assay, it was observed that Mms1-Myc cells showed a slight growth defect compared to untagged cells on 0.01% MMS containing plates, but they were not as inviable as *mms1* cells (Annex Figure 5). Furthermore, western blot analysis confirmed the correct size (171 kDa (Mimura et al, 2010), Myc-tag accounts for ~ 16 kDa) and expression of Myc-tagged Mms1 (Annex Figure 4F).

ChIP-seq was performed with asynchronous Mms1-Myc yeast cultures. Western input and eluate controls were taken, but Mms1 was not detectable in any sample. However, western analysis

was performed with another protein (Myc-tagged DNA Pol2) (Annex Figure 4G). In western input and eluate ChIP-seq control samples of Myc-tagged DNA Pol2 protein bands were detected at above 170 kDa (272 kDa expected) and the protein was considered stable during the analysis (positive control, which was obtained from regular protein lysis of Myc-tagged Pol2, also showed some degradation). Due to the fact that ChIP-seq of Myc-tagged Mms1 was performed in the same way as of Myc-tagged Pol2, it is assumed that Mms1 was also stable.

Using MACS 2.0 (Zhang et al, 2008) 71 chromosomal binding sites for Mms1 were identified. 20 binding sites were found in mitochondrial DNA (see Annex Table 13 for identified binding regions). Because of the interest of this thesis, only the 71 chromosomal binding sites were examined further.

Then, Mms1 binding sites were compared to annotated genomic features (γ-H2A binding sites, genes, centromeres, repeats, ARS), to further elucidate to which regions Mms1 binds. With this analysis, it was discovered that only 13 of the binding regions overlap with repeats and six with ARS within a window of ± 400 bp (maximal shearing size, Annex Figure 1B) and none of the binding sites overlapped with a centromere (Annex Table 14). More overlap was found with genes (70 of 71) and with γ-H2A binding sites (30 of 71) (see Annex Table 14 for the one region that did not overlap with a gene and regions that overlapped with γ-H2A binding sites). Tests for significance should be performed for overlap with genes and γ-H2A binding sites. However, at least the overlap with genes probably is highly significant.

Next, MEME-based motif elicitation (Bailey et al, 2009) was performed to look for a consensus binding motif (BM) among Mms1 binding sites. By this, three BMs were discovered (Figure 20A-C). The most significant BM (Figure 20A) (e-value: 2.5 e-66) was G-rich. Half of these BM sites contain a G4 motif with two guanines in the G-tract (G4tract2) either directly in the motif site or in the motif site plus flanking region (see Annex Table 15 for some examples). 23 of the 71 binding sites (32.4%) contain this BM at least once. Interestingly, John A. Capra discovered as part of a collaboration that the average GC-content of the Mms1 binding sites was 50%, which is significantly higher than expected ($p < 0.001$) from the average GC content of the *S. cerevisiae*

genome (~38% GC). The second most significant BM is also GC-rich, while the third is AT-rich (Figure 20B+C). The second and third BMs predominantly occurred in mitochondrial DNA. Therefore, they were skipped from further analysis. The G-rich nature of the Mms1 BM (Figure 20A) suggested that some Mms1 binding sites might have the potential to form G4 structures. Due to the fact that the MEME-based search cannot be used to search for G4 motifs, a script that discovers G4 motifs (Capra et al, 2010) was applied to all 71 chromosomal binding sites. With this script, it was discovered that 11 of the 71 binding sites (15.5%) contain a G4 motif of the consensus sequence $GGG(N)_{\leq 25}GGG(N)_{\leq 25}GGG(N)_{\leq 25}GGG$ (G4tract3) (Annex Table 16). A G4 motif with the same consensus sequence was used before to analyze the genome-wide occurrence of G4 motifs in *S. cerevisiae* (Capra et al, 2010). Because the BM contained many GG di-nucleotides (Figure 20A) and G4 structures with two G-tetrads were also shown to fold (Chambers et al, 2015; Qin et al, 2015; Siddiqui-Jain et al, 2002; Hazel et al, 2004), it was searched for a G4tract2 motif ($GG(N)_{\leq 7}GG(N)_{\leq 7}GG(N)_{\leq 7}GG$) among Mms1 binding sites as well. With this search, it was revealed that 61 of the 71 binding sites (86%) contain a G4tract2 motif (Annex Table 17). However, John A. Capra elucidated that there were not significantly more G4tract3 or G4tract2 motifs among the binding sites than expected by chance when taking into account the GC-richness of the binding sites.

To further examine the binding characteristics of Mms1, especially the potential contribution of G4 motifs, ChIP-qPCR experiments were performed with the Myc-tagged Mms1 strain. In western eluate conventional ChIP control samples of Myc-tagged Mms1 protein bands were detected at around 171 kDa as expected and the protein was stable during the analysis (Annex Figure 4H). For the qPCR, twelve different regions in the yeast genome were selected. Out of these twelve regions, three regions were selected from ChIP-seq as Mms1 binding sites and seven were chosen based on a G4tract2 occurrence. In addition, based on the ChIP-seq data, two negative control regions were selected.

In Figure 20D binding of Mms1 is depicted as enrichment over untagged control. Here and in the following ChIP experiments, three-fold enrichment over the untagged control was considered as

binding. With this analysis ChIP-seq data was confirmed. As expected Mms1 bound to the three binding regions (BR) selected by ChIP-seq, among which were two regions containing the MEME specific BM (Chr VII_{BM}, X_{BR}, XI_{BM}) (Figure 20A+D, Annex Table 15). Regions that contain the BM are marked with BM, while regions that Mms1 bound to, but do not contain the BM are marked with BR. Interestingly, also the regions Chr VI_{BR}, IX_{BR}, XIa_{BR}, XIb_{BR}, $XIII_{BR}$ and XV_{BR} were identified as Mms1 binding regions that were not identified earlier by ChIP-seq (Figure 20D, Annex Table 13). Using the cut-off of three-fold enrichment over untagged control, Mms1 did not bind to the region on Chr I, XIV and an additional region from Chr XIII. Due to the lack of significant Mms1 binding during ChIP-seq and/or ChIP-qPCR those three regions are considered as negative controls (NC) in the following analyses (Chr I_{NC}, $XIII_{NC}$, XIV_{NC}) (Figure 20D). To find out why Mms1 binds to some regions and not to others, the tested regions were analyzed further. This analysis revealed that the regions Mms1 bound to (Chr VI_{BR}, VII_{BM}, IX_{BR}, X_{BR}, XIa_{BR}, XIb_{BR}, XI_{BM}, $XIII_{BR}$, XV_{BR}) contain one or many G4tract2 or even G4tract3 motifs (see Annex Table 18 for details). However, this was also the case for the negative control regions Chr I_{NC} and $XIII_{NC}$.

Taking into account the distance to the next ARS and the strand location of the G4tract2 motif, it was discovered that all Mms1 binding regions contain a G4tract2 motif with a mean loop length smaller than eight nucleotides located on the lagging strand, while the negative control regions have either not such a G2tract2 motif or non on the lagging strand (Annex Table 19).

To determine whether the G-rich binding regions, G4tract2, can fold into G4 structures *in vitro*, CD on three of the 71 binding regions of Mms1 was performed. These measurements confirmed that the G4tract2 motifs from the binding regions Chr VII_{BM}, XIa_{BR} (mean loop length <eight nt, lagging strand) and XV_{BR} (mean loop length <eight nt, leading strand) folded into G4 structures *in vitro* (Figure 20E), while the G4tract2 from the negative control regions Chr I_{NC} (mean loop length >eight nt, leading strand) and $XIII_{NC}$ (mean loop length >eight nt, lagging strand) did not (Figure 20F).

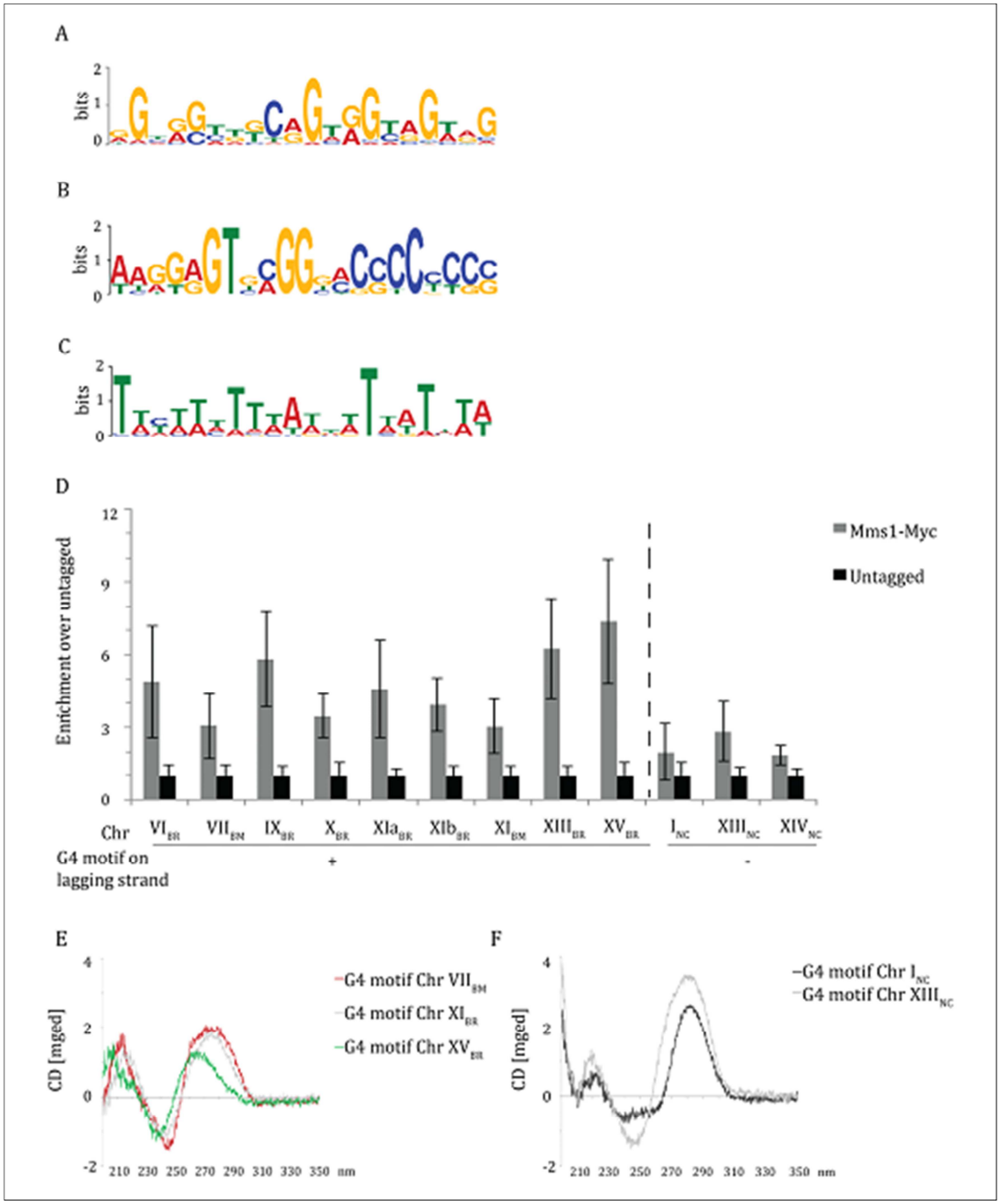


Figure 20: Mms1 binds to G-rich regions *in vivo*. (A-C) BMs of Mms1 identified by MEME-ChIP search. (D) Conventional ChIP was performed and the association of Myc-tagged Mms1 was analyzed by qPCR using primer pairs for the shown regions. Plotted is the enrichment over untagged control as mean value ± SD from at least three independent experiments. Nine regions contain a G4tract2 motif on the lagging strand (Chr VI_{BR}, VII_{BM}, IX_{BR}, X_{BR}, XIa_{BR}, XIb_{BR}, XI_{BM}, $XIII_{BR}$ and XV_{BR}), two regions only contain such a G4 motif on the leading strand (Chr I_{NC}, $XIII_{NC}$) and the region ChrXIV_{NC} contains no G4 motif at all. At three of the tested regions peaks were called by MACS (ChrVII_{BM}, X_{BR}, XI_{BM}) (For more information see Annex Table 18). (E+F) Oligodeoxynucleotides of some G4 motifs in the binding and no binding regions were folded *in vitro*. CD measurements were used to determine possible G4 formation. Shown are the CD [mdeg] values for the lagging strand G4 motifs from Chr VII_{BM} (D, red) and Chr XI_{BR} (D, grey) and the leading strand G4 motif from Chr XV_{BR} (D, green) as well as leading strand G4 motif from Chr I_{NC} (E, grey) and lagging strand G4 motif from Chr $XIII_{NC}$ (E, black).

These measurements were performed before it was analyzed that the lagging strand G4 motifs seem to be required for Mms1 binding. Hence, the CD measurements were repeated with the lagging strand G4 motifs of Chr I_{NC} and XV_{BR}. As expected, Chr I_{NC} (mean loop length >eight nt, lagging strand) did not fold into a G4 structure (Annex Figure 6). Unexpectedly, also the region Chr XV_{BR} (mean loop length <eight nt, lagging strand) did not fold into a structure (Annex Figure 6). Nevertheless, most of these result further supported the observation that Mms1 binds to G-rich/G4 motifs, which harbor two or three guanines in the G-tract (Annex Table 15-17).

5.6.2 Mms1 binds independently of Rtt101 and Mms22

Because Mms1 is a component of the $Rtt101^{Mms1/Mms22}$ and $Rtt101^{Mms1/Crt10}$ ubiquitin ligase (Zaidi et al, 2008), one of the aims of this thesis was to determine if the other ligase components are required for Mms1 binding to G-rich/G4 motif regions. So far the only process Crt10 was shown to be involved in is RNR gene expression (Fu & Xiao, 2006), while Mms22 and Mms1 together with Rtt101 ($Rtt101^{Mms1/Mms22}$) promote replication fork progression and HR at stalled replication forks (Duro et al, 2008; Luke et al, 2006; Vaisica et al, 2011; Zaidi et al, 2008). Hence, it was hypothesized that at the Mms1 binding regions identified in this thesis most likely the $Rtt101^{Mms1/Mms22}$ complex is active. In a former publication it was hypothesized that Mms22 is the DNA interacting protein of the $Rtt101^{Mms1/Mms22}$ ligase (Zaidi et al, 2008). Therefore, it was elucidated if Mms1 binding is dependent on Rtt101 or Mms22.

For this, Mms1 binding was analyzed in *rtt101* and *mms22* cells using ChIP-qPCR experiments. It was observed with this analysis that binding of Mms1 did not depend on Rtt101, but interestingly binding was significantly enriched in *mms22* cells (Figure 21A). At all tested regions (Chr VI_{BR}, VII_{BM}, IX_{BR}, XIa_{BR}, XI_{BM}, $XIII_{BR}$, XV_{BR}, I_{NC}, $XIII_{NC}$ and XIV_{NC}) at least a two-fold enhanced binding of Mms1 was observed in *mms22* compared to wild type cells. As show in Figure 21A, this was even the case for former non-Mms1 binding sites (NC). This data indicates that Mms22 and Rtt101 are not required for Mms1 binding at the G-rich/G4 motifs. Furthermore it indicates that Mms22 prevents Mms1 from binding to these regions.

Next, it was examined whether the observed difference in binding could be explained by an altered Mms1 protein level. Western blot analysis of Myc-tagged Mms1 was performed (Annex Figure 7A) and the protein levels of Mms1 in wild type, *rtt101*, and *mms22* cells was quantified using Hsp60 as a reference protein. Mms22 had no influence of the protein level of Mms1, but the Mms1 protein level significantly increased in *rtt101* cells (~four-fold) (Figure 21B).

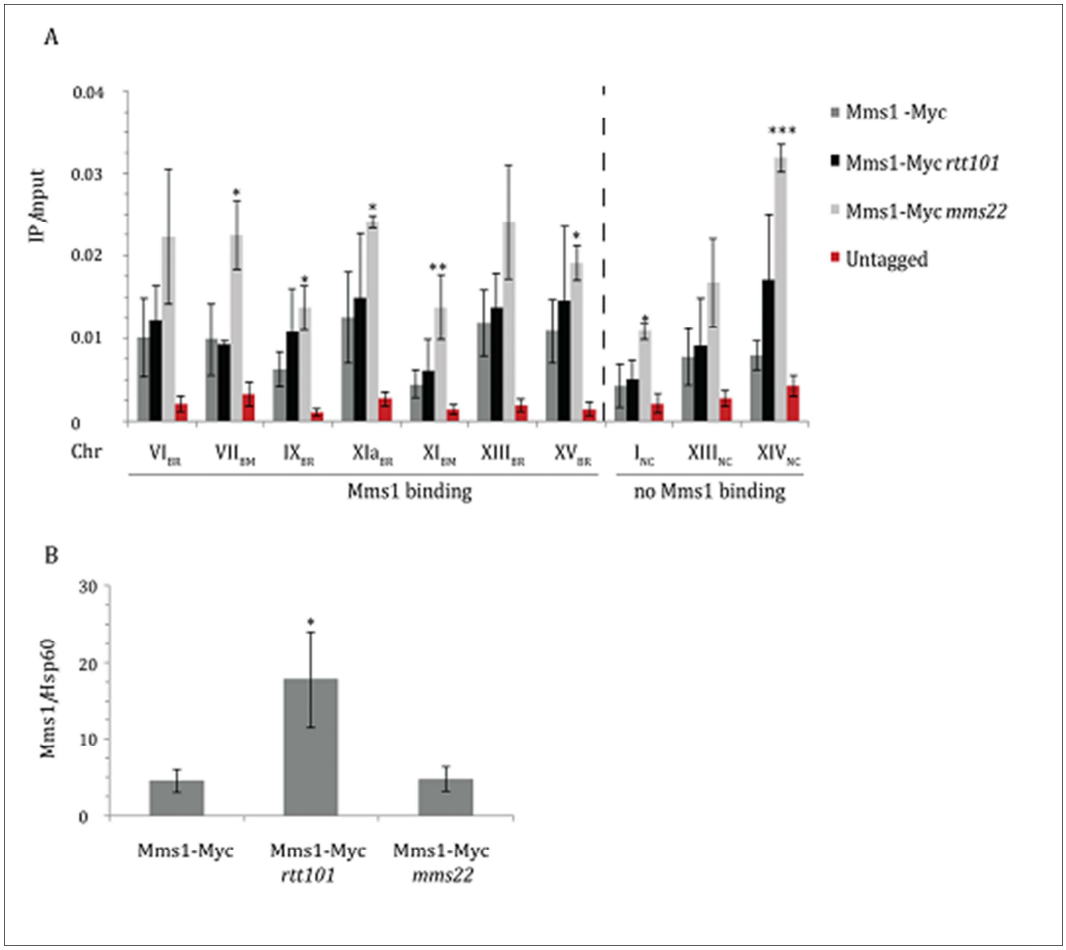


Figure 21: Binding of Mms1 is not dependent on Rtt101 and Mms22. (A) ChIP was performed and the binding of Myc-tagged Mms1 in wild type (dark grey), *rtt101* (black) and *mms22* (light grey) cells was analyzed by qPCR using primer pairs for the shown regions. As a control the results for the untagged strain are displayed (red). Plotted are the IP/input values as means ± SD from at least three independent experiments. For details about the tested regions see Figure 20D and Annex Table 18. Statistical significance compared to Myc-tagged Mms1 wild type cells was determined by student's T-test. *: p≤0.05, **: p≤0.01, ***: p≤0.001. (B) Myc-tagged Mms1 protein levels were quantified in wild type, *rtt101* and *mms22* cells after western blot analysis using HIsp60 (Abcam) as a reference protein. Mean Myc-tagged Mms1 levels normalized to Hsp60 ± SD from three independent experiments are plotted. Statistical significance compared to Myc-tagged Mms1 wild type cells was determined by student's T-test. *: p≤0.05.

5.6.3 Mms1 levels are highest in G1 phase and it binds throughout cell cycle to its target regions

Towards further elucidating the function of Mms1 at G-rich/G4 motifs, it was analyzed in which cell cycle phase Mms1 binds to its target regions. First the protein levels of Mms1 in different cell cycle phases were determined. For this yeast cells expressing Myc-tagged Mms1 were arrested in G1, S, and G2 phase (α factor, HU, and nocadozole, respectively, were used to arrest the cells). FACS analysis confirmed the cell cycle arrests (Figure 22A) and western blot analysis directed against endogenous Myc-tagged Mms1 was utilized to determine the protein levels in each phase (Annex Figure 7B). The highest levels of Mms1 were observed in G1 phase, but Mms1 was also detectable in S and G2 phase (5.3-fold less Mms1 than in G1) (Figure 22B).

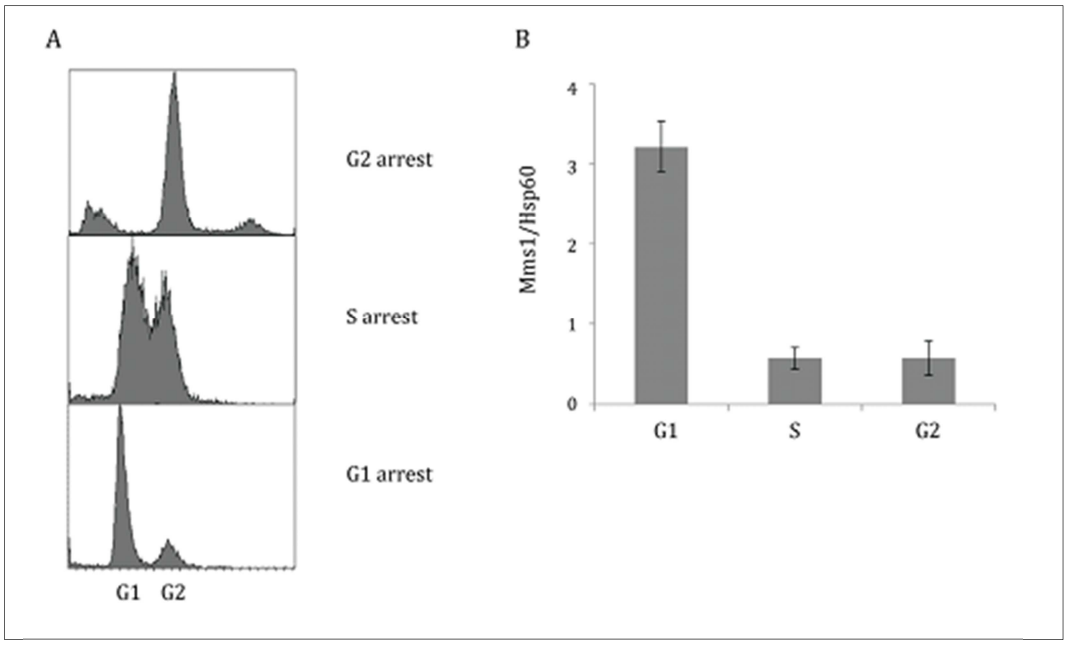


Figure 22: The highest Mms1 protein level is observed in G1 phase. (A) Displayed are the results from FACS analysis after cells were arrested in G1, S or G2 phase. Cells were arrested in G1, S and G2 phase after treatment with α-factor, HU or nocadozole respectively. (B) Myc-tagged Mms1 protein levels were determined in G1, S and G2 phase through western blot analysis and quantification using Hsp60 (Abcam) as a reference protein. Mean Myc-tagged Mms1 levels normalized to Hsp60 ± SD from three independent experiments are plotted.

To address the question when Mms1 binds to its target regions, first yeast cells expressing both HA-tagged DNA Pol2 and Myc-tagged Mms1 were synchronized. DNA Polymerase 2 (DNA Pol2) is the catalytic subunit of DNA polymerase ε with a molecular weight of 256 kDa (Hamatake et al, 1990; Morrison et al, 1990). As done in a former study (Paeschke et al, 2011) the binding of

Mms1 compared to DNA Pol2 was determined to investigate if Mms1 binds prior or after the replication machinery. The expression and correct size of above 180 kDa (259 kDa expected) of HA-tagged Pol2 was confirmed by western blot analysis. The positive control of another HA-tagged Pol2 strain showed the same additional band below 180 kDa (Annex Figure 4I). The arrest of cells in G1 phase and the timely progression through S phase after the release from G1 was confirmed by FACS analysis (Annex Figure 8). Then the binding properties of both proteins throughout the cell cycle were analyzed by ChIP-qPCR. At each time point, Mms1 and DNA Pol2 binding was monitored (Annex Figure 9). The S phase was reached at 30-45 min after release (Annex Figure 8), which correlates with the highest observed DNA Pol2 signal (Annex Figure 9). The results for three tested sites are displayed in Annex Figure 9. Here the IP/input value for each triplicate is shown separately. The binding profile of Mms1 highly differs among the triplicates (Annex Figure 9). Also, the IP/input values are not consistent between the triplicates, even not in case of the Pol2 binding, which would lead to a very high standard deviation if mean values were plotted. A reason for this might be small differences in cell synchrony. Additionally, ChIP of Mms1 might be difficult, especially in synchronized cells, because of the low abundancy of this protein (Kulak et al, 2014). Therefore, no conclusion can be made about a preferred binding of Mms1 to its target sites in any phase of the cell cycle.

Because it was not clear from synchronous ChIP data in which phase of the cell cycle Mms1 binds to its target regions, ChIP-qPCR of Myc-tagged Mms1 cells arrested in G1, S and G2 phase was performed (α factor, HU, and nocadozole, respectively, were used again for the arrests, see Annex Figure 10 for FACS analysis). qPCR at seven Mms1 target regions Chr VI_{BR}, IX_{BR}, X_{BR}, XIa_{BR}, $ChrXI_{BM}$, $XIII_{BR}$ and XV_{BR} demonstrated that Mms1 binds throughout the cell cycle to its target sites (Figure 23). This is also in line with the results from synchronized ChIP (Annex Figure 9), where no consistent peak of Mms1 binding was observed throughout the cell cycle.

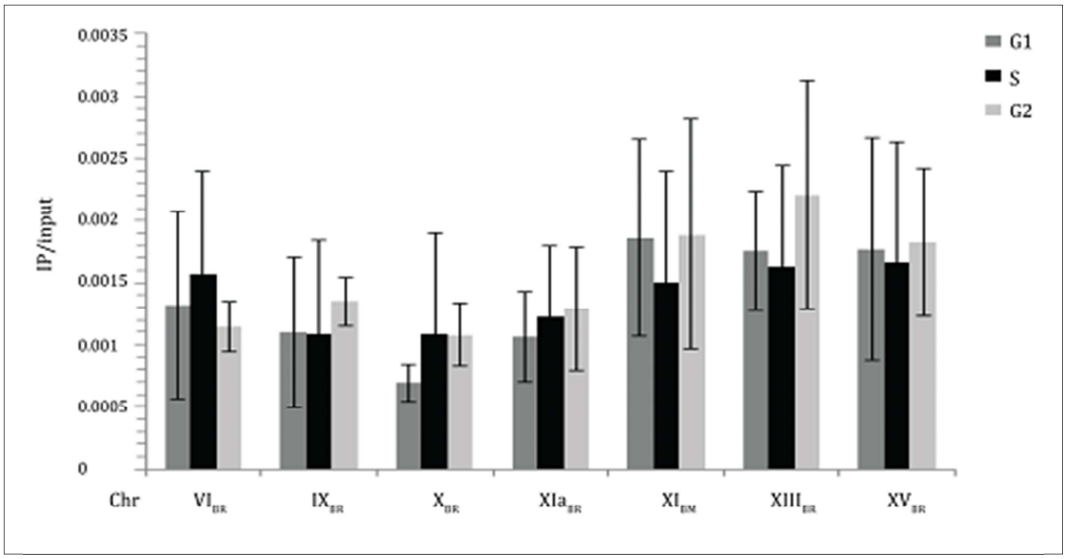


Figure 23: Mms1 binds in all cell cycle phases to its target sites. ChIP was performed of cells arrested in G1, S and G2 phase. Plotted are IP/input values as means ± SD from at least three independent experiments. For details about the tested regions see Figure 20D and Annex Table 18.

5.6.4 Mms1 has an influence on DNA replication fork progression

Former work had shown that DNA replication is slowed in *mms1* cells after MMS or HU treatment, which stalls replication forks (Vaisica et al, 2011; Zaidi et al, 2008). Due to this strong connection of Mms1 function to DNA replication, it was determined if replication forks stall at Mms1 binding sites without addition of drugs. As done previously (Azvolinsky et al, 2006; Paeschke et al, 2011), it was assumed that replication forks move more slowly at regions with elevated DNA Pol2 levels. Therefore, ChIP-qPCR of wild type and *mms1* cells that express endogenous Myc-tagged DNA Pol2 was performed. DNA Pol2-Myc was expressed and a protein band was observed above 170 kDa (272 kDa expected) (Annex Figure 4J).

DNA Pol2 levels were plotted over the untagged control (Figure 24). In wild type cells, binding of DNA Pol2 was observed at three Mms1 binding regions (Chr VI_{BR}, $XIII_{BR}$, and XV_{BR}), as well as one Mms1 negative control region (Chr $XIII_{NC}$) (Figure 24).

Interestingly, DNA Pol2 associated significantly (1.8- to 2.5-fold) more with the regions Chr VI_{BR}, X_{BR}, XIb_{BR}, and XV_{BR} in *mms1* cells compared to wild type cells. The association of DNA Pol2 with the region Chr $XIII_{BR}$ was also enhanced (1.6-fold) in *mms1* cells compared to wild type cells, but this enhancement was not significant (p=0.16). In wild type cells, DNA Pol2 did not associate

with the regions Chr X_{BR}, XIb_{BR}, and XIV_{NC} (enrichment/untagged below three-times), but in *mms1* cells DNA Pol2 binding was observed and elevated at the Mms1 binding regions Chr X_{BR} and XIb_{BR}, but as expected not at the negative control region Chr XIV_{NC} (Figure 24).

In summary, at all Mms1 binding regions DNA Pol2 levels were enhanced in the absence of Mms1, indicating that replication fork progression is impeded at these regions if Mms1 is not present. At the sites Chr $XIII_{NC}$ and XIV_{NC}, DNA Pol2 association did not depend on Mms1, which correlates with the fact that Mms1 did not bind to those regions (see Figure 20D).

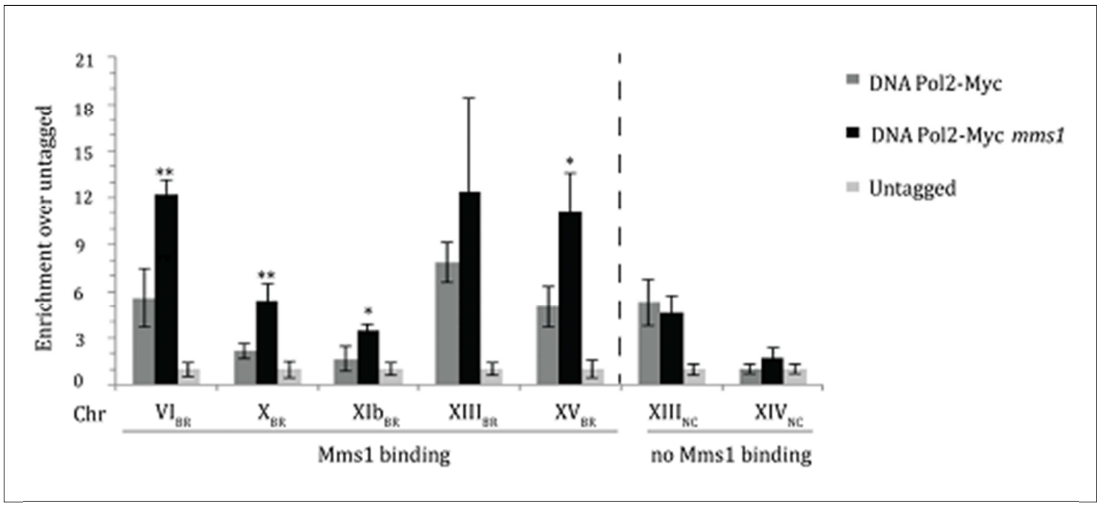


Figure 24: Association of DNA Pol2 is enhanced in *mms1* cells. ChIP of wild type (grey) and *mms1* (black) Myc-tagged DNA Pol2 cells was performed and the binding of DNA Pol2 was determined by qPCR using primer pairs for the shown regions. Plotted is the enrichment over untagged control as mean value ± SD from at least three independent experiments. The results of the untagged strain are shown as a control (light grey). For details about the tested regions see Figure 20D and Annex Table 18. Statistical significance compared to Myc-tagged DNA Pol2 wild type cells was determined by student's T-test. *: p≤0.05, **: p≤0.01.

5.6.5 Mms1 does not influence Mre11 binding

Former studies demonstrated that Mms1 is necessary for HR at stalled replication forks (Duro et al, 2008). Recently, scientists revealed that HR proteins, such as Rad51 and BRCA1, are important to regulate HR at G4 structures during DNA replication (Zimmer et al, 2016). Additionally, it was elucidated in yeast that breakage near a stalled fork induced recombination (Lopes et al, 2011; Paeschke et al, 2011). Hence, in the following experiment it was determined if Mms1 binding to its target regions also causes recruitment of HR factors due to stalled forks. Therefore, it was examined whether Mms1 is required for binding of Mre11, a component of the

MRX complex involved in HR (reviewed in (Heyer et al, 2010)). ChIP of endogenous Myc-tagged Mre11 in wild type and *mms1* cells was performed. Binding of Mre11 at some Mms1 binding sites (Figure 20D) was analyzed by qPCR, and the IP/input values are displayed in Figure 25. It was discovered that binding of Mre11 to all tested Mms1 binding regions was independent of Mms1 (Figure 25). This indicates that Mms1 does not recruit HR factors to the target sites identified in this thesis.

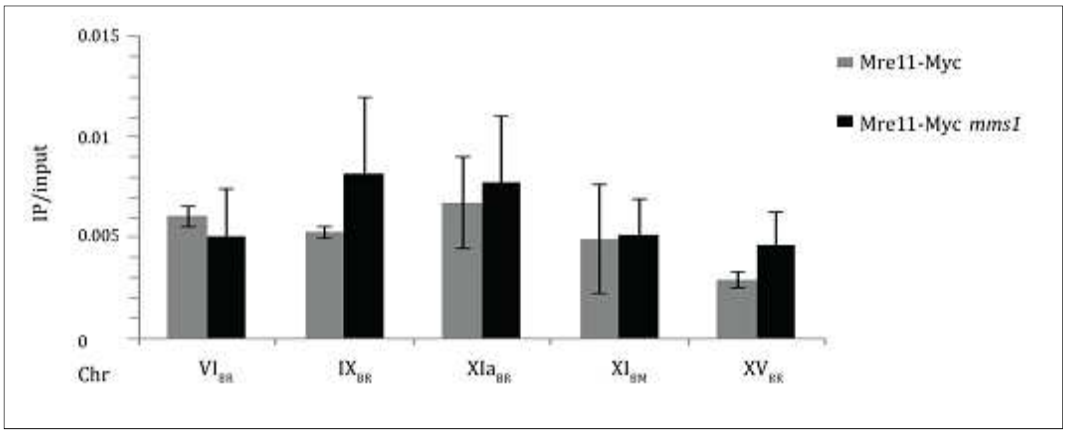


Figure 25: Binding of Mre11 is independent of Mms1. ChIP of wild type (grey) and *mms1* (black) Myc-tagged Mre11 cells was performed and the binding of Mre11 was determined by qPCR using primer pairs for the shown regions. Plotted are the IP/input values ± SD from at least three independent experiments. For details about the tested regions see Figure 20D and Annex Table 18.

5.6.6 Pif1 helicase does not influence Mms1 binding

Pif1 DNA helicase was shown to unwind G4 structures during DNA replication and by this to promote genome integrity (Lopes et al, 2011; Paeschke et al, 2011). Binding of DNA Pol2 to G4 motifs was increased in *pif1-m2* mutants (Paeschke et al, 2011) and *in vitro* assays revealed that DNA replication is stalled at G4 structures (Eddy et al, 2015; Howell et al, 1996; Weitzmann et al, 1996; Woodford et al, 1994). Additionally the Rtt101$^{Mm1/Mms22}$ ubiquitin ligase was required for fork progression after fork stalling (Luke et al, 2006; Vaisica et al, 2011; Zaidi et al, 2008). Therefore, it was examined whether the absence of Pif1 and the consequent replication fork slowing leads to the recruitment and enhanced association of Mms1. For this the association of Mms1 with G-rich/G4 motifs was determined in the absence of Pif1.

ChIP-qPCR with endogenous Myc-tagged Mms1 wild type and *pif1-m2* cells was performed. IP/input values are shown for all the tested regions (Chr VII_{BM}, IX_{BR}, XIa_{BR}, XI_{BM}, $XIII_{BR}$, XV_{BR}, I_{NC} and $XIII_{NC}$) in Figure 26. Note, all of the tested regions were previously identified as Pif1 binding sites (Paeschke et al, 2011). It was discovered that the association of Mms1 with the tested regions was not altered in *pif1-m2* compared to wild type cells (Figure 26). As expected, Mms1 binding to non-Mms1 binding regions (Chr I_{NC}+$XIII_{NC}$) did not change in *pif1-m2* cells (Figure 26).

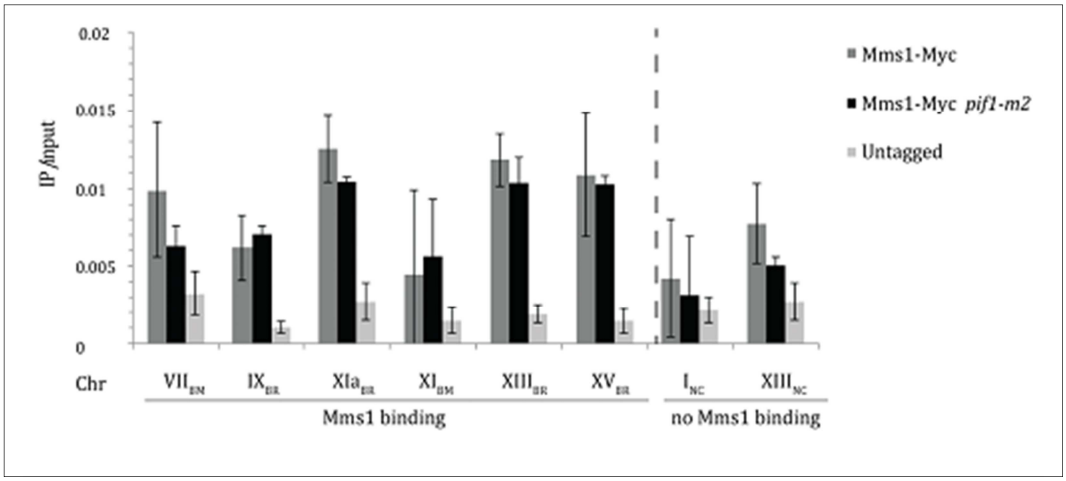


Figure 26: Binding of Mms1 occurs independently of Pif1. ChIP of wild type (grey) and *pif1-m2* (black) Myc-tagged Mms1 cells was performed and the binding of Mms1 was determined by qPCR using primer pairs for the shown regions. Plotted are the IP/input values ± SD from at least three independent experiments. The results from the untagged strain are shown as a control (light grey). For details about the tested regions see Figure 20D and Annex Table 18.

5.6.7 Pif1 binding at Mms1 target sites is dependent on Mms1

Pif1 binds at the end of S phase to G4 motifs genome-wide (Paeschke et al, 2011). In this thesis it was revealed that Mms1 associates throughout the cell cycle with G-rich/G4 motifs. In the absence of Pif1 (Lopes et al, 2011; Paeschke et al, 2011) as well as in *mms1* cells (this thesis) the replication fork pauses at such sites. Hence, it was hypothesized that Mms1 promotes binding of Pif1 to G4 motifs/structures. To confirm this hypothesis, it was first examined if the genome-wide Pif1 binding sites (Paeschke et al, 2011) overlap with Mms1 binding sites (Annex Table 20). This computational analysis was performed by John A. Capra as part of collaboration. 38 of the 71 chromosomal Mms1 binding regions (~ 54%) overlapped with Pif1 binding sites (Annex Table 20). John A. Capra discovered that this is significantly more overlap than expected if these

sites were randomly distributed across the genome (P = 0.001). To further examine the connection of Pif1 and Mms1, ChIP-qPCR was performed using Myc-tagged Pif1 wild type and *mms1* cells. These ChIP-qPCR experiments were performed by Katrin Paeschke. If the hypothesis that Mms1 promotes Pif1 binding is correct, binding of Pif1 should decrease in *mms1* compared to wild type cells.

Indeed, association of Pif1 with Mms1 binding sites was more than two-fold decreased in *mms1* compared to wild type cells (Figure 27). In contrast, association of Pif1 with three different control regions did not change upon *MMS1* deletion. As control regions two Pif1 binding regions, one telomeric region (Chr $VI_{Tel\_VI\text{-}R}$) (Phillips et al, 2015) and the RFB at the rDNA (Chr XII_{rDNA}) (Ivessa et al, 2000), as well as one non Pif1 binding site (tRNA) (Chr VI_{tRNA}) (Paeschke et al, 2011) were used (Figure 27).

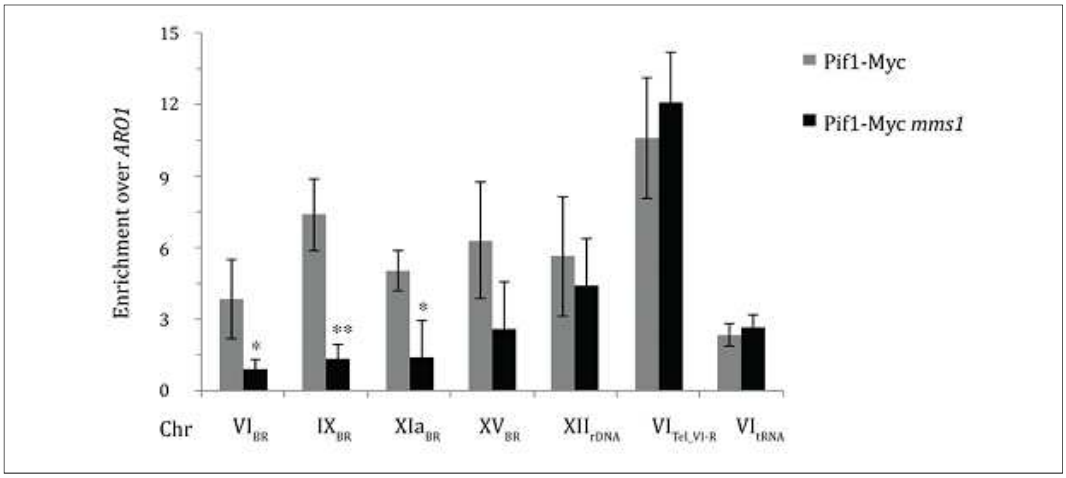


Figure 27: Binding of Pif1 depends on Mms1. ChIP of wild type (grey) and *mms1* (black) Myc-tagged Pif1 cells was performed and the binding of Pif1 was determined by qPCR using primer pairs for the shown regions. Plotted are mean values as enrichment over *ARO1* ± SD from at least three independent experiments. For details about the tested regions see Figure 20D and Annex Table 18. Statistical significance compared to wild type cells was determined by student's T-test. *: p≤0.05, **: p≤0.01. This experiment was performed by Katrin Paeschke.

5.6.8 Mms1 maintains genome stability

It was elucidated that G4 motifs overlap with mitotic DSB sites in *S. cerevisiae* (Capra et al, 2010) and that G4 motifs can lead to breaks, if Pif1 is absent (Paeschke et al, 2011; Piazza et al, 2015; Ribeyre et al, 2009). As demonstrated in this thesis, replication fork progression is slowed at G-rich/G4 motif regions in *mms1* cells (Figure 24). Hence, it was tested if replication fork

progression is impeded due to G4 structures in *mms1* cells, which hampers genome integrity. In order to examine the impact of Mms1, and also of the other two ligase components Rtt101 and Mms22, on genome stability at specific loci, the GCR assay was performed again.

For details about the assay see sections 4.4 and 5.4. For this assay, a G4 motif from Chr I (G4-*LEU2*), one G-rich region from Chr I (GR-*LEU2*), one non-G-rich region from Chr VII (NG-*LEU2*) and the sequence of the *LEU2* marker was inserted into the yeast genome at the *PRB1* locus (Figure 18, see Annex Table 12 for detailed information of regions and sequences and for sequencing results that confirm the correct sequences). The *LEU2* marker was inserted, because *LEU2* was used as a selection marker when inserting the different regions and the *LEU2* gene contains many G4tract2 motifs on the lagging strand that are potential Mms1 target sites based on ChIP and ChIP-seq data. The G4 motif region (G4-*LEU2*) contains an additional G4tract3 on the leading strand.

The GCR rates of wild type, *mms1*, *rtt101* and *mms22* cells are plotted as fold-enrichment over wild type without insert in Figure 28A. Deletions of *mms1* caused an enhanced (3.3-fold over wild type without insert) GCR rate. This was also the case in *rtt101* cells (1.9-fold enrichment) and *mms22* cells (2.9-fold enrichment) (Figure 28A).

In the following figures (24B-E) the GCR rate is plotted as fold-enrichment over the respective strain without insert. In wild type cells, the GCR rate did not increase significantly if a specific region was inserted (Figure 28B). In *mms1* cells, insertion of the *LEU2* marker caused a significant 2.7-fold increase of the GCR rate compared to *mms1* cells without insert. A similar fold increase was observed if the GR-*LEU2* (3.8-fold) or NG-*LEU2* (2.5-fold) insert was present.

In case of G4-*LEU2*, the enhancement was 5.8-fold. The GCR rate was significantly increased if the G4-*LEU2* was present compared to the *LEU2* marker (1.7-fold) (Figure 28C). This was not true for the G-rich and non-G-rich inserts. As noted previously the *LEU2* gene contains G4tract2 motifs on the lagging strand, therefore increased GCR rates were expected for all inserts. Interestingly, insertion of an additional G4 motif (G4tract3) increased the GCR rate in *mms1* cells even further.

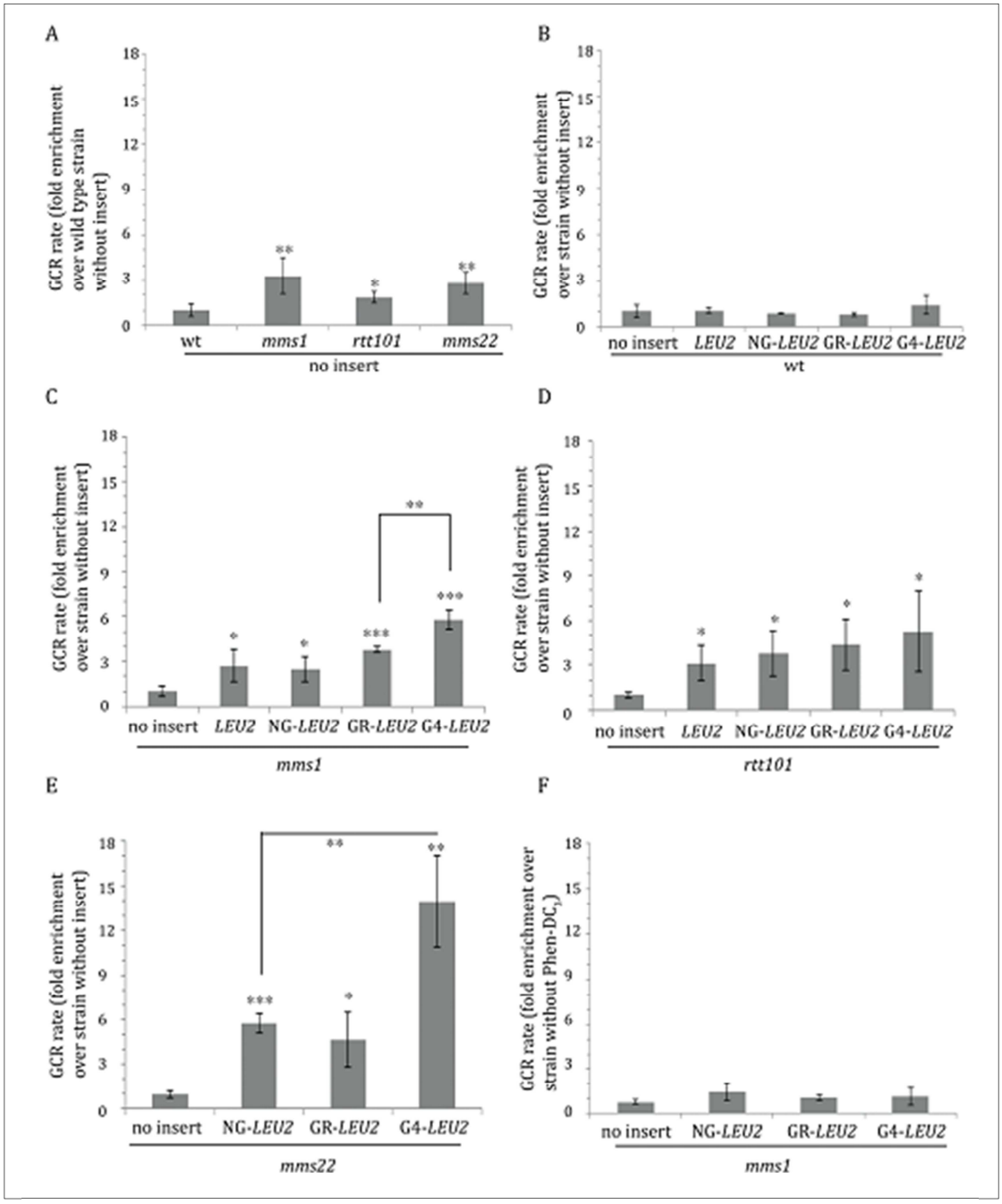


Figure 28: Mms1 and Mms22 are required for genome stability, especially at G4 motifs. (A) The GCR rate was determined for wild type (wt), *mms1*, *rtt101* and *mms22* cells without insert. Shown are mean values from at least three independent experiments ± SD as fold enrichment over wild type without insert. Statistical significance compared to wild type strain was determined by student's T-test. *: p≤0.05, **:p≤0.01. (B-E) The GCR rate was determined for wild type (wt) (B), *mms1* (C), *rtt101* (D) and *mms22* (E) cells with inserts. Inserted sequences were *LEU2* marker, non-G-rich Chr VII (NG-*LEU2*), G-rich Chr I (GR-*LEU2*) and G4 Chr I (G4-*LEU2*) (Annex Table 12). Shown are mean values from at least three independent experiments ± SD as fold enrichment over respective strain without insert. Statistical significance compared to no insert strain was determined by student's T-test. *: p≤0.05, **:p≤0.01, ***:p≤0.001. Further statistical significance is noted in the figure. This was also determined by student's T-test. **:p≤0.01. (F) The GCR rate was determined for *mms1* cells after Phen-DC_3 treatment. Shown are mean values from at least three independent experiments ± SD as fold enrichment over strain without Phen-DC_3 treatment.

In *rtt101* cells, all insertions caused a similar significant increase of the GCR rate (*LEU2* marker: 3.1-fold; NG-*LEU2*: 3.8-fold, GR-*LEU2*: 4.4-fold; G4-*LEU2*: 5.2-fold) compared to *rtt101* cells without inserts. Here, the GCR rate was not significantly higher if a G4tract3 motif was present compared to the other inserts (Figure 28D).

In *mms22* cells, insertion of GR-*LEU2*, NG-*LEU2* and G4-*LEU2* lead to a significantly enhanced GCR rate (5.7-fold, 4.7-fold, 14-fold respectively) compared to *mms22* cells without insert. Also in this case, insertion of G4-*LEU2* (G4tract3) caused a significantly higher GCR rate than the NG-*LEU2* (2.5-fold) or GR-*LEU2* insert (three-fold) (Figure 28E).

To confirm that especially a G4 structure causes an enhanced GCR rate in *mms1* cells, the GCR assay was performed using the G4 stabilizing ligand Phen-DC_3, which mimics the phenotype of *pif1* cells (De Cian et al, 2007; Piazza et al, 2010). Unexpectedly, treatment of *mms1* cells with Phen-DC_3 did not cause an elevated GCR rate in the presence of a G4tract3 motif compared to the strain without Phen-DC_3 treatment (Figure 28F).

To get more insight into the effect of *mms1*, *rtt101* and *mms22* deletion on genome stability, multiplex PCRs were performed. During the multiplex PCR, different regions of the left arm of Chr V are amplified (Figure 29A) to determine the location of GCR events and to indicate what kind of GCR event might have happened (Bochman et al, 2014; Paeschke et al, 2013). Those GCR events could be telomere additions, deletions, insertions or mutations (Bochman et al, 2014; Paeschke et al, 2013; Piazza et al, 2012). Before the GCR events occur, all expected six products were detected in wild type, *mms1*, *rtt101* and *mms22* cells. This was the case for all tested colonies (Figure 29+30, Annex Figure 11). After the GCR event, product number one is expected to be lost due to the growth of the colonies on FOA/CAN containing plates. The product number six needs to be present at all times, because this product lies within an essential gene (*PCM1*).

In the wild type and deletion strains without an insert no product number one could be detected in the tested colonies after the GCR events, except for one colony in each wild type, *mms1* and *rtt101* cells (Figure 29B-E).

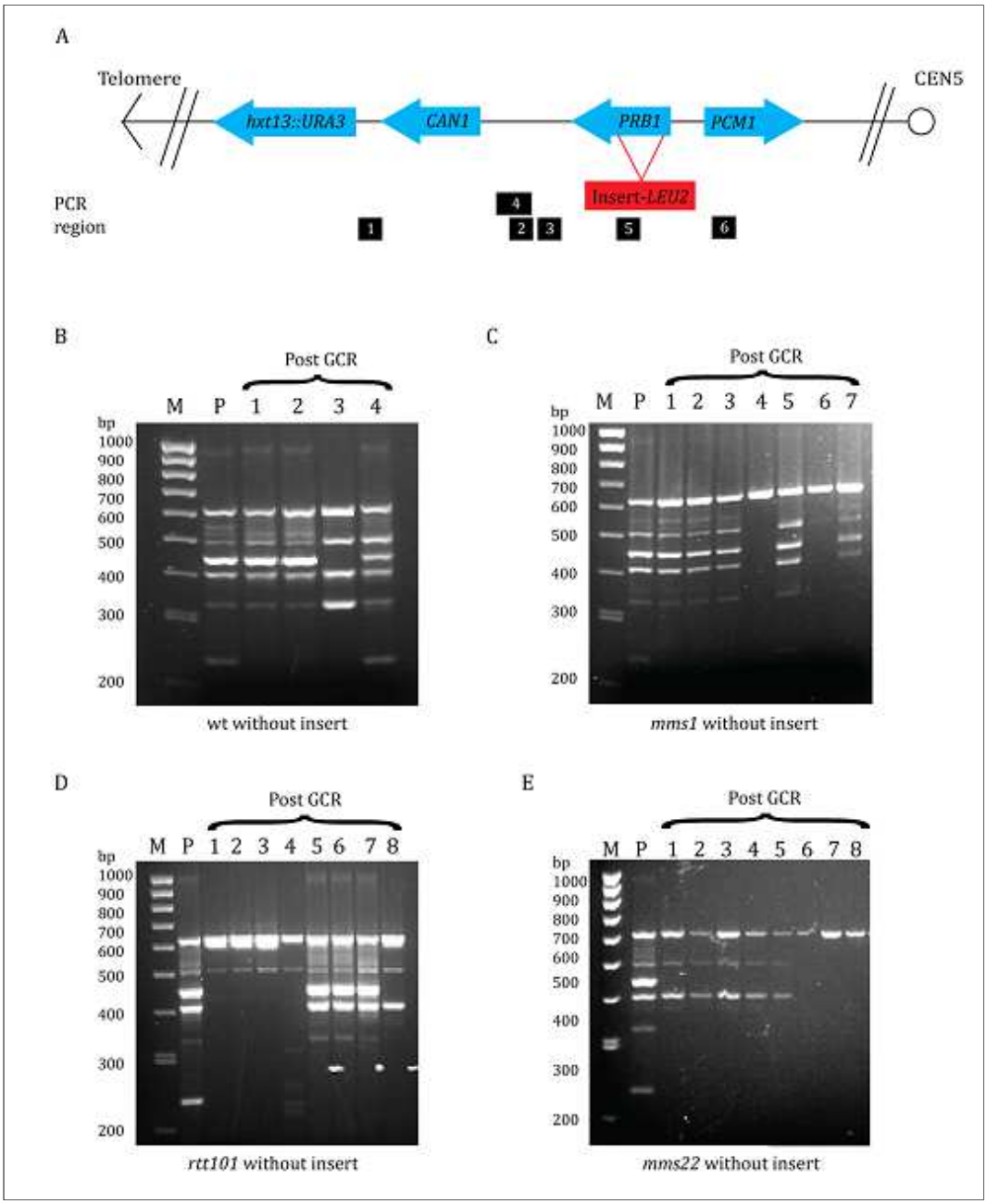


Figure 29: Multiplex PCRs of cells without inserted sequences. (A) Schematic showing regions amplified during multiplex PCR on left arm of Chr V. For further information on left arm of Chr V and its use in the GCR assay see Figure 18. (B-E) Multiplex PCR of wild type (wt) (B), *mms1* (C), *rtt101* (D) and *mms22* (E) cells without insert. Multiplex PCR was performed with yeast genomic (g) DNA before (P) and after the GCR assay (Post GCR). 4-8 individual colonies were tested after the GCR assay.

The occurrence of the products two-five varied between different tested colonies within one strain. For some colonies none of the products two-five could be detected while in others all four were present (Figure 29B-E). After insertion of a specific region (G4-*LEU2*, GR-*LEU2* and NG-*LEU2*), only products number five and six could be detected in wild type, *mms1*, *rtt101* and

mms22 cells (Figure 30 and Annex Figure 11), except for one colony in *rtt101* cells were the additional product number three was present (Annex Figure 11).

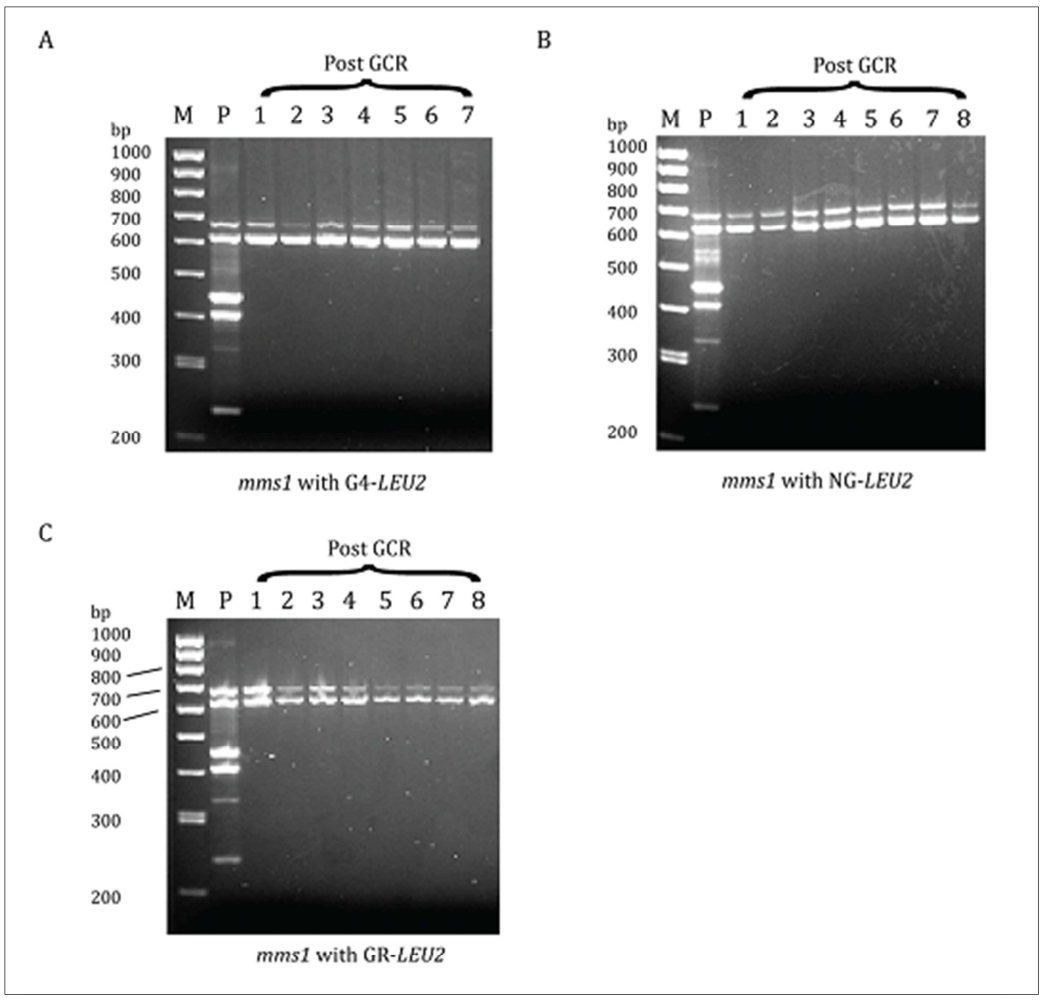


Figure 30: Multiplex PCR of *mms1* cells with inserted sequences. (A-C) Multiplex PCR of *mms1* cells with inserted G4-*LEU2* (A), NG-*LEU2* (B) and GR-*LEU2* (C). Multiplex PCR was performed with yeast gDNA before (P) and after the GCR assay (Post GCR). 7-8 individual colonies were tested after the GCR assay.

5.6.9 γ-H2A levels are decreased at Mms1 binding sites

As observed in this thesis, Mms1 is required for genome stability and replication fork progression at G4 motifs (Figure 18+24+28C). G4 motifs can lead to DSBs if they are not properly regulated (Crabbe et al, 2004; Koole et al, 2014; London et al, 2008; Lopes et al, 2011; Paeschke et al, 2011; Piazza et al, 2015; Piazza et al, 2012; Ribeyre et al, 2009; Sabouri et al, 2014; Zimmer et al, 2016) and in response to DSBs γ-H2A levels increase upon phosphorylation

of H2A (Downs et al, 2000; Shroff et al, 2004). Therefore, it was assumed that γ-H2A levels accumulate at Mms1 binding sites in *mms1* cells.

However, as depicted in Figure 19 γ-H2A levels decreased in *MMS1* deleted cells. Because this was unexpected, γ-H2A ChIP was repeated at identified Mms1 binding sites. Tested were six Mms1 BR identified in this thesis (Figure 20D), three NC sites identified in this thesis (Figure 20D), one positive γ-H2A control region (Capra et al, 2010) and one negative γ-H2A control region (Capra et al, 2010). Strikingly, γ-H2A levels were reduced (2.1-4.6-fold) at all tested Mms1 binding regions in *mms1* cells (Figure 31).

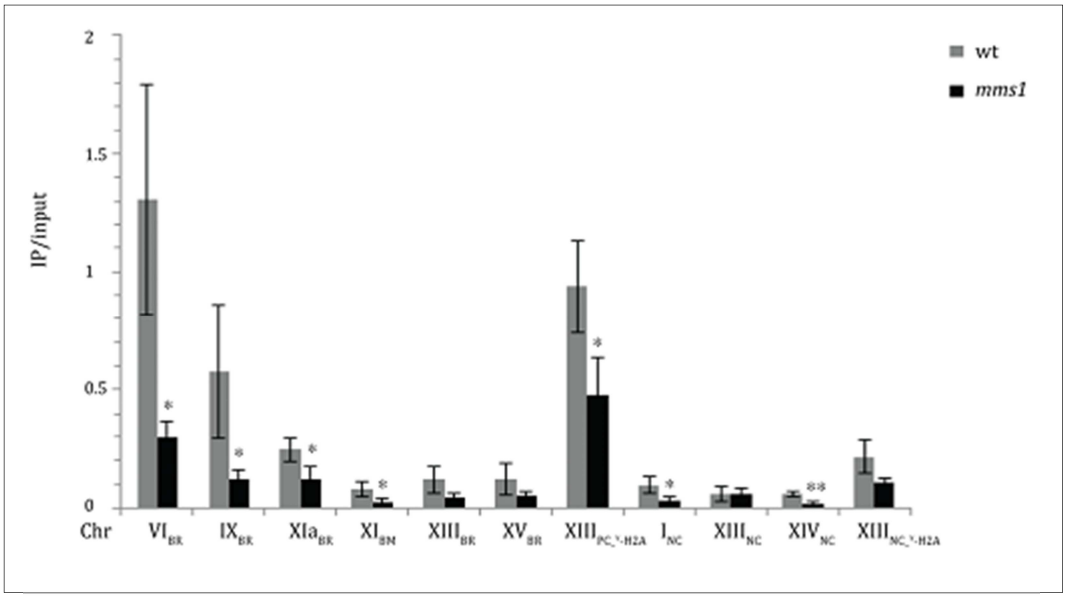


Figure 31: γ-H2A levels decrease upon deletion of *MMS1*. ChIP was performed and the binding of endogenous γ-H2A in wild type (wt) (grey) and *mms1* (black) cells was analyzed by qPCR using primer pairs for the shown regions. As control regions, the enrichment at high (Chr $XIII_{PC\_\gamma\text{-}H2A}$) or low (Chr $XIII_{NC\_\gamma\text{-}H2A}$) γ-H2A binding sites was determined. Plotted are the IP/input values as means ± SD from three independent experiments. For more information about the tested regions see Figure 20D and Annex Table 18. Statistical significance compared to wild type cells was determined by student's T-test. *: p≤0.05, **: p≤0.01.

This decrease was significant at the BR Chr VI_{BR}, IX_{BR}, XIa_{BR} and XV_{BR}, but also at the γ-H2A positive control region (Chr $XIII_{PC\_\gamma\text{-}H2A}$) (Capra et al, 2010), which contains a G2tract2 on the lagging strand and hence it is assumed that Mms1 also binds to that region. Unexpectedly, γ-H2A levels were also reduced (~ three-fold) at two negative control regions (Chr I_{NC}+XIV_{NC}), but the γ-H2A levels at these regions were less than at the γ-H2A negative control region (Chr $XIII_{NC\_\gamma\text{-}H2A}$) (Capra et al, 2010) (Figure 31).

Interestingly, western analysis demonstrated that the overall γ-H2A levels did not change upon *MMS1* deletion (Annex Figure 12). Additionally, ChIP-qPCR experiments of histone H3 were performed to elucidate whether Mms1 influences the binding or position of histones in general. This analysis revealed that binding of histone H3 was independent of Mms1 at all tested regions (Annex Figure 13). Tested were four Mms1 binding sites (Chr VI_{BR}, IX_{BR}, XIa_{BR} and XV_{BR}) and two non-Mms1 binding sites (Chr I_{NC} and XIV_{NC}).

These results suggest that Mms1 specifically influences γ-H2A levels at distinct positions.

5.6.10 Co-IP experiments only revealed Rtt101 as an Mms1 interaction partner

Due to the unexpected fact that Mms1 binding to G-rich/G4 motifs was independent of the other two ligase components Rtt101 and Mms22 (Figure 21), that so far it was hypothesized that Mms22 is the DNA interacting protein (Zaidi et al, 2008) and that no direct interaction of Mms1 with DNA has been demonstrated to my knowledge, a Co-IP of endogenously tagged Mms1 cells was performed in order to clarify the binding and function of Mms1 at its binding sites. This experiment was performed in triplicates. Due to the fact that the ChIP experiment of Mms1 in the absence of Pif1 had not been performed at the time of this analysis, the Co-IP was performed with wild type and *pif1-m2* cells. First one replicate was analyzed by MS and then the other two replicates were analyzed by MS. Therefore, the first experiment was evaluated separately from the second and third ("duplicates") experiment. MS analysis, quantification and figure construction were performed by the group of Andreas Schlosser. Because the results of wild type cells mostly did not differ from those of *pif1-m2* cells, only the results of wild type cells are described in more detail (see Annex Figure 14 for a graphical representation of the results from the second and third analysis and Table 20+21 for complete results). As expected, Mms1 was the most prominent protein in Myc-tagged Mms1 wild type cells with 49, 45 and 54 unique peptides in the first, second and third analysis respectively (Annex Table 21+22). Zero or only one unique peptide of Mms1 was identified in the untagged control. Only Rtt101 was identified as a strong and significant interaction partner of Mms1 (Annex Table 21+22). Rtt101 was identified in all three analyses with 22, 21 and 29 unique peptides in Myc-tagged Mms1 wild type cells, while it

was not found or only with one peptide in the untagged control (Annex Table 21+22). Crt10, which is a component of the $Rtt101^{Mms1/Crt10}$ ubiquitin ligase (Zaidi et al, 2008), was identified in the second and third analysis with two and seven unique peptides in Myc-tagged Mms1 wild type cells respectively, while zero peptides were found in the untagged control (Annex Table 22). However, statistical analysis performed by the group of Andreas Schlosser demonstrated that this interaction was not significant (Annex Table 22, signum value=0). Mms22 could not be confirmed as an interaction partner in any of the three analyses. Interestingly, in the first analysis, the protein Dna2 was identified as a significant interaction partner of Mms1 in *pif1-m2* cells (Annex Table 21). It was demonstrated that mammalian Dna2 nuclease cleaves telomeric G4 structures *in vitro* and that Dna2 is required for telomere maintenance in mammalian cells (Lin et al, 2013). Therefore, ChIP-qPCR experiments were performed with endogenously HA-tagged Dna2 cells in the presence and absence of Mms1. Dna2 is a 3´-5´ DNA helicase and a nuclease with a molecular weight of 172 kDa (Budd & Campbell, 1995). Dna2-HA was expressed in wild type and *mms1* cells and a protein band was observed above 180 kDa (188 kDa expected) (Annex Figure 4K). Four Mms1 binding regions were tested. Binding of Dna2 was independent of Mms1 at three (Chr IX_{BR}, $XIII_{BR}$ and XV_{BR}) of the four regions, while the binding was significantly enhanced (4.4-fold) at one (Chr VII_{BM}) region (Annex Figure 15). Due to the fact that Dna2 was only discovered as a significant interaction partner in *pif1-m2* cells (Annex Table 21), ChIP-qPCR of Dna2 should be performed in *mms1 pif1-m2* double mutants. Due to the fact that Mms1 supported Pif1 binding (Figure 27), it might be possible that Dna2 can compensate for Pif1 in *pif1-m2* cells. That Dna2 might function at non-telomeric G4 motifs in yeast is supported by the identification of Dna2 as a G4 interacting protein in the fourth pull-down experiment completed for this thesis (Annex Table 9). Otherwise, no more significant Mms1 binding partners (signum > 0), which would be of interest for this thesis, were identified (Annex Table 21+22).

6 Discussion

First, more general results will be discussed before the results on individual proteins are discussed in detail.

Due to the fact that the position and the nucleotide composition of G4 motifs are evolutionarily conserved in *S. cerevisiae* across *sensu stricto* species as well as among human populations (Capra et al, 2010; Nakken et al, 2009) and many proteins were demonstrated to bind to G4 structures to influence biological processes and support genome integrity (Crabbe et al, 2004; Koole et al, 2014; Kruisselbrink et al, 2008; London et al, 2008; Lopes et al, 2011; Paeschke et al, 2013; Paeschke et al, 2011; Paeschke et al, 2008; Paeschke et al, 2005; Piazza et al, 2015; Piazza et al, 2012; Ribeyre et al, 2009; Sabouri et al, 2014; Sarkies et al, 2012; Sarkies et al, 2010; Schiavone et al, 2014; Zimmer et al, 2016), it was expected that a G4 structure binds specifically to other proteins than linear G-rich DNA *in vitro*. This assumption could be confirmed (Figure 14).

Performing pull-down experiments with subsequent MS analysis, many proteins were identified in this thesis that uniquely bound to a G4 structure compared to a linear G-rich control (Annex Table 4+5+6+8+9). Of the proteins further studied in this thesis (Dot1, Mgs1, Mms1, Tel1), only Mgs1 was identified in more than one pull-down experiment (Annex Table 5+6+9). However, 224 unique G4 binding proteins were identified in at least two independent pull-down experiments (thanks to Stefan Juranek for this evaluation) (Annex Table 10), which suggests that the method as such works reproducibly. It might be possible that Dot1, Mms1 and Tel1 are artefacts, because they were only identified in one pull-down experiment. Hence, further *in vivo* studies were performed in this thesis to clarify a possible function. Strikingly, the G4 structure unwinding helicases Pif1 and Sgs1 (Paeschke et al, 2013; Ribeyre et al, 2009; Sanders, 2010; Sun et al, 1999; Wallgren et al, 2016), were predominantly identified in the mutated G4 sample (Annex Table 5+6+9). However, Sgs1 was less efficient than Pif1 in unwinding G4 structures *in vitro* and it had a stronger preference for Y structures than for G4 structures (Paeschke et al, 2013). The identification of Pif1 predominantly in the mutated G4 sample is more striking.

However, in one pull-down experiment with wild type cells the difference in peptide counts was not very big (13 in G4, 24 in mutated G4, Annex Table 9). In the other pull-down experiment with *pif1-m2* cells Pif1 was identified in the mutated G4 with 16 peptides and with two peptides in the G4 (Annex Table 5+6). In all pull-down experiments whole cell lysates were used. Therefore it is possible that the mitochondrial form of Pif1 bound here. However, Pif1 from mitochondria should not have different binding properties than nuclear Pif1 *in vitro*, because they are isoforms (Schulz & Zakian, 1994). It might be possible that Pif1, and also Sgs1 as well as Rrm3, unwound and fell off the G4 structure during the pull-down experiment, which made it difficult to analyze their binding to G4 structures. If the analyses would be repeated, helicase deficient instead of helicase deleted cells could be used. That the samples were flipped during the analyses is highly unlikely, because this then had to have happened more than once to receive the above results.

Mre11 and Ku70, which were identified as G4 binding proteins *in vitro* (Cogoi et al, 2008; Ghosal & Muniyappa, 2005), were not identified in any of the pull-down experiments completed for this thesis. It is possible that the G4 motif used in this thesis is not a specific target of Mre11 and Ku70.

Van Hacht and colleagues published a list of proteins that preferentially bound to G4 structures in their pull-down experiment (von Hacht et al, 2014). Many ribosomal proteins were identified that preferentially bound to G4 structures in both the published screen (von Hacht et al, 2014) and the screens completed for this thesis (Annex Table 4+5+6+8+9), although the exact proteins identified varied. The different observations might be due to the different organisms used (human vs. *S. cerevisiae*) and also due to different G4 motifs used. Even in the published study, the authors could not confirm all G4 binding proteins when another G4 motif was used (von Hacht et al, 2014).

Interestingly, the mismatch repair proteins Msh1, Msh2, Msh3 and Msh6 were predominantly found in the mutated G4 (44, 50, 45, 37 peptides respectively) compared to the G4 (four, six, seven, three peptides respectively) sample using *pif1-m2* cells (Annex Table 5+6). It was

demonstrated that the guanines in the G-tract are replaced by other nucleotides in *pif1-m2* cells (Table 1) (Paeschke et al, 2011). Thus, mismatch repair proteins might bind to these incorrectly inserted nucleotides to re-establish the G4 motif. In this case, it can be imagined that GCR rates would be lower in strains with deleted mismatch-repair proteins and *pif1-m2* mutations than in only *pif1-m2* mutants, because if mismatch repair proteins are present and re-establish the G4 motif, the G4 motifs can again induce genome instability in the next round of replication. If the mismatch repair proteins are deleted, a G4 motif can only induce genome instability in one round of replication. Also, ChIP experiments could be performed to examine whether mismatch repair proteins have a preference for mutated G4 motifs.

To conclude, the published data (von Hacht et al, 2014) and the results from the pull-down experiments performed in this thesis demonstrate that many proteins exist that specifically bind to G4 structures *in vitro* and that different G4 structures might be bound by different proteins. Also, it is possible that also RNA G4 interacting proteins were identified in the screens of this thesis, because RNA and DNA G4 structures do not differ much *in vitro* and whole cell lysates were used. This makes the study of G4 structures, especially in *vivo* difficult, because various proteins might only bind to G4 structures of a specific morphology as well as to RNA or DNA G4 structures. In my opinion, the pull-down experiments and MS analyses generally worked, but other studies need to be performed to confirm the results. Due to the high number of proteins that can bind to G4 structures, it is expected, in my opinion, that not all proteins can be reproducibly identified as G4 binding proteins in all MS analyses, especially if the proteins are low abundant. Also, it is possible that some identified proteins in this thesis do not directly bind to DNA, but indirectly through protein-protein interaction. To clarify, if the identified proteins directly bind to G4 structures *in vitro*, they should be expressed and purified and binding studies should be performed. In this case, also many G4 structures of different topology could be tested, as well as other controls (dsDNA, bubble, etc.), to clarify the binding properties of the proteins.

6.1 Dot1, Ku70, Mgs1, Mre11 and Tel1

6.1.1 Binding of Dot1, Ku70, Mgs1, Mre11 and Tel1 to G4 motifs *in vivo*

G4 structures were shown to cause replication fork slow-down and DSBs in the absence of Pif1 helicase (Lopes et al, 2011; Paeschke et al, 2011; Piazza et al, 2015; Piazza et al, 2012; Ribeyre et al, 2009; Sabouri et al, 2014). Both Dot1 and Tel1 play a role in DSB repair, while Mgs1 is required for replication fork progression (Branzei et al, 2002a; Conde et al, 2009; Hishida et al, 2002; Shroff et al, 2004). Hence it was assumed, that these three proteins, identified as G4 structure binding proteins *in vitro* (Annex Table 5+6+9) also bind to G4 structures *in vivo*.

Unexpectedly, they did not preferentially bind to G4 motifs compared to control regions (G-rich/non-G-rich) *in vivo* (Figure 16). A reason for this might be an unspecific binding of these proteins to G4 structures during the pull-down experiments. This might especially be true for Dot1 and Tel1, because these proteins were only identified in one of the four pull-down experiments (Annex Table 5). The very high number of DNA interacting proteins identified in the MS analyses (Annex Table 4+5+6+8+9) might suggest that not all proteins specifically and directly bind to DNA.

In more detail, Tel1 did not preferentially bind to the five tested G4 motifs and the two telomeric regions, compared to a G-rich control region (Figure 16+17). A problem with the tagged protein can be excluded, because western analysis confirmed the expression of Myc-tagged Tel1 (Annex Figure 4A). In former publications it was demonstrated that Tel1 prefers short telomeres as binding targets. In these publications, telomere shortening was induced (e.g. by deleting telomerase subunits) (Hector et al, 2007; Sabourin et al, 2007). If wild type length telomeres were tested, Tel1 only associated slightly with telomeres at some time points during cell cycle (Sabourin et al, 2007). The preference of Tel1 towards short telomeres can be a reason why no enhanced binding to a telomere compared to a G-rich control region was observed in this thesis. Because the published data is plotted as fold enrichment over *ARO1* (Hector et al, 2007; Sabourin et al, 2007) and in this thesis IP/input values are plotted, it cannot be concluded whether the values are comparable. Also, Sabourin and colleagues studied the binding in

synchronized cells (Sabourin et al, 2007), which in general leads to other IP/input values than in asynchronous cells, which were used in this thesis.

Previous studies revealed that Ku70 and Mre11 interact with G4 structures *in vitro* (Cogoi et al, 2008 ; Ghosal & Muniyappa, 2005) and associate with telomeric G4 motifs *in vivo* (Fisher et al, 2004; McGee et al, 2010; Sabourin et al, 2007; Takata et al, 2005). Also, former studies elucidated a connection between G4 induced genome instability and the HR pathway (Lopes et al, 2011; Piazza et al, 2012; Ribeyre et al, 2009; Zimmer et al, 2016). Additionally, G4 motifs are enriched genome-wide at mitotic DSBs and can cause DSBs (Capra et al, 2010; Hershman et al, 2008; Crabbe et al, 2004; Koole et al, 2014; London et al, 2008; Lopes et al, 2011; Paeschke et al, 2011; Piazza et al, 2015; Piazza et al, 2012; Ribeyre et al, 2009; Rodriguez et al, 2012; Sabouri et al, 2014; Zimmer et al, 2016). Therefore, it was hypothesized that the HR-protein Mre11 and the NHEJ-protein Ku70 bind to G4 structures *in vivo* in order to enable the repair of G4 induced damage. Unexpectedly neither Ku70 nor Mre11 preferentially bound to the tested non-telomeric G4 motifs compared to control regions in this thesis (Figure 16). However, for Ku70, one reason could be that it also did not preferentially bind to G4 structures compared to double-stranded DNA in a former study (Cogoi et al, 2008).

It was previously demonstrated that the binding of the Ku70/80 complex was enriched at two wild type telomeres, Chr VII_{Tel_VII-L} and Chr VI_{Tel_VI-R}, compared to a control region (*ARO1*) (McGee et al, 2010; Sabourin et al, 2007). Association of the Ku70/80 complex with telomere Chr VI_{Tel_VI-R} could be confirmed in this thesis (Figure 17B), although exact reproducibility could not be confirmed due to different normalizing. Sabourin and colleagues showed an equal enrichment of the Ku70/80 complex at telomere Chr VII_{Tel_VII-L} and Chr VI_{Tel_VI-R} (Sabourin et al, 2007). However, in this thesis a 60 times less binding to telomere Chr VII_{Tel_VII-L} compared to Chr VI_{Tel_VI-R} was observed (Figure 17A+B). Due to the dynamic nature of G4 structure formation it is possible that the telomere Chr VII_{Tel_VII-L} did not fold into a G4 structure during the experiments performed in this thesis, although this option is improbable, because the experiment was performed with three biological replicates and it is very unlikely that the G4

structure did not fold in any of those. The small standard deviation argues against the option that the G4 structure folded in some but not in all replicates (Figure 17A+B). Another reason why Ku70 did not bind to Chr $VII_{Tel\_VII\text{-}L}$ in this thesis could be the smaller DNA fragment sizes in this thesis (100-500 bp, Annex Figure 1A) compared to the fragment sizes in the publication (100-1000 bp, (Sabourin et al, 2007)). The qPCR product at Chr $VII_{Tel\_VII\text{-}L}$ has a size of 215 bp, which is still in the range of DNA fragment sizes in this thesis, but it might be possible, that a too high percentage of Chr $VII_{Tel\_VII\text{-}L}$ fragments were too small to be amplified. Organism diversities can be excluded as a reason for the different findings, because in this thesis and in the publication (Sabourin et al, 2007) *S. cerevisiae* was used.

Mre11 only associated slightly more with the wild type telomere Chr $VII_{Tel\_VII\text{-}L}$ compared to a G-rich control region, but did not associate with the wild type telomere Chr $VI_{Tel\_VI\text{-}R}$ (Figure 17C). This is in line with a preference of Mre11 for short telomeres compared to wild type telomeres (McGee et al, 2010). However, the data in this thesis cannot be directly compared to the published data (McGee et al, 2010), due to different normalizing, due to the fact that no G-rich and non-G-rich control region was tested in the publication and due to the fact that synchronized instead of asynchronous cells were used in the publication.

A problem with the tagged proteins can be excluded, because western analysis confirmed the expression and stability of Myc-tagged Ku70 and Mre11 (Annex Figure 4A+C+D).

To conclude, neither the *in vitro* G4 interacting proteins Dot1, Mgs1 and Tel1 identified in this thesis (Annex Table 5+6+9) nor the previously identified *in vitro* G4 interacting proteins Ku70 and Mre11 (Cogoi et al, 2008 ; Ghosal & Muniyappa, 2005) preferentially bound to the non-telomeric G4 motifs tested in this thesis compared to a G-rich control region (Figure 16). In my opinion, ChIP data of this thesis is reliable, because western analysis confirmed the stability of two exemplary proteins (Annex Figure 4C+D+H) and standard deviation most of the time is quite low. However, it is possible that the proteins might bind to other non-telomeric G4 motifs that were not tested here. Therefore, more ChIP experiments should be performed to clarify if the proteins bind to G4 motifs *in vivo*. ChIP-qPCR could be performed at the tested regions using

an untagged strain as a control (it is possible that an enrichment at G4 motifs compared to control regions is detectable if the enrichment over untagged control is plotted), ChIP-seq experiments could be performed with the proteins to study their binding genome-wide or, because the proteins are implicated in DNA repair, ChIP could be performed at HO-induced DSB sites (HO induction at G4 motif compared to control region).

6.1.2 Binding of Dot1, Ku70, Mgs1, Mre11 and Tel1 to G4 motifs in *pif1-m2* cells

It was hypothesized that the repair proteins Dot1, Ku70, Mgs1, Mre11 and Tel1 might only bind to non-telomeric G4 motifs *in vivo* in the absence of Pif1 helicase, because DNA damage at G4 regions was more pronounced in the absence of Pif1 (Lopes et al, 2011; Paeschke et al, 2011; Piazza et al, 2015; Piazza et al, 2012; Ribeyre et al, 2009; Sabouri et al, 2014).

However, this assumption could not be confirmed in this thesis (Figure 17). Just the contrary, Ku70 binding to most of the five tested non-telomeric G4 motifs was even significantly (2.1-3.5-fold) reduced in *pif1-m2* compared to wild type cells (Figure 17A), while the binding of Dot1, Mgs1, Mre11 and Tel1 to those sites was mostly Pif1-independent (Figure 17C-F). Interestingly, binding of the telomeric binding proteins Ku70, Mre11 and Tel1 (Fisher et al, 2004; Hector et al, 2007; McGee et al, 2010; Sabourin et al, 2007; Takata et al, 2005) to two telomeres (except Tel1 only at telomere Chr $VI_{Tel\_VI\text{-}R}$) was significantly (2.4-16.7-fold for Ku70, 2.8-15.7-fold for Mre11, 3.5-fold for Tel1) decreased in *pif1-m2* mutants (Figure 17). These result were unexpected as Pif1 was shown to unwind G4 structures *in vitro* (Ribeyre et al, 2009; Sanders, 2010; Wallgren et al, 2016) and it was hypothesized that G4 binding proteins bind stronger to G4 structures in the absence of unwinding helicases. Due to the fact that Mre11 and Tel1 bound preferentially to short telomeres (Hector et al, 2007; McGee et al, 2010; Sabourin et al, 2007) and that in the absence of Pif1 telomere length increases (Schulz & Zakian, 1994), binding of both proteins might be disfavoured in *pif1-m2* cells. An explanation for reduced Ku70 binding in *pif1-m2* cells might be that Pif1 supports Ku70 binding. This has to my knowledge not been tested and this would be interesting to find out.

To conclude, binding of Dot1, Mgs1, Mre11 and Tel1 to the five non-telomeric G4 motifs tested was independent of the helicase Pif1, while binding of Ku70 to most of those sites significantly decreased (2.1-3.5-fold) (Figure 17). This suggests that the proteins Dot1, Ku70, Mgs1, Mre11 and Tel1 are not more recruited to G4 motifs in the absence of Pif1 than they are in wild type cells, at least not to the G4 motifs tested here. This could indicate that, during the experiment performed in this thesis, the tested G4 motifs did not cause DNA damage or replication fork stalling in *pif1-m2* cells, which would require the recruitment of repair proteins or, and I think this is more likely, the tested proteins generally do not bind to the here tested G4 motifs *in vivo*. However, to confirm this assumption, ChIP-qPCR experiments should be performed with untagged cells as a control. All of this of course does not exclude that the proteins bind to other G4 motifs *in vivo*.

6.1.3 Connection of Dot1, Ku70, Mgs1, Mre11 and Tel1 to genome stability at G4 motifs

If the investigated proteins (Dot1, Ku70, Mgs1, Mre11 and Tel1) play a role in maintaining genome integrity at G4 structures *in vivo*, their deletion should cause DNA damage. Therefore, two different assays (γ-H2A ChIP-qPCR and GCR assay) were performed in order to study this assumption. Due to the fact that increased γ-H2A levels were observed at G4 motifs in the absence of helicase Pif1 or HR proteins (Paeschke et al, 2011; Sabouri et al, 2014; Zimmer et al, 2016), γ-H2A levels were determined at five G4 motifs and three control regions in the presence or absence of repair proteins to find out whether these proteins might be involved in the repair of G4 induced DSBs. Four of the tested G4 motifs were the same as in Figure 16+17, while the G4 motif from Chr $IX_{G4tract3}$ was replaced by the G4 motif from Chr $XIII_{G4tract3}$, because Chr $XIII_{G4tract3}$ was located at a γ-H2A binding site identified by (Capra et al, 2010).

Neither deletion of *DOT1*, *KU70*, *MRE11*, *MGS1* nor *TEL1* caused a consistent effect on γ-H2A levels at any of the five tested G4 motifs (Figure 19). This is in line with the inconsistent and non-preferred binding of these proteins to G4 motifs (Figure 16+17) tested in this thesis and suggests that, under the experimental conditions used in this thesis, the proteins do not bind to the tested G4 motifs and do not influence γ-H2A levels at those sites.

Also, using the GCR assay, only deletion of *MRE11*, but not of *DOT1*, *KU70*, *MGS1* nor of *TEL1* caused a G4 dependent increase of the GCR rate (Figure 18). In a previous publication, deletion of *MRE11* caused a 628-fold and deletion of *KU70* a three-fold enrichment of the GCR rate over wild type cells (Chen & Kolodner, 1999). Strikingly, in this thesis, the GCR rate increased only 18-fold in *mre11* compared to wild type cells and 1.3-fold in *ku70* cells. The GCR rate of *mre11* cells in this thesis was around three-fold lower than previously published and the GCR rate of *ku70* cells was around five-times higher than published. Also, the GCR rate of wild type cells without an insert was around ten-fold higher than published (Chen & Kolodner, 1999). These differences might be due to a different method used for GCR rate calculation or due to small differences in experimental conditions. Additionally, a recent study identified almost 200 "genome instability suppressing" (GIS) genes in *S. cerevisiae* using GCR strains without inserts or with inserted repeats (Putnam et al, 2016). The inserts they used were short repeated sequences with homology to Ty1 and Ty2 retrotransposons ± DSF1-HXT13 segmental duplication (low copy repeat) with homology to sequences in Chr IV, X and XIV and a Ty912 insert, which can lead to HR with other Ty retrotransposons in the genome (Putnam et al, 2016). Among the GIS genes were Mre11, Dot1 and Ku70, but not Mgs1. Also, in this thesis it was observed that Mgs1 does not influence genome stability in the no insert strain (Figure 18). In the publication, Mre11 was needed for genome integrity in at least three different strains, while Dot1 and Ku70 were only necessary for genome stability in the no insert strain (Putnam et al, 2016). The requirement of Mre11 and of Dot1 for genome stability in the no insert strain was confirmed in this thesis (Figure 18). However, deletion of *KU70* did not significantly increase the GCR rate of the no insert strain compared to the wild type strain (Figure 18), but was identified as a GIS gene using the GCR strain without insert (Putnam et al, 2016). A reason for this might be the difference in experimental settings and evaluation procedures. In the publication the authors isolated three independent spore clones per strain and let them grow on YPD plates at 30°C for 2 d. Then they were replica-plated on FOA/CAN plates. After growing (no information could be found on the duration), they applied a scoring system. For example a score of one resembled 1-5 colonies on

the FOA/CAN plates. The average score of the spore clones per strain was used to determine if deletion of a specific protein causes an enhanced GCR rate (Putnam et al, 2016). The GCR rates determined in this thesis for strains with inserted G4 motifs cannot be compared to the GCR scores observed for strains with inserted sequences in the publication (Putnam et al, 2016), because they did not insert G4tract3 motifs.

To conclude, deletion of *DOT1*, *KU70*, *MGS1* or *TEL1* neither had a consistent effect on γ-H2A levels at the tested G4 motifs nor at the GCR rate in the presence of G4 motifs. Deletion of *MRE11* did not affect γ-H2A levels at the tested G4 motifs, but increased the GCR rate in the presence of a G4 motif. However, it is possible that Dot1, Ku70, Mgs1 and Tel1 would affect the GCR rate, if other G4 motifs were present or if the G4 motif was located on the other strand (the G4 motif tested in this thesis was located on the leading strand). Therefore, more GCR assays could be performed (other G4 motifs, G4 motif on the lagging strand). However, in my opinion, Mre11 would be the most promising candidate to study its function on genome stability at G4 structures in more detail. In this case, the GCR assay should be performed with control inserts (G-rich/non-G-rich) to confirm that the observed effect is G4tract3 specific.

6.2 Mms1, Mms22, Rtt101

6.2.1 Mms1 binds to G-rich/G4 motifs *in vivo* and this binding is independent of Rtt101 and Mms22

The proteins Mms1 and Rtt101 were identified as *in vitro* G4 structure binding proteins via a pull-down experiment and MS analysis (Figure 15, Annex Table 5). Due to the better peptide count of Mms1 compared to Rtt101, Mms1 was investigated further.

Mms1 and Rtt101 form ubiquitin ligases (Zaidi et al, 2008) and all components of the $Rtt101^{Mms1/Mms22}$ ligase were observed to promote replication fork progression in the presence of fork stalling agents (Luke et al, 2006; Vaisica et al, 2011; Zaidi et al, 2008). Additionally, G4 structures can cause replication fork stalling in the absence of Pif1 (Lopes et al, 2011; Paeschke et al, 2011; Sabouri et al, 2014). Therefore, it was postulated that the $Rtt101^{Mms1/Mms22}$ ubiquitin ligase recognizes G4 structures *in vivo* and thereby promotes replication fork progression. For

this, binding of Mms1 was analyzed genome-wide by ChIP-seq. John A. Capra discovered as part of a collaboration that the average GC-content of the Mms1 binding sites was 50%, which is significantly higher than expected ($p < 0.001$) from the average GC content of the *S. cerevisiae* genome (~38% GC). Analysis of Mms1 binding regions identified a specific G-rich BM with many GG dinucleotides, which has the potential to form G4 structures *in vitro* (Figure 20A+E, Annex Table 15). Indeed, 61 of the 71 binding sites contain a G4tract2 motif (G4 motif with two guanines in the G-tract). John A. Capra found that this is significantly more overlap than expected to occur at random ($P < 0.001$). However, it is not significantly greater than expected by chance when considering the high GC content of the Mms1 binding sites.

Due to the Mms1 binding motif and due to the high occurrence of G4tract2 motifs within Mms1 binding sites, it was considered that a G4tract2 motif is important for Mms1 binding. G4tract2 motifs were also shown to fold into DNA G4 structures *in vitro* (Chambers et al, 2015; Qin et al, 2015; Siddiqui-Jain et al, 2002; Hazel et al, 2004) and analyses completed for this thesis revealed that G4tract2 motifs located in BRs of Mms1 form G4 structures *in vitro* while those located in NC regions did not (Figure 20E+F, Annex Figure 6). One exception was the lagging strand G4 motif located in the BR Chr XV_{BR}, which did not fold into a G4 structure (Annex Figure 6). This experiment should be repeated to confirm the unexpected result. Further analyses and ChIP-qPCR experiments further elucidated that the G4tract2 motifs Mms1 bound to are located on the lagging strand (Figure 20D, Annex Table 19). Due to the low abundance of Mms1 (Kulak et al, 2014), the dynamic nature of G4 structure formation in the cell and the parameters by which MACS (Zhang et al, 2008) defines binding sites, it was assumed that Mms1 binds to more regions than identified in ChIP-seq. ChIP-qPCR experiments (Figure 20D) supported this hypothesis. Here, Mms1 associated strongly with G4tract2 and G4tract3 motifs that were not discovered by ChIP-seq (Figure 20D (Chr VI_{BR}, IX_{BR}, XIa_{BR}, XIb_{BR}, $XIII_{BR}$, XV_{BR}), Annex Table 13).

If Mms1 promotes replication fork progression at G-rich/G4tract2 in a similar manner as at other stalled forks (Vaisica et al, 2011; Vejrup-Hansen et al, 2011), binding should depend on the other ubiquitin ligase ($Rtt101^{Mms1/Mms22}$) partners. However, in ChIP-qPCR experiments binding

of Mms1 occurred independently of Mms22 and Rtt101. This indicates that Mms1 either binds to the region itself or via a so far unknown interaction partner (Figure 21). This is in contrast to former publications where the authors hypothesized that Mms1 binds indirectly to DNA through interaction with Mms22. The different experimental settings used in their analysis (Vaisica et al, 2011; Vejrup-Hansen et al, 2011; Zaidi et al, 2008) compared to the ones used in this thesis might explain the different obtained results. Note, from the analyses performed in this thesis, it cannot be excluded that Mms22 or Rtt101 are also located at Mms1 binding regions. Co-IP of Mms1 followed by MS analysis (MS analysis and evaluation performed by the group of Andreas Schlosser) suggests that there are no further direct Mms1 interaction partners, because only Rtt101 was identified as a significant and relevant binding partner in three independent experiments (Annex Table 21+22). The fact that Mms1 binding to its target sites was even enhanced (≥ two-fold, Figure 21) in *mms22* cells indicates that Mms22 influences Mms1 binding. One possibility might be that Mms22 itself binds to Mms1 binding sites, which would result in less free binding regions for Mms1. Another option is that more free Mms1 molecules are present in the absence of Mms22, because the $Rtt101^{Mms1/Mms22}$ complex cannot be formed and lastly it is possible that Mms22 prevents G4 structure formation, which might be a prerequisite for Mms1 binding.

To conclude, it was revealed in this thesis that Mms1 binds genome-wide to G4tract2 motifs, which can fold into G4 structures *in vitro*. This binding occurred independently of the other two ligase components Rtt101 and Mms22 and no further relevant and significant interaction partners could be identified. In fact, binding of Mms1 was even significantly enhanced in *mms22* cells. This indicates that Mms22 somehow influences Mms1 binding. Because Mms22, and not Mms1, was previously discussed as the DNA binding protein (Zaidi et al, 2008), ChIP experiments could be performed to examine whether Mms22 binds to the same target sites as Mms1 and whether this binding depends on Mms1.

6.2.2 Connection of Mms1 to Pif1 function

In the past, fork pausing at G4 motif regions was detected in the absence of Pif1 DNA helicase (Lopes et al, 2011; Paeschke et al, 2011; Sabouri et al, 2014). Due to the function of Mms1, it was hypothesized its function is somehow connected to Pif1. The first hypothesis was that binding of Mms1 would be enhanced in the absence of Pif1. However, this was not the case (Figure 26). Then, it was investigated by Katrin Paeschke if Mms1 is required for the binding of Pif1. ChIP-qPCR experiments (Figure 27) demonstrated that Pif1 binds three to five-times less to Mms1 binding sites if Mms1 is absent. This was not the case at two other Pif1 binding sites (telomere and rDNA) as well as at one non-Pif1 binding site (tRNA). In my opinion, qPCRs should be performed at sites for which no Mms1, but Pif1 binding was observed as controls. To my knowledge this is the first experimental evidence showing that a specific protein (Mms1) supports Pif1 binding to G-rich/G4 motifs. This model is further supported by the observation that Pif1 and Mms1 binding sites significantly overlap ($P < 0.001$) (Annex Table 20), as determined by John A. Capra. Due to different binding characteristics of Mms1 and Pif1 to G4 motifs, Mms1 has a preference for G4tract2 on the lagging strand, whereas Pif1 binds genome-wide to G4tract3 motifs on both the leading and lagging strand (Paeschke et al, 2011), it is assumed that Mms1 might not support Pif1 binding at all G4 regions and that other proteins or mechanisms will take over here.

To conclude, Mms1 somehow supports binding of Pif1 to Mms1 target sites. Further experiments should be performed to examine whether also other proteins are involved in the promotion of Pif1 binding. Western analysis could be performed to clarify whether Mms1 and Pif1 interact with each other. To identify other proteins that might be involved could be quite difficult, because Co-IP experiments were performed in this thesis to identify novel Mms1-interacting partners. However, in three independent experiments only Rtt101 was identified as a significant and relevant binding partner (Annex Table 21+22). However, it would be interesting to examine if also the other two ligase components (Rtt101 and Mms22) promote Pif1 binding.

6.2.3 Replication fork stalls and genome instability is increased in *mms1* cells

Due to the fact that Pif1 is involved in replication fork progression at G4 motifs (Lopes et al, 2011; Paeschke et al, 2011; Sabouri et al, 2014), replication should stall in the absence of Mms1, if Pif1 function at some G4 motifs requires Mms1. Using levels of DNA Pol 2 occupancy, this hypothesis was confirmed (Figure 24). Strikingly, a former study demonstrated that the association of Polε is decreased in *mms1* cells (Vaisica et al, 2011). A reason for this might be that fork stalling was induced by HU in the publication (Vaisica et al, 2011) and that specific regions were investigated in this thesis. Previous studies showed that the checkpoint is not activated in *mms1* cells (Zaidi et al, 2008). This indicates that the observed replication fork stalling occurs at specific locations rather than due to DNA damage (Lopes et al, 2001; Minca & Kowalski, 2011; Sogo et al, 2002).

If there is less Pif1 binding in *mms1* cells, genome rearrangements should also increase in the absence of Mms1. Additionally, enhanced γ-H2A levels should be observed, because those were elevated in *pif1-m2* cells (Paeschke et al, 2011; Sabouri et al, 2014). Indeed, using the GCR assay, it was found that genome integrity is more hampered in *mms1* cells than in wild type cells (Figure 18+28A). In case of an inserted *LEU2* marker, GR-*LEU2* or NG-*LEU2* insert, the GCR rate further increased (2.5- to 3.8-fold) in *mms1* cells compared to the no insert strain (Figure 28C). A reason for this is probably the occurrence of G4tract2 motifs on the lagging strand within *LEU2* marker sequence, which were identified as Mms1 binding sites (Figure 20D). Strikingly, if a G4tract3 motif was inserted in addition to the *LEU2* marker, the GCR rate was further significantly enhanced (1.7-fold) in *mms1* cells (Figure 28C). Due to the fact that Mms1 target regions form G4 structures *in vitro* (Figure 20E) and a G4tract3 motif significantly enhanced the GCR rate (Figure 28C), it is assumed that most likely G4 structures cause Mms1 binding and downstream effects and not just the G-richness. This is further supported by the connection to Pif1, which specifically binds to G4 structures *in vitro* and *in vivo* (Lopes et al, 2011; Paeschke et al, 2013; Paeschke et al, 2011; Ribeyre et al, 2009). Unexpectedly, treatment of cells with the G4 stabilizing ligand Phen-DC_3 did not further increase the GCR rate (Figure 28F). However, it is

possible that Phen-DC_3 is not specific for the G4 motif used in this GCR assay. In a former study, only Phen-DC_3, but not another G4 stabilizing ligand N-methyl-mesoporphyrin (NMM), influenced genome stability at G4 motifs (Piazza et al, 2010). The authors assumed that the ligands were specific for different G4 topologies (Piazza et al, 2010). Deletion of the other two ubiquitin ligase components *RTT101* and *MMS22* caused opposite effects in the presence of a G4tract3 motif. While the GCR rate was independent of the inserted sequence in *rtt101* cells, a G4tract3 specific and dependent enhancement (up to three-fold) of the GCR rate compared to the other inserts was detected in *mms22* cells (Figure 28D+E). In general, increase of the GCR rate was higher in *mms22* than in *mms1* cells in the presence of an inserted sequence (Figure 28C+E). When also taken into account that Mms1 bound stronger to its target sites in *mms22* compared to wild type cells (Figure 21), it might be possible that Mms22 is the direct and preferred DNA binding protein in wild type cells, as it was also suggested in a previous study (Zaidi et al, 2008). From the ChIP experiments performed in this thesis it cannot be concluded whether the whole ubiquitin ligase complex has an influence on the processes (DNA Pol2 binding, Pif1 binding etc.) or if Mms1 alone fulfills the function. Furthermore it cannot be concluded which component of the ubiquitin ligase binds *in vivo*. Further experiments should be performed to clarify this (see conclusion). Also, it might be possible that the binding targets identified for Mms1 in this thesis are actually direct binding targets of Mms22. ChIP experiments of Mms22 could provide an answer to this question. An explanation for the fact that only Mms1 and Rtt101, but not Mms22, were identified in the pull-down experiment could be that Mms22 might be even less abundant than Mms1 (to my knowledge, the number of Mms22 molecules/cell has not been investigated so far), because it has only been identified as a component of one ubiquitin ligase (Zaidi et al, 2008).

If the ubiquitin ligase $Rtt101^{Mms1/Mms22}$ is involved in replication fork progression at specific DNA sequences, deletion of either component should result in deletions at those target sequences. This assumption was confirmed by performing multiplex PCRs at the left arm of Chr V following GCR events (Figure 30, Annex Figure 11). In the presence of an insert that contains Mms1

binding targets (G4tract2 on lagging strand, all inserts analyzed), DNA sequences downstream of the inserted sequence get lost during the GCR event in *mms1* cells (Figure 30). For Rtt101 and Mms22 it needs to be confirmed that those proteins target the same G4tract2 as Mms1, but the fact that all three proteins are part of the same ubiquitin ligase and that the absence of Rtt101 and Mms22 also leads to a deletion of DNA sequences downstream of the inserted sequence (Annex Figure 11D-I), indicates that they target the same sites as Mms1. However, the fact that deletion of *MMS1* and *MMS22*, but not deletion of *RTT101* caused a G4tract3 dependent increase of the GCR rate argues for an additional ubiquitin ligase-independent function of Mms1/Mms22. Unexpectedly, also in wild type cells, DNA sequences downstream of the inserted sequence get lost during the GCR event (Annex Figure 11A-C). This was not the case in a former publication (Paeschke et al, 2013). These differences might have happened by chance, because very rarely colonies grew on FOA/CAN plates in wild type cells. Thus, the results might be different if more colonies were tested. The requirement of Mms1 and Rtt101 for genome integrity was also demonstrated in a recent study, in which Mms1 and Rtt101 were identified as GIS genes in cells without insert as well as in cells with inserted sequences homologous to Ty retrotransposons (Putnam et al, 2016).

Unexpectedly, γ-H2A levels did not increase in *mms1* cells, in which less Pif1 binding is observed (Figure 19+31). Just the contrary: γ-H2A levels were significantly reduced at most Mms1 binding regions (Figure 31). The fact that the overall γ-H2A levels were independent of Mms1 (Annex Figure 12), suggests that the effect of Mms1 on γ-H2A levels is site specific. Also, ChIP of histone H3 indicates that Mms1 does not have a general effect on histones and their position (Annex Figure 13). As it was elucidated that Mms1 already binds to its target sites in G1 phase (Figure 23), it might be possible that Mms1 is required for the recruitment of repair proteins to its binding sites (either directly or through ubiquitination of targets) that then phosphorylate H2A. Further studies need to be conducted to confirm this assumption. It was already shown in this thesis that Mre11, which plays a role in the phosphorylation of H2A through recruitment of Tel1 (Nakada et al, 2003a), binds independently of Mms1 (Figure 25). Interesting candidates to study

could be Mec1 and Tel1, because those are involved in the phosphorylation of γ-H2A (Downs et al, 2000; Shroff et al, 2004) and were identified as G4 structure binding proteins *in vitro* in this thesis (Annex Table 5). ChIP experiments of Tel1 and Mec1 could be performed in the absence and presence of Mms1. Phosphorylation of γ-H2A is required for repair (Downs et al, 2000; Shroff et al, 2004). If γ-H2A levels decrease upon *MMS1* deletion, it is assumed that no repair can occur. This might also explain the observed increased GCR rate in *mms1* cells, especially in the presence of Mms1 binding regions, and the observation that the DNA sequences downstream of the binding region get lost during the GCR event (Figure 28+30). Due to the GCR results it can be excluded, in my opinion, that a reduced occurrence of DSBs is the explanation for reduced γ-H2A levels. Due to the observed function of Rtt101 and Mms1, as well as their homologues in humans (CUL4-DDB1), on the nucleosome assembly through ubiquitination of H3K56ac (Han et al, 2013) and due to the function of CUL4-DDB1 in H3 methylation *in vivo*, it is also imaginable that Mms1 (perhaps together with Rtt101/Mms22) functions in the remodeling of histones to create the required environment for DSB repair and γ-H2A phosphorylation.

The observation that γ-H2A levels also decreased in *mms1* cells at non-Mms1 binding sites might be an artefact, because the γ-H2A levels at Chr I_{NC} and Chr XIV_{NC} were lower than at the negative control region (Chr $XIII_{NC\_\gamma\text{-}H2A}$); hence the effect is probably negligible.

To conclude, replication fork progression was impeded at Mms1 binding sites (G4tract2 motifs) in *mms1* cells and genome instability was specifically enhanced in the presence of a G4tract3 motif in *mms1* and *mms22* cells. Strikingly, γ-H2A levels decreased in *mms1* cells. Due to the fact that deletion of *RTT101* did not specifically affect genome stability at a G4tract3 and that the GCR rates were much higher if Mms1 or Mms22 were deleted, it can be assumed that Mms1 and Mms22 possess a function that is independent from their function as an ubiquitin ligase component. Because deletion of *MMS22* caused a higher GCR rate than deletion of *MMS1*, GCR assays should be performed in *mms1 mms22* double mutants to find out whether Mms1 can compensate for Mms22. Also, because conventional ChIP experiments indicate that Mms1 specifically binds to G4tract2 motifs on the lagging strand, the GCR assays should be performed

in strains with a G4tract2 motif located on the leading strand. To further elucidate how Mms1 might influence γ-H2A levels, ChIP-qPCR of tagged Mec1 and Tel1 could be performed in the presence and absence of Mms1 to find out whether Mms1 influences γ-H2A through the recruitment of Mec1/Tel1. To study if Mms1 plays a role in histone remodeling it could be tested if it supports Dot1 (implicated in methylation of histone H3, which is required for DSB repair (Conde et al, 2009)) binding. Dot1 was identified as a G4 interacting protein *in vitro* in this thesis (Annex Table 5), but did not bind stronger to the G4tract3 motifs tested in this thesis than to a G-rich control region (Figure 16). However, this G-rich control region was also an Mms1 binding site (Figure 20D). Also, western blot analysis could be performed to investigate if Mms1 is involved in methylation of histone H3 on K79. Additionally, MS analysis could be performed to investigate whether Mms1 influences the posttranslational modifications of histones. However, this analysis is probably very laboratory extensive.

6.2.4 Mms1 does not influence binding of Mre11 to its binding sites

Due to the fact that Mms1 was required for HR at stalled forks (Duro et al, 2008) and that G4 structures can impede replication fork progression (Crabbe et al, 2004; Lopes et al, 2011; Paeschke et al, 2011; Sabouri et al, 2014; Sarkies et al, 2010; Schiavone et al, 2014), as also indicated in this thesis (Figure 24), it was assumed that Mms1 recruits the HR factor Mre11 to its binding sites. Surprisingly, binding of Mre11 was not altered in *mms1* compared to wild type cells. This indicates that Mms1 does not recruit the MRX complex to its target sites and is consequently not responsible for initiation of HR (Figure 25). This is also in line with the observed Pif1-independence of Mre11 binding at Mms1 binding sites (Chr $VI_{G4tract3}$, $IX_{G4tract3}$, XI_{NG}, $XIII_{GR}$) (Figure 17), because if Mms1 supports Pif1 binding (Figure 27) and deletion of *MMS1* has no influence on the binding of Mre11, then also Pif1 should have no influence. Due to the fact that replication fork stalling was enhanced in *mms1* cells (Figure 24) and Mms1 supports Pif1 binding (Figure 27), it could also be assumed that Mre11 is more enriched at Mms1 binding sites in *mms1* cells, due to prolonged stalling which would increase the requirement of Mre11. However, also this was not the case. This could indicate that G4 motifs only cause fork stalling,

but not fork collapse, in *mms1* cells, because Mre11 was only recruited to collapsed forks in a previous study (Lisby et al, 2004).

To conclude, absence of Mms1 did not change binding of Mre11 to Mms1 binding sites (G4tract2) in this thesis, which indicates that Mms1 does not induce HR at its binding sites. However, it is possible that Mms1 would recruit Mre11 to its binding sites if the fork is collapsed. To create a collapsed fork, Mec1 could be deleted as done previously (Lisby et al, 2004) and Mre11 binding could be tested in the presence and absence of Mms1.

6.2.5 Model of Mms1 function at G-rich/G4 motifs *in vivo*

Due to the fact that Mms1 protein levels are highest in G1 phase (Figure 22), that binding is present throughout the cell cycle at the G-rich/G4tract2 motifs (Figure 23) and that a recent publication observed G4 structure formation in duplex DNA *in vivo* (Lam et al, 2013), it is speculated for this thesis that G4 structures already form prior to S phase. This is consistent with the requirement of Mms1 for Pif1 binding, which occurred at the end of the S phase (Paeschke et al, 2011). Note, Mms1 is also part of the ubiquitin ligase $Rtt101^{Mms1/Crt10}$ and Crt10 was required for the expression of the genes *RNR2* and *RNR3* (Fu & Xiao, 2006; Zaidi et al, 2008). *RNR1*, which encodes another component of the RNR enzyme, was shown to be maximally expressed in S phase, while *RNR2* expression was relatively constant during cell cycle (Elledge & Davis, 1990; Koc et al, 2003). Hence, the observed high levels of Mms1 in G1 phase might rather be due to the function of the $Rtt101^{Mms1/Mms22}$ than of the $Rtt101^{Mms1/Crt10}$ ligase. This supports the assumption that Mms1 already binds in G1 phase to the target sites identified in this thesis.

With the data received in this thesis and with published data, two models about the function of Mms1 at G4 motifs can be envisaged (Figure 32). In these models, Mms1 binds throughout the cell cycle to G-rich regions with a strong potential to fold into G4 structures. From the data of this thesis is not clear whether Mms1 binds directly to G4 structures or via other proteins (e.g. Rtt101/Mms22) and it cannot be concluded if G4 structures form in G1 or early S phase and whether Mms1 binds to the G4 motifs or if it binds to formed G4 structures. However, once Mms1 binds to G-rich/G4 regions it supports Pif1 binding (Model A), either directly or via

interaction with other proteins. Due to previous findings that Pif1 binding is strongest at the end of S phase (Paeschke et al, 2011) it is assumed that Pif1 also binds at the end of S phase in this model. Furthermore, in accordance with published data (Paeschke et al, 2011) it is hypothesized that Pif1 unwinds the structured DNA region. Hence replication fork progression and genome integrity are maintained. In the absence of Mms1, binding of Pif1 is reduced, which results in replication fork pausing and an accumulation of genome instability (Model A). In model B, due to the fact that Mms1 is part of ubiquitin ligases (Zaidi et al, 2008), it is assumed that Mms1 binds together with Rtt101/Mms22 and induces in the ubiquitination of one or multiple subunits of the replisome to enable fork restart downstream of the formed G4 structure. This model is supported by the finding that the human homologues of Rtt101 and Mms1 were implicated in the ubiquitination of Pol δ (Lee et al, 2014; Zhang et al, 2013). However, in my opinion, this model (B) does not represent the whole function of Mms1 at G4 motifs, because GCR results observed in this thesis (Figure 28) indicate that Mms1 has a specific function at G4 motifs, which is independent from Rtt101 and literature has demonstrated so far that ubiquitin ligases Mms1 is a component of also contain Rtt101 (Zaidi et al, 2008). However, it is possible that Mms1 induces both pathways, because they are not contradictory and could happen during the same cell cycle. In model A (Paeschke et al, 2011) as well as in model B, gap repair at the G4 structure is required at the end of S phase.

Previous publications have linked G4 structures formed during DNA replication with PRR (Sarkies et al, 2012; Sarkies et al, 2010; Schiavone et al, 2014; Wu & Spies, 2016). However, in my opinion, it is less likely that translesion synthesis is induced by Mms1, because genetic assays in yeast demonstrated that Mms1 is probably not involved in PRR (Hryciw et al, 2002) and Zaidi and colleagues showed that neither Rtt101 nor Mms1 were required for the ubiquitination of PCNA (Zaidi et al, 2008).

To clarify the binding properties of Mms1 and to further support the proposed models, binding studies (electrophoretic mobility shift assay (EMSA)) could be performed with purified Mms1 to

examine whether it directly binds to DNA. Also, with a purified protein and CD analysis it could be tested, if Mms1 perhaps promotes the formation of G4 structures.

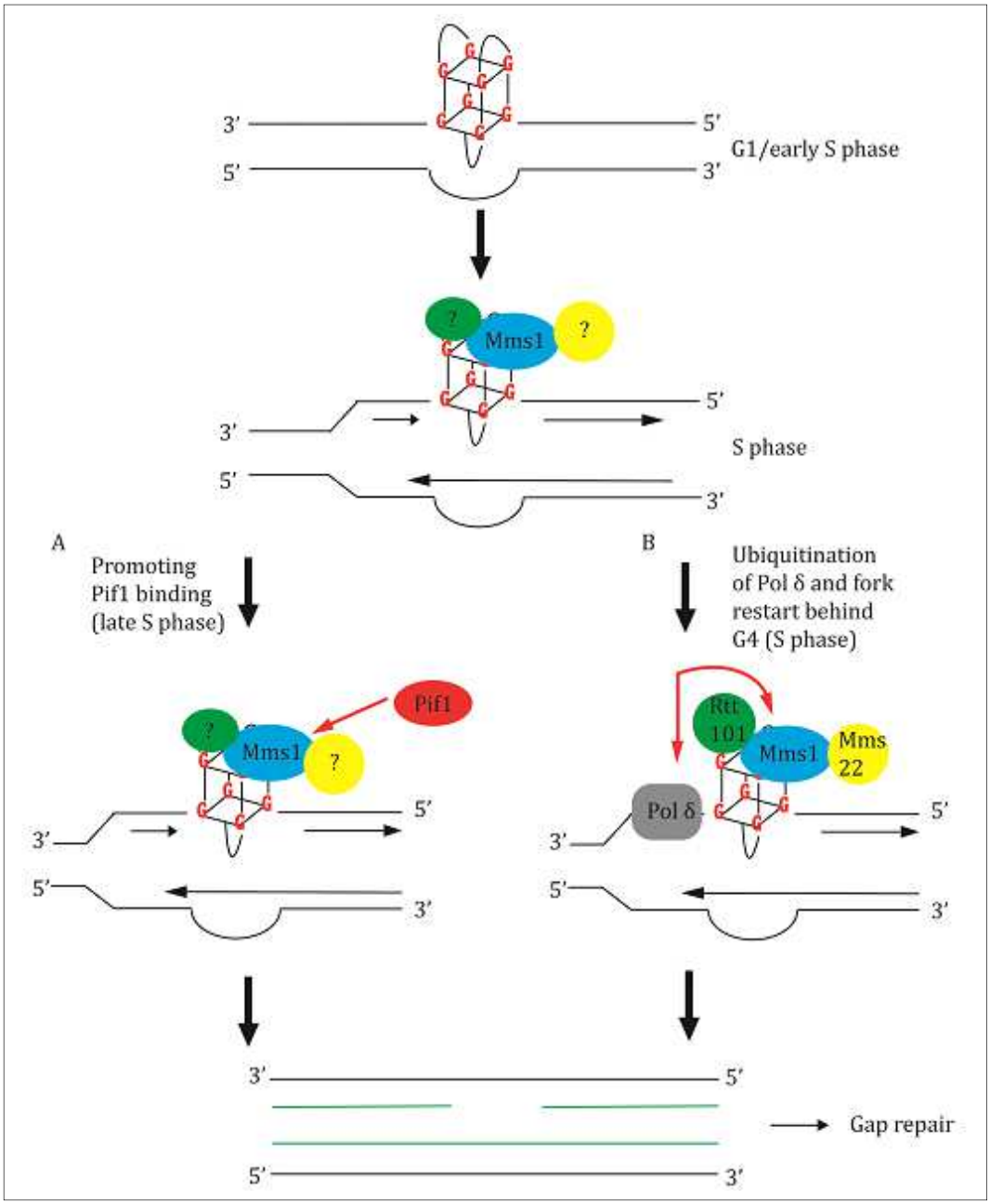


Figure 32: Mechanistic models of Mms1 function at G4 structures. Mms1 binds to G-rich/G4 motifs in G1 or S phase. In model A, it may either bind alone or in complex with other proteins, while in model B Rtt101/Mms22 have to be present. Then Mms1 might support Pif1 binding (Model A). Due to previous findings that Pif1 binding is strongest at the end of S phase (Paeschke et al, 2011) it is assumed that Pif1 also binds at the end of S phase in this model. Furthermore, in accordance with published data (Paeschke et al, 2011) it is hypothesized that Pif1 unwinds the structured DNA region. In model B, it is assumed (taking into account the observations in human cells (Lee et al, 2014; Zhang et al, 2013) that Mms1 (together with Rtt101/Mms22) is involved in the ubiquitination of replisome components (e.g. Pol δ) to enable fork restart. In model A (Paeschke et al, 2011) as well as in model B, gap repair at the G4 structure is required at the end of S phase.

To study whether G4 structures form *in vivo* in yeast and in which phase of the cell cycle they form is not possible so far, because to my knowledge, there are no antibodies which allow for the specific and sensitive detection of G4 structures in yeast cells. To identify other proteins that might be involved, e.g. to support Pif1 binding, could be quite difficult, because Co-IP experiments were performed in this thesis to identify novel Mms1-interacting partners. However, in three independent experiments only Rtt101 was identified as a significant and relevant binding partner (Annex Table 21+22).

As observed in this thesis, Mms1 can bind independently of Rtt101 and Mms22 to DNA. However, it should be tested whether the function of Mms1 at G4tract2 and G4tract3 motifs (support of Pif1 binding, promotion of fork progression, etc.) depends on the other two ligase components. For example, DNA Pol2 ChIP could be performed in *rtt101* cells to examine if deletion of *RTT101* has the same effect as deletion of *MMS1*. In addition to that, ChIP-qPCR of Mms22 can provide an answer whether Mms22 is the preferred DNA interacting protein. Also, it could be tested, if the $Rtt101^{Mms1/Mms22}$ ubiquitin ligase is involved in ubiquitination of the DNA polymerase to enable fork restart. With western analysis it can be determined whether the DNA polymerase in ubiquitinated and with 2D gels it can be examined whether the fork is restarted. Note, a possible dependency of Mms1 on Crt10 was not tested in this thesis, but due to the fact that the only discovered function of Crt10 so far lies in RNR gene transcription (Fu & Xiao, 2006) it is unlikely that it plays a role at the replication fork. However, DNA Pol2 ChIP could be performed to clarify if Crt10 has a function in replication fork progression.

6.2.6 Comparison of Mms1 functions to the functions of its human homologue DDB1

In humans, homologues of Mms1 and Rtt101 exist (DDB1 and CUL4 respectively (Zaidi et al, 2008)). This makes studying those two proteins highly interesting, because it suggests that they possess a conserved function and it suggests that functions of those proteins that are unraveled in yeast might also be valid for humans. Also, DDB1 in humans is linked to a human disease (XP type E) through interaction with DDB2 (Wittschieben et al, 2005). However, there are some contradictory functions reported in the literature. In humans, DDB1 (yeast: Mms1) forms a

complex with DDB2. This complex is connected to human health, because its *in vitro* binding capacity to UV induced damage was decreased if DDB2 contained the same mutations that also cause the phenotype in patients suffering from XP type E, a disease in which the NER pathway is defective (Wittschieben et al, 2005). Further studies revealed that the human DDB1-DDB2-CUL4-ROC1 E3-ligase complex plays a role in GG NER *in vitro* and *in vivo* (Groisman et al, 2003; Wakasugi et al, 2009). Just the contrary, genetic analyses in yeast indicate that Mms1 has no function in NER (Araki et al, 2003; Hryciw et al, 2002; Vejrup-Hansen et al, 2011). To my knowledge, Mms22 and Rtt101 have not been tested on their function in NER, which would be interesting to test in the future. However, to my knowledge, the human homologue of Mms22 has not been unraveled so far. Therefore, there is so far no indication that the functions of DDB2 and Mms22 are connected. Additionally, depletion of DDB1 and CUL4 caused checkpoint activation and depletion of DDB1 increased γ-H2AX foci in human cells (Lovejoy et al, 2006). However, deletion of *MMS1* and *RTT101* did not cause checkpoint activation in undisturbed yeast cells (Luke et al, 2006; Zaidi et al, 2008) and in this thesis reduced γ-H2A levels were observed in *mms1* cells (Figure 31). Mms1, Mms22 and Rtt101 were demonstrated to be required for replication fork progression after replication fork stalling (Luke et al, 2006; Vaisica et al, 2011; Zaidi et al, 2008), which has not been tested for the human homologues to my knowledge. In this thesis, it was discovered that Mms1 influences replication fork progression even in unperturbed cells, probably through binding to G4 motifs/structures and the promotion of Pif1 binding. Therefore, it would be interesting to test whether DDB1 plays a role in replication fork progression and what the target sites of this protein are. The observation that the CUL4-DDB1^{Cdt2} complex ubiquitinated p21 (subunit of Pol δ) *in vitro and in vivo*, supports the hypothesis that also CUL4/DDB1 might function in replication fork progression. To my knowledge DDB1 has also not been connected to G4 structures. Hence, it should be examined if DDB1 binds to G4 structures *in vitro* (EMSA). These studies could provide evidence if the effects and functions that where observed in this thesis are conserved from yeast to humans. A conserved mechanism would highly enhance the importance of the data obtained in this thesis.

7 Conclusion and outlook

To conclude, the repair proteins Dot1, Ku70, Mgs1, Mms1 (+Rtt101/Mms22), Mre11 and Tel1 were investigated in this thesis on their regulatory function at G4 structures. However, from the results obtained in this thesis, models can only be proposed for the function of Mms1 at G4 motifs/G4 structures *in vivo*. Dot1, Mgs1 and Tel1 were identified in a pull-down experiment and subsequent MS analysis as G4 interacting proteins *in vitro* (Annex Table 5+6+9), but those proteins showed no preferred binding towards any of the five tested G4 motifs compared to a G-rich control region *in vivo* (Figure 16), even not in the absence of the helicase Pif1 (Figure 17). Also, no consistent effects of those proteins on γ-H2A levels at G4 motifs were observed (Figure 19). Additionally, deletion of *DOT1*, *MGS1* and *TEL1* did not influence the GCR rate in a G4-dependent manner (Figure 18).

For the *in vitro* G4 interacting proteins Ku70 and Mre11 (Cogoi et al, 2008; Ghosal & Muniyappa, 2005) similar *in vivo* results were obtained. Both proteins did not preferentially bind to the G4 motifs tested in this thesis compared to a G-rich control region *in vivo* (Figure 16), even not in the absence of Pif1 (Figure 17) and only Ku70 caused a consistent effect on γ-H2A levels at G4 motifs (Figure 19). Those decreased significantly at three of five tested G4 motifs in *ku70* cells. Using the GCR assay, only deletion of *MRE11* caused a G4-dependent increase of genome instability (Figure 18).

It is possible that Dot1, Mgs1 and Tel1 are no direct interaction partners of non-telomeric G4 motifs *in vivo*. This may be especially the case for Dot1 and Tel1, because they were only identified in one pull-down experiment (Annex Table 5). Due to the fact that DNA and RNA G4 structures do not differ much *in vitro*, it might also be possible that those proteins possess a function at RNA G4 motifs *in vivo*. However, it is also imaginable that these three proteins, as well as Ku70 and Mre11, bind to other G4 motifs *in vivo* than those tested in this thesis or under specific circumstances (e.g. after inducing DNA damage (UV, HO-induced DSB) or after induction of fork stalling (HU)). Also, performing ChIP-qPCR of an untagged strain could provide more insight into the binding properties of the proteins. To clarify whether the proteins might have a

function at G4 structures, EMSAs could be performed (already done with Mre11 (Ghosal & Muniyappa, 2005), but data could be extended) to examine their direct and specific binding to G4 structures. This might especially be relevant for Ku70, because this protein was also found to bind double-stranded DNA (Cogoi et al, 2008). Then, ChIP-seq experiments could be performed to investigate the binding of those proteins genome-wide. In my opinion, Mre11 is the most promising candidate to do this study, because of its binding to G4 structures *in vitro* (Ghosal & Muniyappa, 2005), due to its effect on the GCR rate (Figure 18) and due to the fact that G4 structures were connected to HR in previous publications (Lopes et al, 2011; Piazza et al, 2012; Ribeyre et al, 2009; Zimmer et al, 2016). Also, more GCR assays could be performed using other G4 motifs or inserting the G4 motif on the other strand. However, in my opinion, in the case of Mre11, GCR assays should first be performed with an inserted control sequence to show that the effect of Mre11 on the GCR rate is G4-dependent. Due to the fact that studies indicate that Mgs1 and Sgs1 function in parallel pathways (Branzei et al, 2002b; Hayashi et al, 2008; Hishida et al, 2001; Kawabe et al, 2001) and that Sgs1 unwinds G4 structures *in vitro* (Paeschke et al, 2013; Sun et al, 1999) it might be interesting to investigate Mgs1 binding in *sgs1* cells.

The ubiquitin ligase component Mms1 and its ligase interaction partner Rtt101 were identified as G4 interacting proteins *in vitro* in this thesis (Annex Table 5). Although they were only discovered in one pull-down experiment, it could be demonstrated for Mms1 that it binds *in vivo* to G4tract2 motifs on the lagging strand of replication (Figure 20A+D, Annex Table 19). Additionally, replication fork progression and genome stability at G4 motifs were affected in *mms1* cells (Figure 24+28). Furthermore, Mms1 bound to G4 motifs in all phases during the cell cycle and it supported binding of Pif1 to its target sites (Figure 23+27).

Due to the fact that Mms22 (another ligase partner) was assumed to be the DNA interacting protein (Zaidi et al, 2008), that binding of Mms1 to its target sites increased in *mms22* cells (Figure 21) and that deletion of *MMS22* caused an even higher G4-dependent GCR rate than deletion of *MMS1* (Figure 28), it should be investigated whether Mms22 binds to the same target sites as Mms1 and whether this binding depends on Mms1. Additionally, GCR assays could be

performed in *mms1 mms22* double mutants to investigate if Mms1 can compensate for Mms22. It might also be interesting to perform binding studies of purified Mms1 and Mms22 in order to find out which protein directly binds to G4 structures *in vitro*. However, this might be difficult, because both proteins have a high molecular weight (~160 kDa (Mimura et al, 2010)) and are low expressed (Kulak et al, 2014) (As far as I know no molecule numbers have been published for Mms22). To my knowledge both proteins have not been purified so far. Perhaps an alternative could be to overexpress the proteins in yeast and to perform EMSAs with this lysate.

Due to the fact that deletion of *MMS1* and *MMS22*, but not deletion of *RTT101*, caused an enhanced GCR rate that was due to the presence of a G4tract3 motif, it is highly likely that the ubiquitin ligase might function at G4tract2 motifs, but that Mms1 and Mms22 have an additional ubiquitin ligase-independent function at G4tract3 motifs. Further GCR assays with other inserted G4tract3 motifs should be performed to verify this hypothesis. Also, it would be interesting to study the effect of Mms1, Mms22 and Rtt101 on the GCR rate, if the G4tract2 motif is located on the leading strand of replication, because ChIP experiments (Figure 20D) hint towards the assumption that Mms1 only binds to lagging strand G4 motifs. To support the hypothesis that the observed effect is G4 structure specific, G4 stabilizing ligands could be used. However, Phen-DC_3 treatment did not enhance the GCR rate in *mms1* cells (Figure 28). This might be due to the hypothesis that a G4 stabilizing ligand might be specific for a distinct G4 topology (Piazza et al, 2010). Therefore, other ligands as for example TMPyP4 could be tested.

Such ligands could also be used to investigate viability of cells (growth curve, spot test) with deleted proteins compared to wild type cells. Growth defects in the absence of proteins can then indicate a G4 specific function of these proteins. Of course, it is also possible to perform all of the suggested experiments in the presence of G4 stabilizing ligands to enhance G4 stability. However, it was suggested that different G4 stabilizing ligands might be specific for different G4 topologies (Piazza et al, 2010) and they might also influence protein binding or genome stability on their own. This should be considered when evaluating those experiments.

In one model (section 6.2.5) it was suggested that Mms1 supports Pif1 function. This function could also depend on Rtt101 and Mms22, because Mms1 also supported binding at G4tract2 motifs. Therefore, it should be tested if Rtt101 and Mms22 affect Pif1 binding, too.

In the second model (section 6.2.5), it was proposed that Mms1 is required for the ubiquitination of replisome components to enable fork restart. To confirm this, it could be tested if the ubiquitin ligase ubiquitinates replisome components (e.g. Pol δ) and if it is required for replication fork restart (to study this 2D gels could be performed).

Finally, it would be interesting to find out if the homologous proteins in humans (CUL4/DDB1) are required for the same processes.

8 References

Adrian M, Ang DJ, Lech CJ, Heddi B, Nicolas A, Phan AT (2014) Structure and conformational dynamics of a stacked dimeric G-quadruplex formed by the human CEB1 minisatellite. *J Am Chem Soc* **136:** 6297-6305

Aguilera A, Gomez-Gonzalez B (2008) Genome instability: a mechanistic view of its causes and consequences. *Nature reviews Genetics* **9:** 204-217

Alekseev S, Luijsterburg MS, Pines A, Geverts B, Mari PO, Giglia-Mari G, Lans H, Houtsmuller AB, Mullenders LH, Hoeijmakers JH, Vermeulen W (2008) Cellular concentrations of DDB2 regulate dynamic binding of DDB1 at UV-induced DNA damage. *Molecular and cellular biology* **28:** 7402-7413

Allen JB, Zhou Z, Siede W, Friedberg EC, Elledge SJ (1994) The SAD1/RAD53 protein kinase controls multiple checkpoints and DNA damage-induced transcription in yeast. *Genes Dev* **8:** 2401-2415.

Amberg DC, Burke D, Strathern JN, Burke D, Cold Spring Harbor Laboratory (2005) *Methods in yeast genetics : a Cold Spring Harbor Laboratory course manual*, 2005 edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Angers S, Li T, Yi X, MacCoss MJ, Moon RT, Zheng N (2006) Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature* **443:** 590-593

Araki Y, Kawasaki Y, Sasanuma H, Tye BK, Sugino A (2003) Budding yeast mcm10/dna43 mutant requires a novel repair pathway for viability. *Genes to cells : devoted to molecular & cellular mechanisms* **8:** 465-480

Arora A, Dutkiewicz M, Scaria V, Hariharan M, Maiti S, Kurreck J (2008) Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *Rna* **14:** 1290-1296

Azvolinsky A, Dunaway S, Torres J, Bessler J, Zakian VA (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, Lieb J, Zakian V (2009) Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae. *Molecular cell* **34:** 722-734

Bachrati CZ, Hickson ID (2006) Analysis of the DNA unwinding activity of RecQ family helicases. *Methods in enzymology* **409:** 86-100

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic acids research* **37:** W202-208

Bang I (1910) Untersuchungen über die Guanylsäure. *Biochem Z* **26:** 293–231

Bartholdy B, Mukhopadhyay R, Lajugie J, Aladjem MI, Bouhassira EE (2015) Allele-specific analysis of DNA replication origins in mammalian cells. *Nat Commun* **6:** 7051

Beaudoin JD, Perreault JP (2010) 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic acids research* **38:** 7022-7036

Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, Marin JM, Lemaitre JM (2012) Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat Struct Mol Biol* **19:** 837-844

Bhasikuttan AC, Mohanty J (2015) Targeting G-quadruplex structures with extrinsic fluorogenic dyes: promising fluorescence sensors. *Chem Commun (Camb)* **51:** 7581-7597

Biffi G, Tannahill D, McCafferty J, Balasubramanian S (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature chemistry* **5:** 182-186

Bochman ML, Paeschke K, Chan A, Zakian VA (2014) Hrq1, a homolog of the human RecQ4 helicase, acts catalytically and structurally to promote genome integrity. *Cell reports* **6:** 346-356

Bochman ML, Paeschke K, Zakian VA (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nature reviews Genetics* **13:** 770-780

Bostelman LJ, Keller AM, Albrecht AM, Arat A, Thompson JS (2007) Methylation of histone H3 lysine-79 by Dot1p plays multiple roles in the response to UV damage in Saccharomyces cerevisiae. *DNA repair* **6:** 383-395

Boulton SJ, Jackson SP (1998) Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. *EMBO J* **17:** 1819-1828

Bourns BD, Alexander MK, Smith AM, Zakian VA (1998) Sir proteins, Rif proteins and Cdc13p bind *Saccharomyces* telomeres i*n vivo*. *Mol Cell Biol* **18:** 5600-5608

Branzei D, Seki M, Onoda F, Enomoto T (2002a) The product of Saccharomyces cerevisiae WHIP/MGS1, a gene related to replication factor C genes, interacts functionally with DNA polymerase delta. *Molecular genetics and genomics : MGG* **268:** 371-386

Branzei D, Seki M, Onoda F, Yagi H, Kawabe Y, Enomoto T (2002b) Characterization of the slow-growth phenotype of S. cerevisiae Whip/Mgs1 Sgs1 double deletion mutants. *DNA repair* **1:** 671-682

Brewer BJ, Fangman WL (1988) A replication fork barrier at the 3' end of yeast ribosomal RNA genes. *Cell* **55:** 637-643

Brewer BJ, Lockshon D, Fangman WL (1992) The arrest of replication forks in the rDNA of yeast occurs independently of transcription. *Cell* **71:** 267-276

Budd ME, Campbell JL (1995) A yeast gene required for DNA replication encodes a protein with homology to DNA helicases. *Proc Natl Acad Sci USA* **92:** 7642-7646.

Cahoon LA, Seifert HS (2009) An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae. *Science* **325:** 764-767

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

Capra JA, Paeschke K, Singh M, Zakian VA (2010) G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. *PLoS Comput Biol* **6:** e1000861

Castillo Bosch P, Segura-Bayona S, Koole W, van Heteren JT, Dewar JM, Tijsterman M, Knipscheer P (2014) FANCJ promotes DNA synthesis through G-quadruplex structures. *EMBO J* **33:** 2521-2533

Cayrou C, Ballester B, Peiffer I, Fenouil R, Coulombe P, Andrau JC, van Helden J, Mechali M (2015) The chromatin environment shapes DNA replication origin organization and defines origin classes. *Genome Res* **25:** 1873-1885

Cayrou C, Coulombe P, Puy A, Rialle S, Kaplan N, Segal E, Mechali M (2012) New insights into replication origin characteristics in metazoans. *Cell Cycle* **11:** 658-667

Cayrou C, Coulombe P, Vigneron A, Stanojcic S, Ganier O, Peiffer I, Rivals E, Puy A, Laurent-Chabalier S, Desprat R, Mechali M (2011) Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res* **21:** 1438-1449

Chai W, Sfeir AJ, Hoshiyama H, Shay JW, Wright WE (2006) The involvement of the Mre11/Rad50/Nbs1 complex in the generation of G-overhangs at human telomeres. *EMBO Rep* **7:** 225-230

Chaires JB, Buscaglia R (2013) *Quadruplex nucleic acids*, Heidelberg u.a.: Springer.

Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP, Balasubramanian S (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nature biotechnology* **33:** 877-881

Chen C, Kolodner RD (1999) Gross chromosomal rearrangements in *Saccharomyces cerevisiae* replication and recombination defective mutants. *Nat Genet* **23:** 81-85

Chen L, Trujillo K, Ramos W, Sung P, Tomkinson AE (2001) Promotion of Dnl4-catalyzed DNA end-joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 complexes. *Molecular cell* **8:** 1105-1115

Cheung I, Schertzer M, Rose A, Lansdorp PM (2002) Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA. *Nature genetics* **31:** 405-409

Chung WJ, Heddi B, Schmitt E, Lim KW, Mechulam Y, Phan AT (2015) Structure of a left-handed DNA G-quadruplex. *Proceedings of the National Academy of Sciences of the United States of America* **112:** 2729-2733

Ciechanover A (2005) Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nature reviews Molecular cell biology* **6:** 79-87

Clark GR, Pytel PD, Squire CJ (2012) The high-resolution crystal structure of a parallel intermolecular DNA G-4 quadruplex/drug complex employing syn glycosyl linkages. *Nucleic acids research* **40:** 5731-5738

Clerici M, Mantiero D, Guerini I, Lucchini G, Longhese MP (2008) The Yku70-Yku80 complex contributes to regulate double-strand break processing and checkpoint activation during the cell cycle. *EMBO Rep* **9:** 810-818

Cogoi S, Paramasivam M, Spolaore B, Xodo LE (2008) Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins. *Nucleic acids research* **36:** 3765-3780

Conde F, Refolio E, Cordon-Preciado V, Cortes-Ledesma F, Aragon L, Aguilera A, San-Segundo PA (2009) The Dot1 histone methyltransferase and the Rad9 checkpoint adaptor contribute to cohesin-dependent double-strand break repair by sister chromatid recombination in Saccharomyces cerevisiae. *Genetics* **182:** 437-446

Conde F, San-Segundo PA (2008) Role of Dot1 in the response to alkylating DNA damage in Saccharomyces cerevisiae: regulation of DNA damage tolerance by the error-prone polymerases Polzeta/Rev1. *Genetics* **179:** 1197-1210

Cortes-Ledesma F, Aguilera A (2006) Double-strand breaks arising by replication through a nick are repaired by cohesin-dependent sister-chromatid exchange. *EMBO Rep* **7:** 919-926

Cotterill S. (1999) Eukaryotic DNA replication a practical approach. *Practical approach series 199.* Oxford University Press, Oxford ; New York, pp. 1 online resource (xxii, 281 p.).

Coutavas E, Ren M, Oppenheim JD, D'Eustachio P, Rush MG (1993) Characterization of proteins that interact with the cell-cycle regulatory protein Ran/TC4. *Nature* **366:** 585-587

Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & cellular proteomics : MCP* **13:** 2513-2526

Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26:** 1367-1372

Crabbe L, Verdun RE, Haggblom CI, Karlseder J (2004) Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. *Science* **306:** 1951-1953

Crnugelj M, Hud NV, Plavec J (2002) The solution structure of d(G(4)T(4)G(3))(2): a bimolecular G-quadruplex with a novel fold. *Journal of molecular biology* **320:** 911-924

De Cian A, Delemos E, Mergny JL, Teulade-Fichou MP, Monchaud D (2007) Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J Am Chem Soc* **129:** 1856-1857

De Piccoli G, Cortes-Ledesma F, Ira G, Torres-Rosell J, Uhle S, Farmer S, Hwang JY, Machin F, Ceschia A, McAleenan A, Cordon-Preciado V, Clemente-Blanco A, Vilella-Mitjana F, Ullal P, Jarmuz A, Leitao B, Bressan D, Dotiwala F, Papusha A, Zhao X, Myung K, Haber JE, Aguilera A, Aragon L (2006) Smc5-Smc6 mediate DNA double-strand-break repair by promoting sister-chromatid recombination. *Nat Cell Biol* **8:** 1032-1034

Deng H, Braunlin WH (1995) Duplex to quadruplex equilibrium of the self-complementary oligonucleotide d(GGGGCCCC). *Biopolymers* **35:** 677-681

Downs JA, Lowndes NF, Jackson SP (2000) A role for Saccharomyces cerevisiae histone H2A in DNA repair. *Nature* **408:** 1001-1004

Dunn MJ, International Electrophoresis Society (1986) *Electrophoresis '86 proceedings of the fifth Meeting of the International Electrophoresis Society; London, (9th-12th September), 1986*, Weinheim Deerfield Beach - Fla.: VCH.

Duro E, Vaisica JA, Brown GW, Rouse J (2008) Budding yeast Mms22 and Mms1 regulate homologous recombination induced by replisome blockage. *DNA repair* **7:** 811-818

Eddy J, Maizels N (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic acids research* **34:** 3887-3896

Eddy S, Ketkar A, Zafar MK, Maddukuri L, Choi JY, Eoff RL (2014) Human Rev1 polymerase disrupts G-quadruplex DNA. *Nucleic acids research* **42:** 3272-3285

Eddy S, Maddukuri L, Ketkar A, Zafar MK, Henninger EE, Pursell ZF, Eoff RL (2015) Evidence for the kinetic partitioning of polymerase activity on G-quadruplex DNA. *Biochemistry* **54:** 3218-3230

Elledge SJ, Davis RW (1990) Two genes differentially regulated in the cell cycle and by DNA-damaging agents encode alternative regulatory subunits of ribonucleotide reductase. *Genes Dev* **4:** 740-751

Emili A (1998) *MEC1*-dependent phosphorylation of Rad9p in response to DNA damage. *Molecular cell* **2:** 183-189

Fang G, Cech TR (1993) The beta subunit of Oxytricha telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell* **74:** 875-885

Feldmann H, Winnacker EL (1993) A putative homologue of the human autoantigen Ku from *Saccharomyces cerevisiae*. *J Biol Chem* **268:** 12895-12900

Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K, Zhang Y (2002) Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr Biol* **12:** 1052-1058

Fernando H, Rodriguez R, Balasubramanian S (2008) Selective recognition of a DNA G-quadruplex by an engineered antibody. *Biochemistry* **47:** 9365-9371

Fisher T, Taggart A, Zakian V (2004) Cell cycle-dependent regulation of yeast telomerase by Ku. *Nat Struct Mol Biol* **11:** 1198-1205

Foulk MS, Urban JM, Casella C, Gerbi SA (2015) Characterizing and controlling intrinsic biases of lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. *Genome Res* **25:** 725-735

Fry M, Loeb LA (1999) Human Werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *J Biol Chem* **274:** 12797-12802

Fu Y, Xiao W (2006) Identification and characterization of CRT10 as a novel regulator of Saccharomyces cerevisiae ribonucleotide reductase genes. *Nucleic acids research* **34:** 1876-1883

Gellert M, Lipsett MN, Davies DR (1962) Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences of the United States of America* **48:** 2013-2018

Ghosal G, Muniyappa K (2005) Saccharomyces cerevisiae Mre11 is a high-affinity G4 DNA-binding protein and a G-rich DNA-specific endonuclease: implications for replication of telomeric DNA. *Nucleic acids research* **33:** 4692-4703

Ghosal G, Muniyappa K (2007) The characterization of Saccharomyces cerevisiae Mre11/Rad50/Xrs2 complex reveals that Rad50 negatively regulates Mre11 endonucleolytic but not the exonucleolytic activity. *Journal of molecular biology* **372:** 864-882

Giannattasio M, Lazzaro F, Plevani P, Muzi-Falconi M (2005) The DNA damage checkpoint response requires histone H2B ubiquitination by Rad6-Bre1 and H3 methylation by Dot1. *J Biol Chem* **280:** 9879-9886

Gietz RD, Woods RA (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods in enzymology* **350:** 87-96

Giraldo R, Rhodes D (1994) The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J* **13:** 2411-2420

Giraldo R, Suzuki M, Chapman L, Rhodes D (1994) Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study. *Proceedings of the National Academy of Sciences of the United States of America* **91:** 7658-7662

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

Goudsouzian L, Tuzon C, Zakian VA (2006) *S. cerevisiae* Tel1p and Mre11p are required for normal levels of Est1p and Est2p telomere association. *Molecular cell* **24:** 603-610

Gray MD, Shen JC, Kamath-Loeb AS, Blank A, Sopher BL, Martin GM, Oshima J, Loeb LA (1997) The Werner syndrome protein is a DNA helicase. *Nature genetics* **17:** 100-103

Greenfeder SA, Newlon CS (1992) Replication forks pause at yeast centromeres. *Mol Cell Biol* **12:** 4056-4066

Greenwell PW, Kronmal SL, Porter SE, Gassenhuber J, Obermaier B, Petes TD (1995) TEL1, a gene involved in controlling telomere length in *S. cerevisiae*, is homologous to the human ataxia telangiectasia gene. *Cell* **82:** 823-829

Griffiths AJF (2012) *Introduction to genetic analysis*, Internat. 10. edn. New York: Freeman u.a.

Groisman R, Polanowska J, Kuraoka I, Sawada J, Saijo M, Drapkin R, Kisselev AF, Tanaka K, Nakatani Y (2003) The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. *Cell* **113:** 357-367

Guedin A, Gros J, Alberti P, Mergny JL (2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic acids research* **38:** 7858-7868

Gullo C, Au M, Feng G, Teoh G (2006) The biology of Ku and its potential oncogenic role in cancer. *Biochim Biophys Acta* **1765:** 223-234

Hall BM, Ma CX, Liang P, Singh KK (2009) Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics* **25:** 1564-1565

Hamatake RK, Hasegawa H, Clark AB, Bebenek K, Kunkel TA, Sugino A (1990) Purification and characterization of DNA polymerase II from the yeast Saccharomyces cerevisiae. Identification of the catalytic core and a possible holoenzyme form of the enzyme. *J Biol Chem* **265:** 4072-4083

Han J, Zhang H, Zhang H, Wang Z, Zhou H, Zhang Z (2013) A Cul4 E3 ubiquitin ligase regulates histone hand-off during nucleosome assembly. *Cell* **155:** 817-829

Hatzakis E, Okamoto K, Yang D (2010) Thermodynamic stability and folding kinetics of the major G-quadruplex and its loop isomers formed in the nuclease hypersensitive element in the human c-Myc promoter: effect of loops and flanking segments on the stability of parallel-stranded intramolecular G-quadruplexes. *Biochemistry* **49:** 9152-9160

Hayashi T, Seki M, Inoue E, Yoshimura A, Kusa Y, Tada S, Enomoto T (2008) Vertebrate WRNIP1 and BLM are required for efficient maintenance of genome stability. *Genes & genetic systems* **83:** 95-100

Hazel P, Huppert J, Balasubramanian S, Neidle S (2004) Loop-length-dependent folding of G-quadruplexes. *J Am Chem Soc* **126:** 16405-16415

Hector R, Shtofman R, Ray A, Chen B, Nyun T, Berkner K, KW R (2007) Tel1p preferentially associates with short telomeres to stimulate their elongation. *Molecular cell* **27:** 851-858

Hefferin ML, Tomkinson AE (2005) Mechanism of DNA double-strand break repair by non-homologous end joining. *DNA repair* **4:** 639-648

Heitzer E, Tomlinson I (2014) Replicative DNA polymerase mutations in cancer. *Current opinion in genetics & development* **24:** 107-113

Henderson A, Wu Y, Huang YC, Chavez EA, Platt J, Johnson FB, Brosh RM, Jr., Sen D, Lansdorp PM (2014) Detection of G-quadruplex DNA in mammalian cells. *Nucleic acids research* **42:** 860-869

Hershman SG, Chen Q, Lee JY, Kozak ML, Yue P, Wang LS, Johnson FB (2008) Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic acids research* **36:** 144-156

Heyer WD, Ehmsen KT, Liu J (2010) Regulation of homologous recombination in eukaryotes. *Annual review of genetics* **44:** 113-139

Higa LA, Wu M, Ye T, Kobayashi R, Sun H, Zhang H (2006) CUL4-DDB1 ubiquitin ligase interacts with multiple WD40-repeat proteins and regulates histone methylation. *Nat Cell Biol* **8:** 1277-1283

Hishida T, Iwasaki H, Ohno T, Morishita T, Shinagawa H (2001) A yeast gene, MGS1, encoding a DNA-dependent AAA(+) ATPase is required to maintain genome stability. *Proceedings of the National Academy of Sciences of the United States of America* **98:** 8283-8289

Hishida T, Ohno T, Iwasaki H, Shinagawa H (2002) Saccharomyces cerevisiae MGS1 is essential in strains deficient in the RAD6-dependent DNA damage tolerance pathway. *EMBO J* **21:** 2019-2029

Hishida T, Ohya T, Kubota Y, Kamada Y, Shinagawa H (2006) Functional and physical interaction of yeast Mgs1 with PCNA: impact on RAD6-dependent DNA damage tolerance. *Molecular and cellular biology* **26:** 5509-5517

Howell RM, Woodford KJ, Weitzmann MN, Usdin K (1996) The chicken beta-globin gene promoter forms a novel "cinched" tetrahelical structure. *J Biol Chem* **271:** 5208-5214

Hryciw T, Tang M, Fontanie T, Xiao W (2002) MMS1 protects against replication-dependent DNA damage in Saccharomyces cerevisiae. *Molecular genetics and genomics : MGG* **266:** 848-857

Huber MD, Duquette ML, Shiels JC, Maizels N (2006) A conserved G4 DNA binding domain in RecQ family helicases. *Journal of molecular biology* **358:** 1071-1080

Huppert JL (2008) Hunting G-quadruplexes. *Biochimie* **90:** 1140-1148

Huppert JL, Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. *Nucleic acids research* **33:** 2908-2916

Huppert JL, Balasubramanian S (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic acids research* **35:** 406-413

Huppert JL, Bugaut A, Kumari S, Balasubramanian S (2008) G-quadruplexes: the beginning and end of UTRs. *Nucleic acids research* **36:** 6260-6268

International Human Genome Sequencing C (2004) Finishing the euchromatic sequence of the human genome. *Nature* **431:** 931-945

Ivessa AS, Lenzmeier BA, Bessler JB, Goudsouzian LK, Schnakenberg SL, Zakian VA (2003) The *Saccharomyces cerevisiae* helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. *Mol Cell* **12:** 1525-1536

Ivessa AS, Zhou J-Q, Schulz VP, Monson EM, Zakian VA (2002) *Saccharomyces* Rrm3p, a 5' to 3' DNA helicase that promotes replication fork progression through telomeric and sub-telomeric DNA. *Genes Dev* **16:** 1383-1396

Ivessa AS, Zhou J-Q, Zakian VA (2000) The *Saccharomyces* Pif1p DNA helicase and the highly related Rrm3p have opposite effects on replication fork progression in ribosomal DNA. *Cell* **100:** 479–489

Javaheri A, Wysocki R, Jobin-Robitaille O, Altaf M, Cote J, Kron SJ (2006) Yeast G1 DNA damage checkpoint regulation by H2A phosphorylation is independent of chromatin remodeling. *Proceedings of the National Academy of Sciences of the United States of America* **103:** 13771-13776

Jeggo PA, Pearl LH, Carr AM (2016) DNA repair, genome stability and cancer: a historical perspective. *Nature reviews Cancer* **16:** 35-42

Jin R, Gaffney BL, Wang C, Jones RA, Breslauer KJ (1992) Thermodynamics and structure of a DNA tetraplex: a spectroscopic and calorimetric study of the tetramolecular complexes of d(TG3T) and d(TG3T2G3T). *Proceedings of the National Academy of Sciences of the United States of America* **89:** 8832-8836

Johnson JE, Cao K, Ryvkin P, Wang LS, Johnson FB (2010) Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic acids research* **38:** 1114-1122

Johzuka K, Ogawa H (1995) Interaction of Mre11 and Rad50: two proteins required for DNA repair and meiosis-specific double-strand break formation in *Saccharomyces cerevisiae*. *Genetics* **139:** 1521-1532

Jutras BL, Verma A, Stevenson B (2012) Identification of novel DNA-binding proteins using DNA-affinity chromatography/pull down. *Current protocols in microbiology* **Chapter 1:** Unit1F 1

Kamath-Loeb AS, Loeb LA, Johansson E, Burgers PM, Fry M (2001) Interactions between the Werner syndrome helicase and DNA polymerase delta specifically facilitate copying of tetraplex and hairpin structures of the d(CGG)n trinucleotide repeat sequence. *J Biol Chem* **276:** 16439-16446

Kang C, Zhang X, Ratliff R, Moyzis R, Rich A (1992) Crystal structure of four-stranded Oxytricha telomeric DNA. *Nature* **356:** 126-131

Katou Y, Kanoh Y, Bando M, Noguchi H, Tanaka H, Ashikari T, Sugimoto K, Shirahige K (2003) S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. *Nature* **424:** 1078-1083

Kawabe Y, Branzei D, Hayashi T, Suzuki H, Masuko T, Onoda F, Heo SJ, Ikeda H, Shimamoto A, Furuichi Y, Seki M, Enomoto T (2001) A novel protein interacts with the Werner's syndrome gene product physically and functionally. *J Biol Chem* **276:** 20364-20369

Kobayashi T (2003) The replication fork barrier site forms a unique structure with Fob1p and inhibits the replication fork. *Mol Cell Biol* **23:** 9178-9188

Kobayashi T, Horiuchi T (1996) A yeast gene product, Fob1 protein, required for both replication fork blocking and recombinational hotspot activities. *Genes to cells : devoted to molecular & cellular mechanisms* **1:** 465-474

Koc A, Wheeler LJ, Mathews CK, Merrill GF (2003) Replication-independent MCB gene induction and deoxyribonucleotide accumulation at G1/S in Saccharomyces cerevisiae. *J Biol Chem* **278:** 9345-9352

Koole W, van Schendel R, Karambelas AE, van Heteren JT, Okihara KL, Tijsterman M (2014) A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nat Commun* **5:** 3216

Kruisselbrink E, Guryev V, Brouwer K, Pontier DB, Cuppen E, Tijsterman M (2008) Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective C. elegans. *Current biology : CB* **18:** 900-905

Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature methods* **11:** 319-324

Kumar N, Sahoo B, Varun KA, Maiti S, Maiti S (2008) Effect of loop length variation on quadruplex-Watson Crick duplex competition. *Nucleic acids research* **36:** 4433-4442

Kumari S, Bugaut A, Huppert JL, Balasubramanian S (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nature chemical biology* **3:** 218-221

Lacoste N, Utley RT, Hunter JM, Poirier GG, Cote J (2002) Disruptor of telomeric silencing-1 is a chromatin-specific histone H3 methyltransferase. *J Biol Chem* **277:** 30421-30424

Lahaye A, Stahl H, Thines-Sempoux D, Foury F (1991) PIF1: a DNA helicase in yeast mitochondria. *EMBO J* **10:** 997-1007

Lam EY, Beraldi D, Tannahill D, Balasubramanian S (2013) G-quadruplex structures are stable and detectable in human genomic DNA. *Nat Commun* **4:** 1796

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10:** R25

Lee MY, Zhang S, Lin SH, Wang X, Darzynkiewicz Z, Zhang Z, Lee EY (2014) The tail that wags the dog: p12, the smallest subunit of DNA polymerase delta, is degraded by ubiquitin ligases in response to DNA damage and during cell cycle progression. *Cell Cycle* **13:** 23-31

Lin W, Sampathi S, Dai H, Liu C, Zhou M, Hu J, Huang Q, Campbell J, Shin-Ya K, Zheng L, Chai W, Shen B (2013) Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. *EMBO J* **32:** 1425-1439

Linskens MH, Huberman JA (1988) Organization of replication of ribosomal DNA in Saccharomyces cerevisiae. *Molecular and cellular biology* **8:** 4927-4935

Lipps HJ, Gruissem W, Prescott DM (1982) Higher order DNA structure in macronuclear chromatin of the hypotrichous ciliate *Oxytricha nova*. *Proc Natl Acad Sci USA* **79:** 2495-2499

Lipps HJ, Rhodes D (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol* **19:** 414-422

Lisby M, Barlow JH, Burgess RC, Rothstein R (2004) Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. *Cell* **118:** 699-713

Liu Z, Lee A, Gilbert W (1995) Gene disruption of a G4-DNA-dependent nuclease in yeast leads to cellular senescence and telomere shortening. *Proc Natl Acad Sci USA* **92:** 6002-6006

Llorente B, Symington LS (2004) The Mre11 nuclease is not required for 5' to 3' resection at multiple HO-induced double-strand breaks. *Molecular and cellular biology* **24:** 9682-9694

Lodish HF (2013) *Molecular cell biology*, 7. edn. New York, NY: Freeman u.a.

London TB, Barber LJ, Mosedale G, Kelly GP, Balasubramanian S, Hickson ID, Boulton SJ, Hiom K (2008) FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *J Biol Chem* **283:** 36132-36139

Lopes J, Piazza A, Bermejo R, Kriegsman B, Colosio A, Teulade-Fichou MP, Foiani M, Nicolas A (2011) G-quadruplex-induced instability during leading-strand replication. *The EMBO journal* **30:** 4033-4046

Lopes M, Cotta-Ramusino C, Pellicioli A, Liberi G, Plevani P, Muzi-Falconi M, Newlon CS, Foiani M (2001) The DNA replication checkpoint response stabilizes stalled replication forks. *Nature* **412:** 557-561.

Lopes M, Foiani M, Sogo JM (2006) Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. *Molecular cell* **21:** 15-27

Lovejoy CA, Lock K, Yenamandra A, Cortez D (2006) DDB1 maintains genome integrity through regulation of Cdt1. *Molecular and cellular biology* **26:** 7977-7990

Luke B, Versini G, Jaquenoud M, Zaidi IW, Kurz T, Pintard L, Pasero P, Peter M (2006) The cullin Rtt101p promotes replication fork progression through damaged DNA and natural pause sites. *Current biology : CB* **16:** 786-792

Lustig AJ, Petes TD (1986) Identification of yeast mutants with altered telomere structure. *Proc Natl Acad Sci USA* **83:** 1398-1402

Ma DL, Zhang Z, Wang M, Lu L, Zhong HJ, Leung CH (2015) Recent Developments in G-Quadruplex Probes. *Chemistry & biology* **22:** 812-828

Maizels N, Gray LT (2013) The G4 genome. *PLoS genetics* **9:** e1003468

Martino F, Kueng S, Robinson P, Tsai-Pflugfelder M, van Leeuwen F, Ziegler M, Cubizolles F, Cockell MM, Rhodes D, Gasser SM (2009) Reconstitution of yeast silent chromatin: multiple contact sites and O-AADPR binding load SIR complexes onto nucleosomes in vitro. *Molecular cell* **33:** 323-334

Masuda T, Tomita M, Ishihama Y (2008) Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *Journal of proteome research* **7:** 731-740

Mateyak M, Zakian V (2006) Human PIF helicase is cell cycle regulated and associates with telomerase. *Cell Cycle* **23:** 2796-2804

McGee JS, Phillips JA, Chan A, Sabourin M, Paeschke K, Zakian VA (2010) Reduced Rif2 and lack of Mec1 target short telomeres for elongation rather than double-strand break repair. *Nature structural & molecular biology* **17:** 1438-1445

Mekmaysy CS, Petraccone L, Garbett NC, Ragazzon PA, Gray R, Trent JO, Chaires JB (2008) Effect of O6-methylguanine on the stability of G-quadruplex DNA. *J Am Chem Soc* **130:** 6710-6711

Mendoza O, Bourdoncle A, Boule JB, Brosh RM, Jr., Mergny JL (2016) G-quadruplexes and helicases. *Nucleic acids research* **44:** 1989-2006

Mimori T, Hardin JA, Steitz JA (1986) Characterization of the DNA-binding protein antigen Ku recognized by autoantibodies from patients with rheumatic disorders. *J Biol Chem* **261:** 2274-2278

Mimura S, Yamaguchi T, Ishii S, Noro E, Katsura T, Obuse C, Kamura T (2010) Cul8/Rtt101 forms a variety of protein complexes that regulate DNA damage response and transcriptional silencing. *J Biol Chem* **285:** 9858-9867

Minca EC, Kowalski D (2011) Replication fork stalling by bulky DNA damage: localization at active origins and checkpoint modulation. *Nucleic acids research* **39:** 2610-2623

Moreau S, Ferguson JR, Symington LS (1999) The nuclease activity of Mre11 is required for meiosis but not for mating type switching, end joining, or telomere maintenance. *Mol Cell Biol* **19:** 556-566

Moreau S, Morgan EA, Symington LS (2001) Overlapping functions of the *Saccharomyces cerevisiae* Mre11, Exo1 and Rad27 nucleases in DNA metabolism. *Genetics* **159:** 1423-1433.

Morris MJ, Negishi Y, Pazsint C, Schonhoft JD, Basu S (2010) An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. *Journal of the American Chemical Society* **132:** 17831-17839

Morrison A, Araki H, Clark AB, Hamatake RK, Sugino A (1990) A third essential DNA polymerase in S. cerevisiae. *Cell* **62:** 1143-1151

Mullen MA, Olson KJ, Dallaire P, Major F, Assmann SM, Bevilacqua PC (2010) RNA G-Quadruplexes in the model plant species Arabidopsis thaliana: prevalence and possible functional roles. *Nucleic acids research* **38:** 8149-8163

Muller S, Kumari S, Rodriguez R, Balasubramanian S (2010) Small-molecule-mediated G-quadruplex isolation from human cells. *Nature chemistry* **2:** 1095-1098

Nakada D, Matsumoto K, Sugimoto K (2003a) ATM-related Tel1 associates with double-strand breaks through an Xrs2-dependent mechanism. *Genes Dev* **17:** 1957-1962

Nakada D, Shimomura T, Matsumoto K, Sugimoto K (2003b) The ATM-related Tel1 protein of Saccharomyces cerevisiae controls a checkpoint response following phleomycin treatment. *Nucleic acids research* **31:** 1715-1724

Nakken S, Rognes T, Hovig E (2009) The disruptive positions in human G-quadruplex motifs are less polymorphic and more conserved than their neutral counterparts. *Nucleic acids research* **37:** 5749-5756

Neidle S (2012) *Therapeutic applications of quadruplex nucleic acids*, London ; Waltham, MA: Elsevier/Academic Press.

Ng HH, Feng Q, Wang H, Erdjument-Bromage H, Tempst P, Zhang Y, Struhl K (2002) Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev* **16:** 1518-1527

Nishitani H, Shiomi Y, Iida H, Michishita M, Takami T, Tsurimoto T (2008) CDK inhibitor p21 is degraded by a proliferating cell nuclear antigen-coupled Cul4-DDB1Cdt2 pathway during S phase and after UV irradiation. *J Biol Chem* **283:** 29045-29052

Paeschke K, Bochman ML, Garcia PD, Cejka P, Friedman KL, Kowalczykowski SC, Zakian VA (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* **497:** 458-462

Paeschke K, Capra JA, Zakian VA (2011) DNA Replication through G-Quadruplex Motifs Is Promoted by the Saccharomyces cerevisiae Pif1 DNA Helicase. *Cell* **145:** 678-691

Paeschke K, Juranek S, Simonsson T, Hempel A, Rhodes D, Lipps HJ (2008) Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates. *Nature structural & molecular biology* **15:** 598-604

Paeschke K, Simonsson T, Postberg J, Rhodes D, Lipps HJ (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nature Struct Mol Biol* **12:** 847-854

Palmbos PL, Daley JM, Wilson TE (2005) Mutations of the Yku80 C terminus and Xrs2 FHA domain specifically block yeast nonhomologous end joining. *Molecular and cellular biology* **25:** 10782-10790

Paulovich AG, Hartwell LH (1995) A checkpoint regulates the rate of progression through S phase in *S. cerevisiae* in response to DNA damage. *Cell* **82:** 841-847

Paulovich AG, Margulies RU, Garvik BM, Hartwell LH (1997) *RAD9*, *RAD17*, and *RAD24* are required for S phase regulation in *Saccharomyces cerevisiae* in response to DNA damage. *Genetics* **145:** 45-62

Petrini JH, Walsh ME, DiMare C, Chen XN, Korenberg JR, Weaver DT (1995) Isolation and characterization of the human MRE11 homologue. *Genomics* **29:** 80-86

Phan AT, Luu KN, Patel DJ (2006) Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K+ solution. *Nucleic acids research* **34:** 5715-5719

Phan AT, Mergny JL (2002) Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic acids research* **30:** 4618-4625

Phan AT, Patel DJ (2003) Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J Am Chem Soc* **125:** 15021-15027

Phillips JA, Chan A, Paeschke K, Zakian VA (2015) The pif1 helicase, a negative regulator of telomerase, acts preferentially at long telomeres. *PLoS genetics* **11:** e1005186

Piazza A, Adrian M, Samazan F, Heddi B, Hamon F, Serero A, Lopes J, Teulade-Fichou MP, Phan AT, Nicolas A (2015) Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. *EMBO J* **34:** 1718-1734

Piazza A, Boule JB, Lopes J, Mingo K, Largy E, Teulade-Fichou MP, Nicolas A (2010) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae. *Nucleic acids research* **38:** 4337-4348

Piazza A, Serero A, Boule JB, Legoix-Ne P, Lopes J, Nicolas A (2012) Stimulation of gross chromosomal rearrangements by the human CEB1 and CEB25 minisatellites in Saccharomyces cerevisiae depends on G-quadruplexes or Cdc13. *PLoS genetics* **8:** e1003033

Postberg J, Tsytlonok M, Sparvoli D, Rhodes D, Lipps HJ (2012) A telomerase-associated RecQ protein-like helicase resolves telomeric G-quadruplex structures during replication. *Gene*

Potts PR, Porteus MH, Yu H (2006) Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks. *EMBO J* **25:** 3377-3388

Putnam CD, Kolodner RD (2010) Determination of gross chromosomal rearrangement rates. *Cold Spring Harbor protocols* **2010:** pdb prot5492

Putnam CD, Srivatsan A, Nene RV, Martinez SL, Clotfelter SP, Bell SN, Somach SB, de Souza JE, Fonseca AF, de Souza SJ, Kolodner RD (2016) A genetic network that suppresses genome rearrangements in Saccharomyces cerevisiae and contains defects in cancers. *Nat Commun* **7:** 11256

Qin M, Chen Z, Luo Q, Wen Y, Zhang N, Jiang H, Yang H (2015) Two-quartet G-quadruplexes formed by DNA sequences containing four contiguous GG runs. *J Phys Chem B* **119:** 3706-3713

Rappsilber J, Ishihama Y, Mann M (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Analytical chemistry* **75:** 663-670

Rawal P, Kummarasetti VB, Ravindran J, Kumar N, Halder K, Sharma R, Mukerji M, Das SK, Chowdhury S (2006) Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation. *Genome Res* **16:** 644-655

Rhodes D, Lipps HJ (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic acids research* **43:** 8627-8637

Ribeyre C, Lopes J, Boulé J, Piazza P, Guédin A, Zakian V, Mergny J-L, Nicolas A (2009) The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences *in vivo*. *PLoS genetics* **5:** e1000475

Ritchie KB, Mallory JC, Petes TD (1999) Interactions of TLC1 (which encodes the RNA subunit of telomerase), TEL1, and MEC1 in regulating telomere length in the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol* **19:** 6065-6075

Rodriguez R, Miller KM, Forment JV, Bradshaw CR, Nikan M, Britton S, Oelschlaegel T, Xhemalce B, Balasubramanian S, Jackson SP (2012) Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nature chemical biology* **8:** 301-310

Rodriguez R, Muller S, Yeoman JA, Trentesaux C, Riou JF, Balasubramanian S (2008) A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J Am Chem Soc* **130:** 15758-15759

Sabouri N, Capra JA, Zakian VA (2014) The essential Schizosaccharomyces pombe Pfh1 DNA helicase promotes fork movement past G-quadruplex motifs to prevent DNA damage. *BMC biology* **12:** 101

Sabourin M, Tuzon C, VA Z (2007) Telomerase and Tel1p preferentially associate with short telomeres in *S. cerevisiae*. *Molecular cell* **27:** 550-561

Sagi J, Renciuk D, Tomasko M, Vorlickova M (2010) Quadruplexes of human telomere DNA analogs designed to contain G:A:G:A, G:G:A:A, and A:A:A:A tetrads. *Biopolymers* **93:** 880-886

Saintome C, Amrane S, Mergny JL, Alberti P (2016) The exception that confirms the rule: a higher-order telomeric G-quadruplex structure more stable in sodium than in potassium. *Nucleic acids research* **44:** 2926-2935

Salazar M, Thompson BD, Kerwin SM, Hurley LH (1996) Thermally induced DNA.RNA hybrid to G-quadruplex transitions: possible implications for telomere synthesis by telomerase. *Biochemistry* **35:** 16110-16115

Sanchez Y, Desany BA, Jones WJ, Liu Q, Wang B, Elledge SJ (1996) Regulation of RAD53 by the ATM-like kinases MEC1 and TEL1 in yeast cell cycle checkpoint pathways. *Science* **271:** 357-360

Sanders CM (2010) Human Pif1 helicase is a G-quadruplex DNA binding protein with G-quadruplex DNA unwinding activity. *The Biochemical journal*

Sang Y, Yan F, Ren X (2015) The role and mechanism of CRL4 E3 ubiquitin ligase in cancer and its potential therapy implications. *Oncotarget* **6:** 42590-42602

Sarkies P, Murat P, Phillips LG, Patel KJ, Balasubramanian S, Sale JE (2012) FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic acids research* **40:** 1485-1498

Sarkies P, Reams C, Simpson LJ, Sale JE (2010) Epigenetic instability due to defective replication of structured DNA. *Molecular cell* **40:** 703-713

Schaffitzel C, Berger I, Postberg J, Hanes J, Lipps HJ, Pluckthun A (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc Natl Acad Sci* **98:** 8572-8577

Schiavone D, Guilbaud G, Murat P, Papadopoulou C, Sarkies P, Prioleau MN, Balasubramanian S, Sale JE (2014) Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *EMBO J* **33:** 2507-2520

Schmidt KH, Pennaneach V, Putnam CD, Kolodner RD (2006) Analysis of gross-chromosomal rearrangements in *Saccharomyces cerevisiae*. *Methods in enzymology* **409:** 462-476

Schulz VP, Zakian VA (1994) The *Saccharomyces PIF1* DNA helicase inhibits telomere elongation and *de novo* telomere formation. *Cell* **76:** 145-155

Sen D, Gilbert W (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334:** 364-366

Sherman F (1991) Getting started with yeast. In *Guide to yeast genetics and molecular biology*, Guthrie C, Fink GR (eds), Vol. 194, pp 3-21. San Diego: Harcourt Brace Jovanovich

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

Shirude PS, Okumus B, Ying L, Ha T, Balasubramanian S (2007) Single-molecule conformational analysis of G-quadruplex formation in the promoter DNA duplex of the proto-oncogene c-kit. *J Am Chem Soc* **129:** 7484-7485

Shroff R, Arbel-Eden A, Pilch D, Ira G, Bonner WM, Petrini JH, Haber JE, Lichten M (2004) Distribution and dynamics of chromatin modification induced by a defined DNA double-strand break. *Current biology : CB* **14:** 1703-1711

Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences of the United States of America* **99:** 11593-11598

Siede W, Friedberg AS, Friedberg EC (1993) RAD9-dependent G1 arrest defines a second checkpoint for damaged DNA in the cell cycle of Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* **90:** 7985-7989

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

Simonsson T, Pecinka P, Kubista M (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic acids research* **26:** 1167-1172

Skolakova P, Bednarova K, Vorlickova M, Sagi J (2010) Quadruplexes of human telomere dG(3)(TTAG(3))(3) sequences containing guanine abasic sites. *Biochemical and biophysical research communications* **399:** 203-208

Smirnov I, Shafer RH (2000) Effect of loop sequence and size on DNA aptamer stability. *Biochemistry* **39:** 1462-1468

Smith FW, Feigon J (1993) Strand orientation in the DNA quadruplex formed from the Oxytricha telomere repeat oligonucleotide d(G4T4G4) in solution. *Biochemistry* **32:** 8682-8692

Smith FW, Lau FW, Feigon J (1994) d(G3T4G3) forms an asymmetric diagonally looped dimeric quadruplex with guanosine 5'-syn-syn-anti and 5'-syn-anti-anti N-glycosidic conformations. *Proceedings of the National Academy of Sciences of the United States of America* **91:** 10546-10550

Smolka MB, Albuquerque CP, Chen SH, Zhou H (2007) Proteome-wide identification of in vivo targets of DNA damage checkpoint kinases. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 10364-10369

Sogo JM, Lopes M, Foiani M (2002) Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science* **297:** 599-602.

Stellwagen AE, Haimberger ZW, Veatch JR, Gottschling DE (2003) Ku interacts with telomerase RNA to promote telomere addition at native and broken chromosome ends. *Genes Dev* **17:** 2384-2395

Strom L, Karlsson C, Lindroos HB, Wedahl S, Katou Y, Shirahige K, Sjogren C (2007) Postreplicative formation of cohesion is required for repair and induced by a single DNA break. *Science* **317:** 242-245

Strom L, Lindroos HB, Shirahige K, Sjogren C (2004) Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair. *Molecular cell* **16:** 1003-1015

Suhasini AN, Brosh RM, Jr. (2013) Disease-causing missense mutations in human DNA helicase disorders. *Mutation research* **752:** 138-152

Sun D, Hurley LH (2009) The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *Journal of medicinal chemistry* **52:** 2863-2874

Sun H, Bennett R, Maizels N (1999) The Saccharomyces cerevisiae Sgs1 helicase efficiently unwinds G-G paired DNAs. *Nucl Acids Res* **27:** 1978-1984

Sun H, Karow J, Hickson I, Maizels N (1998) The Bloom's Syndrome Helicase Unwinds G4 DNA. *J Biol Chem* **273:** 27587-27592

Sun Z, Fay DS, Marini F, Foiani M, Stern DF (1996) Spk1/Rad53 is regulated by Mec1-dependent protein phosphorylation in DNA replication and damage checkpoint pathways. *Genes Dev* **10:** 395-406

Sundquist WI, Klug A (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **342:** 825-829

Sweeney FD, Yang F, Chi A, Shabanowitz J, Hunt DF, Durocher D (2005) Saccharomyces cerevisiae Rad9 acts as a Mec1 adaptor to allow Rad53 activation. *Current biology : CB* **15:** 1364-1375

Symington LS, Gautier J (2011) Double-strand break end resection and repair pathway choice. *Annual review of genetics* **45:** 247-271

Takahashi YH, Schulze JM, Jackson J, Hentrich T, Seidel C, Jaspersen SL, Kobor MS, Shilatifard A (2011) Dot1 and histone H3K79 methylation in natural telomeric and HM silencing. *Molecular cell* **42:** 118-126

Takata H, Kanoh Y, Gunge N, Shirahige K, Matsuura A (2004) Reciprocal association of the budding yeast ATM-related proteins Tel1 and Mec1 with telomeres *in vivo*. *Molecular cell* **14:** 515-522

Takata H, Tanaka Y, Matsuura A (2005) Late S phase-specific recruitment of Mre11 complex triggers hierarchical assembly of telomere replication proteins in *Saccharomyces cerevisiae*. *Molecular cell* **17:** 573-583

Tercero JA, Diffley JF (2001) Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint. *Nature* **412:** 553-557.

Todd AK, Johnston M, Neidle S (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic acids research* **33:** 2901-2907

Toh GW, O'Shaughnessy AM, Jimeno S, Dobbie IM, Grenon M, Maffini S, O'Rorke A, Lowndes NF (2006) Histone H2A phosphorylation and H3 methylation are required for a novel Rad9 DSB repair function following checkpoint activation. *DNA repair* **5:** 693-703

Tsukamoto Y, Kato J, Ikeda H (1996) Hdf1, a yeast Ku-protein homologue, is involved in illegitimate recombination, but not in homologous recombination. *Nucleic acids research* **24:** 2067-2072

Tsukamoto Y, Taggart AKP, Zakian VA (2001) The role of the Mre11-Rad50-Xrs2 complex in telomerase-mediated lengthening of *Saccharomyces cerevisiae* telomeres. *Curr Biol* **11:** 1328-1335

Tuteja N, Tuteja R, Ochem A, Taneja P, Huang NW, Simoncsits A, Susic S, Rahman K, Marusic L, Chen J, et al. (1994) Human DNA helicase II: a novel DNA unwinding enzyme identified as the Ku autoantigen. *EMBO J* **13:** 4991-5001

Ui A, Seki M, Ogiwara H, Lai MS, Yamamoto K, Tada S, Enomoto T (2007) Activation of a novel pathway involving Mms1 and Rad59 in sgs1 cells. *Biochemical and biophysical research communications* **356:** 1031-1037

Usui T, Ohta T, Oshiumi H, Tomizawa J, Ogawa H, Ogawa T (1998) Complex formation and functional versatility of Mre11 of budding yeast in recombination. *Cell* **95:** 705-716

Vaisica JA, Baryshnikova A, Costanzo M, Boone C, Brown GW (2011) Mms1 and Mms22 stabilize the replisome during replication stress. *Molecular biology of the cell* **22:** 2396-2408

Valton AL, Hassan-Zadeh V, Lema I, Boggetto N, Alberti P, Saintome C, Riou JF, Prioleau MN (2014) G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J* **33:** 732-746

Van Driessche B, Tafforeau L, Hentges P, Carr AM, Vandenhaute J (2005) Additional vectors for PCR-based gene tagging in Saccharomyces cerevisiae and Schizosaccharomyces pombe using nourseothricin resistance. *Yeast* **22:** 1061-1068

van Kregten M, Tijsterman M (2014) The repair of G-quadruplex-induced DNA damage. *Experimental cell research* **329:** 178-183

van Leeuwen F, Gafken PR, Gottschling DE (2002) Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* **109:** 745-756

Vejrup-Hansen R, Mizuno K, Miyabe I, Fleck O, Holmberg C, Murray JM, Carr AM, Nielsen O (2011) Schizosaccharomyces pombe Mms1 channels repair of perturbed replication into Rhp51 independent homologous recombination. *DNA repair* **10:** 283-295

von Hacht A, Seifert O, Menger M, Schutze T, Arora A, Konthur Z, Neubauer P, Wagner A, Weise C, Kurreck J (2014) Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic acids research* **42:** 6630-6644

Vummidi BR, Alzeer J, Luedtke NW (2013) Fluorescent probes for G-quadruplex structures. *Chembiochem : a European journal of chemical biology* **14:** 540-558

Wakasugi M, Kasashima H, Fukase Y, Imura M, Imai R, Yamada S, Cleaver JE, Matsunaga T (2009) Physical and functional interaction between DDB and XPA in nucleotide excision repair. *Nucleic acids research* **37:** 516-525

Wallgren M, Mohammad JB, Yan KP, Pourbozorgi-Langroudi P, Ebrahimi M, Sabouri N (2016) G-rich telomeric and ribosomal DNA sequences from the fission yeast genome form stable G-quadruplex DNA structures in vitro and are unwound by the Pfh1 DNA helicase. *Nucleic acids research*

Weinert TA, Hartwell LH (1988) The *RAD9* gene controls the cell cycle response to DNA damage in *Saccharomyces cerevisiae*. *Science* **241:** 317-322

Weinert TA, Kiser GL, Hartwell LH (1994) Mitotic checkpoint genes in budding yeast and the dependence of mitosis on DNA replication and repair. *Genes Dev* **8:** 652-665

Weitzmann MN, Woodford KJ, Usdin K (1996) The development and use of a DNA polymerase arrest assay for the evaluation of parameters affecting intrastrand tetraplex formation. *J Biol Chem* **271:** 20958-20964

Wessel D, Flugge UI (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical biochemistry* **138:** 141-143

Westmoreland JW, Resnick MA (2013) Coincident resection at both ends of random, gamma-induced double-strand breaks requires MRX (MRN), Sae2 (Ctp1), and Mre11-nuclease. *PLoS genetics* **9:** e1003420

Wieland M, Hartig JS (2009) Investigation of mRNA quadruplex formation in Escherichia coli. *Nat Protoc* **4:** 1632-1640

Williamson JR, Raghuraman MK, Cech TR (1989) Monovalent cation-induced structure of telomeric DNA: The G-Quartet model. *Cell* **59:** 871-880

Wittschieben BO, Iwai S, Wood RD (2005) DDB1-DDB2 (xeroderma pigmentosum group E) protein complex recognizes a cyclobutane pyrimidine dimer, mismatches, apurinic/apyrimidinic sites, and compound lesions in DNA. *J Biol Chem* **280:** 39982-39989

Woodford KJ, Howell RM, Usdin K (1994) A novel K(+)-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. *J Biol Chem* **269:** 27029-27035

Wu CG, Spies M (2016) G-quadruplex recognition and remodeling by the FANCJ helicase. *Nucleic acids research*

Wu D, Topper LM, Wilson TE (2008) Recruitment and dissociation of nonhomologous end joining proteins at a DNA double-strand break in Saccharomyces cerevisiae. *Genetics* **178:** 1237-1249

Wu X, Lieber MR (1996) Protein-protein and protein-DNA interaction regions within the DNA end-binding protein Ku70-Ku86. *Molecular and cellular biology* **16:** 5186-5193

Wysocki R, Javaheri A, Allard S, Sha F, Cote J, Kron SJ (2005) Role of Dot1-dependent histone H3 methylation in G1 and S phase DNA damage checkpoint functions of Rad9. *Molecular and cellular biology* **25:** 8430-8443

Yadav VK, Abraham JK, Mani P, Kulshrestha R, Chowdhury S (2008) QuadBase: genome-wide database of G4 DNA--occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. *Nucleic acids research* **36:** D381-385

Yates LR, Campbell PJ (2012) Evolution of the cancer genome. *Nature reviews Genetics* **13:** 795-806

Zaidi IW, Rabut G, Poveda A, Scheel H, Malmstrom J, Ulrich H, Hofmann K, Pasero P, Peter M, Luke B (2008) Rtt101 and Mms1 in budding yeast form a CUL4(DDB1)-like ubiquitin ligase that promotes replication through damaged DNA. *EMBO Rep* **9:** 1034-1040

Zaug AJ, Podell ER, Cech TR (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 10864-10869

Zhang ML, Tong XJ, Fu XH, Zhou BO, Wang J, Liao XH, Li QJ, Shen N, Ding J, Zhou JQ (2010) Yeast telomerase subunit Est1p has guanine quadruplex-promoting activity that is required for telomere elongation. *Nat Struct Mol Biol* **17:** 202-209

Zhang S, Zhao H, Darzynkiewicz Z, Zhou P, Zhang Z, Lee EY, Lee MY (2013) A novel function of CRL4(Cdt2): regulation of the subunit structure of DNA polymerase delta in response to DNA damage and during the S phase. *J Biol Chem* **288:** 29550-29561

Zhang Y, Hefferin ML, Chen L, Shim EY, Tseng HM, Kwon Y, Sung P, Lee SE, Tomkinson AE (2007) Role of Dnl4-Lif1 in nonhomologous end-joining repair complex assembly and suppression of homologous recombination. *Nat Struct Mol Biol* **14:** 639-646

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS (2008) Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9:** R137

Zimmer J, Tacconi EM, Folio C, Badie S, Porru M, Klare K, Tumiati M, Markkanen E, Halder S, Ryan A, Jackson SP, Ramadan K, Kuznetsov SG, Biroccio A, Sale JE, Tarsounas M (2016) Targeting BRCA1 and BRCA2 Deficiencies with G-Quadruplex-Interacting Compounds. *Molecular cell* **61:** 449-460

9 Annex

9.1 Figures

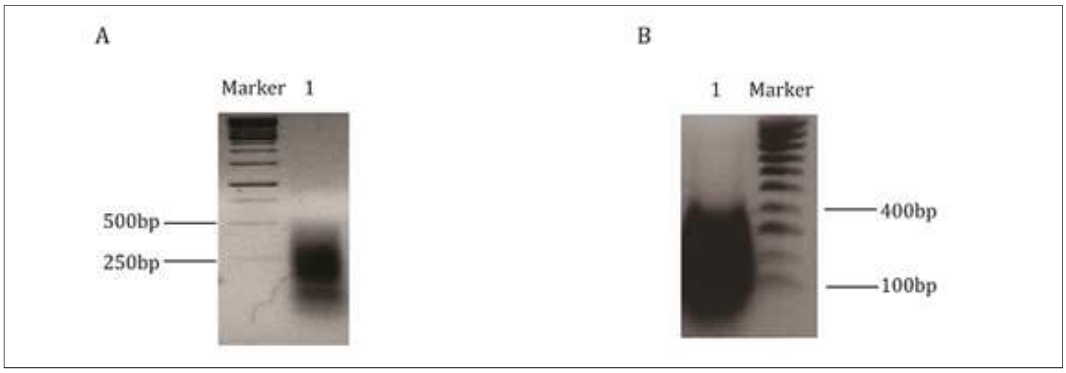


Figure 1: DNA fragment sizes during ChIP and ChIP-seq. (A) DNA fragment sizes after sonication of the Myc-tagged Mms1 strain used for ChIP analysis (1) and DNA ladder (Marker). (B) DNA fragment sizes after sonication of the Myc-tagged Mms1 strain used for ChIP-seq analysis (1) and DNA ladder (Marker).

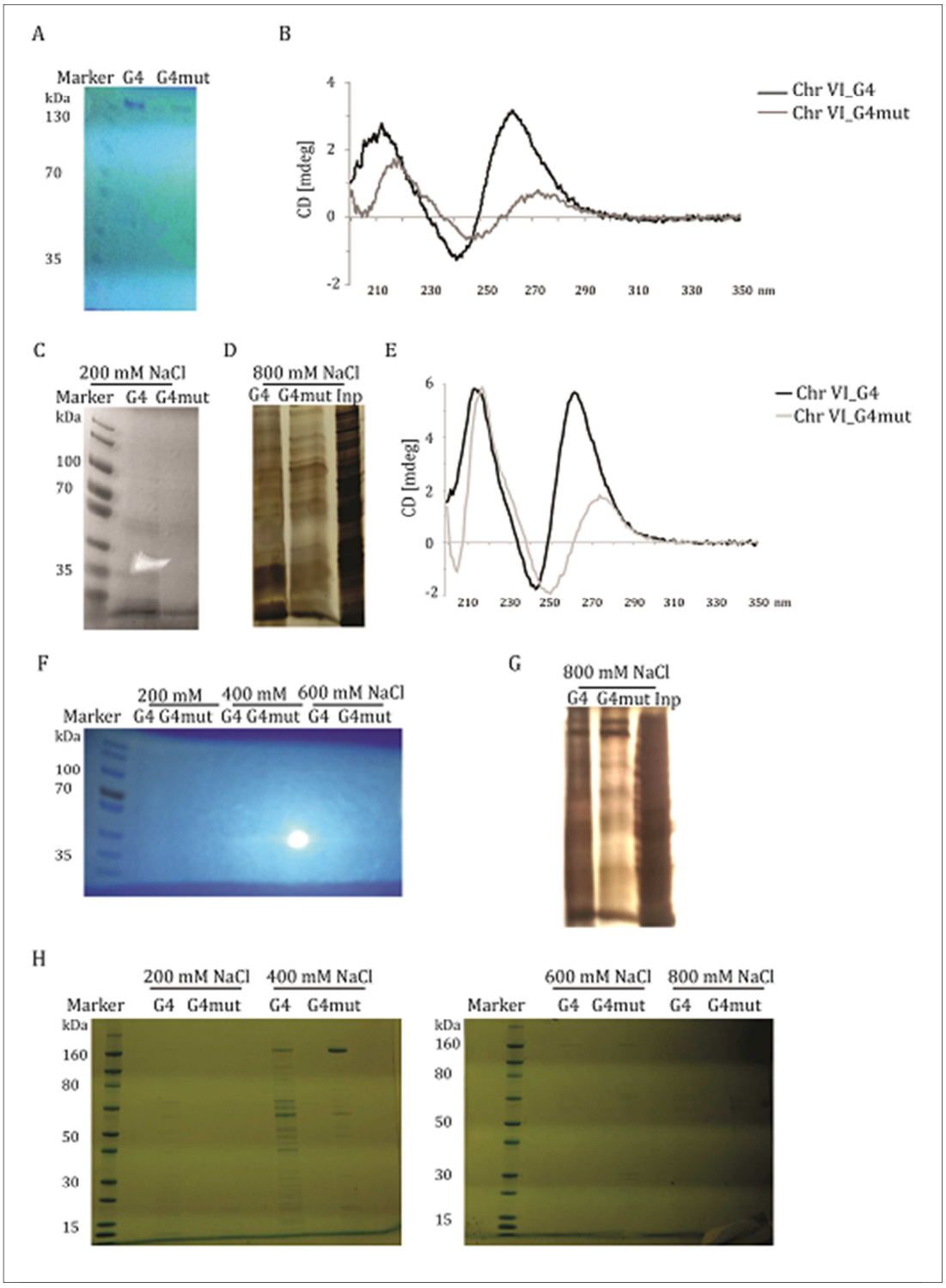


Figure 2: Pull-down experiments. Protein ladder (Marker, Thermo fisher scientific, 26616) is separated on lane 1 of (A+C+F+H). (A) Coomassie staining (Pull-down experiment 1) of separated proteins eluted from G4 or mutated G4 (G4mut) via boiling. (B) CD measurements of the DNA sequences used for pull-down experiment 2. Shown are the CD [mdeg] values. (C+D) Coomassie (C) and silver (D) staining (Pull-down experiment 2) of separated proteins eluted from G4 or G4mut via NaCl. (E) CD measurements of the DNA sequences used for pull-down experiment 3+4. Shown are the CD [mdeg] values. (F+G) Coomassie (F) and silver (G) staining (Pull-down experiment 3) of separated proteins eluted from G4 or G4mut via NaCl. (H) Coomassie staining (Pull-down experiment 4) of separated proteins eluted from G4 or G4mut via NaCl (performed by the group of Andreas Schlosser).

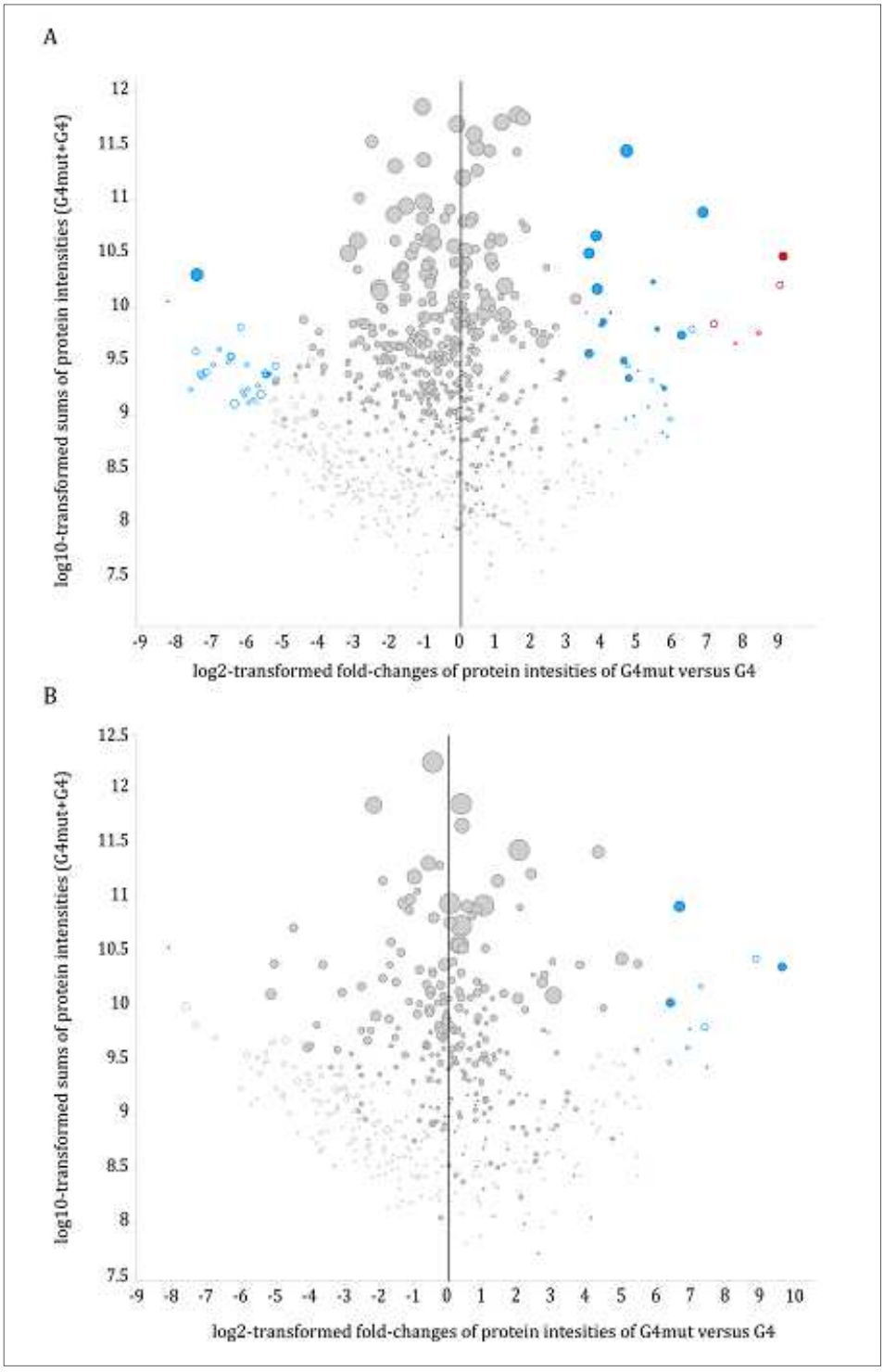


Figure 3: Quantitative analysis of the second pull-down experiment. Enrichment of proteins that bound to the mutated G4 (G4mut) or G4 sample after elution with 400 mM NaCl (A) and 600 mM NaCl (B). Shown are log2-transformed fold-changes of protein intensities G4mut versus G4 and corresponding log10-transformed sums of protein intensities. Only proteins with at least three razor/unique peptides are shown. The size of the circles correlates to the number of razor/unique peptides per protein. Closed circles: intensity values were measured in G4mut and G4 sample, open circles: intensities in one of the samples were imputed by values close to baseline intensities. Red circles: Protein ratios is outside of the 3x interquartile range (IQR) of a local distribution of at least 200 protein ratios, blue circles: protein ratios outside 1.5x IQR, grey circles: not significant (Figure construction and quantification performed by AG Schlosser).

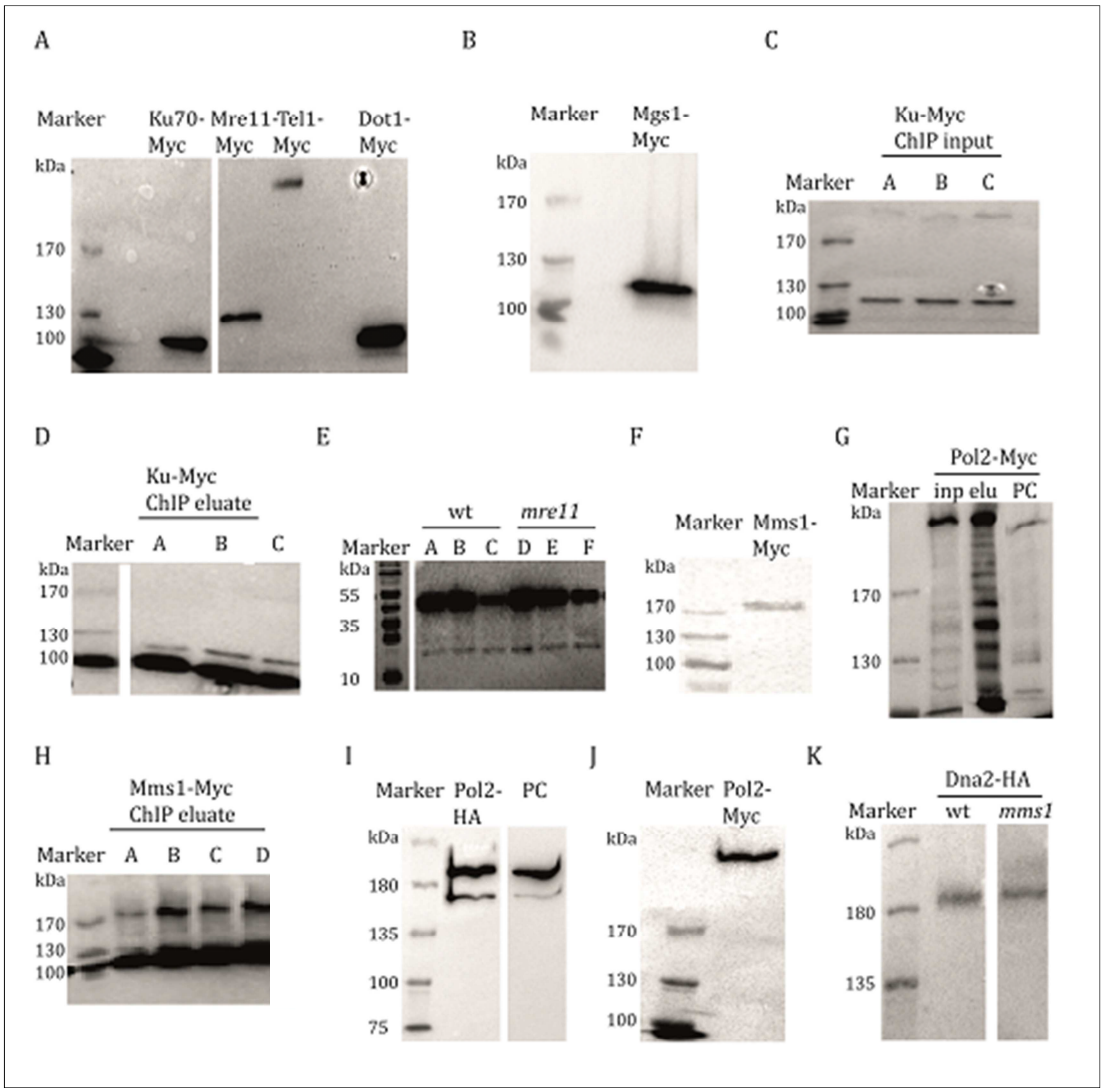


Figure 4: Western of tagged proteins. A Myc-tag counts for ~16 kDa and an HA-tag for ~3 kDa. Protein ladder (Marker) is always shown in lane 1 (Thermo fisher scientific, 26616 or biofroxx, 1123YL500). (A) Western blot analysis of Myc-tagged Ku70, Mre11, Tel1 and Dot1. (B) Western blot analysis of Mgs1-Myc. (C+D) Western input (C) and eluate (D) ChIP control samples (A-C, one for each ChIP replicate) of Myc-tagged Ku70. (E) Western eluate control samples (A-C, one for each ChIP replicate) of γ-H2A in wild type (wt) and *mre11* cells. (F) Western blot analysis of Mms1-Myc. (G) Western input (inp) and eluate (elu) ChIP-seq control samples of Myc-tagged DNA Pol2. The positive control (PC) also showed some degradation. (H) Western eluate ChIP control samples (A-C, one for each ChIP replicate) of Myc-tagged Mms1. (I) Western blot analysis of DNA Pol2-HA. The positive control (PC) showed the same additional band below 180 kDa. (J+K) Western blot analyses of DNA Pol2-Myc (J) and Dna2-HA (K) in wild type (wt) and *mms1* cells.

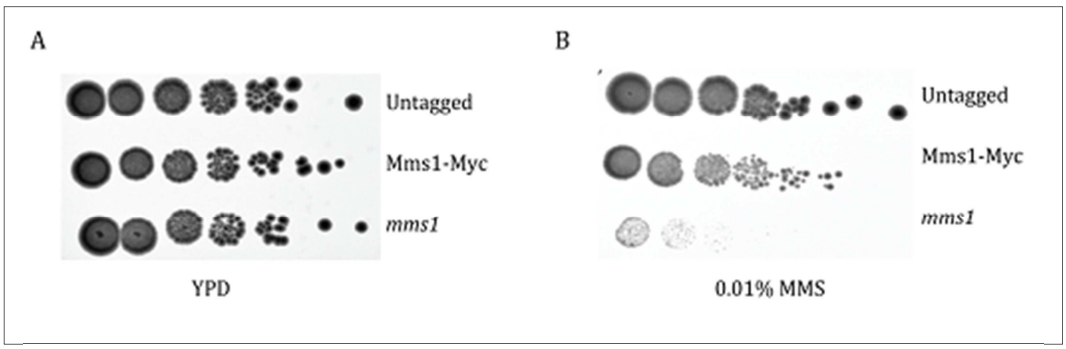


Figure 5: Myc-tagged Mms1 cells show a small growth defect compared to untagged wild type cells on MMS containing media. Serial dilutions of untagged wild type, Myc-tagged Mms1 and *mms1* cells were spotted on YPD (A) and 0.01% MMS (Sigma) (B) containing media. Cells were grown to an OD_{600} 0.8 allowing two doubling times. Then the cell suspension was serially diluted six times ten-fold. The original cell suspension and all dilution steps were spotted on the plate (5 µl per dilution). (A+B) each on the left is the original suspension and on the right the last dilution ($1:10^6$).

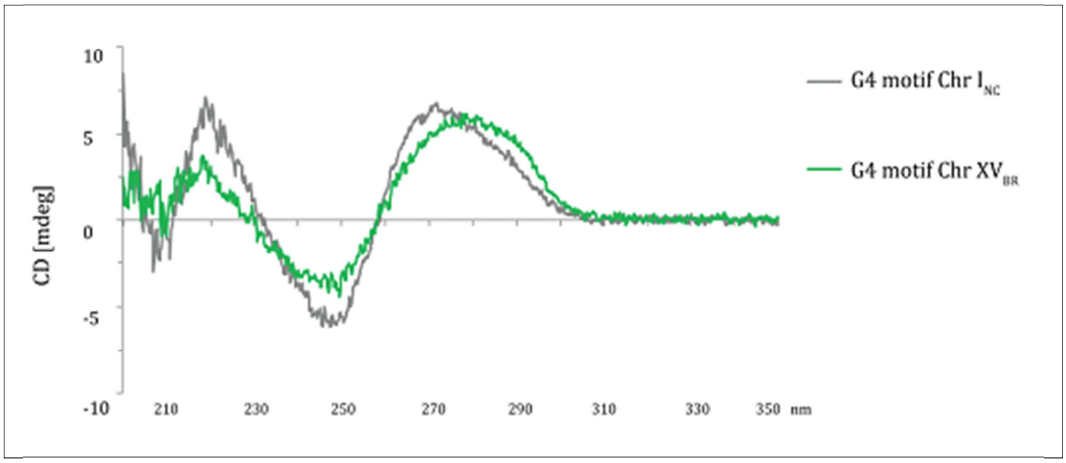


Figure 6: CD measurements. Shown are the CD [mdeg] values of the lagging strand G4 motifs of $ChrI_{NC}$ and XV_{BR}.

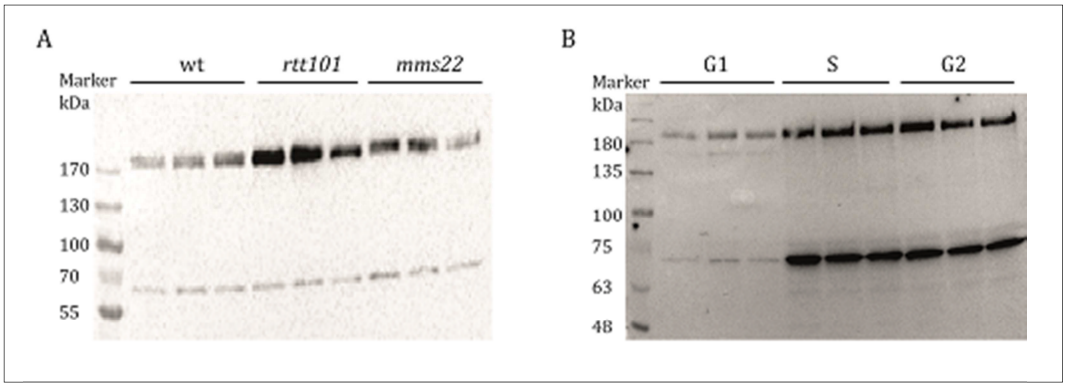


Figure 7: Western blots of Mms1 levels. (A) Mms1-Myc protein bands were detected at around 171 kDa as expected in wild type (wt), *rtt101* and *mms22* cells (each strain in triplicates). For each sample also Hsp60 was detected as a loading control. As expected according to manufactures protocol (Abcam), the Hsp60 protein band was detected at around 60 kDa. (B) Mms1-Myc protein bands were detected above 180 kDa (a little higher than expected) in G1, S and G2 arrested cells (each strain in triplicates). For each sample also Hsp60 was detected as a loading control. This protein band was also detected at a higher molecular weight than expected (~70 kDa instead of 61 kDa).

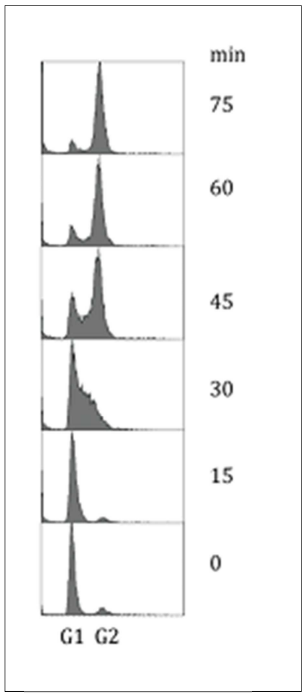


Figure 8: FACS analysis to confirm arrest of cells in G1 and progression through cell cycle. At time point 0 min cells were arrested in G1 phase and progressed synchronously through S phase (30-45 min) into G2 phase (60 min).

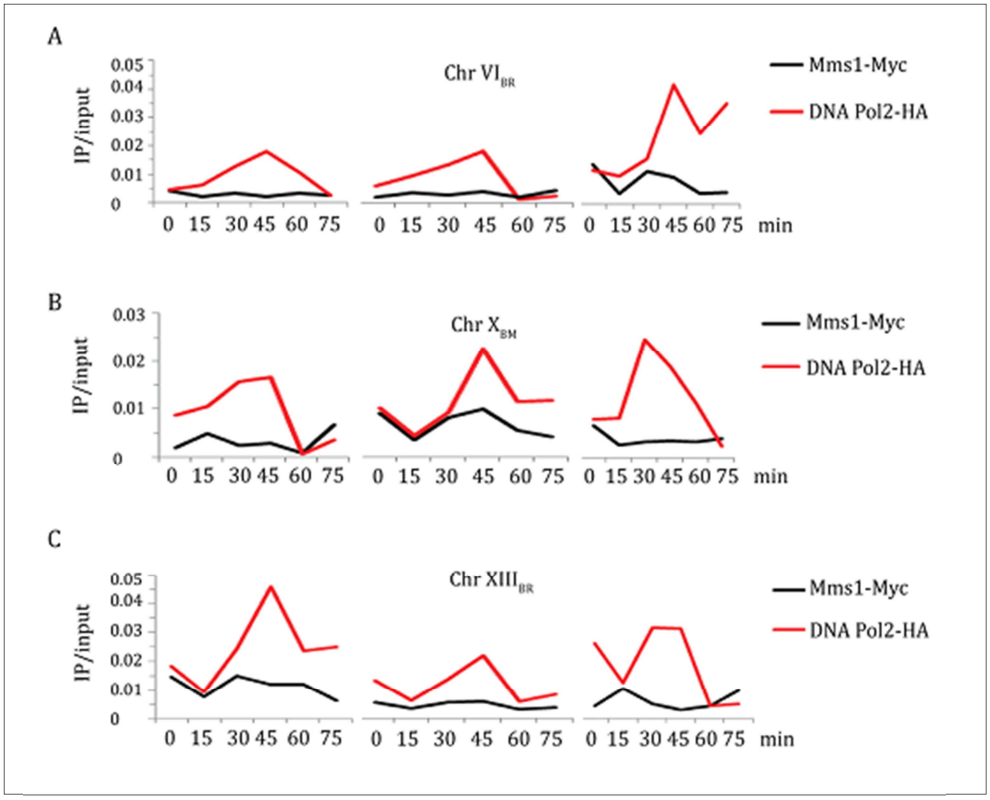


Figure 9: Results of ChIP with synchronized yeast cells. (A-C) Cells were arrested in G1 and released into S phase. Then the cells moved synchronously through the rest of the cell cycle. Samples were taken for ChIP-qPCR analysis at all indicated time points. This analysis was performed with a strain in which both Mms1 and DNA Pol2 were endogenously tagged (Mms1-Myc and DNA Pol2-HA). Using the protein specific antibody, binding of the respective protein at all indicated time points was analyzed. Binding profile of Mms1 is shown in black and the binding profile of DNA Pol2 is shown in red. Analyzed were the Mms1 binding regions Chr VI_{BR} (A), Chr X_{BM} (B) and Chr $XIII_{BR}$ (C). Plotted are IP/input values from three independent experiments. For details about the tested regions see Figure 20D and Annex Table 18.

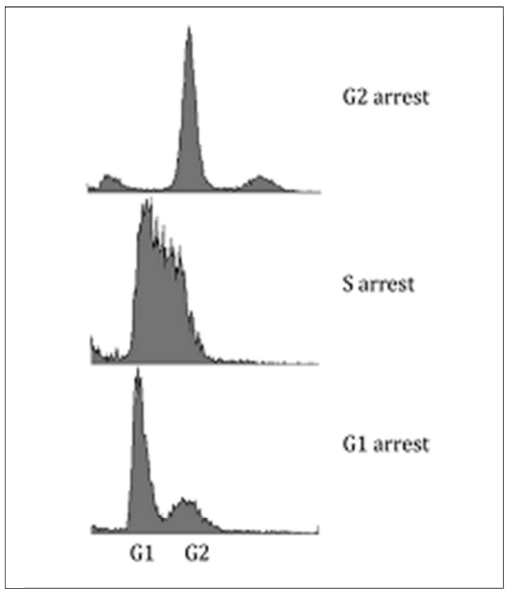


Figure 10: Confirmation of arrests for ChIP analysis. FACS analysis of cells arrested in G1, S and G2 phase was performed to confirm the arrest.

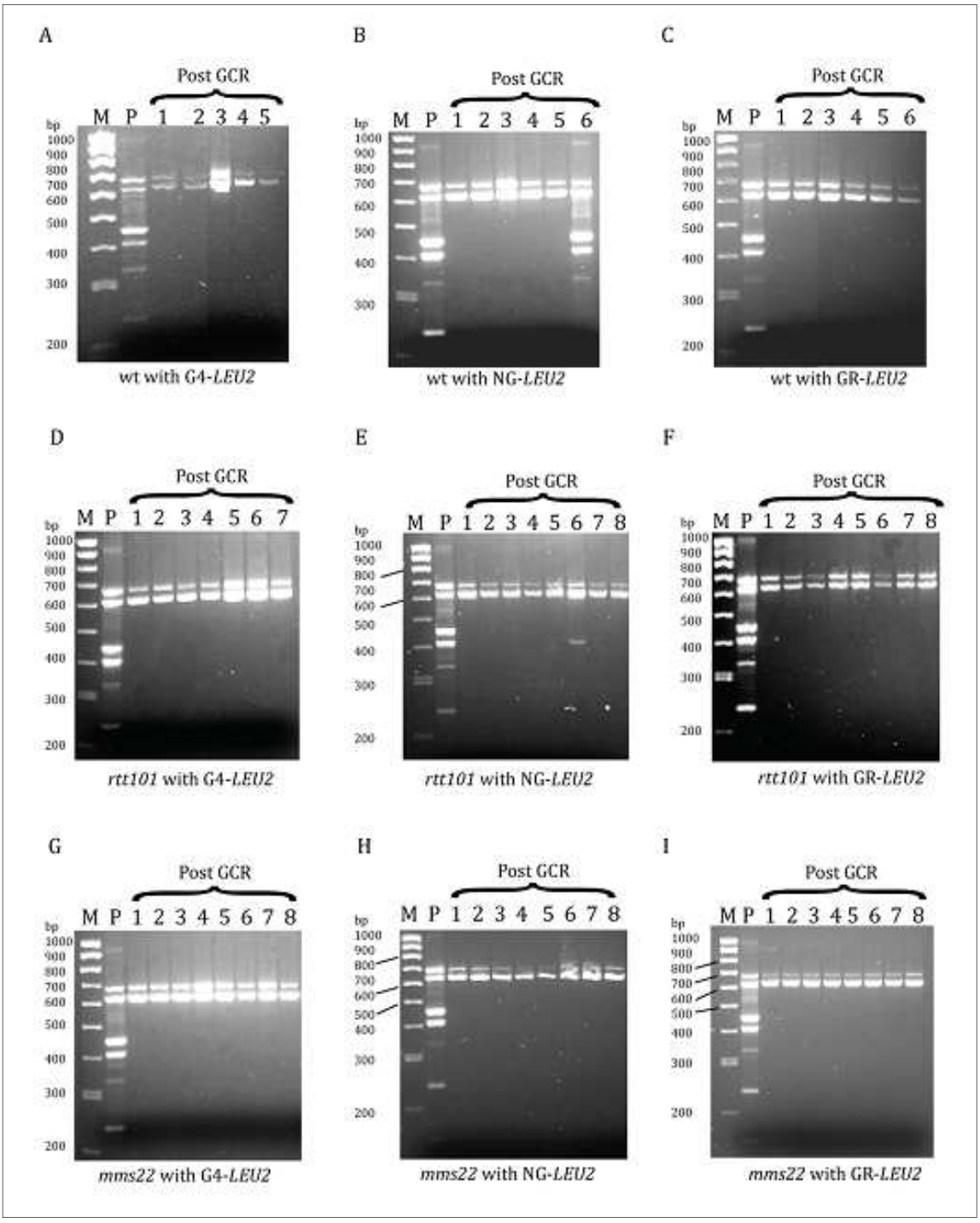


Figure 11: Multiplex PCR of strains with inserted sequences. (A-C) Multiplex PCR of wild type (wt) cells with inserted G4-*LEU2* (A), NG-*LEU2* (B) and GR-*LEU2* (C). (D-F) Multiplex PCR of *rtt101* cells with inserted G4-*LEU2* (D), NG-*LEU2* (E) and GR-*LEU2* (F). (G-I) Multiplex PCR of *mms22* cells with inserted G4-*LEU2* (G), NG-*LEU2* (H) and GR-*LEU2* (I). Multiplex PCR was performed with yeast gDNA before (P) and after the GCR assay (Post GCR). 7-8 individual colonies were tested after the GCR assay.

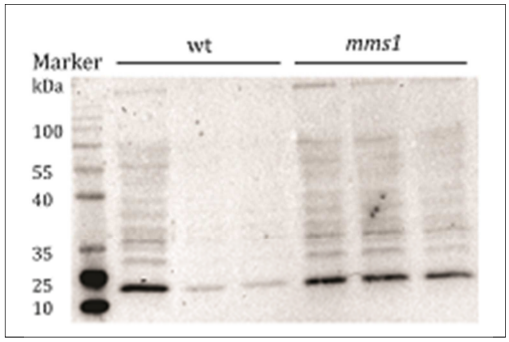


Figure 12: Overall γ-H2A levels. Western blot analysis against endogenous γ-H2A was performed in wild type (wt) and *mms1* cells (each strain in triplicates). The γ-H2A protein band was detected at around 20 kDa (14 kDa expected due to manufactures protocol (Abcam)).

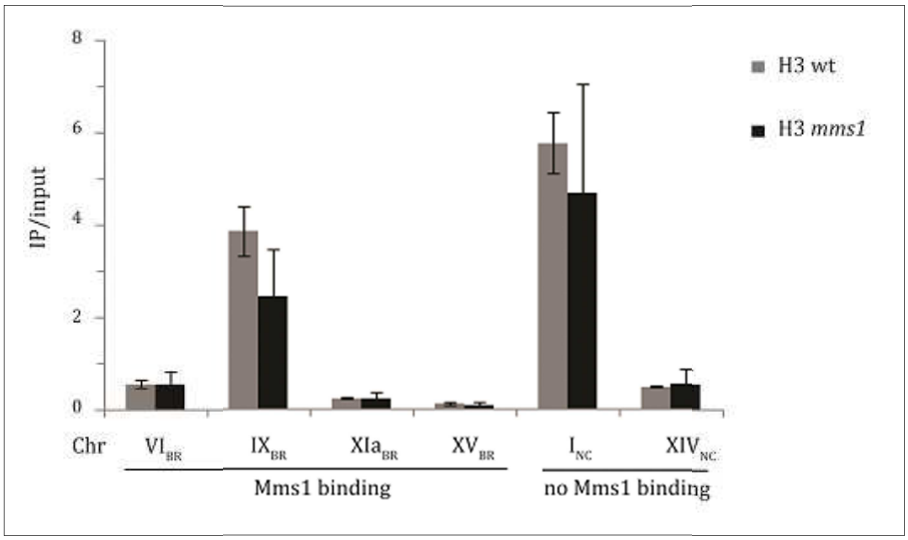


Figure 13: Histone H3 levels are independent of Mms1. ChIP was performed against endogenous histone H3 in wild type (wt) (grey) and *mms1* (black) cells. The association of histone H3 was analyzed by qPCR using primer pairs for the shown regions. Plotted are mean IP/input values ± SD from three independent experiments. For details about the tested regions see Figure 20D and Annex Table 18.

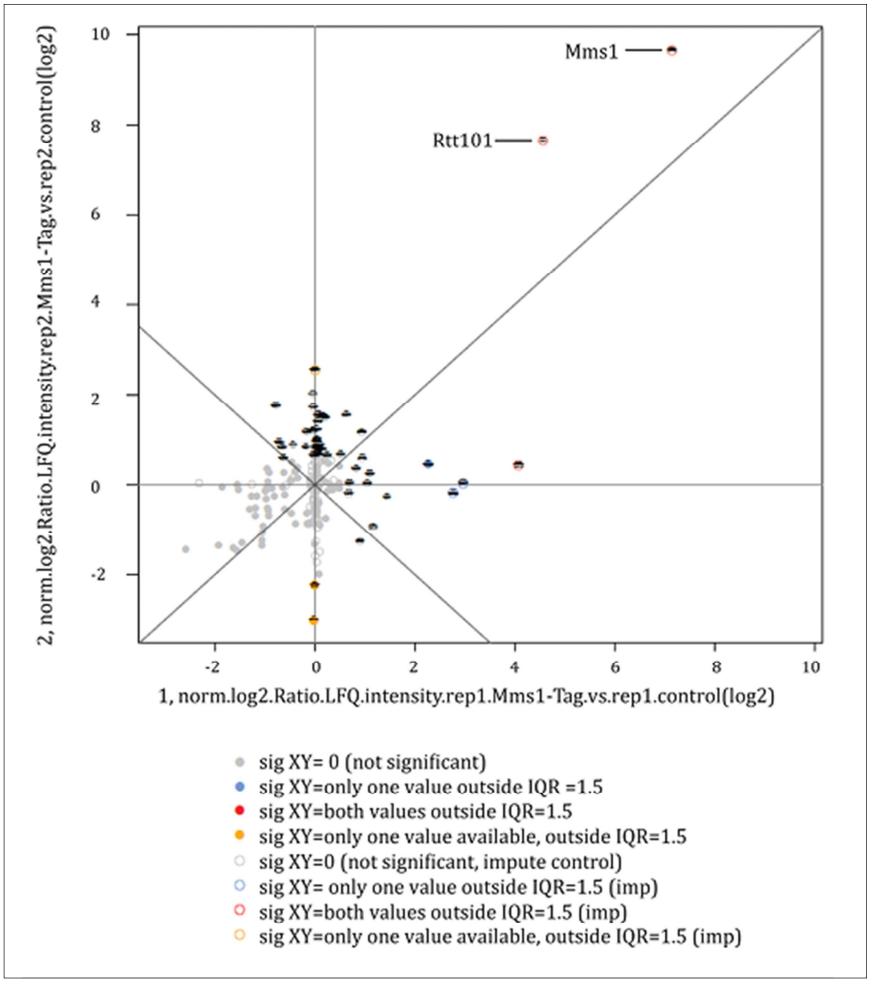


Figure 14: Results of Mms1 Co-IP. Log2-ratios of protein intensities (label-free quantitation, LFQ, MaxQuant) from two replicates are plotted against each other. Missing values in one or both control experiments have been imputed with baseline intensities to allow ratio calculations (open circles). Significance calculations are based on boxplot statistics (grey: not significant (n.s.); blue: one replicate >1.5x interquartile range (IQR), the other n.s.; orange: >1.5xIQR, only data available from one replicate; red: at least >1.5xIQR in two or >3xIQR in one replicate. This MS analysis and evaluation was performed by the group of Andreas Schlosser.

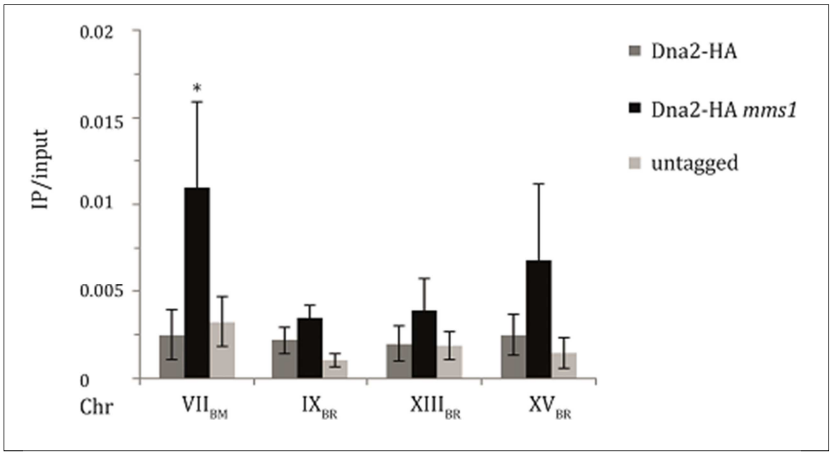


Figure 15: Dna2-ChIP. ChIP was performed with HA-tagged Dna2 in wild type (grey) and *mms1* (black) cells. The results for the untagged strain are shown as a control (light grey). The association of Dna2-HA was analyzed by qPCR using primer pairs for the shown regions. Plotted are mean IP/input values ± SD from three independent experiments. For details about the tested regions see Figure 20D and Annex Table 18. Statistical significance compared to wild type cells was determined by student's T-test. *: $p \leq 0.05$.

9.2 Tables

Table 1: Yeast strains used in this thesis.

Name	Genotype	Created by/published in
W303	*MATa ura3-1 trp1-1 leu2-3,112 his3-11,15 ade2-1 can1-100 ybp1-1 rad5-535*	R. Rothstein
Yph500	*MATα ura3-52 lys2-801 ade2-101 trp1Δ63 his3Δ200 leu2Δ1*	(Sikorski & Hieter, 1989)
KW17	W303a *mre11::TRP1*	Katharina Wanzek
KW20	W303a Pif1-Myc13*::TRP1*	Katrin Paeschke
KW38	W303a *ku70::TRP1*	Katharina Wanzek
KW47	W303a *pif1-m2*	Katharina Wanzek
KW75	Yph500α *hxt13::URA3*	(Paeschke et al, 2013)
KW76	Yph500α *hxt13::URA3 prb1*::$G4_{ChrI}$-*LEU2*	Silvia Götz
KW86	Yph500α *hxt13::URA3 mre11::TRP1*	Katharina Wanzek
KW89	W303a *tel1::TRP1*	Katharina Wanzek
KW91	W303a Ku70-Myc13*::TRP1*	Katharina Wanzek
KW92	W303a *dot1::TRP1*	Katharina Wanzek
KW95	W303a Mre11-Myc13*::TRP1*	Katharina Wanzek
KW96	W303a Tel1-Myc13*::TRP1*	Katharina Wanzek
KW98	Yph500α *hxt13::URA3 ku70::TRP1*	Katharina Wanzek
KW99	Yph500α *hxt13::URA3 prb1*::$G4_{ChrI}$-*LEU2 ku70::TRP1*	Katharina Wanzek
KW100	Yph500α *hxt13::URA3 dot1::TRP1*	Katharina Wanzek
KW102	W303a Dot1-Myc13*::TRP1*	Katharina Wanzek
KW103	Yph500α *hxt13::URA3 prb1*::$G4_{ChrI}$-*LEU2 dot1::TRP1*	Katharina Wanzek
KW104	W303a *mgs1::TRP1*	Katharina Wanzek
KW106	Yph500α *hxt13::URA3 prb1*::$G4_{ChrI}$-*LEU2 mre11::TRP1*	Katharina Wanzek
KW107	Yph500α *hxt13::URA3 mgs1::TRP1*	Julia Wille
KW108	Yph500α *hxt13::URA3 prb1*::$G4_{ChrI}$-*LEU2 mgs1::TRP1*	Julia Wille
KW109	W303a Mms1-Myc13*::TRP1*	Katharina Wanzek
KW110	Yph500α *hxt13::URA3 mms1::TRP1*	Julia Wille
KW111	Yph500α *hxt13::URA3 prb1*::$G4_{ChrI}$-*LEU2 mms1::TRP1*	Julia Wille
KW112	W303a Mgs1-Myc13*::TRP1*	Katharina Wanzek
KW114	W303a Ku70-Myc13*::TRP1 pif1-m2-HIS3*	Katharina Wanzek
KW116	W303a Mre11-Myc13*::TRP1 pif1-m2-HIS3*	Katharina Wanzek
KW117	W303a Tel1-Myc13*::TRP1 pif1-m2-HIS3*	Katharina Wanzek
KW118	W303a Dot1-Myc13*::TRP1 pif1-m2-HIS3*	Katharina Wanzek
KW136	W303a Mms1-Myc13*::TRP1 pif1-m2-HIS3*	Katharina Wanzek
KW137	W303a Mgs1-Myc13*::TRP1 pif1-m2-HIS3*	Katharina Wanzek
KW155	W303a *mms1::TRP1*	Katharina Wanzek
KW159	Yph500α *hxt13::URA3 rtt101::KANMX6*	Katharina Wanzek
KW160	Yph500α *hxt13::URA3 prb1*::$G4_{ChrI}$-*LEU2 rtt101::KANMX6*	Katharina Wanzek
KW163	Yph500α *hxt13::URA3 prb1*::$G4_{ChrI}$-*LEU2 mms22::TRP1*	Katharina Wanzek
KW164	Yph500α *hxt13::URA3 mms22::TRP1*	Katharina Wanzek
KW166	W303a Mms1-Myc13*::TRP1 rtt101::KANMX6*	Katharina Wanzek
KW168	W303a Mms1-Myc13*::TRP1 mms22::KANMX6*	Katharina Wanzek
KW178	W303a Pol2-Myc13*::TRP1*	Katharina Wanzek
KW190	W303a Pol2-Myc13*::TRP1 mms1::HIS3*	Katharina Wanzek
KW200	Yph500α *hxt13::URA3 prb1*::NG_{ChrVII}-*LEU2*	Katharina Wanzek
KW203	Yph500α *hxt13::URA3 prb1*::GR_{ChrI}-*LEU2*	Katharina Wanzek
KW208	Yph500α *hxt13::URA3 mms1::TRP1 prb1*::NG_{ChrVII}-*LEU2*	Katharina Wanzek
KW210	Yph500α *hxt13::URA3 mms22::TRP1 prb1*::NG_{ChrVII}-*LEU2*	Katharina Wanzek
KW212	Yph500α *hxt13::URA3 rtt101::TRP1 prb1*::NG_{ChrVII}-*LEU2*	Katharina Wanzek
KW216	Yph500α *hxt13::URA3 rtt101::TRP1 prb1*::GR_{ChrI}-*LEU2*	Katharina Wanzek
KW217	Yph500α *hxt13::URA3 mms22::TRP1 prb1*::GR_{ChrI}-*LEU2*	Katharina Wanzek
KW220	Yph500α *hxt13::URA3 mms1::TRP1 prb1*::GR_{ChrI}-*LEU2*	Katharina Wanzek
KW231	W303a Mms1-Myc13*::TRP1* Pol2-HA*::HIS3 bar1::KANMX6*	Katharina Wanzek
KW232	W303a Pif1-Myc13*::TRP1 mms1::HIS3*	Katharina Wanzek
KW240	W303a Mre11-Myc13*::TRP1 mms1::HIS3*	Katharina Wanzek
KW256	Yph500α *hxt13::URA3 prb1::LEU2*	Sabrina Bartsch
KW257	Yph500α *hxt13::URA3 rtt101::TRP1 prb1::LEU2*	Katharina Wanzek
KW261	Yph500α *hxt13::URA3 mms1::TRP1 prb1::LEU2*	Katharina Wanzek

Table 2: Oligodeoxynucleotides used for G4 folding and subsequent experiments as well as for proteinbinding studies.

Name	Region	Purpose	Feature	Sequence (5´-3´)
KW33	Chr XI_highγH2A_G4	Protein binding/Pull-down 1	G4	GGCTTTAGCTTGATTGGGCACGGGCACTCAATGGACGGGGTATCCACCCAGCTTGGAAAGGGG
KW40	Chr XI_highH2γA_G4mut	Protein binding/ Pull-down 1	G4mut	GGCTTTAGCTTGATTGGGCACG**T**GCACTCAATGGACG**AT**GTATCCACCCAGCTTGTAAAG**AT**G
KW48	Chr VI_G4	Pull-down 2-4	G4	AAAAAAAAGGGAGTTTCAAAGGGGCAGAATAGTGGGGTTCAGGGG
KW49	Chr VI_G4mut	Pull-down 2-4	G4mut	AAAAAAAAG**A**GAGTTTCAAAG**AT**GCAGAATAGTG**AT**GTTCAG**AT**G
KW336	Chr VII_{BM}	G4 folding and CD	lagging strand G4 motif	GGTCGTGGCGGTAGCCGTGG
KW338	Chr XV_{BR}		leading strand G4 motif	GGCTCAGTGGGTTCAGGGACCACAGG
KW341	Chr XIa_{BR}		lagging strand G4 motif	GGTGGTATCCTGACTGACGGTAAAGG
KW342	Chr $XIII_{NC}$		lagging strand G4 motif	GGTGGAAGCCCCGCCGTAGGCTGACCTATCACCGG
KW343	Chr I_{NC}		leading strand G4 motif	GGCCCTGGTGCGGACTCGTACGTCCTCTTCGG
KW399	Chr I_{NC}		lagging strand G4 motif	GGCACCGTAACATGGCCCACAATAGTTAGGATCGTTATTAACCGGCGCGG
KW400	Chr XV_{BR}		lagging strand G4 motif	GGCACAAGCTCAGGCTCAGGCACAGGCACAGGTGG

Table 3: Multiplex primers and amplified regions. All primer sequences except for KW150 published in (Paeschke et al, 2013).

Name	Orientation	Region	Product length	Sequence (5´-3´)
KW146	For	*PCM1* For	620bp	CGATGAAGGTTGATTACGAGC
KW147	Rev	*PCM1* Internal Rev		GAAGGCTCCTATAAAGAACG
KW148	Rev	IG *SOM1* Rev	with insert: 148+149=673bp without insert: 148+150=508bp	CGAGGTCACGGACACATATACC
KW149	For	MB634		GACACGAAATTACAAAATGGAATATGTTCATAGGGTAGACG
KW150	For	IG *PRB1* For		CGAGTGTAAATAGAGTATTTTCTAAC
KW151	Rev	Internal *CIN8* Rev	415bp	GAAAATCGACATAATAAGAGTAGATTTCC
KW152	For	Internal *CIN8* For		GATTTGCGATAGCGTCGCTGCC
KW153	Rev	Internal *NPR2* Rev2	153+154=341bp 153+155=452bp	CTTAGTTTAGAAATTTTGGCAATG
KW154	For	Inside Hot Spot For		GAAAAGGATACAAAGGATATGAG
KW155	Rev	Outside Hot Spot Rev		GTTTTCTTCGATTTGAAGGTGTTGG
KW156	For	IG before *CAN1* For	242bp	GAGTTTGCTAGATTCATAAAAGCC
KW157	Rev	IG before *CAN1* Rev		CCGATAATGTCTGAGTTAGGTGAG

Table 4: List of proteins identified in the first pull-down experiment (Lysis + washing according to section 4.5.1) (Parameters 20% protein threshold, 20% peptide threshold and minimum of two unique peptide counts). Displayed are the exclusive unique peptide counts of the identified proteins for the G4 and mutated G4 sample.

Protein	G4 (peptides)	mutated G4 (peptides)
ACC1	209	148
ARC1	28	20
COP1	3	1
COY1	2	0
DUR1,2	74	35
HFA1	39	3
PYC1	59	30
PYC1	14	5
PYC2	11	4
RPA43	5	0
RPL10	3	0
RPL3	3	0
TEF1	3	3

Table 5: List of proteins identified in the second pull-down experiment (Lysis + washing according to section 4.3, *pif1-m2* cells, 400 mM elution) (Parameters 20% protein threshold, 20% peptide threshold and minimum of two unique peptide counts). Displayed are the exclusive unique peptide counts of the identified proteins for the G4 and mutated G4 sample.

Protein	G4 (peptides)	mutated G4 (peptides)	Protein	G4 (peptides)	mutated G4 (peptides)	Protein	G4 (peptides)	mutated G4 (peptides)
ACC1	129	59	MIP1	0	5	RPS26A	2	2
ACS2	3	0	MIR1	14	11	RPS27A	4	2
ACT1	23	14	MIS1	89	91	RPS3	28	25
ADE3	4	0	MMS1	12	1	RPS31	1	3
ADH1	8	8	MNE1	0	10	RPS4A	34	26
ADH3	3	0	MNN1	4	1	RPS5	11	12
AL1	0	2	MNN10	3	0	RPS6A	22	17
AIM24	2	0	MNN11	6	3	RPS7A	12	10
AIM46	2	0	MNN5	2	2	RPS7B	4	2
AIR1	2	7	MNN9	3	2	RPS8A	13	10
ALE1	0	2	MOT1	6	19	RPS9B	11	13
ALG6	0	2	MPD1	13	6	RPT1	4	0
ALO1	2	0	MPH1	2	8	RRB1	7	4
AMD1	47	37	MRD1	6	10	RRP1	10	6
APL2	16	3	MRH1	3	5	RRP12	24	13
APL4	10	4	MRM1	5	0	RRP15	4	0
APL6	6	0	MRP1	2	1	RRP3	15	5
APM3	2	0	MRP20	2	0	RRP36	4	2
ARB1	34	26	MRP51	2	1	RRP42	0	2
ARC1	30	20	MRP7	7	1	RRP43	0	3
ARE2	3	0	MRPL1	2	1	RRP46	1	3
ARO1	8	17	MRPL11	4	3	RRP5	135	162
ARO2	3	0	MRPL13	2	0	RRP6	0	4
ARP4	2	2	MRPL3	10	1	RRP7	8	3
ARP5	2	0	MRPL35	2	0	RRP8	5	1
ARX1	6	0	MRPS28	2	0	RRP9	16	12
ASC1	13	9	MRPS5	2	0	RRS1	5	2
ASF1	2	0	MRS1	2	0	RSA4	19	1
ASK10	14	1	MRT4	18	14	RSM22	2	0
ATP1	22	8	MSC6	0	23	RSM24	6	0
ATP2	16	11	MSH1	2	35	RSM7	3	0
ATP3	2	0	MSH2	5	32	RSR1	2	1
ATP4	2	0	MSH3	5	24	RTC5	0	4
BAT1	2	2	MSH6	3	30	RTN1	3	1
BCP1	8	4	MSL5	5	0	RTT101	7	1
BDF1	6	1	MSS116	46	25	RVB1	15	7
BEM2	3	0	MSS18	0	4	RVB2	14	4

BFR1	4	0	MSS4	13	1	SAC1	10	40
BFR2	4	8	MTR4	38	44	SAC7	7	0
BMS1	6	1	MUD1	8	3	SAS10	3	1
BRR2	12	36	MUD2	7	1	SAW1	2	0
BRX1	11	7	MUS81	1	3	SBP1	8	15
BUD2	16	0	MVP1	18	0	SCJ1	2	0
BUD22	5	3	MYO2	34	10	SCP160	33	16
CAM1	13	0	MYO3	3	1	SDA1	4	0
CAT2	16	0	MYO5	19	5	SEA4	22	0
CBF5	37	37	NAB2	2	1	SEC1	2	3
CBP2	3	7	NAB6	1	2	SEC12	0	2
CBP3	11	0	NAM7	32	6	SEC13	2	1
CBR1	12	4	NAM8	6	1	SEC17	0	6
CCA1	2	0	NAM9	7	1	SEC18	3	2
CCE1	13	14	NAN1	28	21	SEC21	10	3
CCR4	5	0	NAP1	8	3	SEC23	6	0
CCT2	28	6	NAT1	2	0	SEC24	7	0
CCT3	33	5	NCE102	0	2	SEC26	16	0
CCT4	24	2	NCL1	7	7	SEC27	23	2
CCT5	15	0	NCP1	2	4	SEC28	3	0
CCT6	20	2	NCS2	3	0	SEC4	4	0
CCT7	20	0	NDE1	22	10	SEC53	6	3
CCT8	25	6	NDI1	4	1	SEC6	2	0
CDC12	7	0	NEW1	19	1	SEC61	3	1
CDC19	25	14	NFS1	14	2	SEC63	10	5
CDC33	7	3	NHP2	5	6	SEC66	2	0
CDC48	7	1	NIP1	30	12	SEH1	7	1
CDC53	10	4	NMD3	16	2	SEN1	43	35
CDC60	6	0	NMT1	2	0	SER33	8	0
CDC73	6	5	NOC2	32	35	SET2	6	1
CFT1	12	0	NOC3	19	8	SGS1	0	28
CFT2	9	0	NOC4	5	5	SHM1	35	37
CHD1	5	0	NOG1	14	10	SHM2	3	1
CHL1	0	16	NOG2	12	4	SKI2	25	7
CHS1	1	5	NOP1	22	21	SKI6	0	2
CIC1	22	15	NOP12	27	14	SLA1	16	5
CKA1	23	12	NOP13	24	10	SLH1	87	48
CKA2	24	10	NOP14	4	5	SLM1	20	8
CKB1	8	2	NOP15	4	4	SLM2	13	0
CKB2	11	6	NOP2	25	13	SLM5	2	1
CLU1	31	0	NOP4	32	18	SLS1	2	0
CMS1	4	4	NOP53	3	0	SMB1	2	1
COP1	31	2	NOP56	37	35	SMC1	7	0
COS2	0	3	NOP58	29	22	SMC3	11	1
CPR6	3	1	NOP6	7	7	SMC5	3	0
CSN12	6	0	NOP7	13	7	SMC6	2	0
CTK1	9	0	NOP8	2	2	SNP1	11	2
CTK3	3	0	NOP9	34	28	SNQ2	2	41
CTR9	35	16	NPL3	10	10	SNU114	34	17
CYS4	15	4	NPR2	2	0	SNU56	13	1
DBP1	4	0	NSA1	15	7	SNU71	3	0
DBP10	39	29	NSA2	4	1	SNX4	3	0
DBP2	64	49	NSR1	19	23	SNX41	2	0
DBP3	34	14	NUG1	9	3	SOF1	23	21
DBP5	6	0	OAC1	6	6	SPB4	2	0
DBP6	18	10	ODC2	4	3	SPC110	16	1
DBP7	24	15	OLA1	26	11	SPC3	2	2
DBP8	11	6	OLE1	6	1	SPO14	9	0
DBP9	36	29	OPT1	0	4	SPT15	0	3
DCW1	0	2	ORC3	2	3	SPT16	41	23
DED1	37	24	ORC4	3	1	SPT5	14	0
DED81	7	0	ORC5	4	3	SRO9	11	7
DEG1	3	0	OSM1	3	0	SRP102	2	0
DHH1	4	0	PAB1	27	23	SRP40	8	5
DHR2	14	3	PAD1	4	4	SRP68	6	1
DIM1	19	7	PAP2	3	6	SRP72	4	1
DIP2	30	19	PCT1	5	0	SSA2	27	18
DIP5	1	5	PDA1	25	20	SSB1	53	34
DIS3	2	12	PDB1	14	9	SSB2	2	1
DNF2	1	11	PDC1	15	9	SSC1	8	6
DOT1	6	0	PDR12	3	1	SSD1	2	2

DPL1	2	1	PDR16	2	1	SSE1	9	1
DPM1	10	5	PDR5	21	49	SSF1	11	9
DPS1	19	4	PDX1	19	19	SSH1	2	2
DRS1	7	0	PEP3	35	0	SSL1	2	0
DSS1	21	0	PEP5	31	0	SSZ1	17	7
DUR1,2	4	0	PET127	7	7	STB4	3	0
DXO1	15	23	PET54	0	9	STM1	10	2
EBP2	10	5	PET9	18	13	STT3	3	0
ECM16	13	25	PFK1	28	10	STT4	4	1
ECM32	26	25	PFK2	13	3	STV1	0	8
ECM33	2	5	PFK26	7	14	SUB1	5	11
EFB1	5	4	PFS2	3	0	SUB2	9	3
EFG1	8	8	PGA3	2	0	SUI2	16	13
EFR3	23	0	PGK1	17	4	SUI3	19	9
EFT1	62	35	PGM1	2	2	SUP35	19	3
ELF1	2	4	PHB1	6	6	SUP45	26	13
ELP2	34	0	PHB2	8	8	SUV3	10	0
ELP3	21	1	PHO8	4	9	SVF1	4	5
EMG1	8	6	PHO84	9	5	SWC4	5	0
ENB1	0	3	PHO86	3	4	TAE2	14	5
ENO2	6	3	PIF1	1	11	TAF1	7	1
ENP1	5	1	PIL1	12	9	TAF5	6	1
ENP2	12	4	PKC1	8	1	TAF6	10	1
EPS1	4	1	PLB1	1	5	TAF7	4	2
ERB1	24	12	PMA1	37	26	TAO3	8	0
ERG1	7	1	PMT1	0	6	TCB1	15	2
ERG11	8	4	PMT2	2	9	TCB2	15	6
ERG2	2	1	PMT4	0	5	TCB3	10	1
ERG27	2	1	PNO1	5	3	TCP1	27	2
ERG3	2	2	POB3	12	8	TDH1	1	4
ERG4	2	0	POL1	1	2	TDH2	2	0
ERG5	9	1	POL2	0	8	TDH3	19	15
ERG6	3	2	POL3	0	6	TEF1	48	27
ERG9	2	0	POL5	11	1	TEF4	20	16
ESF2	3	1	POM152	2	0	TEL1	7	0
ETT1	2	0	POR1	11	10	TEX1	7	7
EXG2	0	3	POX1	8	1	TFC1	7	7
FAA1	9	3	PPN1	3	3	TFC3	0	2
FAA3	12	4	PPZ1	8	1	TFC4	2	0
FAA4	14	11	PRP19	2	1	TFC7	3	4
FAB1	9	0	PRP24	2	7	THI7	0	3
FAP1	9	1	PRP39	25	9	THO2	35	28
FAS1	19	1	PRP40	4	0	THP2	3	7
FAS2	7	0	PRP42	15	8	THS1	14	6
FCJ1	2	0	PRP43	46	34	TIF1	6	1
FEN1	3	1	PRP8	56	19	TIF34	3	1
FKS1	43	40	PRP9	2	0	TIF35	2	0
FMP45	0	2	PRS1	4	0	TIF4631	13	4
FOL2	4	1	PRS3	8	2	TIF5	5	1
FPR3	10	9	PRT1	23	5	TIF6	4	3
FPR4	10	7	PUF6	29	17	TIM50	2	0
FRE1	4	12	PUN1	0	2	TMA19	5	4
FRE4	0	8	PUS1	6	2	TMA46	6	0
FSF1	2	3	PUS4	4	0	TOM70	14	7
FUN12	36	13	PUS7	6	25	TOM71	2	0
FUN30	3	0	PWP1	17	6	TOP1	2	24
GAR1	7	5	PWP2	34	20	TOP2	32	34
GAS1	5	15	PXR1	6	4	TOP3	20	34
GAS3	0	8	PYC1	10	10	TOR1	24	0
GAS5	0	7	PYC2	42	50	TOR2	4	0
GBP2	22	30	RAD52	4	6	TPA1	6	0
GCD1	5	1	RAI1	38	38	TPP1	1	2
GCD10	6	5	RAP1	3	2	TRA1	39	0
GCD11	32	19	RAT1	85	87	TRI1	0	3
GCD14	13	6	RBG1	9	2	TRM1	37	26
GCD2	13	2	RBG2	6	1	TRM2	6	3
GCD6	21	5	RCF2	2	1	TRM44	4	0
GCD7	5	2	RCL1	6	2	TSR1	4	0
GCN1	20	2	RET1	22	42	TY1B-DR1	3	0
GCN20	3	0	REX4	4	1	TY1B-JR2	5	2
GCN3	9	3	RFC1	28	6	TY1B-ML1	6	2

GEP3	3	0	RFC2	27	10	TY1B-ML2	73	36
GGC1	9	6	RFC3	21	11	TY1B-NL2	2	1
GIN4	2	0	RFC4	20	8	TY2B-C	9	2
GLC7	9	5	RFC5	12	8	TYS1	13	4
GOS1	0	4	RGD1	18	1	UBP10	4	0
GPD2	2	1	RHO1	6	4	UBR1	14	0
GPM1	12	6	RIO1	4	2	UBR2	4	0
GRC3	28	24	RIO2	10	1	UGP1	2	0
GRS1	11	2	RIX1	14	5	UME1	2	0
GSF2	5	4	RLI1	51	34	URA2	14	4
GSP1	10	9	RLI1	2	1	URB1	67	47
GUS1	77	44	RLP24	7	2	URB2	41	32
HAS1	37	30	RLP7	12	7	UTP10	39	21
HCA4	36	22	RMD5	2	0	UTP11	0	2
HCS1	17	38	RMI1	2	9	UTP13	19	17
HEK2	3	2	RNH70	8	1	UTP15	25	16
HFA1	34	1	RNR1	8	3	UTP18	8	6
HFD1	6	1	RNT1	7	0	UTP20	43	18
HGH1	10	2	ROD1	2	0	UTP21	14	5
HHO1	10	4	ROK1	27	21	UTP22	63	40
HIS1	12	12	RPA135	77	78	UTP23	13	7
HMG1	13	3	RPA190	119	129	UTP25	8	3
HMO1	1	2	RPA34	11	14	UTP30	10	9
HOC1	3	0	RPA43	12	16	UTP4	36	18
HOS3	20	2	RPA49	29	35	UTP5	10	6
HPR1	20	23	RPB2	14	34	UTP6	16	7
HRB1	18	19	RPB2	2	1	UTP7	11	11
HRR25	18	2	RPB3	1	2	UTP8	22	17
HSC82	12	5	RPB5	10	14	UTP9	13	11
HSL7	20	0	RPC19	7	8	UTR2	0	3
HSP104	23	4	RPC25	1	3	VAM6	32	0
HSP26	3	1	RPC34	2	8	VAN1	2	1
HSP60	3	1	RPC37	1	4	VAS1	2	2
HTS1	2	0	RPC40	20	24	VHT1	0	2
HXT1	0	4	RPC82	11	24	VIP1	65	28
HXT2	1	4	RPF1	5	2	VMA1	11	51
HXT3	2	5	RPF2	9	3	VMA13	0	4
HYP2	3	1	RPG1	41	9	VMA2	5	21
IDP1	10	17	RPL10	23	16	VMA4	0	7
IDP3	0	5	RPL11A	7	5	VMA5	0	4
IFA38	2	5	RPL12A	8	6	VMA6	0	9
IKI3	65	2	RPL13B	12	9	VPH1	1	28
ILV2	5	0	RPL14B	10	8	VPS1	11	26
ILV5	12	3	RPL15A	18	13	VPS16	37	0
IMD2	15	19	RPL15B	3	1	VPS17	2	0
IMD3	24	26	RPL16A	22	20	VPS33	18	0
IMD4	42	47	RPL16B	12	10	VPS41	27	0
IML1	15	1	RPL17B	8	6	VTC1	0	2
IMP4	12	6	RPL18A	6	7	VTC2	1	20
INP52	2	1	RPL19A	19	12	VTC3	3	44
IOC2	3	2	RPL1A	24	14	VTC4	26	49
IOC4	2	2	RPL20A	8	7	WBP1	8	2
IPI3	17	4	RPL21A	7	5	WRS1	4	0
IQG1	2	1	RPL23A	4	2	XRN1	132	132
IRC3	16	5	RPL24A	4	5	YAP1801	2	0
ISW1	5	1	RPL25	6	6	YAP1802	2	0
JEM1	17	7	RPL26B	11	8	YBR096W	1	2
JIP5	24	7	RPL27A	8	7	YCF1	0	3
KAR2	26	16	RPL28	2	4	YCG1	0	3
KEL1	2	0	RPL2A	20	18	YCK2	3	0
KEL3	6	0	RPL3	40	31	YCR016W	6	4
KES1	9	0	RPL30	4	3	YDJ1	20	10
KEX1	0	5	RPL31A	4	1	YDR341C	3	0
KGD1	25	2	RPL32	4	5	YDR514C	5	0
KGD2	13	6	RPL33A	4	4	YEF3	69	47
KIN4	5	0	RPL34B	3	3	YER077C	2	0
KIP3	2	0	RPL35A	9	6	YER184C	1	2
KRE33	26	14	RPL36A	3	4	YGR054W	21	6
KRE6	0	3	RPL43A	5	3	YGR210C	19	0
KRI1	15	5	RPL4A	2	1	YGR250C	1	2
KRR1	14	8	RPL4B	19	19	YGR266W	6	1

KRS1	21	18	RPL5	13	11	YGR283C	4	6
KTI12	2	1	RPL6A	4	3	YHB1	6	0
KTR1	5	1	RPL6B	16	14	YHC1	7	2
KTR3	7	1	RPL7B	28	22	YHM2	9	10
LAS1	24	23	RPL8A	6	5	YHR020W	29	5
LAT1	27	21	RPL8B	36	28	YHR080C	1	3
LCB1	8	3	RPL9A	10	7	YIL108W	0	10
LCB2	8	1	RPN2	16	1	YIL151C	3	0
LCD1	2	0	RPN5	2	1	YJL171C	0	5
LCP5	2	2	RPN6	5	0	YKL187C	0	3
LEM3	0	7	RPN8	3	0	YKR096W	3	1
LHP1	9	6	RPO21	22	49	YKT6	0	4
LOC1	5	2	RPO26	7	7	YLH47	6	2
LPD1	32	35	RPO31	38	60	YLR413W	0	6
LSB3	8	0	RPO41	6	3	YLR419W	30	2
LSP1	5	1	RPP0	18	15	YMC1	2	0
LUC7	2	0	RPS0B	10	5	YML6	6	1
LYS1	4	0	RPS10B	3	1	YMR265C	2	2
LYS20	13	6	RPS11A	8	8	YMR310C	9	5
MAE1	4	3	RPS12	4	4	YNL022C	20	29
MAK11	8	3	RPS13	4	5	YNL181W	2	0
MAK16	9	7	RPS14A	5	1	YNR021W	6	3
MAK21	29	23	RPS15	2	2	YOR093C	14	0
MAK5	32	12	RPS16A	13	8	YOR1	0	20
MCK1	3	0	RPS17A	3	3	YPL088W	0	7
MCM10	9	10	RPS18A	7	6	YPL260W	2	1
MDJ1	7	2	RPS1A	32	24	YPP1	5	3
MDN1	15	0	RPS1B	8	5	YRA1	14	14
MEC1	40	0	RPS2	19	17	YRA2	3	3
MES1	25	4	RPS20	6	6	YRO2	1	2
MEX67	7	1	RPS22A	6	2	YTM1	20	9
MFT1	2	6	RPS23A	4	3	ZUO1	13	4
MGM101	9	13	RPS24A	4	4			
MGS1	15	0	RPS25B	1	2			

Table 6: List of proteins identified in the second pull-down experiment (Lysis + washing according to section 4.3, *pif1-m2* cells, 600 mM elution) (Parameters 20% protein threshold, 20% peptide threshold and minimum of two unique peptide counts). Displayed are the exclusive unique peptide counts of the identified proteins for the G4 and mutated G4 sample.

Protein	G4 (peptides)	mutated G4 (peptides)	Protein	G4 (peptides)	mutated G4 (peptides)	Protein	G4 (peptides)	mutated G4 (peptides)
ACC1	39	3	MRPL10	7	0	RPN11	2	0
ACT1	20	16	MRPL11	7	0	RPN12	2	0
ADH1	8	5	MRPL22	3	0	RPN2	22	1
AIR1	6	0	MRPL35	2	0	RPN3	2	0
AMD1	28	9	MRPL4	2	0	RPN5	2	1
APL2	2	0	MRPL40	4	0	RPN6	5	0
APL4	2	0	MRPL7	5	0	RPN8	2	0
APL6	4	0	MRPS9	2	0	RPN9	3	1
ARB1	18	11	MRT4	9	0	RPO21	12	1
ARC1	15	5	MSB3	2	1	RPO26	1	4
ASC1	9	5	MSH1	2	9	RPO31	20	66
ASK10	8	0	MSH2	1	18	RPP0	13	12
ATP1	8	12	MSH3	2	21	RPP2A	4	2
ATP2	6	10	MSH6	0	7	RPP2B	5	3
BMS1	26	2	MSS116	30	4	RPS0B	4	4
BRR2	4	0	MSS4	18	0	RPS11A	7	10
BRX1	8	7	MTR4	9	11	RPS12	8	5
CAM1	8	0	MYO2	33	4	RPS13	9	7
CBF5	28	26	MYO3	5	1	RPS14A	6	4
CBP3	7	0	MYO5	12	3	RPS15	3	2
CBR1	6	3	NAM7	27	6	RPS16A	13	9
CCE1	4	2	NAN1	20	16	RPS17A	5	6
CCR4	6	0	NAP1	5	2	RPS18A	7	8
CCT2	22	2	NCS2	2	0	RPS19A	2	2

CCT3	22	3	NDE1	9	7	RPS1A	22	19
CCT4	13	1	NFS1	8	0	RPS1B	4	4
CCT5	9	0	NHP2	5	6	RPS2	18	15
CCT6	12	0	NIP1	12	0	RPS20	9	10
CCT7	14	0	NMD3	7	0	RPS22A	5	1
CCT8	15	2	NOC2	19	4	RPS23A	6	5
CDC19	6	5	NOC3	5	0	RPS24A	10	8
CDC48	13	0	NOC4	4	0	RPS25B	4	2
CDC53	3	0	NOG1	11	1	RPS26A	4	2
CFT1	2	0	NOG2	5	0	RPS3	18	20
CHL1	1	8	NOP1	18	16	RPS31	6	6
CIC1	15	7	NOP12	22	14	RPS4A	30	23
CKA1	17	6	NOP13	24	10	RPS5	10	13
CKA2	14	6	NOP14	2	1	RPS6A	24	15
CKB1	9	2	NOP15	4	0	RPS7A	5	5
CKB2	7	3	NOP2	8	0	RPS7B	14	7
CLU1	2	0	NOP4	34	21	RPS8A	14	9
COP1	16	0	NOP56	31	29	RPS9B	24	19
COX2	0	3	NOP58	19	19	RRP1	5	0
COY1	2	2	NOP6	6	2	RRP5	99	130
CTR86	2	0	NOP7	9	5	RRP9	13	5
CTR9	3	1	NOP9	9	2	RRS1	4	0
DBP10	30	7	NPL3	12	9	RSA4	6	0
DBP2	44	36	NSA1	10	3	RVB1	14	6
DBP3	16	2	NSA2	11	0	RVB2	8	1
DBP6	13	3	NSR1	23	20	SAC1	9	34
DBP9	35	5	OAC1	1	2	SAC7	4	0
DED1	28	6	OLA1	12	7	SBP1	4	5
DHR2	6	0	PAB1	15	9	SCP160	12	2
DIP2	25	18	PAD1	4	1	SEC1	1	2
DIP5	0	3	PAP2	8	1	SEC12	0	2
DIS3	0	3	PDA1	3	2	SEC17	0	4
DNF1	0	2	PDB1	4	5	SEC21	5	1
DNF2	0	5	PDC1	7	7	SEC24	4	0
DPM1	2	2	PDR17	4	0	SEC26	7	1
DPS1	2	0	PDR5	4	24	SEC27	15	0
DRS1	7	0	PDX1	1	4	SEC63	0	3
DXO1	20	8	PEP3	11	0	SEN1	55	47
EBP2	6	2	PEP5	4	0	SER3	2	0
ECM32	17	13	PET127	10	3	SGS1	0	11
ECM33	2	4	PET9	6	9	SHM1	18	15
EFB1	4	0	PFK1	13	2	SLA1	6	0
EFG1	6	1	PGK1	5	5	SLH1	28	1
EFR3	17	0	PHB1	2	8	SLM1	13	0
EFT1	37	22	PHB2	1	5	SMB1	2	0
ELP2	22	0	PIF1	1	5	SNP1	2	0
ELP3	8	0	PIL1	10	7	SNQ2	0	9
ENO2	3	6	PMA1	16	18	SNU114	39	3
ERB1	21	6	PMT1	0	5	SOF1	19	10
ERG11	2	2	PMT2	1	5	SPC110	10	1
FAA4	2	1	POL5	2	0	SPO14	13	0
FAP1	2	0	POR1	1	9	SPT16	8	0
FAS1	3	0	PPN1	6	5	SRP40	6	5
FCF2	2	0	PPZ1	2	0	SRP68	2	0
FKS1	8	4	PRP39	3	0	SRP72	2	0
FMP45	0	2	PRP42	9	0	SSA1	2	2
FOB1	5	1	PRP43	39	30	SSA2	15	10
FPR3	8	5	PRP8	34	0	SSB2	39	27
FPR4	6	2	PRT1	11	0	SSF1	7	0
FRE1	0	2	PSD2	9	0	SSZ1	2	4
FUN12	9	4	PUF6	20	11	STM1	2	0
FUN30	2	0	PWP1	11	4	STV1	0	3
FYV7	2	0	PWP2	21	20	SUB1	15	17
GAR1	8	3	PXR1	4	0	SUI2	10	1
GAS1	2	13	PYC1	0	7	SUI3	5	0
GAS3	0	4	PYC2	14	45	SUP35	4	1
GAS5	0	7	RAD16	2	5	SUP45	7	6
GBP2	30	21	RAD52	6	11	SWD2	3	0
GCD11	20	7	RAD7	1	2	TAE2	2	1
GCD14	2	0	RAI1	35	33	TCP1	14	0
GCD2	3	1	RAT1	72	65	TDH3	8	8

GCD6	7	1	RCL1	8	1	TEF1	23	19
GCD7	2	0	RET1	12	50	TEF4	11	12
GCN3	4	0	RFC1	23	2	TEX1	4	4
GIN4	3	0	RFC2	20	7	TFC1	18	12
GIS2	2	4	RFC3	14	9	TFC3	9	3
GLC7	2	1	RFC4	16	6	TFC4	13	3
GPM1	9	4	RFC5	11	2	TFC7	11	6
GRC3	17	9	RIO2	11	1	THO2	8	2
GSP1	5	6	RLI1	37	33	TIF1	2	1
GUS1	26	15	RLP24	3	1	TIF4631	9	1
HAS1	28	10	RLP7	12	6	TIF6	4	2
HCA4	27	5	RMI1	1	6	TOM70	2	0
HCS1	3	9	ROK1	12	2	TOP1	0	25
HEK2	2	0	RPA12	1	3	TOP3	10	21
HGH1	7	1	RPA135	43	61	TRM1	22	1
HHF1	5	5	RPA190	74	73	TY1B-LR2	30	11
HHO1	2	1	RPA34	4	5	TY2B-C	5	1
HHT1	2	3	RPA43	4	7	UBP10	3	0
HIS1	2	0	RPA49	21	18	UBR1	2	0
HOS3	7	0	RPB2	8	0	URA2	8	1
HPR1	6	7	RPB5	3	8	URB1	28	0
HRB1	24	15	RPB8	1	2	URB2	18	1
HRR25	14	0	RPC17	1	8	UTP10	17	6
HSL7	8	0	RPC19	4	4	UTP11	2	0
HSP82	3	2	RPC25	1	3	UTP13	18	17
HTA1	5	2	RPC34	1	13	UTP15	13	9
HTB1	6	6	RPC37	0	5	UTP18	5	2
HXT3	0	3	RPC40	13	15	UTP20	4	0
HYP2	2	0	RPC82	4	33	UTP21	14	4
IKI3	28	0	RPF1	7	0	UTP22	36	11
ILV5	0	2	RPF2	7	2	UTP23	13	0
IMD2	9	9	RPG1	6	1	UTP30	2	1
IMD3	9	15	RPL10	20	12	UTP4	21	18
IMD4	22	27	RPL11A	10	8	UTP5	7	6
IMP3	2	1	RPL12A	9	6	UTP6	17	10
IMP4	10	4	RPL13A	12	8	UTP7	8	6
IPI3	4	0	RPL14B	12	11	UTP8	13	6
IPP1	2	0	RPL15A	13	11	UTP9	10	4
IQG1	2	2	RPL16A	23	19	UTR2	0	2
IRC3	10	0	RPL16B	13	7	VAM6	6	0
JEM1	5	0	RPL17B	10	8	VIP1	6	0
JIP5	3	0	RPL18A	6	6	VMA1	3	33
KAR2	5	2	RPL19A	18	10	VMA13	1	5
KCC4	6	0	RPL1A	16	13	VMA2	2	18
KEX1	0	3	RPL20A	17	12	VMA4	0	3
KGD2	4	0	RPL21A	13	8	VMA5	0	4
KIP3	2	0	RPL23A	4	2	VMA6	0	10
KRE33	5	3	RPL24A	2	1	VPH1	0	22
KRI1	11	1	RPL24B	9	6	VPS16	6	0
KRR1	11	5	RPL25	12	9	VTC1	1	4
KRS1	5	3	RPL26A	2	0	VTC2	0	14
LAS1	16	15	RPL26B	16	10	VTC3	2	39
LAT1	12	12	RPL27A	11	9	VTC4	24	44
LEM3	0	3	RPL28	8	6	WBP1	3	5
LHP1	7	3	RPL2A	14	13	XRN1	134	147
LOC1	2	0	RPL3	25	19	YDJ1	18	7
LPD1	24	20	RPL30	7	3	YDL156W	3	0
LRO1	5	0	RPL31A	11	2	YDR514C	7	0
LSP1	3	0	RPL32	11	7	YEF3	30	28
LYS20	1	3	RPL33A	5	2	YGR210C	5	0
MAK11	13	2	RPL34B	3	4	YGR250C	1	3
MAK16	6	3	RPL35A	13	7	YGR283C	3	1
MAK21	12	2	RPL36B	5	5	YHC1	3	0
MAK5	7	0	RPL37A	3	2	YHM2	4	6
MCM10	8	10	RPL37B	2	1	YHR020W	4	1
MDJ1	3	1	RPL38	2	1	YIL108W	0	4
MGM101	18	14	RPL39	2	0	YJL171C	0	4
MGS1	17	0	RPL42A	7	6	YML6	3	0
MIR1	5	7	RPL43A	8	3	YMR310C	5	2
MIS1	52	39	RPL4A	2	1	YNL022C	3	2
MNN1	2	0	RPL4B	19	13	YOR1	0	14

MNN9	1	3	RPL5	9	10	YPP1	3	0
MNP1	5	1	RPL6A	22	18	YRA1	12	7
MPD1	3	0	RPL6B	8	7	YRA2	4	1
MRH1	0	4	RPL7B	24	16	YRF1-4	0	2
MRM1	3	0	RPL8A	5	5	YRO2	0	2
MRP1	4	0	RPL8B	26	25	YTM1	14	2
MRP7	6	0	RPL9A	8	6	ZUO1	4	1

Table 7: Go term analysis was performed with proteins uniquely identified in the G4 sample after the second pull-down experiment and MS analysis (see Annex Table 5+6). Shown are some go terms, cluster frequency, p-values and the proteins that fall into the specific category.

GO_term	nucleic acid metabolic process	gene expression	DNA damage response	mRNA processing	ribosome biogenesis	DNA replication
Cluster frequency	97 out of 97 genes, 100.0%	80 out of 97 genes, 82.5%	23 out of 97 genes, 23.7%	18 out of 97 genes, 18.6%	22 out of 97 genes, 22.7%	11 out of 97 genes, 11.3%
P-value	1.29e-62	1.57e-27	3.48e-09	4.18e-09	1.21e-05	0.00129
Proteins	ASF1	ASF1	BDF1	CFT1	CAM1	CCR4
	BDF1	BDF1	DOT1	CFT2	DRS1	CHD1
	BFR1	CAM1	FUN30	CTK1	ENP1	FOB1
	CAM1	CCA1	LCD1	LUC7	ESF2	MEC1
	CCA1	CCR4	MCK1	MSL5	FCF2	MGS1
	CCR4	CDC60	MEC1	MUD2	FYV7	MMS1
	CDC48	CFT1	MGS1	NAB2	GEP3	ORC4
	CDC60	CFT2	MMS1	NAM8	IMP3	RAD27
	CFT1	CHD1	RAD27	PFS2	MDN1	RTT101
	CFT2	CTK1	RPN4	PRP19	MRM1	SET2
	CHD1	CTK3	RTT101	PRP40	NOP53	SUV3
	CTK1	DED81	SAW1	PRP9	NSA2	
	CTK3	DEG1	SMC1	SMB1	REX4	
	DED81	DHH1	SMC3	SNU56	RNH70	
	DEG1	DOT1	SMC5	SNU71	RNT1	
	DHH1	DRS1	SMC6	SPT5	RRP15	
	DOT1	ELP2	SPT5	SWD2	RRP8	
	DRS1	ELP3	SSL1	YJR084W	SPB4	
	DSS1	ENP1	SWC4		SPT5	
	ELP2	ESF2	TEL1		SSB2	
	ELP3	FCF2	TOR1		TOR1	
	ENP1	FOB1	TPA1		TSR1	
	ESF2	FUN30	TRA1			
	FCF2	FYV7				
	FOB1	GEP3				
	FUN30	HTS1				
	FYV7	IMP3				
	GEP3	ISW1				
	HTS1	KTI12				
	IMP3	LUC7				
	ISW1	MDN1				
	KTI12	MRM1				
	LCD1	MRS1				
	LUC7	MSL5				
	MCK1	MUD2				
	MDN1	NAB2				
	MEC1	NAM8				
	MGS1	NCS2				
	MMS1	NOP53				
	MRM1	NSA2				
	MRS1	ORC4				
	MSL5	PFS2				
	MUD2	POL5				
	NAB2	PRP19				
	NAM8	PRP40				
	NCS2	PRP9				
	NOP53	PUS4				
	NSA2	REX4				
	ORC4	RNH70				
	PFS2	RNT1				
	POL5	RPB2				

	PRP19	RPN4				
	PRP40	RPT1				
	PRP9	RRP15				
	PUS4	RRP8				
	RAD27	SEC13				
	REX4	SET2				
	RNH70	SLM5				
	RNT1	SMB1				
	RPB2	SNU56				
	RPN4	SNU71				
	RPT1	SPB4				
	RRP15	SPT5				
	RRP8	SSB2				
	RTT101	SSL1				
	SAW1	SUV3				
	SEC13	SWD2				
	SET2	TAF1				
	SLM5	TAF5				
	SMB1	TAF6				
	SMC1	TOR1				
	SMC3	TPA1				
	SMC5	TRA1				
	SMC6	TRM44				
	SNU56	TSR1				
	SNU71	UBP10				
	SPB4	UME1				
	SPT5	WRS1				
	SSB2	YDR341C				
	SSL1	YJR084W				
	SUV3					
	SWC4					
	SWD2					
	TAF1					
	TAF5					
	TAF6					
	TEL1					
	TOR1					
	TPA1					
	TRA1					
	TRM44					
	TSR1					
	UBP10					
	UME1					
	WRS1					
	YDR341C					
	YJR084W					

Table 8: List of proteins identified in the third pull-down experiment (Lysis + washing according to section 4.3, *pif1-m2* cells, 400 mM elution, only G4 sample measured) (Parameters 20% protein threshold, 20% peptide threshold and minimum of two unique peptide counts). Displayed are the exclusive unique peptide counts of the identified proteins for the G4 sample.

Protein	G4 (peptides)	Protein	G4 (peptides)	Protein	G4 (peptides)	Protein	G4 (peptides)
ACT1	5	KRI1	2	RLP24	2	RPS3	14
ADH1	3	KRR1	8	RLP7	5	RPS3A2	13
ARX1	4	LAS1	4	RPA135	17	RPS4A	12
ACS1	8	IPD1	5	RPA190	34	RPS5	10
BCP1	4	MAK11	2	RPA43	4	RPS6A	7
BMS1	2	MAK16	5	RPA49	5	RPS7A	2
BRX1	7	MAK21	9	RPAB3	2	RPS8A	6
BUD22	4	MCM10	3	RPB5	3	RPS9B	13
CBF5	23	MGM101	11	RPC19	2	RRP1	8
CDC19	2	MIR1	2	RPC40	8	RRP12	2
CDC33	4	MIS1	15	RPF1	3	RRP5	54
CIC1	9	MRD1	5	RPF2	2	RRP9	7
CKA1	14	MRT4	4	RPG1	5	RRS1	2
CKA2	12	MUD1	2	RPL10	7	RSA4	2
CKB1	6	NAF1	2	RPL11A	8	RVB1	4

CKB2	12	NAM7	2	RPL12A	4	SBP1	2
CNR1	3	NAN1	15	RPL13B	3	SCP160	4
CTR9	3	NHP2	7	RPL14B	4	SGD1	2
DBP10	10	NIP1	12	RPL15A	7	SHM1	10
DBP2	24	NIP7	3	RPL16B	3	SLH1	3
DBP3	5	NOC2	16	RPL18A	6	SMB1	3
DBP6	4	NOC3	3	RPL19A	4	SNU13	2
DBP9	4	NOG1	7	RPL1A	8	SPT16	5
DIP2	9	NOP1	13	RPL20A	5	SSB1	16
DRS1	3	NOP12	7	RPL21A	3	SSF1	6
DXO1	2	NOP13	4	RPL23A	3	SUB1	4
ECM32	2	NOP15	4	RPL24A	3	SUI2	11
EF2	19	NOP2	13	RPL25	4	SUI3	6
EFG1	3	NOP4	4	RPL27A	8	TAE2	3
EMG1	5	NOP56	18	RPL28	5	TCP1	2
ERB1	8	NOP58	19	RPL2A	6	TDH2	3
ESF1	2	NOP6	5	RPL3	7	TEF1	16
FCF1	2	NOP7	3	RPL30	3	TEF4	7
FKS1	2	NOP9	8	RPL31A	2	TRM1	3
FPR3	5	NPL3	6	RPL32	2	TY1B-H	4
FPR4	2	NSA1	7	RPL5	3	URB1	20
GAR1	3	NSR1	20	RPL6A	6	URB2	11
GBP2	8	OLA1	4	RPL6B	16	UTP10	21
GCD11	18	PET9	7	RPL7A	13	UTP13	8
GCD14	3	PGK1	3	RPL8A	2	UTP15	10
GIS2	5	PMA1	16	RPL8B	16	UTP18	2
GRC3	6	POB3	2	RPL9A	4	UTP21	7
GUS1	12	POR1	4	RPO6	4	UTP22	11
H2B1	4	PPZ1	2	RPPO	11	UTP25	3
H3	4	PRP39	3	RPS0A	3	UTP4	14
H4	4	PRP42	7	RPS11A	9	UTP5	7
HAS1	18	PRP43	18	RPS12	5	UTP6	8
HIS1	8	PRT1	9	RPS13	2	UTP7	2
HRB1	3	PUF6	17	RPS14A	3	UTP8	9
HSP71	3	PWP1	2	RPS15	2	VTC4	8
HTZ1	2	PWP2	14	RPS18A	6	XRN1	48
IMD3	10	RAI1	22	RPS1A	3	YDJ1	3
IMD4	5	RAT1	41	RPS2	8	YEF3	18
IMP4	5	RCL1	4	RPS20	8	YRA2	2
IPI3	2	RL26A	2	RPS26A	2	YTM1	8
KRE33	2	RLI1	13	RPS27A	2		

Table 9: List of proteins identified in the fourth pull-down experiment (Lysis + washing according to section 4.3, wild type cells 400 mM elution) (Parameters 20% protein threshold, 20% peptide threshold and minimum of two unique peptide counts). Displayed are the exclusive unique peptide counts of the identified proteins for the G4 and mutated G4 sample.

Proteins	G4 (peptides)	mutated G4 (peptides)	Proteins	G4 (peptides)	mutated G4 (peptides)	Proteins	G4 (peptides)	mutated G4 (peptides)
ACC1	5	1	MSH1	4	28	RPP2B	3	0
ACT1	16	6	MSH2	7	7	RPS0A	3	0
ADH1	6	0	MSH3	6	0	RPS10A	3	0
APL6	4	0	MSH6	2	3	RPS11A	15	3
APM3	4	0	MTR4	8	0	RPS12	5	3
ARB1	22	16	MUD1	6	0	RPS13	6	2
ARC1	21	11	MYO1	5	0	RPS14B	10	7
ARO2	2	0	MYO2	11	6	RPS15	4	2
ASC1	8	1	MYO5	3	0	RPS16A	13	12
ASK10	3	0	NAB2	2	4	RPS17A	5	2
ATP1	5	3	NAF1	4	0	RPS18A	13	8
ATP2	5	4	NAM7	3	0	RPS19A	8	4
BCP1	7	0	NAM8	5	0	RPS1A	16	3
BDF1	11	0	NAN1	23	1	RPS1B	5	2
BFR1	0	3	NAP1	2	0	RPS2	15	6
BFR2	11	0	NCE102	1	3	RPS20	17	10
BMS1	17	0	NET1	4	0	RPS22A	4	0

BRR2	2	0	NFS1	4	0	RPS23A	5	1
BRX1	12	1	NHP2	7	2	RPS24A	12	4
BUD21	2	0	NHP6A	0	2	RPS25B	7	1
BUD22	10	0	NHP6B	0	2	RPS26A	5	4
CBF5	32	25	NIP1	18	0	RPS27A	5	1
CBR1	2	0	NIP7	4	0	RPS28A	3	0
CCE1	8	13	NOB1	2	0	RPS3	19	2
CCT2	18	0	NOC2	38	9	RPS30A	8	8
CCT3	21	0	NOC3	6	0	RPS31	7	3
CCT4	8	0	NOC4	9	0	RPS4A	18	10
CCT5	11	0	NOG1	13	0	RPS5	12	9
CCT6	5	0	NOG2	3	0	RPS6A	14	8
CCT7	15	0	NOP1	22	4	RPS7A	4	1
CCT8	14	0	NOP10	10	10	RPS7B	11	0
CDC19	6	0	NOP12	27	0	RPS8A	8	0
CDC48	5	0	NOP13	24	0	RPS9A	2	1
CHL1	4	14	NOP14	19	0	RPS9B	16	15
CIC1	31	6	NOP15	3	0	RRM3	1	13
CKA1	23	10	NOP16	7	0	RRP1	9	0
CKA2	21	1	NOP2	19	0	RRP14	0	2
CKB1	12	1	NOP4	31	13	RRP15	10	0
CKB2	19	2	NOP56	38	9	RRP3	8	0
CMS1	4	0	NOP58	26	8	RRP36	4	0
CSL4	3	0	NOP6	8	0	RRP42	4	0
CTA1	0	9	NOP7	14	1	RRP43	4	0
CTR9	4	0	NOP8	7	0	RRP45	3	0
CYS3	2	0	NOP9	12	0	RRP46	3	0
DBP10	10	2	NPL3	16	8	RRP5	128	179
DBP2	52	39	NSA1	8	0	RRP7	5	0
DBP3	7	0	NSR1	36	30	RRP8	0	3
DBP6	23	1	OLA1	7	0	RRP9	19	0
DBP8	12	0	PAB1	8	5	RRS1	3	0
DBP9	31	8	PAD1	2	0	RRT14	3	0
DED1	13	12	PDA1	8	0	RSA3	5	0
DED81	3	0	PDB1	5	0	RVB1	6	0
DEG1	0	6	PDC1	9	5	SAC1	0	16
DHR2	15	0	PDR5	0	18	SAS10	8	3
DIM1	4	1	PET9	4	0	SBP1	10	3
DIP2	34	2	PGK1	13	1	SCJ1	2	0
DIS3	18	0	PIF1	13	24	SCP160	4	1
DNA2	4	0	PIL1	8	7	SEA4	2	0
DRE2	2	0	PMA1	16	11	SEC1	5	0
DXO1	10	6	PNO1	9	0	SEC27	4	0
EBP2	17	1	POB3	9	0	SEC61	2	0
ECM16	14	16	POL5	2	0	SEH1	3	0
ECM32	18	26	POR1	3	3	SEN1	14	27
ECM33	0	4	POX1	6	2	SGD1	7	0
EFB1	2	0	PPN1	6	1	SGS1	0	16
EFG1	2	0	PPZ1	2	0	SHM1	3	31
EFR3	2	0	PRP24	0	8	SKI6	2	0
EFT1	21	3	PRP39	5	0	SLA1	5	4
EGD2	4	1	PRP42	8	0	SLH1	19	13
ELC1	2	0	PRP43	39	10	SLM1	2	2
ELF1	5	0	PRP8	9	0	SMB1	5	0
ELP3	5	0	PRT1	10	0	SMD1	3	0
EMG1	7	0	PUF6	17	0	SMD2	2	0
ENO2	8	1	PWP1	14	0	SNP1	6	0
ENP1	4	0	PWP2	37	0	SNU114	15	0
ENP2	9	0	PXR1	10	0	SNU13	10	1
ERB1	32	2	PYC2	7	19	SNU23	2	0
ERV46	2	0	RAD52	0	4	SNU56	2	0
ESF1	17	0	RAD7	6	0	SNU71	5	0
ESF2	4	0	RAI1	25	36	SOF1	21	1
FAF1	6	0	RAT1	74	67	SPB4	2	0
FBA1	2	1	RCL1	11	0	SPC110	20	1
FCF1	7	0	RET1	7	0	SPS19	0	4
FCF2	6	0	REX4	2	2	SPT15	1	3
FKS1	0	8	RFC2	7	0	SPT16	37	0
FOX2	0	4	RFC3	3	0	SRP40	16	3
FPR3	9	3	RFC4	3	0	SSA1	2	0
FPR4	14	0	RGD1	2	0	SSA2	14	4

FRE1	0	2	RHO1	3	0	SSB1	34	22
FUN12	15	7	RIM1	6	4	SSF1	10	0
FYV7	2	1	RIO1	5	0	SSZ1	2	0
GAR1	6	2	RIO2	6	6	STE20	0	2
GAS1	0	11	RLI1	52	40	STM1	2	1
GAS5	0	5	RLP24	4	0	SUB1	6	10
GBP2	24	10	RLP7	10	0	SUI2	6	0
GCD10	8	0	RMI1	2	6	SUI3	4	0
GCD11	11	0	ROK1	24	0	SUP45	2	0
GCD14	13	3	RPA12	6	2	SVF1	3	0
GCD2	2	0	RPA135	42	26	TAE2	3	0
GCD6	5	0	RPA190	53	35	TAF1	0	5
GCN2	2	0	RPA34	8	4	TCP1	16	0
GIS2	5	6	RPA43	9	3	TDH2	6	2
GLN1	2	0	RPA49	24	14	TEF1	26	13
GPM1	2	0	RPB10	4	2	TEF4	14	3
GRC3	6	0	RPB2	2	0	TFC1	4	3
GRX4	4	0	RPB5	5	3	TFC7	3	0
GSP1	3	0	RPB8	3	2	TIF1	2	0
GUS1	22	15	RPC10	2	1	TIF34	4	0
HAS1	23	2	RPC17	2	0	TIF35	4	0
HCA4	20	2	RPC19	5	2	TIF6	4	0
HCS1	2	22	RPC40	12	8	TMA19	2	0
HEK2	10	6	RPF1	5	0	TOP2	2	0
HHO1	0	14	RPF2	5	0	TOP3	23	30
HHT1	2	0	RPG1	15	0	TPM1	4	0
HIF1	2	0	RPL10	20	9	TRM1	9	5
HIS1	11	0	RPL11A	17	10	TY1B-LR2	16	11
HRB1	26	12	RPL12A	7	3	URB1	30	0
HSP82	3	0	RPL13B	15	1	URB2	30	0
HTA1	5	2	RPL14B	16	8	UTP10	25	0
HTB2	11	11	RPL15A	13	0	UTP11	6	0
IDP1	0	6	RPL16A	11	2	UTP13	23	3
IDP3	0	26	RPL16B	3	1	UTP14	10	0
IKI3	10	0	RPL17B	10	4	UTP15	16	2
IMD2	6	8	RPL18A	4	0	UTP18	12	2
IMD3	22	22	RPL19A	6	5	UTP20	15	0
IMD4	12	12	RPL1A	15	3	UTP21	17	1
IMP3	10	0	RPL20A	12	1	UTP22	41	11
IMP4	21	0	RPL21A	8	2	UTP23	6	0
JIP5	11	0	RPL23A	6	2	UTP25	7	0
KRE33	25	0	RPL24B	9	5	UTP30	6	0
KRI1	7	0	RPL25	14	2	UTP4	31	1
KRR1	18	0	RPL26A	3	4	UTP5	19	0
KTR1	2	1	RPL26B	20	28	UTP6	25	0
KTR3	3	0	RPL27A	16	4	UTP7	26	4
LAS1	6	0	RPL28	5	0	UTP8	18	0
LAT1	11	0	RPL29	2	4	UTP9	19	1
LCP5	5	0	RPL2A	17	0	VMA1	0	9
LEM3	0	2	RPL3	20	0	VPH1	0	10
LHP1	5	0	RPL30	8	0	VPS33	2	0
LOC1	4	4	RPL31A	9	2	VTC1	1	3
LPD1	10	1	RPL32	8	0	VTC2	0	4
LSP1	2	1	RPL33A	10	0	VTC3	0	23
LYS20	4	0	RPL34B	11	1	VTC4	19	36
MAK11	18	0	RPL35A	9	13	XRN1	105	126
MAK16	11	2	RPL36B	13	0	YBL028C	2	4
MAK21	27	1	RPL37A	3	3	YCR016W	14	0
MAK5	3	0	RPL38	4	0	YCR087C-A	2	0
MCM10	5	9	RPL42A	3	2	YDJ1	18	6
MGM101	12	14	RPL43A	6	0	YDL156W	2	0
MGS1	6	0	RPL4A	11	4	YEF3	25	3
MIS1	30	31	RPL5	11	0	YGR283C	13	0
MLC1	4	0	RPL6A	8	3	YHC1	5	0
MLC2	2	0	RPL6B	19	8	YHR020W	3	0
MNN11	3	0	RPL7A	20	7	YIL096C	2	1
MNP1	2	0	RPL7B	3	2	YJL171C	0	2
MOS1	2	0	RPL8A	5	4	YMR310C	10	0
MOT1	2	3	RPL8B	29	14	YNK1	3	0
MPH1	1	2	RPL9B	7	1	YNL022C	4	11
MPP10	27	0	RPN6	2	0	YNL050C	3	0

MRD1	8	1	RPO26	3	1	YRA1	13	18
MRH1	1	2	RPO31	3	0	YRA2	5	2
MRPL20	3	0	RPP0	14	2	YTM1	14	1
MRT4	10	0	RPP1B	2	0			
MSC6	0	7	RPP2A	2	1			

Table 10: Proteins identified in at least two pull-down experiments. Parameters 20% protein threshold, 20% peptide threshold and minimum of two unique peptide counts were used for evaluation. Displayed are the proteins identified uniquely in the G4 sample in at least two of the pull-down experiments number 2-4 (Thanks to Stefan Juranek for this evaluation).

Protein	Pull-down (#)	Protein	Pull-down (#)	Protein	Pull-down (#)
ACT1	2,3,4	PUF6	2,3,4	YEF3	2,3,4
ADH1	2,3,4	PWP1	2,3,4	YRA2	2,3,4
ASC1	2,3,4	PWP2	2,3,4	YTM1	2,3,4
BCP1	2,3,4	RAT1	2,3,4	ACC1	2,4
BMS1	2,3,4	RCL1	2,3,4	APL6	2,4
BRX1	2,3,4	RLI1	2,3,4	ARB1	2,4
BUD22	2,3,4	RLP24	2,3,4	ARC1	2,4
CBF5	2,3,4	RLP7	2,3,4	ASK10	2,4
CDC19	2,3,4	RPA49	2,3,4	CBR1	2,4
CIC1	2,3,4	RPF1	2,3,4	CCT2	2,4
CKA1	2,3,4	RPF2	2,3,4	CCT3	2,4
CKA2	2,3,4	RPG1	2,3,4	CCT4	2,4
CKB1	2,3,4	RPL10	2,3,4	CCT5	2,4
CKB2	2,3,4	RPL12A	2,3,4	CCT6	2,4
CTR9	2,3,4	RPL13B	2,3,4	CCT7	2,4
DBP10	2,3,4	RPL14B	2,3,4	CCT8	2,4
DBP2	2,3,4	RPL15A	2,3,4	CDC48	2,4
DBP3	2,3,4	RPL16B	2,3,4	DHR2	2,4
DBP6	2,3,4	RPL1A	2,3,4	EBP2	2,4
DBP9	2,3,4	RPL20A	2,3,4	EFR3	2,4
DIP2	2,3,4	RPL21A	2,3,4	EFT1	2,4
DXO1	2,3,4	RPL23A	2,3,4	ELP3	2,4
EFG1	2,3,4	RPL25	2,3,4	FUN12	2,4
EMG1	2,3,4	RPL27A	2,3,4	GCD2	2,4
ERB1	2,3,4	RPL28	2,3,4	GCD6	2,4
FPR3	2,3,4	RPL2A	2,3,4	GPM1	2,4
FPR4	2,3,4	RPL3	2,3,4	HCA4	2,4
GAR1	2,3,4	RPL30	2,3,4	IKI3	2,4
GBP2	2,3,4	RPL31A	2,3,4	JIP5	2,4
GCD11	2,3,4	RPL32	2,3,4	LHP1	2,4
GCD14	2,3,4	RPL5	2,3,4	MAK5	2,4
GRC3	2,3,4	RPL6A	2,3,4	MGS1	2,4
GUS1	2,3,4	RPL6B	2,3,4	MYO2	2,4
HAS1	2,3,4	RPL8B	2,3,4	MYO5	2,4
HIS1	2,3,4	RPS12	2,3,4	NAP1	2,4
HRB1	2,3,4	RPS13	2,3,4	NFS1	2,4
IMP4	2,3,4	RPS1A	2,3,4	NOG2	2,4
KRE33	2,3,4	RPS2	2,3,4	PAB1	2,4
KRI1	2,3,4	RPS27A	2,3,4	POL5	2,4
KRR1	2,3,4	RPS3	2,3,4	PPZ1	2,4
LPD1	2,3,4	RPS6A	2,3,4	PRP8	2,4
MAK11	2,3,4	RPS7A	2,3,4	PXR1	2,4
MAK16	2,3,4	RPS8A	2,3,4	RFC2	2,4
MAK21	2,3,4	RRP1	2,3,4	RFC3	2,4
MRT4	2,3,4	RRP9	2,3,4	RFC4	2,4
MUD1	2,3,4	RRS1	2,3,4	ROK1	2,4
NAM7	2,3,4	RVB1	2,3,4	RPL11A	2,4
NAN1	2,3,4	SCP160	2,3,4	RPL16A	2,4
NIP1	2,3,4	SLH1	2,3,4	RPL17B	2,4
NOC2	2,3,4	SMB1	2,3,4	RPL43A	2,4
NOC3	2,3,4	SPT16	2,3,4	RPN6	2,4
NOG1	2,3,4	SSB1	2,3,4	RPS22A	2,4
NOP1	2,3,4	SSF1	2,3,4	RPS4A	2,4
NOP12	2,3,4	SUI2	2,3,4	RPS7B	2,4
NOP13	2,3,4	SUI3	2,3,4	SEC27	2,4

NOP15	2,3,4	TAE2	2,3,4	SNP1	2,4
NOP2	2,3,4	TCP1	2,3,4	SNU114	2,4
NOP4	2,3,4	TDH2	2,3,4	SOF1	2,4
NOP56	2,3,4	TEF1	2,3,4	SPC110	2,4
NOP58	2,3,4	TEF4	2,3,4	SSA2	2,4
NOP6	2,3,4	URB1	2,3,4	TRM1	2,4
NOP7	2,3,4	URB2	2,3,4	UTP20	2,4
NOP9	2,3,4	UTP10	2,3,4	UTP23	2,4
NPL3	2,3,4	UTP13	2,3,4	UTP9	2,4
NSA1	2,3,4	UTP15	2,3,4	YHC1	2,4
NSR1	2,3,4	UTP18	2,3,4	YHR020W	2,4
OLA1	2,3,4	UTP21	2,3,4	YMR310C	2,4
PET9	2,3,4	UTP22	2,3,4	DRS1	2,3
PGK1	2,3,4	UTP25	2,3,4	FKS1	2,3
PMA1	2,3,4	UTP4	2,3,4	IPI3	2,3
POB3	2,3,4	UTP5	2,3,4	RPL19A	2,3
PRP39	2,3,4	UTP6	2,3,4	RPL9A	2,3
PRP42	2,3,4	UTP7	2,3,4	RPS14A	2,3
PRP43	2,3,4	UTP8	2,3,4	RSA4	2,3
PRT1	2,3,4	YDJ1	2,3,4		

Table 11: Primers used for qPCR.

Name	Orientation	Region (Chr)	Feature	Sequence (5´-3´)
KW57	For	$XII_{G4tract3}$	G4 rDNA	TGTCCGTACCAGTTCTAAG
KW58	Rev			TCTCCAAGGTGAACAGCCTC
KW59	For	$XI_{G4tract3}$	G4, high y-H2A	AATCCCGTCGCTATGCTC
KW60	Rev			CTCCCGGTCTGTTATTTTC
KW61	For	$VI_{G4tract3}/VI_{BR}$	G4, high y-H2A/Mms1 binding region (BR)	TGCATAGTTCTTAGGTCTTC
KW62	Rev			GTATAGCAGTGACGCGTG
KW71	For	$XIII_{PC\_\gamma H2A}$	high y-H2A, no G4	CAACGAAGCTATGCTGATC
KW72	Rev			ATGACGATTCCAGGTAGAC
KW77	For	$XIII_{NC\_\gamma H2A}$	no y-H2A, no G4	GAGGACGAAACGATTGATG
KW78	Rev			AGATAATGAGCCACGGTAC
KW81	For	$XV_{PC\_\gamma H2A}$	high y-H2A, no G4	TCAACAGCTCGCGGCTC
KW82	Rev			TGGCACAAGTGTTTGCTAC
KW99	For	$XVI_{G4tract3}$	G4, no y-H2A	ACCAGCCACAGCACTTAG
KW100	Rev			GGTGCCACTGCGGAAAG
KW160	For	$IX_{G4tract3}/IX_{BR}$	G4, no y-H2A/Mms1 BR	AGAGTCTTTGGCACTGTTG
KW161	Rev			ATTATCCCTTAATGGCCTAC
KW167	For	VII_{Tel_VII-L}	Tel-VII-L (McGee et al, 2010)	TGATATGTGTTACGCAGAATAC
KW168	Rev			TGAGAAGCACCGCAATG
KW169	For	VI_{Tel_VI-R}	Tel-VI-R (McGee et al, 2010)	ATCATTGAGGATCTATAATC
KW170	Rev			CTTCACTCCATTGCG
KW195	For	$XIII_{G4tract3}/XIII_{NC}$	G4, high y-H2A/Mms1 NC	GCTTCAGCCTGGGGTAAC
KW196	Rev			GGCACCATTAGATTCACCAC
KW278	For	XIb_{BR}	no Pol2 binding	ACTAGGTCTCTTAGCTCTC
KW279	Rev			TTTTGAACACGTTCTACGAG
KW322	For	I_{NC}	NC ChIPseq	TCGTATACATGCGGAGTAG
KW323	Rev			GTTACCACAGAATTGAACTG
KW328	For	X_{BR}	ChIPseq peakcall	CACAAACACATAAACACATAC
KW329	Rev			CGGATTTCGCATAGTTGTC
KW377	For	XIV_{NC}	NC ChIPseq	AGTGATTGTGCCGTTATAAC
KW378	Rev			CGGTTCGCACTACGATAC
KW382	For	VII_{BM}	Mms1 MEME motif	AGTCTAATCTAACTGGTCTG
KW383	Rev			GCCAAGAAGGCTCTAGAC
KW391	For	XI_{BM}	Mms1 MEME motif	CAGTATGAAATTATCCGCTC
KW392	Rev			CACTATGGTGGACAGCTG
KP230	For	XI_{NG}/XIa_{BR}	non-G-rich region/Mms1 BR (Paeschke et al, 2011)	GGCAACGATAGAACCAATTC
KP230	Rev			GCAACCATTATACCATCTCC
KP233	For	$XIII_{GR}/XIII_{BR}$	G-rich region/Mms1 BR (Paeschke et al, 2011)	CCAAACCAGACCAACCATTG
KP233	Rev			TGCTGACCACAACGAACC
rDNA	Rev	XII_{rDNA}	rDNA (RFB) (Paeschke et al, 2011)	AAGATGGGTTGAAAGAGAAGGG
	For			TCATATCAAAGGCATGTCCTGT
ARO	For	IV_{ARO}	*ARO1* (McGee et al, 2010)	TCGTTACAAGGTGATG
	Rev			AATAGCGGCAACAAC
tRNA	For	VI_{tRNA}	tRNA-Ala (Azvolinsky et al, 2009)	GGAAAGATTGTACGGGAAATGG
	Rev			GCTAATGAAACTACTAATGTCTTGC

Table 12: Inserts used in the GCR assay. All inserts contain a G4tract2 on the lagging strand (CC-dinucleotides are marked grey). The G4-*LEU2* contains an additional G4tract3 on the lagging strand. The actual NG-*LEU2* and GR-*LEU2* inserts are located on the leading strand.

Insert	Orientation of G4tract2 motifs	Sequence (5´-3´)	Sequencing result (5´-3´) (Primer KP354fw: CGCGCAATTAACCCTCACTA)
G4 motif (Chr I) (G4-*LEU2*)	lagging strand	CCCAACAATTATCTCAAA ATTCCCCCAATTCTCATC AGTAACACCCCACCCC	>KW111_primer_KP354fw TATCCCGCGGTGGCGGCCGCTCTAGATGTAGAATTGCAGATTCCCTT TTATGGATTCCTAAATCCTCGAGAAGAACTTCTAGTATATCTACGTA CCTAATATTATTGCCTTATTAAAAATGGAATCCCAACAATTATCTCA AAATTCCCCCAATTCTCATCAGTAACACCCCACCCCGTATTACTTTT ACCGTGATGAAGATTGGCATCGTTACTTTCTAAACGTAGGACGTGCG GAATGACAAAACCATCAGCAGTGTCACGATCTCTCCAGTCACGGATC CCCCGGGCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACC TCGAGGGGGGGCCCGGTACCCAATTCGCCCTATAGTGAGTCGTATTA CGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGGGGGGA
Non-G-rich (Chr VII) (NG-*LEU2*)	lagging strand	CTAATCTTTCAGCGTTGT AAATGTTGGTACCCAAAC CCAATTGTCTACAAGTTT CCTTAGC	>KW200_primer_KP354fw TGGATTACTTTAACGTTAACTCCGCTTGGCGGCCGCTCTGGAGAAAC TGTAATGTCACCACCGCCTCCTAGACCTAATCTTTCACGGTTTTAAA CGTTGGTACCCAAACCCAATTGCCTACAAGTCTCCTTACCAATACCA ACAGCATCACCGTGATCCCCCGGGCTGCAGGAATTTTTATCAAGCTT ATCGATACCGTCGACCTCCAGGGGGGGCCCGGTACCCAATTCGCCCT ATAGTGAGTCGTATTACGCGCGCTCTCTGGCCGTCGGTTTACAACGT CTTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGC ACATCCCCCTTTCGCCAGCTGGCGTATTACCGAAGAGGCCTGCACCG ATCGCCCTTCCCAACAGTTGCGCAGCCTGT
G-rich (Chr I) (GR-*LEU2*)	lagging strand	ATGGTGGTCATCTCAGTA GATGTAGAGGTGAAAGTA CCGGTCCATGGCTCGGT	>KW203_primer_KP354fw TCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCCACC GCGGTGGCGGCCGCTCTAGATGACCGTTGGTTCCAGTGATGGTGGTC ATCTCAGTAGATGTAGAGGTAAAAGTACCAGTCCATGGTTCAGTGGT GGTTGTAACCAAACCTTCACTAGTTGGGGTTCTGATAACAATCACAG TTTCGTCAGTTGGTTGACCGTTGGTACCGGTGACGGTGGTCATCTAA AGTACCAAAAGTACCAGTCCATGGTTCAGTGGTGGTTGTAACCAAAC CTTCACTGGTTGGGGATCCCCCGGGCTGCGGGAATTCGATATCAAGC TTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAATTCGCC CTATAGTGAGTCGTATACGCGCGCGGACAGGGAAAT

Table 13: Binding regions of Mms1 identified by MACS peakcall.

Peaks from MACS 2.0									
Region	Start	End	Length	Abs_summit	Pileup	-LOG10(pvalue)	Fold_enrichment	-LOG10(pvalue)	Name
Chr I	112581	112785	205	112690	33.00	625.095	247.853	327.508	peakcall_mms1_default_extsize180_peak_1
Chr I	130082	130284	203	130242	40.00	527.558	208.622	249.880	peakcall_mms1_default_extsize180_peak_2
Chr I	191966	192177	212	192031	40.00	809.461	266.000	481.361	peakcall_mms1_default_extsize180_peak_3
Chr II	193643	193822	180	193769	31.00	645.572	260.302	344.464	peakcall_mms1_default_extsize180_peak_4
Chr II	215789	215983	195	215843	34.00	685.749	258.984	377.494	peakcall_mms1_default_extsize180_peak_5
Chr II	463750	464101	352	464057	28.00	749.299	303.780	429.304	peakcall_mms1_default_extsize180_peak_6
Chr III	123535	123722	188	123612	47.00	997.933	280.550	642.703	peakcall_mms1_default_extsize180_peak_7
Chr IV	427404	427648	245	427568	41.00	1.393.092	395.679	984.699	peakcall_mms1_default_extsize180_peak_8

Chr IV	461798	462449	652	462293	77.00	1.555.914	285.887	1.121.153	peakcall_mms1_default_extsize180_peak_9
Chr IV	1049063	1049304	242	1049164	27.00	488.485	232.904	219.689	peakcall_mms1_default_extsize180_peak_10
Chr IV	1239631	1239824	194	1239696	44.00	927.112	276.729	582.947	peakcall_mms1_default_extsize180_peak_11
Chr IV	1251077	1251389	313	1251202	45.00	1.238.366	336.579	853.083	peakcall_mms1_default_extsize180_peak_12
Chr V	40470	41119	650	40735	38.00	1.001.155	317.243	645.062	peakcall_mms1_default_extsize180_peak_13
Chr V	42007	42474	468	42233	42.00	1.040.526	307.755	680.290	peakcall_mms1_default_extsize180_peak_14
Chr V	43223	43945	723	43695	72.00	1.981.364	356.086	1.493.835	peakcall_mms1_default_extsize180_peak_15
Chr V	312025	312220	196	312084	44.00	625.972	219.505	328.202	peakcall_mms1_default_extsize180_peak_16
Chr V	335545	335734	190	335617	40.00	405.728	184.715	160.186	peakcall_mms1_default_extsize180_peak_17
Chr V	396418	396611	194	396505	34.00	816.523	291.130	487.064	peakcall_mms1_default_extsize180_peak_18
Chr V	442044	442465	422	442270	75.00	1.264.826	254.802	876.649	peakcall_mms1_default_extsize180_peak_19
Chr V	574898	575199	302	575071	60.00	1.073.506	258.619	708.690	peakcall_mms1_default_extsize180_peak_20
Chr VI	224947	225231	285	225020	44.00	816.422	255.176	487.064	peakcall_mms1_default_extsize180_peak_21
Chr VII	567185	567405	221	567287	38.00	718.636	253.025	403.215	peakcall_mms1_default_extsize180_peak_22
Chr VII	700603	700830	228	700704	50.00	1.771.911	424.219	1.300.422	peakcall_mms1_default_extsize180_peak_23
Chr VII	794326	794519	194	794426	38.00	1.242.680	377.673	856.834	peakcall_mms1_default_extsize180_peak_24
Chr VII	806450	806663	214	806581	36.00	1.082.030	348.574	716.266	peakcall_mms1_default_extsize180_peak_25
Chr VII	1049535	1049759	225	1049724	41.00	787.711	258.281	461.973	peakcall_mms1_default_extsize180_peak_26
Chr VII	1088959	1089156	198	1089058	56.00	883.253	238.573	545.299	peakcall_mms1_default_extsize180_peak_27
Chr VIII	189408	189652	245	189581	38.00	502.145	207.392	229.984	peakcall_mms1_default_extsize180_peak_28
Chr VIII	215841	216046	206	215971	47.00	799.402	244.240	471.951	peakcall_mms1_default_extsize180_peak_29
Chr IX	54151	54474	324	54323	38.00	657.242	239.832	353.795	peakcall_mms1_default_extsize180_peak_30
Chr IX	334128	334307	180	334294	31.00	484.923	219.695	217.488	peakcall_mms1_default_extsize180_peak_31
Chr IX	392158	392481	324	392309	26.00	537.260	250.760	257.480	peakcall_mms1_default_extsize180_peak_32
Chr X	121001	121236	236	121088	43.00	613.254	219.159	318.162	peakcall_mms1_default_extsize180_peak_

									33
Chr X	391941	392163	223	392038	45.00	663.878	224.383	359.497	peakcall_mms1_default_extsize180_peak_34
Chr X	400014	400260	247	400116	35.00	798.738	282.325	471.355	peakcall_mms1_default_extsize180_peak_35
Chr XI	99804	99985	182	99895	47.00	1.158.282	311.415	782.455	peakcall_mms1_default_extsize180_peak_36
Chr XI	380959	381322	364	381207	31.00	711.099	277.528	397.450	peakcall_mms1_default_extsize180_peak_37
Chr XI	519606	519853	248	519720	50.00	1.436.254	353.430	1.020.517	peakcall_mms1_default_extsize180_peak_38
Chr XII	84189	84385	197	84328	35.00	787.982	279.721	461.973	peakcall_mms1_default_extsize180_peak_39
Chr XII	369707	369916	210	369834	55.00	1.022.171	262.313	664.327	peakcall_mms1_default_extsize180_peak_40
Chr XII	451471	451833	363	451826	350.00	425.587	123.512	173.564	peakcall_mms1_default_extsize180_peak_41
Chr XII	452032	452232	201	452047	354.00	365.961	120.950	132.597	peakcall_mms1_default_extsize180_peak_42
Chr XII	455226	455521	296	455258	357.00	393.367	121.972	151.503	peakcall_mms1_default_extsize180_peak_43
Chr XII	455722	456041	320	455824	358.00	528.897	127.085	251.040	peakcall_mms1_default_extsize180_peak_44
Chr XII	458653	459006	354	458788	352.00	427.775	123.523	175.395	peakcall_mms1_default_extsize180_peak_45
Chr XII	459111	459290	180	459285	355.00	374.987	121.291	138.398	peakcall_mms1_default_extsize180_peak_46
Chr XII	464854	465115	262	464975	355.00	515.768	126.738	240.564	peakcall_mms1_default_extsize180_peak_47
Chr XII	467535	468132	598	467560	354.00	373.050	121.244	137.317	peakcall_mms1_default_extsize180_peak_48
Chr XII	704512	704746	235	704580	40.00	682.404	239.637	374.708	peakcall_mms1_default_extsize180_peak_49
Chr XII	904036	904280	245	904110	32.00	472.551	214.098	208.308	peakcall_mms1_default_extsize180_peak_50
Chr XII	921241	921442	202	921330	39.00	959.489	302.822	609.679	peakcall_mms1_default_extsize180_peak_51
Chr XII	1074797	1075077	281	1074850	52.00	591.819	200.487	300.473	peakcall_mms1_default_extsize180_peak_52
Chr XIII	1732	1917	186	1804	47.00	630.240	214.612	331.825	peakcall_mms1_default_extsize180_peak_53
Chr XIII	16259	16578	320	16428	55.00	1.609.468	363.317	1.165.234	peakcall_mms1_default_extsize180_peak_54
Chr XIV	494008	494211	204	494082	30.00	498.671	225.984	227.759	peakcall_mms1_default_extsize180_peak_55
Chr XIV	547087	547297	211	547204	42.00	767.811	251.326	444.554	peakcall_mms1_default_extsize180_peak_56
Chr XIV	619049	619366	318	619157	37.00	561.935	222.102	276.947	peakcall_mms1_default_extsize180_peak_57

Chr XIV	631801	631983	183	631871	42.00	600.525	218.798	307.301	peakcall_mms1_default_extsize180_peak_58
Chr XIV	652076	652255	180	652157	34.00	669.706	255.143	363.903	peakcall_mms1_default_extsize180_peak_59
Chr XV	31004	31286	283	31151	41.00	400.039	182.258	155.643	peakcall_mms1_default_extsize180_peak_60
Chr XV	159995	160358	364	160118	63.00	943.633	234.574	596.621	peakcall_mms1_default_extsize180_peak_61
Chr XV	480109	480435	327	480246	39.00	597.875	224.877	305.507	peakcall_mms1_default_extsize180_peak_62
Chr XV	588163	588351	189	588225	31.00	600.297	248.641	307.168	peakcall_mms1_default_extsize180_peak_63
Chr XV	710140	710326	187	710244	45.00	1.570.881	411.661	1.133.412	peakcall_mms1_default_extsize180_peak_64
Chr XV	855670	855950	281	855717	20.00	418.960	239.090	169.052	peakcall_mms1_default_extsize180_peak_65
Chr XV	1060687	1061023	337	1060915	40.00	1.470.885	423.391	1.049.152	peakcall_mms1_default_extsize180_peak_66
Chr XV	1091100	1091289	190	1091222	47.00	742.374	234.139	423.706	peakcall_mms1_default_extsize180_peak_67
Chr XVI	338814	339003	190	338904	54.00	752.677	222.313	432.364	peakcall_mms1_default_extsize180_peak_68
Chr XVI	536587	536770	184	536716	34.00	1.409.377	460.031	999.169	peakcall_mms1_default_extsize180_peak_69
Chr XVI	922314	922518	205	922398	30.00	675.542	272.462	368.757	peakcall_mms1_default_extsize180_peak_70
Chr XVI	945944	946241	298	945944	68.00	373.205	155.965	137.412	peakcall_mms1_default_extsize180_peak_71
Chr M	2468	2658	191	2619	40.00	1.144.746	341.038	769.971	peakcall_mms1_default_extsize180_peak_72
Chr M	6034	6275	242	6176	55.00	1.609.468	363.317	1.165.234	peakcall_mms1_default_extsize180_peak_73
Chr M	15086	15328	243	15173	20.00	590.802	302.739	299.602	peakcall_mms1_default_extsize180_peak_74
Chr M	18608	18863	256	18768	28.00	910.884	354.172	568.783	peakcall_mms1_default_extsize180_peak_75
Chr M	19248	19506	259	19354	20.00	577.220	297.502	288.352	peakcall_mms1_default_extsize180_peak_76
Chr M	24795	24974	180	24909	24.00	480.983	242.098	214.714	peakcall_mms1_default_extsize180_peak_77
Chr M	28760	29077	318	28904	17.00	982.868	522.979	630.414	peakcall_mms1_default_extsize180_peak_78
Chr M	36684	36965	282	36858	13.00	988.256	609.433	634.195	peakcall_mms1_default_extsize180_peak_79
Chr M	37940	38414	475	38264	19.00	1.601.831	831.931	1.158.768	peakcall_mms1_default_extsize180_peak_80
Chr M	40345	40537	193	40405	35.00	1.058.635	348.621	695.164	peakcall_mms1_default_extsize180_peak_81
Chr M	40815	41221	407	40936	21.00	1.318.506	620.844	921.853	peakcall_mms1_default_extsize180_peak_

									82
Chr M	41558	42373	816	41825	19.00	1.382.073	702.586	975.087	peakcall_mms1_default_extsize180_peak_83
Chr M	43258	43661	404	43513	26.00	841.561	346.917	508.972	peakcall_mms1_default_extsize180_peak_84
Chr M	44088	44272	185	44225	18.00	690.562	362.646	381.355	peakcall_mms1_default_extsize180_peak_85
Chr M	50660	50926	267	50781	50.00	2.070.018	493.880	1.566.621	peakcall_mms1_default_extsize180_peak_86
Chr M	53592	53813	222	53693	66.00	610.028	187.350	315.245	peakcall_mms1_default_extsize180_peak_87
Chr M	68518	68770	253	68639	30.00	775.677	300.202	451.627	peakcall_mms1_default_extsize180_peak_88
Chr M	83903	84362	460	84229	49.00	1.705.155	415.901	1.242.801	peakcall_mms1_default_extsize180_peak_89
Chr M	84631	84909	279	84761	26.00	1.287.038	515.339	894.333	peakcall_mms1_default_extsize180_peak_90
Chr M	85491	85769	279	85636	44.00	1.176.328	328.041	798.103	peakcall_mms1_default_extsize180_peak_91

Table 14: Oberlap of Mms1 binding regions (BR) identified by MACS 2.0 with selected genome features.

Region	BR start	BR end	Feature	Feature start	Feature end
ChrV	312024	312220	long_terminal_repeat	312274	312445
ChrV	442043	442465	long_terminal_repeat	442734	443072
ChrV	574897	575199	telomere	569599	576874
ChrVII	567184	567405	LTR_retrotransposon	561843	567760
ChrVII	567184	567405	long_terminal_repeat	567427	567760
ChrVII	700602	700830	long_terminal_repeat	701104	701350
ChrVII	1088958	1089156	telomere	1083635	1090940
ChrXII	1074796	1075077	telomere	1064281	1078177
ChrXIII	1731	1917	telomere	1	6344
ChrXIV	547086	547297	long_terminal_repeat	546423	546714
ChrXIV	547086	547297	long_terminal_repeat	546738	547077
ChrXIV	631800	631983	long_terminal_repeat	631998	632114
ChrXIV	631800	631983	long_terminal_repeat	632119	632489
ChrXV	710139	710326	long_terminal_repeat	709691	710023
ChrXV	710139	710326	long_terminal_repeat	710024	710107
ChrXV	710139	710326	LTR_retrotransposon	704063	710023
ChrXV	1091099	1091289	telomere	1083922	1091291
ChrXV	1091099	1091289	telomeric_repeat	1091274	1091291
ChrXVI	945943	946241	telomere	942396	948010
ChrIV	461797	462449	ARS	462567	462622
ChrIV	461797	462449	ARS_consensus_sequence	462596	462606
ChrV	442043	442465	ARS	442416	442735
ChrXII	458652	459006	ARS	458991	459097
ChrXII	458652	459006	ARS_consensus_sequence	459078	459088
ChrXII	459110	459290	ARS	458991	459097
ChrXII	459110	459290	ARS_consensus_sequence	459078	459088
ChrXII	467534	468132	ARS	468128	468234
ChrXII	467534	468132	ARS_consensus_sequence	468215	468225
ChrXII	921240	921442	ARS	919800	920994
ChrXII	84188	84385	no overlap with a gene		
ChrI	112580	112785	γ-H2A binding site	112482	112538
ChrI	112580	112785	γ-H2A binding site	112637	112686
ChrI	112580	112785	γ-H2A binding site	112789	112846

ChrI	112580	112785	γ-H2A binding site	113089	113144
ChrI	130081	130284	γ-H2A binding site	129699	129755
ChrI	130081	130284	γ-H2A binding site	130007	130061
ChrI	130081	130284	γ-H2A binding site	130157	130213
ChrI	130081	130284	γ-H2A binding site	130328	130378
ChrI	191965	192177	γ-H2A binding site	191693	191737
ChrI	191965	192177	γ-H2A binding site	191840	191884
ChrI	191965	192177	γ-H2A binding site	192001	192045
ChrI	191965	192177	γ-H2A binding site	192237	192296
ChrII	215788	215983	γ-H2A binding site	215548	215605
ChrII	215788	215983	γ-H2A binding site	215780	215839
ChrII	463749	464101	γ-H2A binding site	463480	463539
ChrII	463749	464101	γ-H2A binding site	463657	463711
ChrII	463749	464101	γ-H2A binding site	463877	463933
ChrII	463749	464101	γ-H2A binding site	464048	464092
ChrII	463749	464101	γ-H2A binding site	464259	464313
ChrIV	461797	462449	γ-H2A binding site	461832	461889
ChrIV	461797	462449	γ-H2A binding site	462216	462274
ChrIV	461797	462449	γ-H2A binding site	462820	462879
ChrIV	1049062	1049304	γ-H2A binding site	1048983	1049040
ChrIV	1251076	1251389	γ-H2A binding site	1251093	1251137
ChrIV	1251076	1251389	γ-H2A binding site	1251270	1251314
ChrIV	1251076	1251389	γ-H2A binding site	1251441	1251492
ChrV	40469	41119	γ-H2A binding site	40763	40811
ChrV	40469	41119	γ-H2A binding site	40939	40991
ChrV	40469	41119	γ-H2A binding site	41108	41166
ChrV	42006	42474	γ-H2A binding site	41740	41795
ChrV	42006	42474	γ-H2A binding site	42707	42766
ChrV	43222	43945	γ-H2A binding site	43554	43610
ChrV	43222	43945	γ-H2A binding site	43733	43788
ChrV	43222	43945	γ-H2A binding site	43917	43973
ChrV	43222	43945	γ-H2A binding site	44081	44134
ChrVII	1049534	1049759	γ-H2A binding site	1049121	1049180
ChrVIII	189407	189652	γ-H2A binding site	189224	189270
ChrVIII	189407	189652	γ-H2A binding site	189404	189453
ChrIX	54150	54474	γ-H2A binding site	54041	54085
ChrIX	54150	54474	γ-H2A binding site	54205	54249
ChrIX	54150	54474	γ-H2A binding site	54596	54651
ChrIX	54150	54474	γ-H2A binding site	54853	54900
ChrIX	334127	334307	γ-H2A binding site	333911	333968
ChrIX	334127	334307	γ-H2A binding site	334068	334119
ChrIX	334127	334307	γ-H2A binding site	334273	334321
ChrIX	392157	392481	γ-H2A binding site	392709	392764
ChrX	121000	121236	γ-H2A binding site	120875	120922
ChrX	400013	400260	γ-H2A binding site	399626	399681
ChrX	400013	400260	γ-H2A binding site	400141	400199
ChrX	400013	400260	γ-H2A binding site	400350	400409
ChrXI	380958	381322	γ-H2A binding site	380732	380781
ChrXI	380958	381322	γ-H2A binding site	380898	380955
ChrXI	380958	381322	γ-H2A binding site	381335	381385
ChrXI	519605	519853	γ-H2A binding site	519471	519524
ChrXII	451470	451833	γ-H2A binding site	451118	451177
ChrXII	921240	921442	γ-H2A binding site	921227	921286
ChrXIV	619048	619366	γ-H2A binding site	618975	619034
ChrXIV	619048	619366	γ-H2A binding site	619202	619258
ChrXIV	652075	652255	γ-H2A binding site	652637	652691
ChrXV	31003	31286	γ-H2A binding site	30665	30714
ChrXV	31003	31286	γ-H2A binding site	30846	30900
ChrXV	31003	31286	γ-H2A binding site	31135	31181
ChrXV	31003	31286	γ-H2A binding site	31631	31690
ChrXV	480108	480435	γ-H2A binding site	480399	480451
ChrXV	480108	480435	γ-H2A binding site	480582	480637
ChrXV	480108	480435	γ-H2A binding site	480828	480875
ChrXV	588162	588351	γ-H2A binding site	587930	587989
ChrXV	855669	855950	γ-H2A binding site	855916	855974
ChrXV	1060686	1061023	γ-H2A binding site	1060719	1060774
ChrXVI	922313	922518	γ-H2A binding site	922247	922296
ChrXVI	922313	922518	γ-H2A binding site	922595	922649
ChrXVI	922313	922518	γ-H2A binding site	922777	922836

Table 15: Examples of Mms1 BMs that also resemble a G4tract2 motif. All guanines that can be used for G4 formation are marked red.

Location	5´flanking region	BM	3´flanking region
Chr VII:806449-806663	T**GG**TTTT**GG**T	**GG**TCGT**GG**C**GG**T**GG**TCGT**GG**	C**GG**TAACCGT
Chr XIII:1731-1917	TGTT**GG**AGTC	**GG**TACTTTC**GG**T**GG**TAGTAG	CACTAGTGTT
Chr IV:427403-427648	T**GG**TTCCAGA	**GG**C**GG**CTTC**GG**T**GG**TAGA**GG**	C**GG**TTCTCGT
Chr VII:806449-806663	**GG**T	**GG**TCGT**GG**C**GG**TAGCCGT**GG**	TTTT**GG**T**GG**T
Chr X:400013-400260	T**GG**A**GG**TGCT	**GG**A**GG**TGTT**GG**A**GG**TACT**GG**	A**GG**TATTGAA
Chr XII:904035-904280	CCTT	**GG**T**GG**AAGAAGAAGA**GG**AAG	AAGA**GG**ATTC
Chr XI:519605-519853	CAGCAGCCGC	CGA**GG**TAGC**GG**T**GG**T**GG**CAG	C
Chr II:215788-215983	CAGAAGCA**GG**	AGC**GG**CTGCA**GG**AGTAGCTG	CA**GG**AGCTTC

Table 16: Identified G4tract3 motifs (consensus sequence $GGG(N)_{\leq 25}GGG(N)_{\leq 25}GGG(N)_{\leq 25}GGG$) within Mms1 chromosomal binding regions.

Region peak	Start	End	Region $G4_{tract3}$ motif	Start	End	Strand location
Chr II	193642	193822	Chr II	193198	193248	C
Chr V	42006	42474	Chr V	42254	42287	C
Chr V	335544	335734	Chr V	335624	335711	W
Chr V	335544	335734	Chr V	335996	336059	W
Chr V	396417	396611	Chr V	396518	396583	W
Chr X	391940	392163	Chr X	391690	391737	C
Chr X	400013	400260	Chr X	399981	400036	W
Chr XII	452031	452232	Chr XII	452560	452609	C
Chr XII	459110	459290	Chr XII	459596	459645	C
Chr XIII	16258	16578	Chr XIII	16793	16877	C
Chr XV	588162	588351	Chr XV	588156	588197	C
Chr XV	1091099	1091289	Chr XV	1091091	1091186	W

Table 17: Identified G4tract2 motifs (consensus sequence $GG(N)_{\leq 7}GG(N)_{\leq 7}GG(N)_{\leq 7}GG$) within Mms1 chromosomal binding regions.

Region peak	Start	End	Region $G4_{tract2}$ motif	Start	End	Strand location
Chr I	112580	112785	Chr I	112715	112739	W
Chr I	112580	112785	Chr I	112710	112742	C
Chr I	112580	112785	Chr I	113034	113050	C
Chr I	130081	130284	Chr I	130405	130426	C
Chr I	130081	130284	Chr I	130468	130492	C
Chr I	191965	192177	Chr I	191858	191885	W
Chr I	191965	192177	Chr I	191730	191752	C
Chr II	193642	193822	Chr II	193782	193804	W
Chr II	193642	193822	Chr II	193239	193259	C
Chr II	193642	193822	Chr II	193838	193868	C
Chr II	215788	215983	Chr II	215764	215791	W
Chr II	215788	215983	Chr II	216310	216338	W
Chr II	463749	464101	Chr II	463614	463640	W
Chr II	463749	464101	Chr II	463663	463684	W
Chr II	463749	464101	Chr II	463756	463791	W
Chr II	463749	464101	Chr II	464469	464488	W
Chr III	123534	123722	Chr III	123596	123615	C
Chr III	123534	123722	Chr III	123630	123651	C
Chr IV	427403	427648	Chr IV	427378	427619	W
Chr IV	427403	427648	Chr IV	427815	427836	W
Chr IV	461797	462449	Chr IV	461760	461787	W
Chr IV	461797	462449	Chr IV	462086	462106	W
Chr IV	1049062	1049304	Chr IV	1048742	1048753	W
Chr IV	1049062	1049304	Chr IV	1049093	1049110	W
Chr IV	1239630	1239824	Chr IV	1239628	1239648	W
Chr IV	1251076	1251389	Chr IV	1251103	1251124	W
Chr IV	1251076	1251389	Chr IV	1251373	1251385	C
Chr V	40469	41119	Chr V	41260	41274	W
Chr V	40469	41119	Chr V	41238	41255	C

Chr V	42006	42474	Chr V	41859	41871	W
Chr V	42006	42474	Chr V	42602	42618	W
Chr V	42006	42474	Chr V	42104	42125	C
Chr V	42006	42474	Chr V	42186	42211	C
Chr V	43222	43945	Chr V	43383	43404	W
Chr V	43222	43945	Chr V	43795	43811	W
Chr V	335544	335734	Chr V	335612	335640	W
Chr V	335544	335734	Chr V	335821	335844	W
Chr V	335544	335734	Chr V	335956	335976	W
Chr V	335544	335734	Chr V	335996	336059	W
Chr V	335544	335734	Chr V	335562	335579	C
Chr V	335544	335734	Chr V	335698	335718	C
Chr V	396417	396611	Chr V	396013	396035	W
Chr V	442043	442465	Chr V	441987	442015	W
Chr V	442043	442465	Chr V	442241	442257	W
Chr V	442043	442465	Chr V	442274	442306	W
Chr V	442043	442465	Chr V	442365	442378	W
Chr V	574897	575199	Chr V	575161	575181	C
Chr V	574897	575199	Chr V	575197	575217	C
Chr VI	224946	225231	Chr VI	225042	225063	W
Chr VI	224946	225231	Chr VI	225073	225096	W
Chr VI	224946	225231	Chr VI	225398	225425	W
Chr VI	224946	225231	Chr VI	225055	225071	C
Chr VI	224946	225231	Chr VI	225459	225481	C
Chr VII	567184	567405	Chr VII	567174	567193	W
Chr VII	567184	567405	Chr VII	567246	567278	W
Chr VII	700602	700830	Chr VII	700674	700712	W
Chr VII	806449	806663	Chr VII	806751	806773	W
Chr VII	806449	806663	Chr VII	806505	806606	C
Chr VII	1049534	1049759	Chr VII	1049655	1049672	W
Chr VII	1049534	1049759	Chr VII	1049748	1049768	W
Chr VII	1088958	1089156	Chr VII	1089227	1089247	C
Chr VII	1088958	1089156	Chr VII	1089263	1089283	C
Chr VIII	189407	189652	Chr VIII	189556	189578	W
Chr VIII	189407	189652	Chr VIII	190019	190034	C
Chr VIII	215840	216046	Chr VIII	216208	216227	W
Chr IX	54150	54474	Chr IX	54096	54119	W
Chr IX	54150	54474	Chr IX	54132	54167	W
Chr IX	54150	54474	Chr IX	54340	54357	W
Chr IX	54150	54474	Chr IX	54381	54405	W
Chr IX	54150	54474	Chr IX	54391	54422	C
Chr IX	54150	54474	Chr IX	54732	54753	C
Chr IX	334127	334307	Chr IX	334432	334449	C
Chr IX	392157	392481	Chr IX	391749	391767	W
Chr IX	392157	392481	Chr IX	391956	391974	W
Chr IX	392157	392481	Chr IX	392001	392019	W
Chr IX	392157	392481	Chr IX	392127	392145	W
Chr IX	392157	392481	Chr IX	392208	392226	W
Chr IX	392157	392481	Chr IX	392289	392307	W
Chr IX	392157	392481	Chr IX	392478	392496	W
Chr IX	392157	392481	Chr IX	392676	392694	W
Chr IX	392157	392481	Chr IX	392841	392862	W
Chr IX	392157	392481	Chr IX	392880	392909	W
Chr X	121000	121236	Chr X	120593	120619	C
Chr X	121000	121236	Chr X	120725	120745	C
Chr X	121000	121236	Chr X	121358	121377	C
Chr X	391940	392163	Chr X	391531	391557	C
Chr X	391940	392163	Chr X	391989	392008	C
Chr X	400013	400260	Chr X	399807	399826	W
Chr X	400013	400260	Chr X	399921	399935	W
Chr X	400013	400260	Chr X	399987	400019	W
Chr X	400013	400260	Chr X	400049	400067	W
Chr X	400013	400260	Chr X	400119	400175	W
Chr X	400013	400260	Chr X	400206	400238	W
Chr X	400013	400260	Chr X	400251	400343	W
Chr X	400013	400260	Chr X	400621	400647	W
Chr XI	99803	99985	Chr XI	99538	99562	W
Chr XI	99803	99985	Chr XI	99857	99875	W
Chr XI	99803	99985	Chr XI	99826	99872	C
Chr XI	380958	381322	Chr XI	380538	380562	W
Chr XI	380958	381322	Chr XI	381207	381227	W

Chr XI	519605	519853	Chr XI	519909	519935	W
Chr XI	519605	519853	Chr XI	519557	519593	C
Chr XI	519605	519853	Chr XI	519676	519697	C
Chr XI	519605	519853	Chr XI	519752	519832	C
Chr XII	84188	84385	Chr XII	83853	83870	W
Chr XII	84188	84385	Chr XII	84187	84212	W
Chr XII	84188	84385	Chr XII	84234	84256	W
Chr XII	84188	84385	Chr XII	84291	84342	W
Chr XII	369706	369916	Chr XII	369862	369893	W
Chr XII	369706	369916	Chr XII	369918	369935	W
Chr XII	451470	451833	Chr XII	451745	451766	C
Chr XII	451470	451833	Chr XII	451927	451953	C
Chr XII	452031	452232	Chr XII	451745	451766	C
Chr XII	452031	452232	Chr XII	451927	451953	C
Chr XII	452031	452232	Chr XII	452557	452580	C
Chr XII	458652	459006	Chr XII	458670	458693	W
Chr XII	458652	459006	Chr XII	458962	458983	C
Chr XII	458652	459006	Chr XII	459100	459126	C
Chr XII	458652	459006	Chr XII	459355	459368	C
Chr XII	459110	459290	Chr XII	458962	458983	C
Chr XII	459110	459290	Chr XII	459100	459126	C
Chr XII	459110	459290	Chr XII	459355	459368	C
Chr XII	459110	459290	Chr XII	459424	459437	C
Chr XII	459110	459290	Chr XII	459590	459620	C
Chr XII	467534	468132	Chr XII	467807	467830	W
Chr XII	467534	468132	Chr XII	468099	468120	C
Chr XII	467534	468132	Chr XII	468237	468263	C
Chr XII	467534	468132	Chr XII	468492	468505	C
Chr XII	704511	704746	Chr XII	704811	704831	C
Chr XII	904035	904280	Chr XII	904168	904188	C
Chr XII	904035	904280	Chr XII	904198	904215	C
Chr XII	904035	904280	Chr XII	904285	904312	C
Chr XII	921240	921442	Chr XII	921139	921165	W
Chr XII	921240	921442	Chr XII	921179	921215	W
Chr XII	921240	921442	Chr XII	921270	921288	C
Chr XII	921240	921442	Chr XII	921338	921360	C
Chr XII	1074796	1075077	Chr XII	1074732	1074746	W
Chr XII	1074796	1075077	Chr XII	1074924	1074959	W
Chr XIII	1731	1917	Chr XIII	1771	1791	W
Chr XIII	1731	1917	Chr XIII	1807	1827	W
Chr XIII	16258	16578	Chr XIII	16439	16453	C
Chr XIII	16258	16578	Chr XIII	16477	16497	C
Chr XIII	16258	16578	Chr XIII	16793	16865	C
Chr XIII	16258	16578	Chr XIII	16375	16403	C
Chr XIV	547086	547297	Chr XIV	547553	547589	W
Chr XIV	547086	547297	Chr XIV	546902	546923	C
Chr XIV	619048	619366	Chr XIV	619068	619145	C
Chr XIV	619048	619366	Chr XIV	619257	619328	C
Chr XIV	631800	631983	Chr XIV	631401	631416	W
Chr XIV	631800	631983	Chr XIV	631865	631884	C
Chr XIV	652075	652255	Chr XIV	652159	652192	C
Chr XV	31003	31286	Chr XV	31207	31230	W
Chr XV	31003	31286	Chr XV	31345	31368	W
Chr XV	159994	160358	Chr XV	160141	160164	W
Chr XV	159994	160358	Chr XV	160501	160522	W
Chr XV	159994	160358	Chr XV	159972	159998	C
Chr XV	159994	160358	Chr XV	160035	160063	C
Chr XV	159994	160358	Chr XV	160382	160402	C
Chr XV	480108	480435	Chr XV	480274	480289	W
Chr XV	480108	480435	Chr XV	479703	479723	C
Chr XV	480108	480435	Chr XV	479940	479964	C
Chr XV	588162	588351	Chr XV	588136	588180	C
Chr XV	588162	588351	Chr XV	588222	588245	C
Chr XV	855669	855950	Chr XV	855515	855531	W
Chr XV	855669	855950	Chr XV	855689	855745	C
Chr XV	855669	855950	Chr XV	855909	855927	C
Chr XV	1060686	1061023	Chr XV	1060827	1060853	W
Chr XV	1060686	1061023	Chr XV	1060971	1060991	W
Chr XVI	922313	922518	Chr XVI	922297	922314	W
Chr XVI	922313	922518	Chr XVI	921929	921951	C
Chr XVI	945943	946241	Chr XVI	946229	946249	C

Table 18: More details about regions investigated by qPCR for Mms1 binding.

Region (Chr)	Start	End	Gatract3 motif (Capra et al, 2010)	G4tract2 motif on leading strand	G4tract2 motif on lagging strand	MACS Peakcall
I_{NC}	61257	61473	+	+	-	-
VI_{BR}	255397	255624	+	+	+	-
VII_{BM}	806633	806696	-	+	+	+
IX_{BR}	356233	356408	+	+	+	-
X_{BR}	391792	391919	+	+	+	+
XIa_{BR}	142007	142159	-	+	+	-
XIb_{BR}	503681	503933	-	-	+	-
XI_{BM}	519523	519599	+	+	+	+
$XIII_{BR}$	672879	673048	-	+	+	-
$XIII_{NC}$	250596	250670	+	+	-	-
XIV_{NC}	88913	89103	-	-	-	-
XV_{BR}	318817	318975	+	+	+	-

Table 19: Characteristics of G4 motifs within qPCR regions tested for Mms1 binding. Strand location and distance to next ARS were used to elucidate if the G4 motif is replicated on the leading or lagging strand.

Region	Location region		Location G4		Strand	Sequence G4 (5´-3´)	Location next ARS	
(Chr)	Start	End	Start	End	location		Start	End
			60974	61018	W	GGGCAGCATCTCCGTTGGATTGTTGTGCATGGCCAGTGTCTTGG		
			61437	61493	W	GGTGAACGAGTGGGGACAGTTCAATTCTGTGGTAACAAGGCCACAATTGGTGGTGG		
			61618	61662	W	GGGTTCACGATGTCTAGGTTGAATAGCGAGGGTCGCCCCGTGGG		
			61683	61712	W	GGGTGGGAACGGCGACGGAACCGCGCCGG		
			61836	61864	W	GGAGGCAGGCTGGGCTTTTTTCGACGGG		
I_{NC}	61257	61473	61703	61753	C	CCGCGCCGGTTAATAACGATCCTAACTATTGTGGGCCATGTTACGGTGCC	70300	70469
			255328	255355	W	GGTCCTAAGGTACCAAAATCCGGGGGGG		
			255498	255548	W	GGGGCACACGTGCGGGAGTTTCAAAGGGGCAGAATAGTGGGGTTCAGGGG		
			255819	255849	W	GGTAAGACCAGGTGCAAGGAGAATACTGGG		
			254976	255047	C	CCATCAATTCCTTGGGCACATCAGCCATGGAACCCTTTCTAGCCTGTGGTTTCTTTGGACCTAAATGAACC		
			255319	255349	C	CCTCTCAACGGTCCTAAGGTACCAAAATCC		
VI_{BR}	255397	255624	255710	255757	C	CCGGTTTATTTCCAACCGGGAAATAAATTATTCCTAATAAAATTTCC	256277	256431
			806736	806773	W	GGTTGTTCAGTTTCTGGATGTGTTGGGGATACGGACGG		
			806484	806606	C	CCAGAACCAGATGGTCTGAAACCGCCACGGCCACCGCGAGCGCCACCACGACCACCGAATCCACGGTTACCGCCACGACCACCGCCACGACCACCAAAACCACGGCTACCGCCACGACCACC		
VII_{BM}	806633	806696	807069	807109	C	CCAATGACACCACCGATGTGTTCGAATTCCTTCTTCAACC	834491	834734
			356335	356403	W	GGAGACTGATTTGGAGGGTACGGTGGGTAATAAGGGAAGGTATCGGGATTGGGGTAGGCCATTAAGGG		
			356572	356600	C	CCGTAGCCTTTTGGTGTTCCCGTATTCC		
IX_{BR}	356233	356408	356762	356806	C	CCTGTGGAGTGCCCTCGATAGATAGTTTACCCACAAGTTCATCC	357160	357396
X_{BR}	391792	391919	391294	391323	W	GGTTTCGGGCCAGATTCATGGCCCTGTGG	375706	376227

			391333	391368	C	CCTACAGACAAAAAACCGTTACGTCCCGCCTCACC		
			391531	391557	C	CCGCGCAAGCCAGATCCAAGCACGCC		
			391854	391877	C	CCACTATTCAGCGCCGTCCGCCC		
			391916	391968	C	CCGGAGCAACGGGCAACCGTTTGGGGAAAGACCACACCCACGCGCGATCGCC		
			391989	392008	C	CCCCACACCCAGACCTCCC		
			392394	392448	C	CCAATTACCATGCCTAAAGAGACCCCTTCCAAAGCTGCTGCCGATGCATTGTCC		
			141891	141942	W	GGTAAAAAGTATCCTGGTCACCGATGGCCAAGTTACCTTCTGGGGTGATGG		
			141974	142008	W	GGTGGAGGACCATCGAATTGGAATTGTCTGTTGG		
			142127	142155	W	GGAGCTGGAGCTGGAGATGGTATAATGG		
			142174	142238	W	GGCTTGGATTTGGCCATCAGTGATTTGGGAGACTGGAGCAACAGTAGTATTGGTGGTGGCTTGG		
			142428	142487	W	GGGAGACGGCTGCAGCGGTAGTCTTAGCTGAGGTAGTCTTGGTGGTGGCTTGGATTTGG		
			142576	142595	W	GGTTTTGGTAGTGGCTTGG		
			142627	142666	W	GGCTCTCTTAGCCTTGGAGGAAGCAGTAGTGGCAATTGG		
			141903	141949	C	CCTGGTCACCGATGGCCAAGTTACCTTCTGGGGTGATGGACCAACC		
XIa_{BR}	142007	142159	142019	142055	C	CCAATTCTACCTTTACCGTCAGTCAGGATACCACCC	153020	153135
			503309	503359	W	GGAGAATCGAATCCTGGCATTAGTGGGATAAGAATAAAGGCTTTCCAAGG		
			503654	503686	W	GGCGGGGTTAGTAATGGAGTTAACCTACTAGG		
			503529	503560	C	CCGTTTCCTTTAGCCCAAGAGACCAGTTCCC		
			504100	504138	C	CCAAAGACTATCTGTCCCAACAATATTCCCCACATTCC		
XIb_{BR}	503681	503933	504279	504335	C	CCTTTGCCATTAAGAGGGCCTCGCTTAACGACCAAAAATAATCACGCCCATCTACC	517017	517265
			519028	519058	W	GGTGCCTGGATCTCGAGGCCGCGGCACTGG		
			519142	519185	W	GGGCAAGTAGGTCTTTCTGCACGGCCCGGCCCGGGTCGTGCGG		
			519699	519719	W	GGCGGCTGCTGATGGAAAGG		
			519889	519957	W	GGACAAGGTGCTGTCTCCTTGGACTGGTTAGGTCTAGGCGGCTGGGCTTCCATCATGGACATGAACGG		
			519018	519099	C	CCTAGTTCCTGGTGCCTGGATCTCGAGGCCGCGGCACTGGAAAAGCCCTTTCTTTTCCAGATCGGGAAACCTAATGAGTCC		
			519264	519304	C	CCATCTAATGTGTTTCCTTCTCGAGACCTCGGCGTCTCCC		
			519557	519637	C	CCTCCACCGCCGTCTTGGCCGCTCCAGCTGTCCACCATAGTGACAACCACCACCACAACGACAAGCGTGCCGTTGTCACC		
			519676	519697	C	CCAGCTGCCACCACCGCTACC		
			519737	519832	C	CCACCACTACTTTGTCCTCGACTGCCGCCGCCGCTACTACCTCTGCCGCCGCCTCTTCTTCCTCCTCTTCCTCTTCCTCCTCTTCCTCTTCTTCC		
			519863	519889	C	CCATTTCCTGTTCTGATTTCCCATCC		
XI_{BM}	519523	519599	520009	520063	C	CCAGGTTACGCTAAGACCCAATGGCCTTCTGAACAACCTTCCGATGGTAGATCC	517017	517265
			250258	250297	W	GGTTAATAGACGTGGTAGAGTTCCACAGGCCAGAATGGG		
$XIII_{NC}$	250596	250670	250327	250417	W	GGCCACTGGCGGTATTGCAGCAGGTGCTGCGGCTACCTCTTCTGGTCT	263063	263296

						TAGCGGTGGTATGACACCAGGATG GAGCTCCTTCGATGGTGG		
			250442	250732	W	GGAGGCTCAGGTGGTGGCGGTGTC TCCTCATGGGGTGGTGCTTCCACT TGGGGTGGCCAAGGTAATGGAGGT GCATCCGCTTGGGGCGGTGCTGGC GGCGGTGCCTCAGCTTGGGGCGGC CAAGGTACTGGTGCTACTTCTACT TGGGGTGGTGCTTCAGCCTGGGGT AACAAATCAAGTTGGGGCGGTGCA TCCACTTGGGCGTCGGGTGGTGAA TCTAATGGTGCCATGTCTACTTGG GGTGGTACCGGTGATAGGTCAGCC TACGGCGGGGCTTCCACCTGGGGA GG		
			250760	250822	W	GGCGGAGCTTCTGCATGGGGTAAC CAAGACGATGGAAATAGGTCTGCT TGGAACAACCAAGG		
			250838	250861	W	GGTGGTAACAGTACATGGGGAGG		
			250280	250331	C	CCACAGGCCAGAATGGGCCCAAGT TACGTCAGTGCCCCAAGAAACATG GCC		
			250689	250724	C	CCGGTGATAGGTCAGCCTACGGCG GGGCTTCCACC		
			673216	673259	W	GGTGATGGGGTGTCTGGAGTGGAT TCCAAAACGGCATTAACGG		
			673509	673530	W	GGAAGTGGCTGGCTTCTCTGG		
			672461	672483	C	CCAAATAGCCTGAGTTACCACC		
			672802	672848	C	CCGATTTCAGAACCACCACGGGAT TGCCAACCCAACATATCCTTCC		
$XIII_{BR}$	672879	673048	672878	672894	C	CCAAACCAGACCAACC	649309	649552
			88727	88775	W	GGGTTTTAATGTGGTAAACAAGAT GGCCCTTACGGGGCTCTTAGTGGG		
XIV_{NC}	88913	89103	88907	88950	C	CCCTAAGTGATTGTGCCGTTATAA CTTCCATTCGGGTGATACC	89531	89804
			318436	318471	W	GGCACAAGCTCAGGCTCAGGCACA GGCACAGGTGG		
			319281	319318	W	GGGTTGGACATCCTGTATGGGCTT CCGAAGATTCGGG		
			318523	318565	C	CCATCCTTCTAACCGAGGTATTCC ACAGCAAAACTTGCCTCC		
			318581	318614	C	CCTCCAACAAACGGTTCGACCGTA CATGAAGCC		
			318638	318661	C	CCACCCCACTTCATGCCCTTACC		
XV_{BR}	318817	318975	318863	318902	C	CCTAGCATTGTGCCTGTGGTCCCT GAACCCACTGAGCCC	309359	309925

Table 20: Overlap of Mms1 binding regions with Pif1 binding regions.

mms1_peak_Chr	mms1_peak_start	mms1_end	pif1_peak_Chr	pif1_peak_start_coord	pif1_peak_end_coord
Chr I	112580	112785	Chr I	111548	114298
Chr I	130081	130284	Chr I	129548	131298
Chr I	191965	192177	Chr I	189298	193548
Chr II	463749	464101	Chr II	463082	464332
Chr IV	461797	462449	Chr IV	461450	461700
Chr IV	1239630	1239824	Chr IV	1239450	1240200
Chr IV	1251076	1251389	Chr IV	1250700	1251700
Chr V	40469	41119	Chr V	40562	41562
Chr V	42006	42474	Chr V	40562	41562
Chr V	42006	42474	Chr V	42812	43562
Chr V	43222	43945	Chr V	42812	43562
Chr V	335544	335734	Chr V	333062	336812
Chr VI	224946	225231	Chr VI	224415	225415
Chr VII	567184	567405	Chr VII	567541	569041
Chr VII	700602	700830	Chr VII	700916	701166
Chr VII	806449	806663	Chr VII	806666	806916
Chr VIII	189407	189652	Chr VIII	188763	190263
Chr IX	54150	54474	Chr IX	51865	55115
Chr IX	334127	334307	Chr IX	333615	334365
Chr IX	392157	392481	Chr IX	389365	391365

Chr IX	392157	392481	Chr IX	392365	393115
Chr X	121000	121236	Chr X	120811	121311
Chr X	400013	400260	Chr X	399561	401061
Chr XI	380958	381322	Chr XI	380319	382069
Chr XI	519605	519853	Chr XI	519444	519694
Chr XI	519605	519853	Chr XI	520319	521319
Chr XII	84188	84385	Chr XII	84818	86318
Chr XII	458652	459006	Chr XII	459318	460818
Chr XII	459110	459290	Chr XII	459318	460818
Chr XII	904035	904280	Chr XII	902568	903068
Chr XII	904035	904280	Chr XII	903568	903818
Chr XII	921240	921442	Chr XII	920943	921193
Chr XIV	494007	494211	Chr XIV	493616	494366
Chr XIV	547086	547297	Chr XIV	547241	547491
Chr XIV	652075	652255	Chr XIV	650366	651366
Chr XV	31003	31286	Chr XV	28945	31445
Chr XV	159994	160358	Chr XV	159945	160695
Chr XV	480108	480435	Chr XV	479695	480445
Chr XV	480108	480435	Chr XV	481070	481320
Chr XV	710139	710326	Chr XV	709570	709820
Chr XV	1060686	1061023	Chr XV	1060570	1060820
Chr XVI	536586	536770	Chr XVI	536543	536793
Chr XVI	922313	922518	Chr XVI	921918	922918

Table 21: Proteins identified in the first Co-IP. Shown are proteins identified in the untagged, Mms1-Myc tagged wild type and Mms1-Myc tagged *pif1-m2* cells. Displayed are number of unique peptides of the proteins, normalized log2 ratios of LFQ intensities of the proteins for the different strains and signum values (determines significance, 0=not significant) (MS analysis and evaluation was performed by AG Schlosser).

	Unique peptides			Norm.log2.Ratio.LFQ.intensity		Signum	
Protein	un-tagged	Mms1-Myc wild type	Mms1-Myc *pif1-m2*	Mms1-Myc *pif1-m2*_ vs_untagged	Mms1. Myc.wildtype_ vs_untagged	Mms1-Myc *pif1-m2*_ vs_untagged	Mms1. Myc.wild type_ vs_untagged
ACC1	22	19	15	-0.568643594	0.155583961	0	0
ACS2	10	11	10	0.070328883	-0.012592059	0	0
ACT1	18	18	17	0.068654997	-0.004152203	0	0
ADH1	16	13	14	-0.256102992	-0.428367925	0	0
ADK1	7	6	6	-0.045302475	-0.196225399	0	0
AFG3	11	10	7	-0.814851328	-0.33602234	0	0
AHA1	2	2	2	-0.01299116	-0.103047706	0	0
AHP1	3	2	3	-0.69795286	-0.109145448	0	0
ALA1	6	7	3	-0.683768762	-0.020889446	0	0
ALD4	9	6	7	-0.202398283	-0.211284868	0	0
ALD6	15	12	12	0.134358082	-0.241751075	0	0
ALO1	4	5	4	-0.178936615	0.617499766	0	0
ARC1	5	6	6	0.011777708	-0.046458781	0	0
ARF1;ARF2	8	8	6	-1.304190839	-0.05332309	1	0
ARG1	0	1	1	3.05147505	0	2	0
ARG5,6	4	3	3	0.012962604	0	0	0
ARO1	3	4	4	0.804459397	0	0	0
ARO4	8	7	7	0.110387588	0.547072767	0	0
ARX1	3	3	4	-0.215780978	0.255682448	0	0
ASC1	8	8	8	0.601498289	-0.061530293	0	0
ASN1	3	1	1	0.256324477	-0.150012549	0	0
ASN2	2	1	2	0.107453784	0	0	0
ATP1	13	14	12	-0.353863713	-0.226839495	0	0
ATP2	11	11	12	-0.443763585	-0.264235358	0	0
BAT1	7	7	9	0.558876665	0.219954835	0	0
BBC1	1	2	2	1.97992335	0	1	0
BFR1	4	4	5	-0.295663082	0.432310351	0	0
BGL2	2	1	2	0.657530735	0	0	0
BMH1	3	3	3	-0.27297131	0.01771012	0	0
BMH2	2	2	2	0.28242628	0.799133724	0	0
CAM1	2	2	3	-0.238267237	-0.21894399	0	0

CBF1	2	2	2	1.635212563	0	1	0
CBF5	7	6	6	0.160240765	-0.500032682	0	0
CBR1	5	3	4	-0.268532471	-0.06417888	0	0
CCT2	3	4	4	-0.392979176	-0.19556825	0	0
CCT3	6	6	5	-0.504099573	0.260005118	0	0
CCT6	2	4	2	-0.150998413	0.708571757	0	0
CCT7	6	6	5	0.057551274	0.319144962	0	0
CDC19	29	32	31	0.219419775	-0.084794264	0	0
CDC33	4	4	4	-0.475512582	-0.027600826	0	0
CDC48	7	9	8	0.703801791	0.763431033	0	0
CDC60	6	10	8	0.106293005	-0.020703143	0	0
CLU1	1	4	2	3.272036122	0	2	0
COF1	2	2	2	0.651318762	-0.280375108	0	0
CPA2	4	2	4	-0.281629414	0.964019793	0	0
CPR1	4	4	4	-0.82943787	-0.482435816	0	0
CPR6	7	9	7	0.235680994	0.147963862	0	0
CYS3	10	9	9	-0.207285174	-0.006245346	0	0
CYS4	17	16	15	-0.569187123	-0.032710495	0	0
DBP2	7	6	5	-0.291235329	-0.100396588	0	0
DBP3	1	1	2	-0.62417334	0	0	0
DED1	5	5	3	-0.693776478	-0.445090099	0	0
DLD3	2	2	2	-0.016072722	-0.314650276	0	0
DNA2	2	1	1	5.925623662	0	2	0
DPS1	6	2	5	0.798826702	0	0	0
ECM33	2	4	3	-1.777893007	0.118945804	1	0
EFB1	6	5	4	-0.396781906	-0.265287804	0	0
EFT1	29	26	25	-0.059415957	-0.157033907	0	0
EGD1	4	4	3	0.113730265	-0.084129719	0	0
EGD2	5	4	4	-0.153587902	-0.212608888	0	0
EMP24	2	2	2	-0.81153241	-0.155362724	0	0
ENO1	3	3	2	0.224334667	-0.106689452	0	0
ENO2	13	13	14	0.261057911	-0.222543524	0	0
ERG1	4	7	5	0.395402396	0.455214023	0	0
ERG11	8	8	7	-0.321071022	0.338420653	0	0
ERG13	5	4	5	0.302892397	0.271359166	0	0
ERG2	1	2	2	0.159975263	0	0	0
ERG20	2	2	2	3.384557286	0	2	0
ERG25	3	2	2	0.045489284	-0.009544094	0	0
ERG27	6	7	7	-0.000626278	0.804604101	0	0
ERG6	8	9	10	0.263332018	0.647032657	0	0
ERG9	4	4	3	0.475355604	0.809996474	0	0
FAA1	4	5	6	-0.201723444	0.47444611	0	0
FAS1	28	26	23	0.256945028	0.554298191	0	0
FAS2	25	28	24	0.22815145	0.512919642	0	0
FBA1	11	11	10	0.438118065	0.364918373	0	0
FRS1	7	4	7	0.104496632	0.048346644	0	0
FUR1	5	5	5	-0.34572275	0.044469912	0	0
GAR1	2	1	1	3.4524236	0	2	0
GAS1	5	5	5	-0.082678225	0.045944458	0	0
GAS5	3	4	2	0.393160428	0.290542231	0	0
GCD11	6	5	5	0.147386824	0.336061782	0	0
GDH1	6	9	10	0.616837561	0.072259453	0	0
GET3	2	3	3	0.224306464	0.432995276	0	0
GFA1	8	9	9	0.281354852	0.178338673	0	0
GLK1	1	2	2	0.873114994	0.593554803	0	0
GLT1	2	1	1	0.345312533	0	0	0
GLY1	4	3	4	0.408166277	-1.615505116	0	1
GND1	6	5	8	0.234564093	-0.223900697	0	0
GPD2	6	4	5	-0.486266968	0.556742471	0	0
GPM1	13	12	12	0.135013951	-0.3782762	0	0
GPP1	10	8	7	-1.842559462	-0.901890832	1	0
GRS1	6	5	7	0.568500999	-0.127282826	0	0
GSF2	4	4	3	-0.999150427	0.268617011	0	0
GSP1;G	6	6	5	-0.266181839	0.120346011	0	0
GUA1	9	7	7	-0.365618841	-0.262653474	0	0
GUK1	2	2	2	2.574091064	0	2	0
GUS1	19	17	16	0.130461843	0.052557974	0	0
GUT2	3	2	2	0.159742777	-0.02990146	0	0
GVP36	4	3	2	0.210679759	0.222594672	0	0
HHO1	7	8	9	1.74872548	0.730181412	1	0
HIS3	0	0	2	7.090430963	0	2	0

HIS4	2	5	4	0.975658601	2.832372275	0	2
HIS5	1	2	2	0.154706389	0.742826833	0	0
HOM2	2	2	4	0.763950003	0.146679698	0	0
HOM3	4	4	4	-0.015064447	0.370801254	0	0
HOM6	2	1	1	1.148984674	0	0	0
HRP1	3	2	3	0.330959024	0	0	0
HSC82	9	10	6	-0.17857153	0.115993007	0	0
HSP104	11	11	11	-0.528479633	-0.201659427	0	0
HSP26	9	9	8	-0.95716687	-0.577579424	0	0
HSP60	10	10	7	0.058435466	-0.139501619	0	0
HSP82	1	2	2	4.185494175	4.443550755	2	2
HTB2;HTB1	4	3	4	-0.467217609	-0.743263364	0	0
HTS1	1	1	1	0.589418145	0	0	0
HTZ1	1	1	1	1.267939738	0.519588557	0	0
HXK2	8	6	9	-0.305183497	-0.159928269	0	0
HYP2	6	5	4	-0.142481092	-0.443498387	0	0
IDH1	1	2	1	0.763461962	0	0	0
IDP1	2	2	1	-0.455965935	-0.471928362	0	0
IFA38	2	2	2	0.004710582	0.536054382	0	0
ILS1	6	4	6	0.020680626	0.634455096	0	0
ILV1	5	4	3	-0.289081114	0.285806176	0	0
ILV2	8	8	5	-0.490620885	0.182224615	0	0
ILV3	4	2	4	0.662395849	0	0	0
ILV5	17	15	12	0.011785731	-0.128967283	0	0
IMD3	4	3	4	-0.127932333	-0.210034031	0	0
IMD4	3	3	2	-0.014359631	0.075357575	0	0
IPP1	5	4	5	0.028754127	-0.333074052	0	0
KAP123	8	13	11	-0.446211241	-0.236113136	0	0
KAP95	2	1	1	0.32639233	0	0	0
KAR2	7	9	6	0.056559565	0.511051238	0	0
KES1	7	10	9	-0.117574507	0.465042715	0	0
LAT1	6	5	5	1.21474392	0.197801344	1	0
LHS1	3	3	2	-0.704599279	0.276457399	0	0
LIA1	2	1	1	-0.909188853	0	0	0
LYS12	3	3	2	-0.345013476	0.235529047	0	0
LYS21;	8	9	8	0.140800717	0.337730806	0	0
MBF1	3	3	3	0.927672099	-0.18228207	0	0
MES1	9	7	7	0.312656749	0.056716025	0	0
MET17	7	7	7	0.306408409	-0.05622737	0	0
MET6	17	17	19	0.159750074	0.067534851	0	0
MMF1	2	2	2	0.18179208	0	0	0
MMS1	0	52	49	10.63490298	11.03991713	2	2
NEW1	4	2	2	-0.291237771	0.114654939	0	0
NIP1	4	6	5	0.428408	0.601168332	0	0
NOP1	10	7	10	0.555318617	-0.440377698	0	0
NOP56	10	11	13	0.288311203	-0.499167357	0	0
NOP58	5	4	5	0.408612136	-0.029292975	0	0
NPL3	5	6	3	-0.490767559	0.022803857	0	0
OYE2;OYE3	2	2	2	0.656727901	0.584280143	0	0
PAA1	4	3	4	-0.142228457	-0.410835498	0	0
PAB1	16	14	17	-0.168553853	-0.654074318	0	0
PAD1	1	1	3	4.156601746	0	2	0
PDC1	29	28	26	-0.070889252	0.031763442	0	0
PDI1	4	4	2	0.766835462	0.812324739	0	0
PEP4	5	4	4	0.331009349	0.35467653	0	0
PET9	10	10	10	0.677592585	0.444584526	0	0
PFK1	22	24	23	-0.262961754	0.134892679	0	0
PFK2	16	11	13	-0.042813269	0.018303781	0	0
PGI1	8	8	7	0.351274089	0.548832732	0	0
PGK1	28	27	27	0.03219104	0.036506465	0	0
PHO12;PHO11	3	4	3	-0.326961825	0.944790573	0	0
PLB1	1	1	2	-0.632947037	0	0	0
PMA1;PMA2	17	14	7	-1.702040494	-0.947866833	1	0
PMI40	0	1	2	0.575571231	0	0	0
PMT2	3	3	2	-0.804798386	-0.390352517	0	0
PNO1	3	3	3	-0.016071347	-0.140536763	0	0
POR1	8	5	6	-0.577391635	-0.465191103	0	0
PRT1	5	5	5	-0.480624026	-0.010686113	0	0
PSA1	13	11	11	-0.244982666	-0.411389459	0	0
PST2	2	3	3	3.011764735	0	2	0
PSY4	2	2	2	-0.197367011	0.321524024	0	0

PUB1	3	2	2	0.033704792	-0.629620664	0	0
PUP2	2	1	2	-0.137920898	0	0	0
PYC2;PYC1	2	2	2	-0.390110777	-0.325183328	0	0
RHO1	4	5	5	0.222653198	0.37418415	0	0
RHO3	3	2	1	-0.995749863	0	0	0
RLI1	5	3	3	-2.0544516	0	2	0
RNR1	10	10	7	-0.263235348	0.412383305	0	0
RNR2	7	7	8	0.122908078	0.253748949	0	0
RNR4	7	7	8	0.529918937	0.509094799	0	0
RPC40	6	4	5	-0.186529677	-0.163832302	0	0
RPG1	17	15	10	0.067482135	-0.01842324	0	0
RPL12B;RPL12A	4	4	4	-0.997276175	-0.408735131	0	0
RPL13A;RPL13B	3	2	2	-1.431873707	-1.191207167	0	1
RPL14B;RPL14A	5	2	5	1.508002198	-0.022714558	1	0
RPL18B;RPL18A	4	4	4	-0.016260502	-0.577651325	0	0
RPL19B;RPL19A	6	6	4	-0.275314472	-0.796207357	0	0
RPL1B;RPL1A	7	6	9	-0.670448733	-0.710566634	0	0
RPL20B;RPL20A	9	8	7	0.232919499	-0.241604703	0	0
RPL21B	1	1	0	-1.124587484	0.000906961	0	0
RPL22A;RPL22B	0	1	1	1.791406763	0	1	0
RPL26A	2	2	2	0.321925258	2.144700805	0	2
RPL26B	2	3	3	-0.186189132	1.551866644	0	1
RPL27B;RPL27A	8	6	8	0.728211622	-0.543026564	0	0
RPL28	4	4	4	-0.409338658	-0.030335565	0	0
RPL31B;RPL31A	4	4	4	-0.678467878	0.365734599	0	0
RPL32	6	5	5	0.126930097	0.338621864	0	0
RPL33A	2	2	2	0.58855901	-0.333004385	0	0
RPL35B;RPL35A	5	5	4	-0.470816353	0.609367378	0	0
RPL36A;RPL36B	4	4	4	0.799696442	0.091473959	0	0
RPL39	3	3	3	0.015758019	0.431875352	0	0
RPL6A	3	2	3	0.106689449	-0.400854138	0	0
RPL6B	5	5	5	0.327745702	0.115518326	0	0
RPL7B;RPL7A	5	5	5	0.576993642	-0.244812995	0	0
RPL9B;RPL9A	7	5	6	-0.063017275	0.1660132	0	0
RPN11	3	3	3	1.017336162	0	0	0
RPN5	3	4	3	-0.006574395	0.982514145	0	0
RPN9	3	3	2	2.169608192	-1.537288354	2	1
RPP0	1	2	2	1.292793217	1.218645706	0	1
RPS0B;RPS0A	4	5	3	-1.359100597	0.383614835	1	0
RPS11B;RPS11A	9	9	8	-0.446718859	-0.233866876	0	0
RPS12	6	6	6	0.199576914	-0.183086719	0	0
RPS13	2	3	3	2.006865091	1.951053509	2	2
RPS14A;RPS14B	3	3	2	-1.288056594	-0.792770797	0	0
RPS16B;RPS16A	7	7	6	-0.753648241	-0.049463801	0	0
RPS17B;RPS17A	5	5	5	-0.458238826	-0.0493234	0	0
RPS18B;RPS18A	8	7	5	-1.221985385	-0.819597731	1	0
RPS1B	3	2	2	-0.650417943	-0.385578853	0	0
RPS2	6	5	6	0.286915314	-0.070545984	0	0
RPS20	5	5	5	0.20620548	0.770835199	0	0
RPS22B;RPS22A	6	7	7	-0.202207439	0.755346651	0	0
RPS25A;RPS25B	3	3	3	-2.172544973	-1.710113102	2	1
RPS3	9	8	7	-1.065962862	-0.372738443	0	0
RPS31	5	3	4	0.176221758	-0.01628026	0	0
RPS4B;RPS4A	8	7	6	-0.445786003	0.097574711	0	0
RPS5	10	10	8	0.317718282	0.575532136	0	0
RPS7A	7	7	7	-0.288860211	0.438200846	0	0
RPS9B;RPS9A	8	8	8	-0.420323265	-0.04491447	0	0
RPT1	4	5	5	0.939410225	1.146854378	0	1
RPT5	3	4	2	-0.410670563	0	0	0
RRP3	2	2	2	-0.445484416	-0.225185551	0	0
RTN1	4	4	3	-0.163179931	0.324968425	0	0
RTT101	0	22	20	7.416611583	7.796972052	2	2
RVB1	3	4	2	-0.57612575	0.058944049	0	0
RVB2	1	2	1	-0.770922047	0	0	0
SAC1	1	3	2	1.484163547	0	0	0
SAH1	10	9	10	0.293961587	-0.008175224	0	0
SAM1	5	4	3	-0.240988762	0.250869051	0	0
SAM2	3	2	2	-1.155702018	-0.797487207	1	0
SAR1	3	5	2	-0.54509029	0.179575994	0	0
SCP160	18	17	14	0.083570614	0.157500208	0	0
SDH1	2	2	1	5.588682309	0	2	0

SEC13	3	3	3	-0.144354582	-0.02820485	0	0
SEC14	10	10	8	-0.249249247	0.4129737	0	0
SEC17	3	3	2	-0.402337865	0.534229844	0	0
SEC23	1	0	1	0.55713248	0	0	0
SEC31	2	1	1	2.547454621	0	2	0
SEC4	4	4	4	-0.029481374	0.446214253	0	0
SEC53	8	8	6	-0.902186305	0.122227319	0	0
SEC61	3	3	3	-0.211847463	-0.096626759	0	0
SER33	2	3	1	-2.471493867	0.254949111	2	0
SES1	5	4	4	0.019764666	-0.3028904	0	0
SGT2	5	5	4	0.44343365	0.199639918	0	0
SHM1	6	8	9	1.154498428	0.432588063	1	0
SHM2	5	6	6	-0.046408069	0.378902692	0	0
SIS1	13	14	13	0.206448745	-0.066490212	0	0
SNU13	1	2	2	5.639411231	0	2	0
SRO9	5	3	3	-0.336472093	0.135185739	0	0
SRP14	2	3	3	0.094767387	0.208332559	0	0
SSA1	7	6	5	-0.029373009	0.152520745	0	0
SSA2	8	8	8	0.032277296	0.0501241	0	0
SSB1	2	3	2	-0.219669619	0.030809156	0	0
SSB2	1	2	1	3.738370081	0	2	0
SSC1	12	14	12	0.11821483	0.09962755	0	0
SSE1	16	15	16	0.035977941	-0.185775713	0	0
SSZ1	15	19	13	-0.078800842	0.233672203	0	0
STI1	13	12	12	0.314423309	0.427681763	0	0
SUB2	6	6	7	-0.249719702	-0.027741835	0	0
TAL1	5	6	5	-0.123953537	-0.138633209	0	0
TDH1	1	1	3	1.721108941	0	1	0
TDH2	3	3	3	0.901233864	0.214259168	0	0
TDH3	4	4	3	0.236340028	0.345910372	0	0
TEF1	26	28	25	-0.142054797	0.403266988	0	0
TEF4	13	13	13	-0.17827413	0.364513882	0	0
THR1	6	2	4	-0.447545211	0	0	0
TIF1	15	16	13	-0.447695527	-0.246229171	0	0
TIF3	2	2	2	-0.123527618	-0.674840605	0	0
TKL1	8	7	9	0.069475624	-0.246672565	0	0
TMA19	6	6	6	-0.166833299	-0.337939235	0	0
TPA1	1	2	2	0.164940843	0	0	0
TPI1	10	10	9	-0.099014637	0.155598801	0	0
TRM1	34	33	34	1.040913958	-0.173663354	0	0
TRP2	7	7	5	-0.321489102	-0.025484423	0	0
TRP5	11	15	10	0.28856245	0.440900167	0	0
TRX1;TRX2	1	2	1	-0.049612192	0	0	0
TSA1	5	5	5	-0.03470957	-0.348110737	0	0
TUB2	3	4	4	-0.570758873	-0.039912669	0	0
TY1B-DR4	11	12	9	0.140463055	0.60565125	0	0
TY2B-LR2	1	2	1	3.890950502	0	2	0
TYS1	7	10	8	0.354534364	0.762017191	0	0
URA2	20	21	18	-0.121004969	0.16195433	0	0
URA5	2	1	2	0.822356978	0	0	0
URA6	2	1	1	-1.48913427	0	0	0
URA7	10	8	4	-0.257759814	0.094900623	0	0
UTP7	3	2	2	-1.19495773	-0.927743234	1	0
UTR2	2	2	2	1.165669404	0	0	0
VAS1	13	14	11	-0.041665573	0.255882193	0	0
VMA1	11	10	11	-0.058055586	-0.189200625	0	0
VMA10	2	2	2	-0.814142906	0.50587981	0	0
VMA13	3	5	3	0.391646569	0.341632276	0	0
VMA2	12	13	12	-0.439776773	-0.340281721	0	0
VMA4	4	5	5	0.00644265	0.529814679	0	0
VMA5	6	3	3	-0.290734003	0.103757498	0	0
VPH1	3	3	2	-0.257763239	0	0	0
VPS1	8	8	8	-0.10121683	0.659812144	0	0
VTC2	2	1	2	-1.281090505	0	0	0
VTC3	8	7	4	-0.19122054	1.109277431	0	1
VTC4	8	4	3	-1.044787194	-0.768076869	0	0
WBP1	2	4	2	1.061444721	0	0	0
WRS1	1	2	2	-1.270838009	0	0	0
WTM1	3	3	3	0.334487161	-0.538262637	0	0
YAR1	2	1	1	0.018397691	0	0	0
YBT1	4	5	4	0.630338311	0.716586659	0	0

YDJ1	3	4	3	0.098799907	0.106939591	0	0
YDL124W	5	4	4	-0.379378446	-0.532754175	0	0
YDR341C	14	11	10	0.154319669	0.123434666	0	0
YEF3	17	17	17	0.28371835	-0.156599496	0	0
YGR054W	2	3	2	1.972872935	1.358191477	1	1
YHB1	7	7	6	-0.613107421	-0.323367588	0	0
YHR020W	17	17	15	0.213093211	0.326504592	0	0
YIP3	3	4	3	2.316471263	3.062679712	1	2
YMR099C	3	2	2	-0.79127831	0	0	0
YMR226C	3	3	3	0.886666857	-0.10253187	0	0
YOP1	2	2	2	0.324483449	0.710326219	0	0
YPL225W	1	1	2	2.446920424	0	1	0
YPT1	3	6	3	-0.185714327	0.539420252	0	0
YPT32;YPT31	2	2	2	2.066909106	0	1	0
YRA1	7	7	7	0.68806062	0.195502636	0	0
ZUO1	5	6	5	-0.338347767	0.195906878	0	0
ZWF1	3	1	3	-0.06105124	0	0	0

Table 22: Proteins identified in the second and third Co-IP (duplicates). Shown are proteins identified in the untagged, Mms1-Myc tagged wild type and Mms1-Myc tagged *pif1-m2* cells. Displayed are number of unique peptides of the proteins, normalized log2 ratios of LFQ intensities of the proteins for the different strains and signum values (determines significance, 0=not significant) (MS analysis and evaluation was performed by AG Schlosser).

	Unique peptides						Norm.log2.Ratio.LFQ.intensity				Signum			
	Mms1-Myc wild type		Mms1-Myc *pif1-m2*		untagged		Mms1-Myc wildtype_ vs_untagged		Mms1-Myc *pif1-m2*_vs_untagged		Mms1-Myc wildtype_ vs_untagged		Mms1-Myc *pif1-m2*_vs_untagged	
	replicate													
Protein	1	2	1	2	1	2	1	2	1	2	1	2	1	2
ACC1	0	3	0	0	0	2		-0,51			0	0	0	0
ACS2	1	2	0	4	2	3		0,81		1,02	0	0	0	0
ACT1	10	12	9	14	9	13	-1,30	-0,53	-1,11	-0,96	0	0	0	0
ADH1	5	7	9	5	6	13	-0,39	-0,86	0,25	-2,28	0	0	0	0
ADK1	0	2	0	1	0	1		0,84			0	0	0	0
AFG3	7	9	6	8	9	11	-1,27	-0,24	-0,75	0,06	0	0	0	0
ALD6	0	6	1	1	0	8		-0,73			0	0	0	0
ARC1	1	3	1	1	0	3		0,83			0	0	0	0
ARF2;ARF1	6	6	5	4	6	7	-0,15	-0,86	0,00	-0,34	0	0	0	0
ARO4	4	4	4	4	3	4	0,13	0,81	0,01	0,05	0	0	0	0
ASC1	1	5	1	3	2	7		-0,83		-2,79	0	0	0	1
ATP1	8	10	6	10	8	11	0,46	-0,04	0,17	-0,85	0	0	0	0
ATP2	7	10	6	10	6	13	0,12	0,13	-0,49	-0,43	0	0	0	0
BFR1	3	3	1	3	2	1	1,16	-0,93		-1,56	0	0	0	0
BSP1	0	4	0	3	6	9		-3,00		-3,19	0	1	0	1
CDC19	15	24	15	17	12	26	-0,17	-0,65	0,07	-1,71	0	0	0	0
CLA4	1	2	1	3	2	3				-1,38	0	0	0	0
CRT10	2	7	0	9	0	0		2,02		2,34	0	0	0	1
CYS4	2	7	1	4	1	6	2,77	-0,19		0,04	1	0	0	0
CYT1	0	2	1	2	1	2		-0,86		-1,45	0	0	0	0
DIA4	1	2	2	1	1	2		-0,15		-0,68	0	0	0	0
DPM1	1	3	0	1	2	3		-0,74			0	0	0	0
ECM33	3	3	2	3	1	3	0,95	0,59	0,86	0,03	0	0	0	0
EFB1	0	2	0	3	2	3		-0,59		-0,80	0	0	0	0
EFT1	3	14	3	6	2	15	0,10	-0,90	0,19	-2,67	0	0	0	1
EGD2	2	3	1	2	1	2	0,93	1,16		0,54	0	0	0	0
ENO2	8	16	10	11	10	19	-1,17	-0,26	-0,02	-1,98	0	0	0	0
ERG11	1	4	4	2	2	6		-0,39	-0,25	-0,79	0	0	0	0
ERG27	1	3	0	1	0	2		0,98			0	0	0	0
ERG6	2	4	2	3	2	4	-0,08		0,08	-0,50	0	0	0	0
FAA1	2	6	2	3	1	4	-0,04	0,06	-0,24	0,21	0	0	0	0
FAS1	0	2	0	0	0	7		-0,75			0	0	0	0
FBA1	3	5	3	4	4	6	-2,59	-1,46	-1,83	-3,25	0	0	0	1
FUR1	1	3	0	2	1	4		0,12		-0,41	0	0	0	0
GAS1	3	7	4	6	5	7	-0,47	0,08	-0,40	-0,03	0	0	0	0
GAT1	1	3	0	3	2	3		0,50		0,49	0	0	0	0

GFA1	2	7	3	5	2	5		1,19		0,69	0	0	0	0
GPD2	2	3	2	1	3	5		-0,35			0	0	0	0
GPM1	1	6	4	5	4	5		-0,03	-0,34	-1,04	0	0	0	0
GSF2	2	0	1	3	0	3				-0,13	0	0	0	0
GUA1	0	2	0	1	0	4		-0,49			0	0	0	0
GUS1	2	5	2	1	1	5		-0,22			0	0	0	0
HOM3	1	2	0	2	1	2		-0,64		-0,76	0	0	0	0
HSC82;HSP82	8	14	8	9	5	14	0,31	0,01	0,04	-0,97	0	0	0	0
HSP26	4	6	5	4	4	6	-0,47	0,50	-0,54	0,59	0	0	0	0
HTZ1;HTA2;HTA1	1	2	1	1	1	2		-1,57			0	0	0	0
HXK2	0	3	0	1	0	2		-0,30			0	0	0	0
HYP2	1	3	2	2	2	3		-0,57	-0,64	-0,64	0	0	0	0
ILV2	3	9	4	8	4	11	-0,95	-0,51	-0,65	-0,64	0	0	0	0
ILV5	0	6	0	1	1	6		-0,03			0	0	0	0
ILV6	2	1	1	0	0	1	-0,59				0	0	0	0
IPP1	0	2	0	0	0	3		-0,95			0	0	0	0
KAP123	3	3	2	2	2	6	0,90	-1,27			0	0	0	0
KAR2	3	3	3	1	2	2			1,15		0	0	0	0
LAT1	1	3	2	4	2	2		1,02	0,28	0,43	0	0	0	0
LEU1	3	4	2	1	1	3	1,05				0	0	0	0
LHS1	2	2	1	1	3	3	-1,55				0	0	0	0
LYS21;LYS20	1	4	2	2	2	4		0,72	-0,50	-0,42	0	0	0	0
MBF1	1	3	2	3	2	2		1,55	0,45	1,83	0	0	0	0
MDJ1	3	2	1	3	2	2	2,26	0,45	-0,28	0,61	1	0	0	0
MES1	1	2	0	0	0	4		-0,49			0	0	0	0
MET6	0	5	0	0	0	10		-0,80			0	0	0	0
MHR1	1	2	1	2	0	0				0,76	0	0	0	0
MIR1	0	3	1	2	1	4		-0,77			0	0	0	0
MMS1	45	54	44	51	1	1	7,13	9,64	6,97	9,64	2	2	2	2
MRPL39	2	2	1	2	1	2	0,67	-0,19		0,38	0	0	0	0
NDE1	0	1	1	1	1	2		-0,03			0	0	0	0
NOP1	1	6	1	4	1	5		0,50		-0,04	0	0	0	0
NOP56	2	4	4	2	2	5	-0,97	-0,08	0,01	-0,64	0	0	0	0
NOP58	0	3	0	2	0	3		0,66			0	0	0	0
NPL3	3	4	4	3	4	3	-0,18	0,83	-0,20	0,58	0	0	0	0
NSR1	0	3	0	3	2	4		-0,50		-0,59	0	0	0	0
OLE1	0	4	1	3	2	5		-0,38		-0,96	0	0	0	0
PAB1	1	3	0	3	1	3		-0,46		-1,95	0	0	0	0
PDC1	25	25	22	24	23	26	0,05	0,40	0,07	-0,96	0	0	0	0
PET9	11	10	11	11	10	12	-0,35	0,14	-0,38	-0,16	0	0	0	0
PFK1	1	5	3	2	1	5		1,23			0	0	0	0
PFK2	2	6	2	4	1	6	-0,44	0,14	-0,36	-0,48	0	0	0	0
PGK1	0	13	3	5	4	16		-2,23	-0,86	-3,79	0	1	0	1
PHO84	2	1	2	0	2	2	-1,86		-1,66		0	0	0	0
PHO88	1	3	1	3	2	5				-0,49	0	0	0	0
PMA1;PMA2	14	18	15	19	18	21	-0,92	-0,70	-0,74	-0,72	0	0	0	0
PNO1	1	3	3	2	1	4		0,17	1,23		0	0	0	0
POR1	1	3	2	3	2	3		-0,36	1,68	-0,59	0	0	0	0
PST2	1	1	1	0	1	1		-0,12			0	0	0	0
RHO1	0	3	1	3	1	2				2,62	0	0	0	1
RNR1	3	9	4	6	6	9	-0,91	0,28	-1,43	-0,42	0	0	0	0
RNR4	0	3	0	0	0	3		0,40			0	0	0	0
RPL10	9	11	7	11	8	10	-0,60	-0,22	0,06	0,21	0	0	0	0
RPL11B;RPL11A	2	2	2	2	2	2	-1,06	-1,25	-0,84	-0,55	0	0	0	0
RPL12B;RPL12A	4	4	4	2	3	4	-0,09	-0,87	0,09	-0,85	0	0	0	0
RPL13A;RPL13B	2	2	2	1	0	2		2,55			0	1	0	0
RPL14B;RPL14A	2	4	3	1	2	3	0,14	1,53	0,33		0	0	0	0
RPL15B;RPL15A	0	2	1	1	0	1				0,86	0	0	0	0
RPL16B;RPL16A	1	0	0	0	1	0	-2,31				0	0	0	0
RPL17B;RPL17A	0	0	0	1	1	2				-0,40	0	0	0	0
RPL18B;RPL18A	3	4	4	3	3	4	0,18	0,54	0,47	0,48	0	0	0	0
RPL19B;RPL19A	2	5	1	5	2	3	2,96	0,04		0,52	1	0	0	0
RPL1B;RPL1A	0	3	1	2	0	3		-0,08		-0,98	0	0	0	0
RPL20B;RPL20A	2	3	5	4	2	4	4,07	0,43	3,93	0,06	2	0	2	0
RPL21A	1	1	1	1	1	1	-0,62	0,27	-0,83	0,63	0	0	0	0
RPL23B;RPL23A	1	4	2	4	4	4		-1,31	-1,73	-1,31	0	0	0	0
RPL24B;RPL24A	1	2	1	2	1	3				0,18	0	0	0	0
RPL26A	2	3	2	4	3	4	-1,63	-1,43	-2,64	0,01	0	0	1	0
RPL26B	4	4	4	4	4	4	-1,03	-0,92	-2,10	0,02	0	0	0	0
RPL27B;RPL27A	4	4	4	5	5	5	-0,49	0,09	-0,56	0,04	0	0	0	0

RPL28	1	4	3	5	3	4		0,25	-1,26	-0,29	0	0	0	0
RPL2B;RPL2A	1	2	1	3	2	4		0,32		0,18	0	0	0	0
RPL31A;RPL31B	5	3	4	3	4	4	-1,53	-0,11	-0,56	0,66	0	0	0	0
RPL32	5	6	5	6	6	6	-0,96	0,35	-1,25	0,80	0	0	0	0
RPL33A;RPL33B	3	3	3	3	3	3	0,21	1,48	-0,10	0,94	0	0	0	0
RPL35B;RPL35A	9	9	8	11	10	7	-0,72	0,94	-1,07	0,81	0	0	0	0
RPL36A;RPL36B	5	5	3	4	3	4	-0,78	1,75	-0,90	0,80	0	0	0	0
RPL38	2	2	2	3	2	4	-1,93	-1,35	-1,41	-0,61	0	0	0	0
RPL39	5	2	2	3	3	5	0,21	-0,74	-0,94	-1,14	0	0	0	0
RPL4B;RPL4A	3	7	4	4	4	8	0,02	-0,14	-0,94	0,17	0	0	0	0
RPL9A;RPL9B	5	4	5	5	6	4	-0,66	0,83	-0,62	0,55	0	0	0	0
RPS0B;RPS0A	4	7	4	7	4	7	-0,48	-0,63	-0,90	-0,95	0	0	0	0
RPS11B;RPS11A	5	4	6	6	7	8	-0,93	-0,57	-0,30	0,10	0	0	0	0
RPS12	4	5	4	5	4	5	0,35	-0,12	0,14	-0,28	0	0	0	0
RPS13	4	5	3	4	2	4	-0,65	-0,06	-0,43	0,10	0	0	0	0
RPS14A;RPS14B	3	5	4	3	3	5	0,69	0,04	1,03	0,56	0	0	0	0
RPS16B;RPS16A	9	8	9	9	9	10	-0,94	0,07	-0,55	0,14	0	0	0	0
RPS17B;RPS17A	5	5	5	5	5	5	-0,17	1,17	-0,47	0,82	0	0	0	0
RPS18B;RPS18A	6	7	6	7	7	7	-1,46	-1,29	-1,64	-0,94	0	0	0	0
RPS19B;RPS19A	2	3	2	5	4	5	-1,57	-1,47	-1,78	-0,46	0	0	0	0
RPS1B	3	3	3	3	3	3	-0,86	-0,08	-0,84	-0,01	0	0	0	0
RPS2	4	4	3	4	4	4	0,82	0,36	1,69	0,50	0	0	0	0
RPS20	5	5	5	6	5	4	0,24	0,64	0,41	0,77	0	0	0	0
RPS22B;RPS22A	6	4	5	6	6	7	1,09	0,25	0,48	-0,38	0	0	0	0
RPS23B;RPS23A	0	1	2	2	2	4			-0,87	-0,49	0	0	0	0
RPS25A;RPS25B	3	2	1	2	4	3	-1,19	-0,52		-0,70	0	0	0	0
RPS26B;RPS26A	2	3	2	3	2	3	-1,19	-0,69	-1,28	-0,95	0	0	0	0
RPS29A	0	1	0	0	0	1		-0,55			0	0	0	0
RPS3	9	11	8	13	8	12	-0,61	-0,54	-0,08	0,09	0	0	0	0
RPS31	6	7	4	5	6	6	-0,98	-0,30	-1,12	-0,69	0	0	0	0
RPS4B;RPS4A	6	9	8	8	6	7	-0,02	0,65	0,05	0,41	0	0	0	0
RPS5	12	14	11	13	12	14	-0,35	0,37	-0,52	-0,01	0	0	0	0
RPS7A	7	7	6	7	7	7	0,25	-0,15	-0,04	-0,22	0	0	0	0
RPS7B	5	4	3	4	3	4	-0,39	0,30	-0,31	0,38	0	0	0	0
RPS8B;RPS8A	2	5	5	4	3	4	-0,99	-0,15	-0,53	0,03	0	0	0	0
RPS9B;RPS9A	8	8	7	8	7	8	-0,63	0,59	-0,06	0,63	0	0	0	0
RPT1	3	1	4	2	1	4	0,02		-0,10	0,59	0	0	0	0
RPT6	1	2	2	1	1	1			-2,22		0	0	0	0
RTN1	1	1	2	3	2	3			0,31	-0,72	0	0	0	0
RTT101	21	29	21	28	0	1	4,56	7,64	4,02	7,68	2	2	2	2
RUB1	1	2	1	1	0	0				1,56	0	0	0	0
SAM1	2	3	2	2	1	3		1,41	4,73	0,02	0	0	2	0
SAM2	1	3	1	2	1	3	-0,40	0,29	0,01	-0,44	0	0	0	0
SAR1	3	2	2	2	5	4	0,04	0,26		-0,08	0	0	0	0
SCP160	0	3	0	0	0	2		0,29			0	0	0	0
SDH1	3	2	3	2	3	3	-1,06	-1,37	-0,86	-0,38	0	0	0	0
SDS24	0	1	0	1	0	1				-0,46	0	0	0	0
SEC14	2	3	1	3	1	1		1,71		1,58	0	0	0	0
SEC4	3	3	2	2	3	3	-0,03	0,12	-0,32		0	0	0	0
SEC53	1	3	1	1	1	2		-1,72			0	0	0	0
SGT2	1	3	1	2	1	4		0,00			0	0	0	0
SHM1	0	3	0	0	1	0		-0,59			0	0	0	0
SIS1	13	15	12	17	14	15	-0,44	0,88	-0,11	0,97	0	0	0	0
SLI15	0	1	0	1	0	2				0,65	0	0	0	0
SSA1	1	3	1	4	1	4		0,65		0,32	0	0	0	0
SSA2	4	5	5	7	3	6	0,51	0,68	0,24	0,09	0	0	0	0
SSB1;SSB2	13	13	13	11	11	16	0,21	0,08	0,24	-0,80	0	0	0	0
SSC1	1	4	1	3	1	4		0,15		-1,19	0	0	0	0
SSE1;SSE2	1	3	1	2	1	4		0,05			0	0	0	0
SSZ1	0	4	1	3	0	5		0,00		-0,63	0	0	0	0
TDH2	0	3	1	2	2	3		0,67		-1,54	0	0	0	0
TDH3	0	6	1	6	3	7	-1,05	-1,06	-0,62	-2,54	0	0	0	0
TEF1	22	21	18	18	18	20	0,27	0,48	0,01	-0,07	0	0	0	0
TEF4	7	8	8	7	8	8	0,49	-0,07	0,36	-0,24	0	0	0	0
TIF1	2	3	1	2	2	4		0,08			0	0	0	0
TPI1	0	5	0	2	3	7		-2,00		-3,85	0	0	0	1
TRM1	22	21	23	20	20	23	0,09	0,79	0,14	0,46	0	0	0	0
TSA1	0	2	0	2	0	2		0,47			0	0	0	0
TUB2	2	3	2	5	2	4	0,33	0,51		0,12	0	0	0	0
TY1B	7	7	6	8	8	9	0,49	0,40	0,18	0,18	0	0	0	0

TY2B	2	1	1	1	1	2	-1,26				0	0	0	0
TYS1	1	3	2	2	0	3		-0,36			0	0	0	0
URA2	12	14	12	16	11	13	0,10	0,89	0,21	1,04	0	0	0	0
VMA4	1	1	0	0	0	0		-1,26			0	0	0	0
VTC4	1	4	2	1	1	4		-0,67	-1,76		0	0	0	0
WRS1	1	3	0	0	0	2		1,54			0	0	0	0
YBT1	3	5	4	3	3	2	0,48	0,02	0,67	0,10	0	0	0	0
YDJ1	0	2	0	2	1	1				2,11	0	0	0	0
YEF3	6	12	3	10	5	11	1,44	-0,27	0,36	-1,20	0	0	0	0
YHB1	0	3	0	1	1	3		-1,50			0	0	0	0
YHR020W	3	13	4	7	5	13	-0,81	-0,25	0,24	-0,87	0	0	0	0
YPR003C	2	0	0	0	0	0	0,38				0	0	0	0
YRA1	6	6	6	6	6	5	0,63	1,56	1,29	1,79	0	0	0	0

10 Publication

Wanzek K, Capra JA, Paeschke K (2016) Mms1 binds to G-rich regions in *Saccharomyces cerevisiae* and influences replication and genome stability. *EMBO.* Ahead of being submitted.

11 Curriculum Vitae

12 Acknowledgements

Most of all I would like to thank Dr. Katrin Paeschke for offering the interesting PhD project, for her support and valuable discussions during the entire thesis, for performing Pif1-ChIP experiments, for help with other experiments and for providing the opportunity to attend international conferences.

I also want to thank my second supervisor Prof. Hans-Joachim Lipps and my third supervisor Prof. Caroline Kisker for their effort, constructive ideas and support during the thesis, as well as Prof. Alexander Buchberger for serving as chair.

Then, I want to thank Julia Wille for the effort she put into the experiments during her bachelor thesis, some of which were used in this thesis, as well as for her independent and reliable work.

Additionally, I would like to thank John A. Capra, Stefan Juranek and Elmar Wolf for help with and performing some biocomputational analyses.

Furthermore, I would like to thank the group of Andreas Schlosser for MS analysis and evaluation, the groups of Caroline Kisker and Frank Würthner for help with CD analysis as well as the groups of Boris Pfander and Martin Eilers for help with FACS analysis, the groups from the Institute for Biochemistry, especially the group of Katrin Paeschke, and the group of Stefan Gaubatz for experimental devices as well as overall help, and Dr. Katrin Paeschke, Silvia Götz and Sabrina Bartsch for providing some of the yeast strains used in this thesis.

I also thank the GSLS for financial support to attend an international conference.

Additionally, I want to thank Dr. Katrin Paeschke and Amy-Jayne Hutchings for proofreading this thesis.

One special thanks goes to my parents and friends for their motivation, all kind of support, their patience and understanding.

Affidavit

I hereby confirm that my thesis entitled "The investigation of the function of repair proteins at G-quadruplex structures in Saccharomyces cerevisiae revealed that Mms1 promotes genome stability" is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Würzburg, 30.09.16
Place, Date Signature

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation "Die Untersuchung der Funktion von Reparaturproteinen an G-Quadruplex Strukturen in Saccharomyces cerevisiae zeigte, dass Mms1 Genomstabilität fördert" eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg, 30.09.16
Ort, Datum Unterschrift